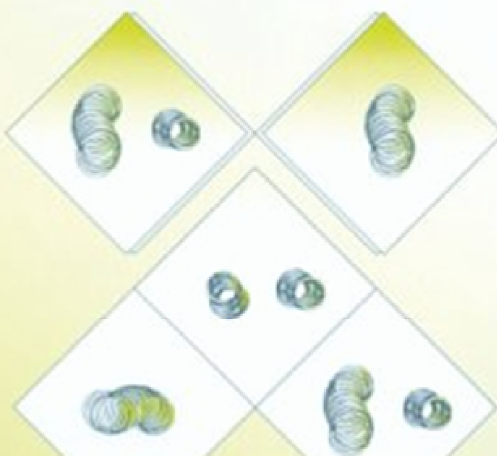
Economic Growth and Distribution

On the Nature and Causes of the Wealth of Nations



Edited by

Neri Salvadori

Economic Growth and Distribution

On the Nature and Causes of the Wealth of Nations

Edited by

Neri Salvadori

Professor of Economics, University of Pisa, Italy

2006

Edward Elgar

Cheltenham, UK • Northampton, MA, USA

Contents

Figures

Tables

Contributors

Michele Bagella, *University of Rome Tor Vergata, Italy*

Nelson H. Barbosa-Filho, *Federal University of Rio de Janeiro, Brasil*

Leonardo Becchetti, *University of Rome Tor Vergata, Italy*

Gabriella Berloffa, *University of Trento, Italy*

Graziella Bertocchi, *University of Modena and Reggio Emilia, Italy*

Amit Bhaduri, *University of Pavia, Italy*

Stefano Caiazza, *University of Rome Tor Vergata, Italy*

Sergio Cesaratto, *University of Siena, Italy*

Maurizio Ciaschini, *University of Macerata, Italy*

Claudio H. Dos Santos, *Levy Institute of Economics, USA*

Oscar De-Juan, *University of Castilla La Mancha, Spain*

Domenico Delli Gatti, *Catholic University of Milan, Italy*

Duncan K. Foley, *New School University, USA*

Mauro Gallegati, *Politechnical University of Marche, Italy*

Oded Galor, *Brown University, USA*

Cecilia García-Peñalosa, *GREQAM and CNRS, France*

Fabio Hideki Ono, *Universidade Federal do Paraná, Brasil*

Ferdinando Meacci, *University of Padua, Italy*

José Luís Oreiro, *Universidade Federal do Paraná, Brasil*

Codrina Rada, *New School University, USA*

Alberto Russo, *Politechnical University of Marche, Italy*

Maria Luigia Segnana, *University of Trento, Italy*

Claudio Socci, *Politechnical University of Marche, Italy*

Lance Taylor, *New School University, USA*

Stephen J. Turnovsky, *University of Washington, USA*

Graham White, *University of Sydney, Australia*

Gennaro Zezza, *University of Cassino, Italy*

Introduction

Neri Salvadori

The interest in the study of economic growth has experienced remarkable ups and downs in the history of economics. It was central in classical political economy from Adam Smith to David Ricardo, and then in the critique of it by Karl Marx, but was moved to the periphery during the so-called 'marginal revolution'. John von Neumann's growth model and Roy Harrod's attempt to generalise Keynes's principle of effective demand to the long run re-ignited an interest in growth theory. Following the publication of papers by Robert Solow and Nicholas Kaldor in the mid 1950s, growth theory became one of the central topics of the economics profession until the early 1970s. After a decade of dormancy, since the mid 1980s, economic growth has once again become a central topic in economic theorizing. The recent 'new' growth theory (NGT) is also called 'endogenous growth theory', since the growth rate is determined accordingly from within the model and is not given as an exogenous variable.

The interaction between economic growth and distribution was the hallmark of classical economic theorizing. After the Second World War this theme experienced a revival, especially within the post-Keynesian, classical, and Marxian schools. With the development of the 'new growth theory', the connection between distribution and growth has become the subject of intensive research. It has been a lens through which the complex interplay of the factors explaining the nature and causes of the wealth of nations has been investigated. Particular attention has been devoted to population growth, structural change, technological progress and (physical, social and human) capital accumulation.

A conference held in Lucca in the summer of 2004 was a forum for the presentation and discussion of different approaches to the issues of growth and distribution, and their theoretical, empirical, historical and methodological implications. This book is the main product of the conference. Other papers will soon appear in a special issue of *Metroeconomica* (2006).[1] The conference was hosted by a research group, and several of the papers elaborated by members of the group were delivered at the conference. The main products of the research group are companion books on *Innovation, Unemployment and Policy in the Theories of Growth and Distribution* (Salvadori and Balducci, 2006) and on *Classical, Neoclassical and Keynesian Views on Growth and Distribution* (Salvadori and Panico, 2006). There is, of course, no overlap among the mentioned publications, which constitute the proceedings of the conference. The papers more directly related to the title of the conference have been inserted in this volume, which shares the same title as the conference itself. The result is that the book analyses the recent developments in the interplay of economic growth and distribution.

The book opens with a chapter by Oded Galor, who proposes a unified theory able to explain both the epoch of Malthusian stagnation, characterizing most of human history, and the contemporary era of modern economic growth. The proposed theory also underlies the driving forces that triggered the transition between these regimes and the associated phenomenon of the Great Divergence in income per capita across countries. It unifies two fundamental approaches regarding the effect of income distribution on the process of development: the classical approach and the Credit Market Imperfection approach. In this way an intertemporal reconciliation between the conflicting viewpoints regarding the effect of inequality on economic growth is provided. The classical viewpoint is interpreted as reflecting the state of the world in early stages of industrialization when physical capital accumulation was the prime engine of economic growth. In contrast, the credit market imperfection approach is interpreted as reflecting later stages of development, when human capital accumulation becomes a prime engine of economic growth and credit constraints are largely binding. The following chapter, by Amit Bhaduri, also seeks to blend elements of the classical tradition with modern theory. Bhaduri focuses on the Keynesian theory of effective demand and the Schumpeterian emphasis on the influence of market structures on technological change. The third chapter, by Ferdinando Meacci is, by contrast, mainly historical: it completes the two previous viewpoints on the classical economists with a reconstruction of Smith's competition-of-capitals doctrine.

The next four chapters introduce heterodox models and comparison among them. Chapter 4, by Duncan Foley and Lance Taylor, describes a heterodox macroeconomic model put together with two explicit aims in mind: 'to set out a benchmark for comparison of heterodox and orthodox approaches to economic growth and income distribution, and to point out similarities shared by a wide range of heterodox models'. Chapter 5, by Gennaro Zezza and Claudio Dos Santos, presents a stock-flow model of growth for a closed economy that encompasses virtually all one-sector post-Keynesian growth models as special cases and uses it to analyse the relationship between growth and the distribution of income in financially sophisticated economies. Chapter 6, by Fabio Hideki Ono and José Luís Oreiro, presents a post-Keynesian growth model in which, on the one hand, the mark-up rate varies in the long-term due to a misalignment between the actual rate and the 'desired' profit rate and, on the other, the capital–output ratio may shift as a result of technological progress. Finally, chapter 7, by Graham White, analyses not only the differences between post-Keynesian and Kaleckian growth theory but also the implications for growth theory flowing from a Sraffian analysis of value and distribution.

Chapters 8 to 11 introduce problems of policy. Chapter 8, by Cecilia García-Peñalosa and Stephen J. Turnovsky, employs a stochastic growth model to analyse the effect of macroeconomic volatility on the relationship between income distribution and growth. In the first part of the chapter, the authors first show how the distribution of income depends upon the initial distribution of capital and the equilibrium labour supply and then find that an increase in volatility raises the mean growth rate and income inequality. The second part of the chapter uses this framework to analyse the design of tax policy to achieve desired growth, distribution and welfare objectives. Chapter 9, by Sergio Cesaratto, brings new insights on the current debates on pension reforms. The chapter seeks to show that even if a reform aiming to create a fully funded pension scheme (based on the accumulation of real assets) is successful at raising the marginal propensity to save, the larger potential saving supply is not necessarily translated into an increased amount of investment. Chapter 10, by Maurizio Ciaschini and Claudio Socci, introduces the income distribution process in a SAM (Social Accounting Matrix) and applies the method to the relationship between income distribution and output change in a region in Italy. A SAM is also used in Chapter 11, by Oscar De-Juan, to build a model which may be relevant both to policy evaluation and growth analysis.

The next two chapters introduce dynamics and business cycles. Chapter 12, by Lance Taylor, Nelson H. Barbosa-Filho and Codrina Rada, outlines an approach to the analysis of cyclical macroeconomic fluctuations in industrialized economies based on low-order systems of differential equations. It combines partial models of both the real and financial sides of the economy into a higher order analytical framework, which may shed light both on observed cycles and their policy. Chapter 13, by Alberto Russo, Domenico Delli Gatti and Mauro Gallegati, suggests a scaling approach to business cycles by developing a heterogeneous interacting agents model that replicates well-known stylized facts of industrial dynamics; agent-based simulations show that power law shifts are a consequence of changes in firms' capital accumulation behaviour due to technological progress and a wage–firm size relationship.

The last (but not least) three chapters introduce institutions into the picture. Chapter 14, by Graziella Bertocchi, illustrates the ongoing research line which adds a historical and institutional dimension to economic growth analysis both at the theoretical and empirical level: it presents applications of this research strategy to the impact of colonization on growth, the extension of the franchise and the welfare state, the evolution of educational systems, the relationship between industrialization and democratization, and international migration. Chapter 15, by Michele Bagella, Leonardo Becchetti and Stefano Caiazza, argues that religious differences among countries are crucial determinants of the evolution of market rules and financial institutions; it shows that a positive link between institutions and growth arises only in those countries whose cultural background allowed them to reach a sufficient degree of institutional development and that the effect of institutions on growth is, for a significant part, exogenous. Chapter 16, by Gabriella Berloffa and Maria Luigia Segnana, questions the views that trade liberalization 'is always good for growth' and that 'growth is always good for the poor' and argues that the problem of poverty reduction cannot be separated from the context in which trade is liberalized.

Almost all the chapters of this book as well as all the papers included in the special issue of *Metroeconomica* have been peer-reviewed (the exceptions are the invited lectures to the conference). I would like to take this opportunity to thank all the referees who contributed to improving the published papers and advised me of their publishability. The following scholars helped me with this task:[2] Syed M. Ahsan (Concordia University, Canada), Fahim Al-Marhubi (Department of Economics, Sultanate of Oman), Nelson H. Barbosa Filho (Federal University of Rio de Janeiro, Brazil), Leonardo Becchetti ('Tor Vergata' University, Rome, Italy), Enrico Bellino (University Cattolica del Sacro Cuore, Milan, Italy), Roland J. Benabou (Princeton University, USA), Amit Bhaduri (University of Pavia, Italy), Claudia Biancotti (Bank of Italy), Harry Bloc (Curtin University of Technology, Australia), Tony Brewer (University of Bristol, UK), Elise S. Brezis (Bar-Ilan University, Israel), Alberto Bucci (Milan University, Italy), Maria Rosaria Carillo ('Parthenope' University, Naples, Italy), Maurizio Ciaschini (Macerata University, Italy), Mario Cimoli (United Nations Economic Commission for Latin America and the Caribbean), Pasquale Commendatore ('Federico II' University, Naples, Italy), Guido Cozzi (Macerata University, Italy), John Cranfield (University of Guelph, Canada), Francesco Daveri (Parma University, Italy), Erik Dietzenbacher (University of Groningen, Netherlands), Francesco Drago (Siena University, Italy) Amitava Dutt (University of Notre Dame, USA), Alessandro Federici ('La Sapienza' University, Rome, Italy), Jesus Felipe (Asian Development Bank, Philippines), Davide Fiaschi (Pisa University, Italy), Franklin M. Fisher

(Massachusetts Institute of Technology, USA), Peter Flaschel (Bielefeld University, Germany), Reto Föllmi (University of Zurich, Switzerland), James K. Galbraith (University of Texas, Austin, USA), Mauro Gallegati (Polytechnic University of Marche, Ancona, Italy), Christian Gehrke (University of Graz, Austria), Wynne Godley (University of Cambridge, UK), Gianluca Grimalda (University of Warwick, UK), Volker Grossman (University of Zurich, Switzerland), André A. Hofman (Economic Development Division of the United Nations Economic Commission for Latin America and the Caribbean), Nazrul Islam (Emory University, USA), Zamir Iqbal (World Bank, USA), Alan Kirman (Paul Cézanne University (Aix-Marseille III) and GREQAM, France), Heinz Kurz (University of Graz, Austria), Christian Lager (University of Graz, Austria), Michael Landesmann (Johannes Kepler University, Linz, Austria), Marc Lavoie (University of Ottawa, Canada), Frederic Lee (University of Missouri-Kansas City, USA), Marco Lippi ('La Sapienza' University at Rome, Italy), Humberto Llavador (Pompeu Fabra University, Spain), Piero Manfredi (Pisa University, Italy), Yishai Maoz (University of Haifa, Israel), Angelo Marano (Tuscia University, Italy), Adalmir A. Marquetti (Pontíficia Universidade Católica do Rio Grande do Sul, Brazil), Nicola Meccheri (Pisa University, Italy), Alfredo Medio (Udine University, Italy), Antonio J.A. Meirelles (State University of Campinas, Brazil), Raffaele Miniaci (Padua University, Italy), Marco Missaglia (Pavia University, Italy), Edward Nell (New School University, USA), Antonella Palumbo ('Roma Tre' University, Italy), Ugo Panizza (Inter-American Development Bank, USA), Chris Papageorgiou (Louisiana State University, USA), Carmelo Parello (Université Catholique de Louvain, Belgium), Manseop Park (Korea University, Korea), Joe Pearlman (London Metropolitan University, UK), Fidel Perez (Alicante University, Spain), Apostolis Philippopoulos (Athens University of Economics & Business, Greece), Mario Pomini (Padua University, Italy), Codrina Rada (New School University, USA), Jaime Ros (University of Notre Dame, USA), J. Barkley Jr. Rosser (James Madison University, USA), Rafaella Sadun (London School of Economics and Political Science, UK), Roger Sandilands (University of Strathclyde, UK), Paul Segal (Oxford Univesity, UK), Mark Setterfield (Trinity College, USA), Rodolfo Signorino (Palermo University, Italy), Peter Skott (University of Massachusetts, Amherst, USA), Serena Sordi (Siena University, Italy), Luca Spataro (Pisa University, Italy), Richard Sturn (University of Graz, Austria), Gilberto Tadeu Lima (São Paulo University, Brazil), Lance Taylor (New School University, USA), Alberto Zazzaro (Politechnical University of Marche, Ancona, Italy), Katia Zhuravskaya (Center for Economic and Financial Research at Moscow, Russia) and Stefania Zotteri (Bank of Italy).

Last but not least, I wish to thank the members of the Scientific Committee of the Lucca Conference who shared with me the responsibility of selecting the papers to be given at the meeting, namely Duncan K. Foley (New School for Social Research, USA), Oded Galor (Brown University, USA, and Hebrew University of Jerusalem, Israel), Heinz D. Kurz (University of Graz, Austria) and Stephen J. Turnovsky (University of Washington, USA). Oded Galor and Heinz D. Kurz also provided advice in choosing referees during the editing of the proceedings.

NOTES

1. The call for papers of the Conference also provided special issues of two other journals: the *European Journal of the History of Economic Thought* and the *Journal of Economic Growth*. Unfortunately, the papers which were considered suitable for these outlets of the conference were too small in number to be able to produce the special issues mentioned. One paper suitable for the *European Journal of the History of Economic Thought* is published in this volume. One paper suitable for the *Journal of Economic Growth* will be published in that journal.
2. The list does not include the scholars asked to review papers which were made available only for the special issue of the *Journal of Economic Growth* by their authors, as these papers were processed directly by the office of the journal.

REFERENCES

Salvadori N. and R. Balducci (2006), *Innovation, Unemployment and Policy in the Theories of Growth and Distribution*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Salvadori N. and C. Panico (2006), *Classical, Neoclassical and Keynesian Views on Growth and Distribution*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

1. Inequality and the process of development

Oded Galor

1.1. INTRODUCTION

The evolution of economies during the major portion of human history was marked by Malthusian stagnation. Technological progress and population growth were minuscule by modern standards and the average growth rate of income per capita in various regions of the world was even slower due to the offsetting effect of population growth on the expansion of resources per capita. In the past two centuries, in contrast, the pace of technological progress increased significantly in association with the process of industrialization. Various regions of the world departed from the Malthusian trap and experienced initially a considerable rise in the growth rates of income per capita and population. Unlike episodes of technological progress in the pre-Industrial Revolution era that failed to generate sustained economic growth, the increasing role of human capital in the production process in the second phase of industrialization ultimately prompted a demographic transition, liberating the gains in productivity from the counterbalancing effects of population growth. The decline in the growth rate of population and the associated enhancement of technological progress and human capital formation paved the way for the emergence of the modern state of sustained economic growth.

The transitions from a Malthusian epoch to a state of sustained economic growth and the related phenomenon of the Great Divergence, have significantly shaped the contemporary world economy.[1] Nevertheless, the distinct qualitative aspects of the growth process during most of human history were virtually ignored in the shaping of growth models, resulting in a growth theory that is consistent with a small fragment of human history.

The inconsistency of exogenous and endogenous growth models with some of the most fundamental features of process of development, has led recently to a search for a unified theory that would unveil the underlying micro-foundations of the growth process in its entirety, capturing the epoch of Malthusian stagnation that characterized most of human history, the contemporary era of modern economic growth, and the underlying driving forces that triggered the recent transition between these regimes and the associated phenomenon of the Great Divergence in income per capita across countries.

The preoccupation of growth theory with the empirical regularities that have characterized the growth process of developed economies in the past century and of less developed economies in the last few decades, has become harder to justify from a scientific viewpoint in light of the existence of vast evidence about qualitatively different empirical regularities that characterized the growth process over most of human existence. It has become evident that in the absence of a unified growth theory that is consistent with the entire process of development, the understanding of the contemporary growth process would be limited and distorted.

The evolution of theories in older scientific disciplines suggests that theories that are founded on the basis of a subset of the existing observations and their driving forces may be attractive in the short run, but non-robust and ultimately non-durable in the long run.[2] The attempts to develop unified theories in physics have been based on the conviction that all physical phenomena should ultimately be explainable by some underlying unity.[3] Similarly, the entire process of development and its basic causes ought to be captured by a unified growth theory.

The transition from stagnation to growth and the associated phenomenon of the great divergence have been the subject of intensive research in the growth literature in recent years.[4] It has been increasingly recognized that the understanding of the contemporary growth process would be fragile and incomplete unless growth theory could be based on proper micro-foundations that would reflect the various qualitative aspects of the growth process and their central driving forces. Moreover, it has become apparent that a comprehensive understanding of the hurdles faced by less developed economies in reaching a state of sustained economic growth would be futile unless the factors that prompted the transition of the currently developed economies into a state of sustained economic growth could be identified and their implications would be modified to account for the differences in the growth structure of less developed economies in an interdependent world.

Imposing the constraint that a single theory should account for the entire intricate process of development and its prime causes in the last thousands of years is a discipline that enhances the viability of growth theory. A unified theory of economic growth reveals the fundamental micro-foundations that are consistent with the process of economic development over the entire course of human history, rather that with the last century only, boosting the confidence in growth theory, its predictions and policy implications.

Moreover, it improves the understanding of the underlying factors that led to the transition from stagnation to growth of the currently developed countries, shedding light on the growth process of the less developed economies.

The establishment of a unified growth theory has been a great intellectual challenge, requiring major methodological innovations in the construction of dynamical systems that could capture the complexity which characterized the evolution of economies from a Malthusian epoch to a state of sustained economic growth. Historical evidence suggests that the transition from the Malthusian epoch to a state of sustained economic growth, rapid as it may appear, was a gradual process and thus could not plausibly be viewed as the outcome of a major exogenous shock that shifted economies from the basin of attraction of the Malthusian epoch into the basin of attraction of the Modern Growth Regime.[5] The simplest methodology for the generation of this phase transition – a major shock in an environment characterized by multiple locally stable equilibria – was therefore not applicable for the generation of the observed transition from stagnation to growth.

An alternative methodology for the observed phase transition was rather difficult to establish since a unified growth theory in which economies take-off gradually but swiftly from an epoch of a stable Malthusian stagnation would necessitate a gradual escape from an absorbing (stable) equilibrium – a contradiction to the essence of a stable equilibrium. Ultimately, however, it has become apparent that the observed rapid, continuous, phase transition would be captured by a single dynamical system, if the set of steady-state equilibria and their stability would be altered qualitatively in the process of development. As proposed in unified growth theory, first advanced by Galor and Weil (2000), during the Malthusian epoch the dynamical system would have to be characterized by a stable Malthusian equilibrium, but ultimately due to the evolution of latent state variables the dynamical system would change qualitatively, the Malthusian equilibrium would vanish endogenously, leaving the arena to the gravitational forces of the emerging Modern Growth Regime, and permitting the economy to take off and to converge to a modern growth steady-state equilibrium.

The observed role of the demographic transition in the shift from the Post-Malthusian Regime to the Sustained Growth Regime and the associated non-monotonic evolution of the relationship between income per capita and population growth added to the complexity of the desirable dynamical system. In order to capture this additional transition unified growth theory had to generate endogenously, in the midst of the process of industrialization, a reversal in the positive Malthusian effect of income on population, providing the reduction in fertility the observed role in the transition to a state of sustained economic growth.

As discussed in Galor (2005 and 2006), unified growth theory explores the fundamental factors that generated the remarkable escape from the Malthusian epoch and their significant for the understanding of the contemporary growth process of developed and less developed economies. It deciphers some of the most fundamental questions that have been shrouded in mystery: what accounts for the epoch of stagnation that characterized most of human history? What is the origin of the sudden spurt in growth rates of output per capita and population? Why had episodes of technological progress in the pre-industrialization era failed to generate sustained economic growth? What was the source of the dramatic reversal in the positive relationship between income per capita and population that existed throughout most of human history? What triggered the demographic transition? Would the transition to a state of sustained economic growth have been feasible without the demographic transition? And, what are the underlying behavioural and technological structures that can simultaneously account for these distinct phases of development and what are their implications for the contemporary growth process of developed and underdeveloped countries?

Moreover, unified growth theory sheds light on the perplexing phenomenon of the Great Divergence in income per capita across regions of the world in the past two centuries: what accounts for the sudden take-off from stagnation to growth in some countries in the world and the persistent stagnation in others? Why has the positive link between income per capita and population growth reversed its course in some economies but not in others? Why have the differences in per capita incomes across countries increased so markedly in the last two centuries? And has the transition to a state of sustained economic growth in advanced economies adversely affected the process of development in less-developed economies?

Unified growth theory suggests that the transition from stagnation to growth is an inevitable by-product of the process of development. The inherent Malthusian interaction between the technology and population accelerated the pace of technological progress, and ultimately brought about an industrial demand for human capital, stimulating human capital formation, and thus further technological progress, and triggering a demographic transition, enabling economies to convert a larger share of the fruits of factor accumulation and technological progress into growth of income per capita. The timing of the transition may differ, however, across countries. Variations in the economic performance across countries and regions reflect initial differences in geographical factors, and historical accidents and their manifestation in variations in income distribution, institutional, demographic and cultural factors, trade patterns, colonial status, and public policy.

This chapter presents a unified approach for the dynamic implications of income inequality on the process of development. This unified theory of Galor and Moav (2004) provides an intertemporal reconciliation for conflicting viewpoints about the effect of inequality on economic growth. Galor and Moav argue that the replacement of physical capital accumulation by human capital accumulation as a prime engine of economic growth altered the qualitative impact of inequality on the process of development. In early stages of industrialization, as physical capital accumulation is a prime source of economic growth, inequality enhances the process of development by channelling resources towards the owners of capital whose marginal propensity to save is higher. In later stages of development, however, as the return to human capital increases due to capital–skill complementarity, human capital becomes the prime engine of economic growth. Since human capital is inherently embodied in humans and its accumulation is larger if it is shared by a larger segment of society, equality, in the presence of credit constraints, stimulates investment in human capital and promotes economic growth. As income further increases, credit constraints gradually diminish, differences in saving rates decline, and the effect of inequality on economic growth becomes insignificant.

Galor and Moav (2004) develop a growth theory that captures the endogenous replacement of physical capital accumulation by human capital accumulation as a prime engine of economic growth in the transition from the Industrial Revolution to modern growth. The proposed theory offers a unified account for the effect of income inequality on the growth process during this transition. It argues that the replacement of physical capital accumulation by human capital accumulation as a prime engine of economic growth changed the qualitative impact of inequality on the process of development. In the early stages of the Industrial Revolution, when physical capital accumulation was the prime source of economic growth, inequality enhanced the process of development by channelling resources towards individuals whose marginal propensity to save is higher. In the later stages of the transition to modern growth, as human capital emerged as a prime engine of economic growth, equality alleviated the adverse effect of credit constraints on human capital accumulation and stimulated the growth process.

The proposed theory unifies two fundamental approaches regarding the effect of income distribution on the process of development: the classical approach and the Credit Market Imperfection approach.[6] The classical approach was originated by Adam Smith (1776) and was further interpreted and developed by Keynes (1920), Lewis (1954), Kaldor (1957), and Bourguignon (1981). According to this approach, saving rates are an increasing function of wealth, and inequality therefore channels resources towards individuals whose marginal propensity to save is higher, increasing aggregate savings and capital accumulation and enhancing the process of development. Strands of the capital market imperfection approach suggests, in contrast, that equality in sufficiently wealthy economies alleviates the adverse effect of credit constraints on investment in human capital and thereby enhances economic growth (Galor and Zeira, 1993).[7]

The proposed unified theory provides an intertemporal reconciliation between the conflicting viewpoints about the effect of inequality on economic growth. It suggests that the classical viewpoint, regarding the positive effect of inequality on the process of development, reflects the state of the world in early stages of industrialization when physical capital accumulation was the prime engine of economic growth. In contrast, the credit market imperfection approach regarding the positive effect of equality on economic growth reflects later stages of development when human capital accumulation becomes a prime engine of economic growth, and credit constraints are largely binding.

The fundamental hypothesis of this research stems from the recognition that human capital accumulation and physical capital accumulation are fundamentally asymmetric. In contrast to physical capital, human capital is inherently embodied in humans and the existence of physiological constraints subjects its accumulation at the individual level to diminishing returns. The aggregate stock of human capital would be therefore larger if its accumulation were widely spread among individuals in society, whereas the aggregate productivity of the stock of physical capital is largely independent of the distribution of its ownership in society.[8] This asymmetry between the accumulation of human and physical capital suggests therefore that as long as credit constraints are largely binding, equality is conducive for human capital accumulation, whereas provided that the marginal propensity to save increases with income, inequality is conducive for physical capital accumulation.

The theory captures the endogenous replacement of physical capital accumulation by human capital accumulation as a prime engine of economic growth in the transition of the currently advanced economies from the Industrial Revolution to modern growth. It captures the historical intensification in the importance of human capital relative to physical capital in the process of development and its significance for the determination of the effect of inequality on economic growth. The model is based on three central elements, in addition to the fundamental asymmetry between human capital and physical capital. The first element captures the central mechanism in the classical approach. The preference structure is designed such that, consistently with empirical evidence, the marginal propensity to save and to bequeath increases with wealth.[9] Hence, consistently with some empirical evidence, inequality has a positive effect on aggregate savings [10]

The second element captures the central mechanism of the credit market imperfection approach. The economy is characterized by credit constraints that, consistently with empirical evidence, undermine investment in human capital.[11] Although, there is no asymmetry in the ability of individuals to borrow for investment in either human capital or physical capital, credit constraints along with the inherent diminishing marginal returns in the production of human capital generate an inefficient investment only in human capital. Given the competitive neoclassical aggregate production structure, the return to physical capital across all individuals and firms is identical, and individuals, therefore, have no incentive to borrow for investment in physical capital.

The third element is designed to capture the increasing importance of human capital in the process of development. Consistently with historical evidence, the economy is characterized by capital–skill complementarity. The accumulation of physical capital increases the demand for human capital and induces human capital accumulation.[12]

1.1. HISTORICAL EVIDENCE

The transition to a state of sustained economic growth in developed as well as less developed regions was accompanied by a rapid process of industrialization. Per Capita Level of Industrialization (measuring per capita volume of industrial production) doubled in the time period 1860–1913 and tripled in the course of the 20th century. Similarly, the per capita level of industrialization in the United States increased six-fold over the years 1860–1913, and tripled along the 20th century. A similar pattern was experienced in Germany, France, Sweden, Switzerland, Belgium and Canada where industrialization increased significantly in the time interval 1860–1913 as well as over the rest of the 20th century. Moreover, less developed economies that made the transition to a state of sustained economic growth in recent decades have experienced a significant increase in industrialization.

The process of industrialization was characterized by a gradual increase in the relative importance of human capital in the production process. The acceleration in the rate of technological progress increased gradually the demand for human capital, inducing individuals to invest in education, and stimulating further technological advancement. Moreover, in developed as well as less developed regions the onset of the process of human capital accumulation preceded the onset of the demographic transition, suggesting that the rise in the demand for human capital in the process of industrialization and the subsequent accumulation of human capital played a significant role in the demographic transition and the transition to a state of sustained economic growth.

In the first phase of the Industrial Revolution, the extensiveness of the provision of public education was not correlated with industrial development and it differed across countries due to political, cultural, social, historical and institutional factors. Human capital had a limited role in the production process and education served religious, social and national goals. In contrast, in the second phase of the Industrial Revolution the demand for skilled labour in the growing industrial sector markedly increased. Human capital formation was designed primarily to satisfy the increasing skill requirements in the process of industrialization.

This transition was characterized by a gradual increase in the importance of the accumulation of human capital relative to physical capital as well as with a sharp decline in fertility rates. In the first phase of the Industrial Revolution (1760–1830), capital accumulation as a fraction of GDP increased significantly whereas literacy rates remained largely unchanged. Skills and literacy requirements were minimal, the state devoted virtually no resources to raise the level of literacy of the masses, and workers developed skills primarily through on-the-job training (Green, 1990; Mokyr, 1990, 1993). Consequently, literacy rates did not increase during the period 1750–1830 (Sanderson, 1995). As argued by Landes (1969, p. 340)

> although certain workers – supervisory and office personal in particular – must be able to read and do the elementary arithmetical operations in order to perform their duties, large share of the work of industry can be performed by illiterates as indeed it was especially in the early days of the Industrial Revolution.

In the second phase of the Industrial Revolution, however, the pace of capital accumulation subsided, the education of the labour force markedly increased and skills became necessary for production. The investment ratio which increased from 6 per cent in 1760 to 11.7 per cent in 1831, remained at around 11 per cent on average in the years 1856–1913 (Crafts, 1985; Matthews et al., 1982). In contrast, the average years of schooling of male in the labour force, that did not change significantly until the 1830s, tripled by the beginning of the 20th century (Matthews et al., 1982, p. 573). The significant rise in the level of income per capita in England as of 1865 was associated with an increase in the standard of living (Voth, 2004), and an increase in school enrolment of 10-year-olds from 40 per cent in 1870 to 100 per cent in 1900. Notably, the reversal of the Malthusian relation between income and population growth during the demographic transition corresponded to an increase in the level of resources invested in each child.

For example, the literacy rate among men, which was stable at around 65 per cent in the first phase of the Industrial Revolution, increased significantly during the second phase, reaching nearly 100 per cent at the end of the 19th century (Clark, 2003). In addition, the proportion of children aged 5 to 14 in primary schools increased significantly from 11 per cent in 1855 to 74 per cent in 1900. A similar pattern is observed in other European societies (Flora et al., 1983). In particular, the proportion of children aged 5 to 14 in primary schools in France increased significantly in the second phase of industrialization, rising from 30 per cent in 1832 to 86 per cent in 1901.

The transition to a state of sustained economic growth in the US, as well, was characterized by a gradual increase in the importance of the accumulation of human capital relative to physical capital. Over the time period 1890–1999, the contribution of human capital accumulation to the growth process in the US nearly doubled whereas the contribution of physical capital declined significantly. Goldin and Katz (2001) show that the rate of growth of educational productivity was 0.29 per cent per year over the 1890–1915 period, accounting for about 11 per cent of the annual growth rate of output per capita over this period.[13] In the period 1915–99, the rate of growth of educational productivity was 0.53 per cent per year accounting for about 20 per cent of the annual growth rate of output per capita over this period. Abramovitz and David (2000) report that the fraction of the growth rate of output per capita that is directly attributed to physical capital accumulation declined from an average of 56 per cent in the years 1800–1890 to 31 per cent in the period 1890–1927 and 21 per cent in the time interval 1929–66.

Evidence about the evolution of the return to human capital over this period is scarce and controversial. It does not indicate that the skill premium increased markedly in Europe over the course of the 19th century (Clark, 2003). One can argue that the lack of clear evidence about the increase in the return to human capital over this period is an indication for the absence of a significant increase in the demand for human capital. This partial equilibrium argument, however, is flawed. The return to human capital is affected by the demand and the supply of human capital. Technological progress in the second phase of the Industrial Revolution brought about an increase in the demand for human capital, and indeed, in the absence of a supply response, one would have expected an increase in the return to human capital. However, the significant increase in schooling that took place in the 19th century, and in particular the introduction of public education that lowered the cost of education, generated a significant increase in the supply of educated workers. Some of this supply response was a direct reaction to the increase in the demand for human capital, and thus may only operate to partially offset the increase in the return to human capital. However, the removal of the adverse effect of credit constraints on the acquisition of human capital (as reflected by the introduction of public education) generated an additional force that increased the supply of educated labour and operated towards a reduction in the return to human capital.[14]

1.3. THE BASIC STRUCTURE OF THE MODEL[15]

Consider an overlapping-generations economy in a process of development. In every period the economy produces a single homogeneous good that can be used for consumption and investment. The good is produced using physical capital and human capital. Output per capita grows over time due to the accumulation of these factors of production. The stock of physical capital in every period is the output produced in the preceding period net of consumption and human capital investment, whereas the level of human capital in every period is the outcome of individuals' education decisions in the preceding period, subject to borrowing constraints.

1.3.1. Production of Final Output

Production occurs within a period according to a neoclassical, constant-returns-to-scale, production technology. The output produced at time t, Y_t, is

$$Y_t = F(K_t,\ H_t) \equiv H_t f(k_t) = AH_t k_t^{\alpha};\ \ k_t \equiv K_t / H_t;\ \ \alpha \in (0,\ 1), \tag{1.1}$$

where K_t and H_t are the quantities of physical capital and human capital (measured in efficiency units) employed in production at time t, and A is the level of technology. The production function, $f(k_t)$, is therefore strictly monotonic increasing, strictly concave satisfying the neoclassical boundary conditions that assure the existence of an interior solution to the producers' profit-maximization problem.

Producers operate in a perfectly competitive environment. Given the wage rate per efficiency unit of labour, w_t, and the rate of return to capital, r_t, producers in period t choose the level of employment of capital, K_t, and the efficiency units of labour, H_t, so as to maximize profits. That is, $\{K_t,\ H_t\} = \arg\max\ [H_t f(k_t) - w_t H_t - r_t K_t]$. The producers' inverse demand for factors of production is therefore

$$\begin{aligned} r_t &= f'(k_t) = \alpha A k_t^{\alpha-1} \equiv r(k_t); \\ w_t &= f(k_t) - f'(k_t)k_t = (1-\alpha)Ak_t^{\alpha} \equiv w(k_t). \end{aligned} \tag{1.2}$$

1.3.2. Individuals

In every period a generation which consists of a continuum of individuals of measure 1 is born. Each individual has a single parent and a single child. Individuals, within as well as across generations, are identical in their preferences and innate abilities. They may differ, however, in their family wealth and thus, due to borrowing constraints, in their investment in human capital.

Individuals live for two periods. In the first period of their lives individuals devote their entire time to the acquisition of human capital. The acquired level of human capital increases if their time investment is supplemented with capital investment in education.[16] In the second period of their lives (adulthood), individuals supply their efficiency units of labour and allocate the resulting wage income, along with their inheritance, between consumption and transfers to their children. The resources devoted to transfers are allocated between an immediate finance of their offspring's expenditure on education and saving for the future wealth of their offspring.

1.3.2.1. Wealth and preferences

In the second period of life, an individual i born in period t (a member i of generation t) supplies the acquired efficiency units of labour, h^i_{t+1}, at the competitive market wage, w_{t+1}. In addition, the individual receives an inheritance of x^i_{t+1}. The individual's second period wealth, I^i_{t+1}, is therefore

$$I^i_{t+1} = w_{t+1}h^i_{t+1} + x^i_{t+1}. \tag{1.3}$$

The individual allocates this wealth between consumption, c^i_{t+1}, and transfers to the offspring, b^i_{t+1}. That is,

$$c^i_{t+1} + b^i_{t+1} \leq I^i_{t+1}. \tag{1.4}$$

The transfer of a member i of generation t, b^i_{t+1}, is allocated between an immediate finance of their offspring's expenditure on education, e^i_{t+1}, and saving, s^i_{t+1}, for the future wealth of their offspring. That is, the saving of a member i of generation t, s^i_{t+1}, is[17]

$$s^i_{t+1} = b^i_{t+1} - e^i_{t+1}. \tag{1.5}$$

The inheritance of a member i of generation t, x^i_{t+1}, is therefore the return on the parental saving, s^i_t.

$$x_{t+1}^i = s_t^i R_{t+1} = (b_t^i - e_t^i) R_{t+1} \tag{1.6}$$

where $R_{t+1} \equiv 1 + r_{t+1} - \delta \equiv R(k_{t+1})$. For simplicity the rate of capital depreciation $\delta = 1$.

Preferences of a member i of generation t are defined over family consumption in period $t+1$, c_{t+1}^i, and the value in period $t+1$ of total transfer to the offspring, b_{t+1}^i (that is, the sum of the immediate finance of the offspring's investment in human capital, e_{t+1}^i, and the saving for the offspring's future wealth, s_{t+1}^i). They are represented by a log-linear utility function that as will become apparent captures the spirit of Kaldorian–Keynesian saving behaviour (that is, the saving rate is an increasing function of wealth),

$$u_t^i = (1-\beta)\log c_{t+1}^i + \beta \log(\bar{\theta} + b_{t+1}^i), \tag{1.7}$$

where $\beta \in (0,\ 1)$ and $\bar{\theta} > 0$.[18]

1.3.2.2. The formation of human capital

In the first period of their lives individuals devote their entire time to the acquisition of human capital. The acquired level of human capital increases if their time investment is supplemented with capital investment in education. However, even in the absence of real expenditure individuals acquire one efficiency unit of labour – basic skills. The number of efficiency units of labour of a member i of generation t in period $t+1$, h_{t+1}^i, is a strictly increasing, strictly concave function of the individual's real expenditure on education in period t, e_t^i.[19]

$$h_{t+1}^i = h(e_t^i), \tag{1.8}$$

where $h(0) = 1$, $\lim_{e_t^i \to 0^+} h'(e_t^i) = \gamma < \infty$, and $\lim_{e_t^i \to \infty} h'(e_t^i) = 0$. As is the case for the production of physical capital (which converts one unit of output into one unit of capital), the slope of the production function of human capital is finite at the origin. This assumption, along with the ability of individuals to supply some minimal level of labour, $h(0)$, regardless of the physical investment in human capital (beyond time), assures that under some market conditions (non-basic) investment in human capital is not optimal.[20] The asymmetry between the accumulation of physical and human capital that is postulated in the chapter is manifested in the larger degree of diminishing marginal productivity in the production of human capital (that is, the strict concavity of $h(e_t^i)$ in contrast to the linearity of the production function of physical capital).

Given that the indirect utility function is a strictly increasing function of the individual's second period wealth, the unconstrained optimal real expenditure on education in every period t, e_t^i, from the viewpoint of individual i of generation t, maximizes the second period wealth, I_{t+1}^i,

$$e_t^i = \arg\max[w_{t+1}h(e_t^i) + (b_t^i - e_t^i)R_{t+1}]. \tag{1.9}$$

Although formally parents are indifferent about the internal allocation of the aggregate intended transfers to their offspring, the allocation of funds to their offspring's education is assumed to be optimal from their offspring's viewpoint.

Hence, as follows from the properties of $h(e_t^i)$, the optimal unconstrained real expenditure on education in every period t, e_t, is unique and identical across members of generation t.

If $R_{t+1} > w_{t+1}\gamma$ then $e_t = 0$, otherwise e_t is given by

$$w_{t+1}h'(e_t) = R_{t+1}. \tag{1.10}$$

Moreover, since $w_{t+1} = w(k_{t+1})$ and $R_{t+1} = R(k_{t+1})$, it follows that $e_t = e(k_{t+1})$.

Given the properties of $f(k_t)$, there exists a unique capital–labour ratio $\widehat{k}$, below which individuals do not invest in human capital (that is, do not acquire non-basic skills). That is, $R(\widehat{k}) = w(\widehat{k})\gamma$, where $\lim_{e_t^i \to 0^+} h'(e_t^i) = \gamma$. As follows from (1.2), $\widehat{k} = \alpha/(1-\alpha)\gamma \equiv \widehat{k}(\gamma) > 0$ where $\widehat{k}'(\gamma) < 0$. Since $R'(k_{t+1}) < 0$, $w'(k_{t+1}) > 0$, and $h''(e_t) < 0$, it follows that the optimal unconstrained real expenditure on education in every period t is a function of the capital–labour ratio in the subsequent period. In particular,

$$e_t = e(k_{t+1}) \begin{cases} = 0 & if \quad k_{t+1} \leq \tilde{k} \\ > 0 & if \quad k_{t+1} > \tilde{k}, \end{cases} \tag{1.11}$$

where $e'(k_{t+1}) > 0$ for $k_{t+1} > \tilde{k}$. Hence, if the capital–labour ratio in the next period is expected to be below $\tilde{k}$, individuals do not acquire non-basic skills.

Suppose that individuals can not borrow. It follows that the expenditure on education of a member i of generation t, e_t^i, is limited by the aggregate transfer, b_t^i, that the individual receives. As follows from (1.10) and the strict concavity of $h(e_t)$, $e_t^i = b_t^i$ if $b_t^i \leq e_t$, whereas $e_t^i = e_t$ if $b_t^i > e_t$. That is, the expenditure on education of a member i of generation t, e_t^i, is

$$e_t^i = \min[e(k_{t+1}),\ b_t^i]. \tag{1.12}$$

where e_t^i is a non-decreasing function of k_{t+1} and b_t^i.

1.3.2.3. Optimal consumption and transfers

A member i of generation t chooses the level of second period consumption, c_{t+1}^i, and a non-negative aggregate level of transfers to the offspring, b_{t+1}^i, so as to maximize the utility function subject to the second period budget constraint (1.4).

Hence the optimal transfer of a member i of generation t is:

$$b_{t+1}^i = b(I_{t+1}^i) \equiv \begin{cases} \beta(I_{t+1}^i - \theta) & \text{if} \quad I_{t+1}^i \geq \theta; \\ \\ 0 & \text{if} \quad I_{t+1}^i \leq \theta, \end{cases} \tag{1.13}$$

where $\theta \equiv \bar{\theta}(1-\beta)/\beta$. As follows from (1.13), the transfer rate b_{t+1}^i / I_{t+1}^i is increasing in I_{t+1}^i. Moreover, as follows from (1.5) and (1.11) the saving of a member i of generation $t-1$, s_t^i, is

$$s_t^i = \begin{cases} b_t^i & \text{if} \quad k_{t+1} \leq \tilde{k}; \\ \\ b_t^i - e_t^i & \text{if} \quad k_{t+1} > \tilde{k}. \end{cases} \tag{1.14}$$

Hence, since b_{t+1}^i / I_{t+1}^i is increasing in I_{t+1}^i, it follows from (1.12) that s_{t+1}^i / I_{t+1}^i is increasing in I_{t+1}^i as well. The transfer function and the implied saving function capture the properties of the Kaldorian–Keynesian saving hypothesis.

1.3.3. Aggregate Physical and Human Capital

Suppose that in period 0 the economy consists of two groups of adult individuals – Rich and Poor. They are identical in their preferences and differ only in their initial capital ownership. The Rich, denoted by R, are a fraction λ of all adult individuals in society, who equally own the entire initial physical capital stock. The Poor, denoted by P, are a fraction $1-\lambda$ of all adult individuals in society, who have no ownership over the initial physical capital stock. Since individuals are ex-ante homogenous within a group, the uniqueness of the solution to their optimization problem assures that their offspring are homogenous as well. Hence, in every period a fraction λ of all adults are homogenous descendents of the Rich, denoted by members of group R, and a fraction $1-\lambda$ are homogenous descendents of the Poor, denoted by members of group P.

The optimization of groups P and R of generations $t-1$ and t in period t, determines the levels of physical capital, K_{t+1}, and human capital, H_{t+1}, in period $t+1$,

$$K_{t+1} = \lambda s_t^R + (1-\lambda)s_t^P = \lambda(b_t^R - e_t^R) + (1-\lambda)(b_t^P - e_t^P), \tag{1.15}$$

where $K_0 > 0$.

$$H_{t+1} = \lambda h(e_t^R) + (1-\lambda)h(e_t^P), \tag{1.16}$$

where in period 0 there is no (non-basic) human capital, that is, $h_0^i = 1$ for all $i = R,\ P$ and thus $H_0 = 1$.

Hence, (1.12) implies that the levels of physical capital, K_{t+1}, and human capital, H_{t+1}, in period $t+1$, are functions of intergenerational transfers in each of the groups, b_t^R and b_t^P, and the capital–labour ratio in the subsequent period, k_{t+1}.

$$\begin{aligned} H_{t+1} &= H(b_t^R,\ b_t^P,\ k_{t+1}); \\ K_{t+1} &= K(b_t^R,\ b_t^P,\ k_{t+1}). \end{aligned} \tag{1.17}$$

where (1.11), (1.12) and $e'(k_{t+1}) \geq 0$, imply that $\partial H_{t+1}/\partial k_{t+1} \geq 0$, $\partial K_{t+1}/\partial k_{t+1} \leq 0$, $H(b_t^R,\ b_t^P,\ 0) = 1$, and $K(b_t^R,\ b_t^P,\ 0) > 0$ for $b_t^R > 0$.

The capital–labour ratio in period $t+1$ is therefore,

$$k_{t+1} = \frac{K(b_t^R,\ b_t^P,\ k_{t+1})}{H(b_t^R,\ b_t^P,\ k_{t+1})}, \tag{1.18}$$

where the initial level of the capital–labour ratio, k_0, is assumed to be

$$k_0 \in (0,\ \tilde{k}). \tag{1.A1}$$

This assumption assures that in the initial stages the rate of return to physical capital is higher than the rate of return to human capital.

As follows from (1.11), this assumption is consistent with the assumption that the initial level of human capital is $H_0 = 1$.

Hence, it follows from (1.18) and the properties of the functions in (1.17) that there exists a continuous single valued function $\kappa(b_t^R,\ b_t^P)$ such that the capital–labour ratio in period $t+1$ is fully determined by the level of transfer of groups R and P in period t.

$$k_{t+1} = \kappa(b_t^R,\ b_t^P),$$

where $\kappa(0,\ 0) = 0$ (since in the absence of transfers and hence savings the capital stock in the subsequent period is zero).

1.3.4. The Evolution of Transfers within Dynasties

The evolution of transfers within each group $i = R,\ P,$ as follows from (1.13), is

$$b_{t+1}^i = \max\{\beta[w_{t+1}h(e_t^i) + (b_t^i - e_t^i)R_{t+1} - \theta],\ 0\}; \qquad i = R,\ P. \tag{1.20}$$

Hence, it follows from (1.12) that

$$b_{t+1}^i = \max\left\{\begin{array}{ll} \beta[w(k_{t+1})h(b_t^i) - \theta] & \text{if} \quad b_t^i \le e(k_{t+1}) \\ & \\ \beta[w(k_{t+1})h(e(k_{t+1})) + (b_t^i - e(k_{t+1}))R(k_{t+1}) - \theta] & \text{if} \quad b_t^i > e(k_{t+1}) \end{array},\ 0\right\}. \tag{1.21}$$

Namely, intergenerational transfers within group i in period $t+1$, b_{t+1}^i are determined by the intergenerational transfers within the group in the proceeding period, as well as the rewards to factors of production, as determined by the capital–labour ratio in the economy, that is,

$$b_{t+1}^i \equiv \phi(b_t^i,\ k_{t+1}). \tag{1.22}$$

Let $\hat{k}$ be the critical level of the capital–labour ratio below which individuals who do not receive transfers from their parents (that is, $b_t^i = 0$ and therefore $h(b_t^i) = 1$) do not transfer income to their offspring. That is, $w(\hat{k}) = \theta.$ As follows from (1.2), $\hat{k} = [\theta/(1-\alpha)A]^{1/\alpha} \equiv \hat{k}(\theta)$ where if $k_{t+1} \le \hat{k}$ then $w(k_{t+1}) \le \theta,$ whereas if $k_{t+1} > \hat{k}$ then $w(k_{t+1}) > \theta.$ Hence, intergenerational transfers within group i in period $t+1$, b_{t+1}^i is positive if and only if $k_{t+1} > \hat{k},$ that is,

$$b_{t+1}^i = \phi(0,\ k_{t+1}) \begin{cases} = 0 & \text{if} \quad k_{t+1} \le \hat{k}; \\ & \\ > 0 & \text{if} \quad k_{t+1} > \hat{k}. \end{cases} \tag{1.23}$$

In order to reduce the number of feasible scenarios for the evolution of the economy, suppose that once wages increase sufficiently such that members of group P transfer resources to their offspring, that is, $k_{t+1} > \hat{k}$, investment in human capital is profitable, that is, $k_{t+1} > \tilde{k}$, That is,

$$\tilde{k} \leq \hat{k}. \tag{1.A2}$$

Note that, since $\hat{k} = \hat{k}(\theta)$ and $\hat{k}'(\theta) > 0$, it follows that for any given γ, there exists θ sufficiently large such that $\tilde{k}(\gamma) \leq \hat{k}(\theta)$.

Let $\tilde{t}+1$ be the first period in which the capital–labour ratio exceeds $\tilde{k}$ (that is, $k_{\tilde{t}+1} > \tilde{k}$). That is, since $k_0 < \tilde{k}$, it follows that $k_{t+1} \leq \tilde{k}$ for all $0 \leq t < t$. Let $\tilde{t}+1$ be the first period in which the capital–labour ratio exceeds $\hat{k}$. That is, $k_{t+1} \leq \hat{k}$ for all $0 \leq t < t$. It follows from Assumption 1.A2 that $\tilde{t} \leq \hat{t}$.

The evolution of transfers within each of the two groups, as follows from the fact that $k_{t+1} = \kappa(b_t^R,\ b_t^P)$, is fully determined by the evolution of transfers within both types of dynasties. Namely,

$$b_{t+1}^i = \phi(b_t^i,\ k_{t+1}) = \phi[b_t^i,\ \kappa(b_t^R,\ b_t^P)] \equiv \psi^i(b_t^R,\ b_t^P); \quad i = R,\ P, \tag{1.24}$$

where the initial transfers of the Rich and the Poor are

$$\begin{aligned} b_0^R &= \max\left\{\beta\left[w(k_0) + \frac{k_0 R(k_0)}{\lambda}\right] - \theta,\ 0\right\} \\ b_0^P &= \max\left\{\beta\left[w(k_0) - \theta\right],\ 0\right\}, \end{aligned} \tag{1.25}$$

noting that the level of human capital of every adult i in period 0 is $h_0^i = 1$, and the entire stock of capital in period 0 is distributed equally among the Rich.

Lemma 1 *The intergenerational transfers of members of group R (the Rich) is higher than that of members of group P (the Poor) in every time period, that is,*

$$b_t^R \geq b_t^P \text{ for all } t.$$

The proof follows from (1.22) noting that $b_0^R \geq b_0^P$.

1.4. THE PROCESS OF DEVELOPMENT

This section analyses the endogenous evolution of the economy from early to mature stages of development. The dynamical system is uniquely determined by the joint evolution of the intergenerational transfers of members of groups P and R. As follows from (1.24), the evolution of the economy is given by the sequence $\{b_t^P, b_t^R\}_{t=0}^{\infty}$ that satisfies in every period

$$b_{t+1}^P = \psi^P(b_t^R, b_t^P);$$
$$b_{t+1}^R = \psi^R(b_t^R, b_t^P), \qquad (1.26)$$

where b_0^P and b_0^R are given by (1.25).

As will become apparent, if additional plausible restrictions are imposed on the basic model, the economy endogenously evolves through two fundamental regimes:

- Regime I: In this early stage of development the rate of return to human capital is lower than the rate of return to physical capital and the process of development is fuelled by capital accumulation.
- Regime II: In these mature stages of development, the rate of return to human capital increases sufficiently so as to induce human capital accumulation, and the process of development is fuelled by human capital as well as physical capital accumulation.

In Regime I, physical capital is scarce and the rate of return to human capital is therefore lower than the rate of return to physical capital. Since there is no incentive for investment in human capital the process of development is fuelled by capital accumulation. The wage rate is lower than the critical level that would enable individuals who do not own any capital to engage in intergenerational transfers (and thus savings). The Poor, therefore, consume their entire wages, they are not engaged in saving, capital accumulation and intergenerational transfers. Their decedents, therefore, are also unable to engage in savings and intergenerational transfers and the Poor are in a temporary steady state equilibrium in which there is neither investment in physical capital nor in human capital. In contrast, the income of the Rich, who own the entire stock of capital in the economy, is sufficiently high, permitting intergenerational transfers and capital accumulation. Intergenerational transfers among the Rich increase over time and the stock of physical capital in the economy, therefore, increases as well. During this regime, physical capital accumulation by the rich raises the wages and therefore the return to human capital and decreases the return to physical capital. However, as long as the rate of return to human capital remains lower than the rate of return to physical capital, the qualitative structure of the economy remains unchanged. That is, the Poor are in a poverty trap, the Rich get richer and the process of development is based solely on physical capital accumulation. Inequality in Regime I increases the wealth of individuals whose marginal propensity to save is higher and consequently increases aggregate savings and capital accumulation and enhances the process of development.

The accumulation of physical capital by the Rich in Regime I raises gradually the rate of return to human capital. Ultimately, the rate of return to human capital is sufficiently high so as to induce human capital accumulation, and the economy enters into Regime II where the process of development is fuelled by human capital accumulation as well as physical capital accumulation.

Regime II is subdivided into three stages. In Stage I, investment in human capital is selective and it is feasible only for the Rich. In Stage II, investment in human capital is universal but it is still sub-optimal due to binding credit constraints and, in Stage III, investment in human capital is optimal since credit constraints are no longer binding.

Stage I (Selective Human Capital Accumulation). In this stage, the capital–labour ratio in the economy is higher than that in Regime I, and although it generates wage rates that justify investment in human capital, these wages are still lower than the critical level that would permit intergenerational transfers for individuals who do not own any capital. Hence, although the rate of return justifies investment in human capital, in the absence of parental support, credit constraints deprives the Poor from this investment. The Poor consume their entire income and they are not engaged in saving and capital accumulation. Their decedents are therefore unable to engage in savings and intergenerational transfers and the Poor remain in a temporary steady state equilibrium in which there is neither investment in physical capital nor in human capital. In contrast, the income of the Rich is sufficiently high, permitting intergenerational transfers and physical capital accumulation as well as human capital accumulation. Intergenerational transfers and the accumulation of physical capital by the Rich gradually rise in Stage I of Regime II, and ultimately the wage rate is sufficiently high so as to permit some investment in human capital by the Poor (that is, the economy enters Stage II of Regime II).

Stage II (Universal Human Capital Accumulation). In this stage, the capital–labour ratio in the economy generates wage rates that permit some investment in human capital by all individuals. In contrast to the Rich, the investment of the poor is constrained by parental wealth and it is therefore sub-optimal. That is, the marginal return on investment in human capital among the Poor is higher than that among the Rich. Equality alleviates the adverse effect of credit constraints on the investment of the Poor in human capital, and has therefore a positive effect on the level of human capital and economic growth. The gradual increase in the wage income of the decedents of the Poor that takes place in Stage II of Regime II, due to a gradual increase in their investment in human capital, makes the credit constraint less binding over time and the aggregate effect of income distribution on the growth process subsides.

Stage III (Unconstrained Investment in Human Capital). In Stage III, credit constraints are non-binding due to the increase in wage income in Stage II, the rate of return to human capital is equalized across groups, and inequality therefore has no effect on economic growth.

1.4.1. Regime I: Physical Capital Accumulation

In this early stage of development the rate of return to human capital is lower than the rate of return to physical capital and the process of development is fuelled by capital accumulation.

Regime I is defined as the time interval $0 \le t < \tilde{t}$. In this early stage of development the capital–labour ratio in period $t+1$, k_{t+1}, which determines the return to investment in human capital in period t, is lower than $\tilde{k}$. The rate of return to human capital is therefore lower than the rate of return to physical capital, and the process of development is fuelled by capital accumulation. As follows from (1.11) the level of real expenditure on education in Regime I is therefore zero and members of both groups acquire only basic skills. That is, $h[e(k_{t+1})] = 1$. Furthermore, since the income of members of group P (the Poor) is lower than the threshold that permits intergenerational transfer there are no intergenerational transfers among dynasties of this group, that is,

$$b_t^P = 0 \quad \text{for} \quad 0 \le t \le \hat{t} \tag{1.27}$$

As follows from (1.15)–(1.29), and (1.26), since $e_t^R = e_t^P = b_t^P = 0$ in the time interval $0 \le t < \tilde{t}$ (where $\tilde{t} \le \hat{t}$ as follows from 1.A2) the capital–labour ratio, k_{t+1}, is determined in Regime I by the intergenerational transfers of members of group R, according to their fraction in the population λ; $k_{t+1} = \kappa(b_t^R, 0) = \lambda b_t^R$ for $0 \le t < \tilde{t}$ (that is, for $k_{t+1} \in (0, \tilde{k})$). Since $b_t^R \in [0, \tilde{b}]$ for $0 \le t < t$,

$$k_{t+1} = \kappa(b_t^R, 0) = \lambda b_t^R \qquad \text{for} \quad b_t^R \in [0, \tilde{b}], \tag{1.28}$$

where $\tilde{b} \equiv \tilde{k} / \lambda = \alpha / [(1-\alpha)\gamma\lambda]$.

THE DYNAMICS OF TRANSFERS

A. Unconditional dynamics

The evolution of the economy in Regime I, as follows from (1.26) and (1.27), is given by

$$\begin{aligned} b_{t+1}^R &= \psi^R(b_t^R,\ 0) = \max\left\{\beta[w(\lambda b_t^R) + b_t^R R(\lambda b_t^R) - \theta],\ 0\right\} \\ b_{t+1}^P &= \psi^P(b_t^R,\ 0) = \max\left\{\beta[w(\lambda b_t^R) - \theta],\ 0] = 0\right\}, \end{aligned} \tag{1.29}$$

for $b_t^R \in [0,\ \tilde{b}]$ where $b_0^P = 0$ and b_0^R is given by (1.25).

In order to assure that the economy would ultimately take off from Regime I to Regime II, it is assumed that the technology is sufficiently productive. That is,

$$A \geq \underline{A} \equiv A(\alpha,\ \gamma,\ \lambda,\ \beta,\ \theta). \tag{1.A3}$$

As depicted in Figure 1.1, the function $\psi^R(b_t^R,\ 0)$ is equal to zero for $b_t^R \leq \underline{b}$, it is increasing and concave for $\underline{b} < b_t^R \leq \tilde{b}$ and it crosses the 45° line once in the interval $\underline{b} < b_t^R \leq \tilde{b}$.

Hence, the dynamical system $\psi^R(b_t^R,\ 0)$, depicted in Figure 1.1, has two steady-state equilibria in the interval $b_t^R \in [0,\ \tilde{b}]$; a locally stable steady-state, $\bar{b} = 0$, and an unstable steady-state, $\bar{b}^u \in (\underline{b}, \tilde{b})$. If $b_t^R < \bar{b}^u$ then the transfers

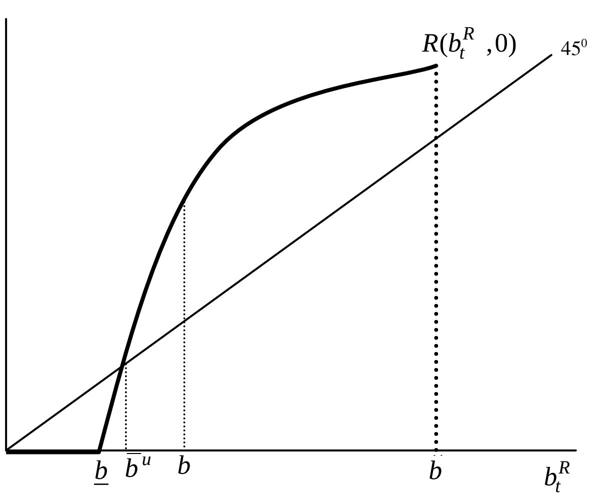


Figure 1.1 The dynamical system in Regime I and Stage I of Regime II

within each dynasty of type R contract over time and the system converges to the steady-state equilibrium $\bar{b}=0$. If $b_t^R > \bar{b}^u$ then the transfers within each dynasty of type R expand over the entire interval $(\bar{b}^u, \tilde{b}]$, crossing into Regime II. To assure that the process of development starts in Regime I and ultimately reaches Regime II, it is assumed that[21]

$$b_0^R \in (\bar{b}^u, \tilde{b}). \tag{1.A4}$$

B. Conditional dynamics

In order to visualize the evolution of the threshold for the departure of members of group P from the zero transfer state, the dynamics of transfers within dynasties is depicted in Figure 1.2(a), for a given k. This conditional dynamical system is given by (1.20). For a given $k \in (0, \tilde{k}]$, the dynamic of transfers within dynasty i, is

$$b_{t+1}^i = \phi(b_t^i;\ k) = \max\{\beta[w(k) + b_t^i R(k) - \theta],\ 0\}. \tag{1.30}$$

Hence, there exist a critical level $b(k)$ below which $\phi(b_t^i;\ k) = 0$ and above which $\phi(b_t^i;\ k)$ is linear in b_t^i, with a slope $\beta R(k) > 1$, that is,

$$\begin{aligned} &\phi(b_t^i;\ k) = 0 && \text{for} \quad 0 \le b_t^i \le b(k); \\ &\partial\phi(b_t^i;\ k)/\partial b_t^i = \beta R(k) > 1 && \text{for} \quad b_t^i > b(k). \end{aligned} \tag{1.31}$$

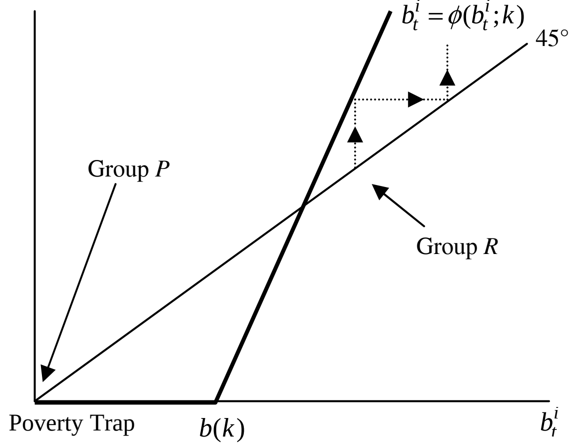


Figure 1.2(a) The conditional dynamical system in Regime I

Note that under Assumption 1.A3 $\beta R(k)>1$. Otherwise $\psi^R(b^R, 0)<b^R$ for $b^R \in (0,\tilde{b}]$, in contradiction to Lemma 1.

As depicted in Figure 1.2(a), in Regime I, members of group P are trapped in a zero transfer temporary steady-state equilibrium, whereas the level of transfers of members of group R increases from generation to generation. As the transfers of members of group R increase the capital–labour ratio increases and the threshold level of transfer, $b(k)$, that enables dynasties of type P to escape the attraction of the no-transfer temporary steady-state equilibrium, eventually declines.

INEQUALITY AND THE DYNAMICS OF OUTPUT PER WORKER

The evolution of output per worker, Y_t, in Regime I, follows from (1.1), (1.2), (1.28) and (1.29). Provided that Assumption 1.A4 is satisfied, output per worker, Y_{t+1}, is

$$Y_{t+1}=A\left[\beta\{\lambda[(1-\alpha)Y_t-\theta]+\alpha Y_t\}\right]^{\alpha}\equiv Y(Y_t), \tag{1.32}$$

where $Y'(Y_t)>0$.

In order to examine the effect of inequality on economic growth, consider two economies (or two alternative initial states of the same economy): a relatively egalitarian economy, E, and a relatively inegalitarian one, U. Suppose that the economies are identical in all respects except for their degree of inequality. Suppose that income in period t is distributed differently between group R and group P in the two economies. That is, the income of members of group i, $(I_t^i)^E$, in the egalitarian economy, E, is

$$\begin{aligned}(I_t^R)^E &= (I_t^R)^U-\varepsilon_t \equiv I^R(I_t^R,\ \varepsilon_t)\\ (I_t^P)^E &= (I_t^P)^U+\frac{\lambda}{1-\lambda}\varepsilon_t \equiv I^P(I_t^P,\ \varepsilon_t),\end{aligned} \tag{1.33}$$

where $\varepsilon_t>0$, is sufficiently small such that: (i) the economy does not depart from its current stage of development, and (ii) the net income of members of group P remains below that of member of group R.

The transfer of member i of generation t to the offspring in economy, E, is therefore

$$(b_t^i)^E=\max\{\beta[I^i(I_t^i,\ \varepsilon_t)-\theta],\ 0\}\equiv b^i(I_t^i,\ \varepsilon_t) \qquad i=P,\ R \tag{1.34}$$

Proposition 1 *(The effect of inequality on economic growth in Regime I.) Consider two economies (or two alternative initial states of the same economy). Suppose that the economies are identical in all respects except for* *their degree of inequality. Under Assumptions 1.A2–1.A4, the less egalitarian economy would be characterized by a superior path of output per worker.*

Proof. See Galor and Moav (2004).

Inequality enhances the process development in Regime I since a higher concentration of wealth among members of group P (the Poor), would increase aggregate consumption, decrease aggregate intergenerational transfers, and thus would slow capital accumulation and the process of development.

Remark 1 *If income is distributed less equally within groups (that is, if additional income groups are created), then it would not affect output per worker as long as the marginal propensity to save remains equal among all sub-groups of each of the original groups (that is, β for group R and 0 for group P). Otherwise, since saving is a convex function of wealth, more inequality would promote economic growth.*

1.4.2. Regime II: Human Capital Accumulation

In these mature stages of development, the rate of return to human capital increases sufficiently so as to induce human capital accumulation, and the process of development is fuelled by human capital as well as physical capital accumulation. In Stages I and II members of group P are credit constrained and their marginal rate of return to investment in human capital is higher than that on physical capital, whereas those marginal rates of returns are equal for members of group R who are not credit constrained. In Stage III all individuals are not credit constrained and the marginal rate of return to investment in human capital is equal to the marginal rate of return on investment in physical capital.

1.4.2.1. Stage I: selective human capital accumulation

Stage I of Regime II is defined as the time interval $\tilde{t} \le t \le \hat{t}$. In this time interval $k_{t+1} \in (\tilde{k}, \hat{k})$ and the marginal rate of return on investment in human capital is higher than the rate of return on investment in physical capital for individuals who are credit constrained (members of group P), whereas those rates of returns are equal for members of group R.[22]

As follows from (1.11) and Lemma 1, $e_t^R > 0$ and $e_t^P = 0$. Hence, given (1.18), it follows that the capital–labour ratio, k_{t+1}, in the interval $k_{t+1} \in (\tilde{k}, \hat{k})$ is determined by the savings of members of group R, as well as their investment in human capital. Namely,

$$k_{t+1} = \frac{\lambda(b_t^R - e(k_{t+1}))}{1-\lambda+\lambda h(e(k_{t+1}))}. \tag{1.35}$$

Since $e'(k_{t+1})>0$, it follows that $k_{t+1}=\kappa(b_t^R,\ 0)$ where $\partial\kappa(b_t^R,\ 0)/\partial b_t^R>0$. Hence, there exist a unique value $\hat{b}$ of the level of b_t^R such that $k_{t+1}=\hat{k}$. That is, $\kappa(\hat{b},\ 0)=\hat{k}$.

THE DYNAMICS OF TRANSFERS

A. Unconditional dynamics

The evolution of the economy in Stage I of Regime II, as follows from (1.24) and (1.26) is given by[23]

$$\begin{aligned} b_{t+1}^R &= \psi^R(b_t^R;\ 0) = \beta[w(k_{t+1})h(e(k_{t+1})) + (b_t^R - e(k_{t+1}))R(k_{t+1}) - \theta]; \\ b_{t+1}^P &= \psi^P(b_t^R;\ 0) = 0, \end{aligned} \tag{1.36}$$

for $b_t^R \in [\frown{b},\ \frown{b}]$.

In order to assure that the process of development does not come to a halt in this pre-mature stage of development (that is, in order to assure that there is no steady-state equilibrium in Stage I of Regime II) it is sufficient that $\beta[w(\lambda\hat{b})+\hat{b}R(\lambda\hat{b})-\theta]>\hat{b}$ – a condition that is satisfied under Assumption 1.A3.[24] This condition assures that if the equation of motion in Regime I would remain in place in Stage I of Regime II, then there is no steady-state in Stage I. This condition is sufficient to assure that given the actual equation of motion in Stage I of Regime II, the system has no steady-state in this stage.

Figure 1.1 depicts the properties of $\psi^R(b_t^R,\ 0)$ over the interval $b_t^R \in [\tilde{b},\hat{b}]$. The transfers within each dynasty of type R expand over the entire interval crossing into Stage II.

B. Conditional dynamics

In order to visualize the evolution of the threshold for the departure of dynasties of type P from the zero transfer state, the dynamics of transfers within dynasties is depicted in Figure 1.2(b) for a given k. This conditional dynamical system is given by (1.22). For a given $k \in (\tilde{k},\hat{k}]$

$$\begin{aligned} b_{t+1}^i &= \max\left\{ \begin{matrix} \beta[w(k)h(b_t^i)-\theta] & \text{if} \quad b_t^i \le e(k) \\ \\ \beta[w(k)h(e(k))+(b_t^i-e(k))R(k)-\theta] & \text{if} \quad b_t^i > e(k) \end{matrix} \quad ,0 \right\} \\ &= \phi(b_t^i,k). \end{aligned} \tag{1.37}$$

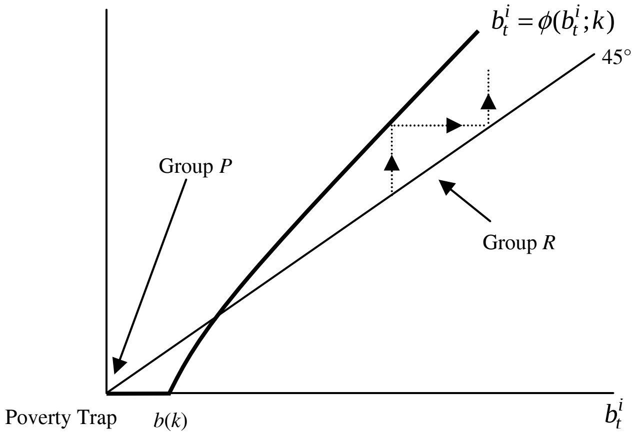


Figure 1.2(b) The conditional dynamical system in Stage I of Regime II

Hence, for a given $k \in (\tilde{k}, \hat{k})$ there exist a critical level $b(k)$ below which $\phi(b_t^i;\, k) = 0$ and above which $\phi(b_t^i;\, k)$ is increasing and concave in b_t^i. In particular,[25]

$$\begin{aligned}
&\frac{\partial \phi(b_t^i;\ k)}{\partial b_t^i} > \beta R(k) > 0 \quad && \textit{for} \quad b(k) < b_t^i < e(k); \\
&\frac{\partial^2 \phi(b_t^i;\ k)}{\partial b_t^{i\,2}} < 0 && \textit{for} \quad b(k) < b_t^i < e(k); \qquad (1.38)\\
&\frac{\partial \phi(b_t^i;\ k)}{\partial b_t^i} = \beta R(k) > 1 && \textit{for} \quad b_t^i \geq e(k).
\end{aligned}$$

Note that $\phi(b_t^i,\, k) > b_t^i$ for all $b^i > \tilde{b}$.

As depicted in Figure 1.2(b), in Stage I of Regime II, members of group P are still trapped in a zero transfer temporary steady-state equilibrium, whereas the level of transfers of members of group R increases from generation to generation. As the transfer of members of group R increases the capital–labour ratio increases and the threshold level of transfer, $b(k)$, that enables members of group P to escape the attraction of the no-transfer temporary steady-state equilibrium, eventually declines and ultimately vanishes as the economy enters Stage III.

Stage I of Regime II is an intermediate stage in which inequality has an ambiguous effect on the rate of economic growth. A lower level of wealth among members of group R, along with a higher level of wealth, but below the threshold θ, among some members of group P, would increase aggregate consumption, decrease aggregate intergenerational transfers, and thus would slow physical and human capital accumulation and the process of development. However a lower level of wealth among members of group R, along with a higher level of wealth, above the threshold θ, among some members of group P, would generate investment in human capital among these individuals, bringing about an increase in the aggregate stock of human capital that can offset the negative effect of equality on the accumulation of physical capital.

1.4.2.2. Stage II: universal human capital investment

Stage II of Regime II is defined as the time interval $\hat{t} < t < t^*$, where t^* is the time period in which the credit constraints are no longer binding for members of group P, that is, $b_{t^*}^P \geq e_{t^*}$. In this time interval, the marginal rate of return on investment in human capital is higher than the marginal rate of return on investment in physical capital for members of group P, whereas these rates of return are equal for members of group R. As established previously once $t > \hat{t}$ the economy exits Stage I of Regime II and enters Stage II of Regime II. In the initial period $k_{\hat{t}+1} > \hat{k}$ and therefore $b_{\hat{t}+1}^P > 0$ and consequently the sequence $\{b_t^R,\ b_t^P\}$ increases monotonically over the time interval $\hat{t} < t < t^*$.

As follows from (1.11), (1.12) and (1.18), in Stage II $e_t^P = b_t^P < e_t$ and $e_t^R = e_t$ and therefore the capital–labour ratio is determined by intergenerational transfers and investment in human capital of both types of individuals.

$$k_{t+1} = \frac{\lambda(b_t^R - e(k_{t+1}))}{(1-\lambda)h(b_t^P) + \lambda h(e(k_{t+1}))}. \tag{1.39}$$

Since $e'(k_{t+1}) > 0$, it follows that $k_{t+1} = \kappa(b_t^R,\ b_t^P)$ where $\partial\kappa(b_t^R,\ b_t^P)/\partial b_t^R > 0$ and $\partial\kappa(b_t^R,\ b_t^P)/\partial b_t^P < 0$.

THE DYNAMICS OF TRANSFERS

A. Unconditional dynamics

The evolution of the economy, in Stage II of Regime II (that is, as long as credit constraints are still binding, $b_t^P < e_t$), as follows from (1.20) and (1.26), is given by

$$\begin{aligned} b_{t+1}^R &= \psi^R(b_t^R,\ b_t^P) = \beta[w(k_{t+1})h(e(k_{t+1})) + (b_t^R - e(k_{t+1}))R(k_{t+1}) - \theta]; \\ b_{t+1}^P &= \psi^P(b_t^R,\ b_t^P) = \max\{\beta[w(k_{t+1})h(b_t^P) - \theta],\ 0\}, \end{aligned} \tag{1.40}$$

where $k_{t+1} = \kappa(b_t^R,\ b_t^P)$.

The unconditional dynamical system in Stage II of Regime II is rather complex and the a sequence of technical results that are presented in the Appendix of Galor and Moav (2004) characterizes the properties of the system In particular, it is shown that intergenerational transfers within the two groups, $(b_t^R,\ b_t^P)$, increase monotonically over time in Stage II of Regime II and the economy necessarily enters into Stage III of Regime II.

B. Conditional dynamics

The evolution of transfers within dynasties is depicted in Figure 1.2(c) for a given $k > \hat{k}$.[26] This conditional dynamical system is given by (1.22). For a given $k > \hat{k}$,

$$b_{t+1}^i = \max\begin{Bmatrix} \beta[w(k)h(b_t^i) - \theta] & \text{if} \quad b_t^i \le e(k) \\ \beta[w(k)h(e(k)) + (b_t^i - e(k))R(k) - \theta] & \text{if} \quad b_t^i > e(k) \end{Bmatrix} \tag{1.41}$$

$$= \phi(b_t^i, k).$$

Hence, for a given $k > \hat{k}$, over the interval $0 < b_t^i < e(k)$, $\phi(b_t^i;\ k)$ is a positive, increasing, and concave function of b_t^i, where

$$\begin{aligned} &\frac{\partial \phi(b_t^i;\ k)}{\partial b_t^i} > \beta R(k) > 0 \qquad \textit{for} \quad 0 < b_t^i < e(k); \\ &\frac{\partial \phi(b_t^i;\ k)}{\partial b_t^i} = \beta R(k) \qquad \textit{for} \quad b_t^i \ge e(k). \end{aligned} \tag{1.42}$$

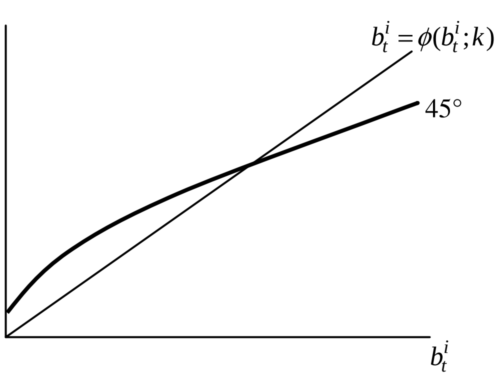


Figure 1.2(c) The conditional dynamical system in Stages II and III of Regime II

Note that for $k > \hat{k}$ it follows that $\phi(b_t^i, k) > b_t^i$ for at least a strictly positive range $b_t^i \in [0, b]$, where $b > \hat{b}$.

As depicted in Figure 1.2(c), in Stage II of Regime II, members of group P depart from the zero transfer temporary equilibrium. The level of transfers of members of group P increases from generation to generation. Eventually members of group P are not credit constrained, that is, $b_t^P \geq e_t$ and the economy endogenously enters into Stage III of Regime II.

INEQUALITY AND THE DYNAMICS OF OUTPUT PER WORKER

Since in Stages II and III of Regime II the income of each individual is greater than θ, it follows from (1.13) that the marginal propensity to transfer is equal to β among all individuals. The aggregate transfers of members of generation t, $\lambda b_t^R + (1-\lambda) b_t^P$, is therefore simply a fraction β of $Y_t - \theta > 0$. That is,

$$\lambda b_t^R + (1-\lambda) b_t^P = \beta(Y_t - \theta). \tag{1.43}$$

The evolution of output per worker, Y_t, in Stage II of Regime II, as follows from (1.1), (1.15), (1.16), noting that $e_t^R = e_t$ and $e_t^P = b_t^P$, is therefore

$$Y_{t+1} = AK_{t+1}^{\alpha} H_{t+1}^{1-\alpha} = A[\beta(Y_t - \theta) - \lambda e_t - (1-\lambda) b_t^P]^{\alpha} [\lambda h(e_t) + (1-\lambda) h(b_t^P)]^{1-\alpha}. \tag{1.44}$$

Since $e_t = \arg\max [w_{t+1} h(e_t) - R_{t+1} e_t] = \arg\max Y_{t+1}$ (and since therefore $\partial Y_{t+1} / \partial e_t = 0$), it follows that

$$Y_{t+1} \equiv Y(Y_t, b_t^P), \tag{1.45}$$

where $\partial Y(Y_t, b_t^P)/\partial Y_t > 0$ and $\partial Y(Y_t, b_t^P)/\partial b_t^P > 0$, noting that as follows from (1.2) and (1.10), $h(b_t^P) > h'(e_t) = \alpha / [(1-\alpha) k_{t+1}]$.

Lemma 2 *Under 1.A2–1.A4, Y_t increases monotonically over Stage II.*

The Lemma follows from (1.43) and the Appendix of Galor and Moav (2004).

Proposition 2 *(The effect of inequality on economic growth in Stage II of Regime II.) Consider two economies (or two alternative initial states of the same economy). Suppose that the economies are identical in all respects except for their degree of inequality. Under Assumptions A2–A4, the more egalitarian economy would be characterized by a superior path of output per worker.*

Proof. See Galor and Moav (2004).

Inequality negatively affects the process development in Stage II of Regime II. A lower concentration of wealth among members of group R and a higher concentration of wealth among member of group P would not affect aggregate consumption, and aggregate intergenerational transfers, but due to liquidity constraints of members of group P would allow for a more efficient allocation of aggregate investment between physical and human capital.

Remark 2 *If income is distributed less equally within groups, then it would not affect the aggregate level of intergenerational transfers as long as the marginal propensity to transfer, β, is equal among all member of the economy. However, an unequal distribution of income among members of group P would generate a less efficient allocation of human capital, due to the liquidity constraints and the concavity of $h(e_t^P)$, and thus would lower the path of output per worker. An unequal distribution among members of group R, as long as all the members of sub-groups of R remain unaffected by credit constraint, will not affect output. If, however, an unequal distribution is associated with some members of sub-groups of R being credit constrained, it would be associated with a lower path of output per worker.*

1.4.2.3. Stage III: Unconstrained investment in human capital

Stage III of Regime II is defined as $t \geq t^*$ where credit constraints are no longer binding (that is, $b_t^R \geq b_t^P \geq e_t$). In this time interval the marginal rate of return on investment in human capital is equal to the marginal rate of return on investment in physical capital for all individuals.

As follows from (1.12), in Stage III of Regime II $e_t^P = e_t^R = e_t$. Hence, given (1.18) and (1.43) it follows that k_{t+1} is given by

$$k_{t+1} = \frac{\beta[Y_t - \theta] - e(k_{t+1})}{h[e(k_{t+1})]}. \tag{1.46}$$

Since $e'(k_{t+1}) > 0$, it follows that $k_{t+1} = k(Y_t)$ where $k'(Y_t) > 0$ and $\lim_{Y_t \to \infty} k_{t+1} = \infty$.

THE DYNAMICS OF TRANSFERS AND OUTPUT PER WORKER

The evolution of the economy in Stage III of Regime II, as follows from (1.24) and (1.26), is given by

$$\begin{aligned} b_{t+1}^R &= \psi^R(b_t^R,\ b_t^P) = \beta[w(k_{t+1})h(e(k_{t+1})) + (b_t^R - e(k_{t+1}))R(k_{t+1}) - \theta]; \\ b_{t+1}^P &= \psi^P(b_t^R,\ b_t^P) = \beta[w(k_{t+1})h(e(k_{t+1})) + (b_t^P - e(k_{t+1}))R(k_{t+1}) - \theta]. \end{aligned} \tag{1.47}$$

The evolution of output per worker, Y_t, in Stage III of Regime II, is independent of the distribution of intergenerational transfers. As follows from (1.1) and (1.43)

$$Y_{t+1} = A[\beta(Y_t - \theta) - e_t]^{\alpha}[h(e_t)]^{1-\alpha}. \tag{1.48}$$

Since $e_t = \arg\max\ Y_{t+1}$, it follows that $\partial Y_{t+1}/\partial e_t = 0$ and therefore

$$Y_{t+1} = Y^{III}(Y_t), \tag{1.49}$$

where $Y^{III'}(Y_t) = \beta\alpha A k_t^{\alpha-1} > 0$, $Y^{III''}(Y_t) < 0$ and $\lim_{Y_t\to\infty}\ Y^{III'}(Y_t) = 0$ since $\lim_{Y_t\to\infty} k_{t+1} = \infty$.

In Stage III of Regime II, Y_t increases monotonically and converges to a unique, *locally stable,* steady-state equilibrium $\bar{Y} > 0$, where intergenerational transfers are positive and equal across all individuals, that is, $\bar{b}^P = \bar{b}^R > 0$.

REDISTRIBUTION AND THE DYNAMICS OF OUTPUT PER WORKER

Proposition 3 *(The effect of inequality on economic growth when credit constraints are no longer binding.) Consider two economies (or two alternative initial states of the same economy). Suppose that the economies are identical in all respects except for their degree of inequality. The two economies would be characterized by an identical path of output per worker.*

The Proposition follows from the fact that Y_{t+1} in (1.43) is independent of the distribution of output per worker in period *t* between the two groups.

Inequality has no effect on the growth process in Stage III of Regime II, since in the absence of credit constraints investment in human capital is optimal and since the marginal propensity to save is equal across individuals.

1.5. INEQUALITY AND DEVELOPMENT

Theorem 1 *Under Assumption 1.A1–1.A4*

(a) In the early stage of development when the process of development is driven by capital accumulation, inequality raises the rate of growth of output per worker over the entire stage.
(b) In the mature stage of development when the process of development is driven by universal human capital accumulation and credit constraints are binding, equality raises the growth rate of output per worker over the entire stage.

The Theorem is a corollary of Propositions 1 and 2 and Remarks 1 and 2.

In the early stage of development inequality is conducive for economic development. In this stage the rate of return to human capital is lower than the rate of return to physical capital and the process of development is fuelled by capital accumulation. Since capital accumulation is the prime engine of growth and since the marginal propensity to save is an increasing function of the individual's wealth, inequality increases aggregate savings and capital accumulation and enhances the process of development. Inequality enhances the process development in Regime I since a transfer of wealth from members of group R to members of group P (who do not save in this stage) would increase aggregate consumption, decrease aggregate intergenerational transfers, and thus would slow capital accumulation and the process of development.

In mature stages of development, the rate of return to human capital increases sufficiently so as to induce human capital accumulation, and the process of development is fuelled by human capital as well as physical capital accumulation. Since human capital is embodied in individuals and each individual's investment is subjected to diminishing marginal returns, the aggregate return to investment in human capital is maximized if the marginal returns are equalized across individuals. Equality therefore alleviates the adverse effect of credit constraints on investment in human capital and promotes economic growth.

1.6. CONCLUDING REMARKS

The described unified theory of Galor and Moav suggests that the role of inequality has changed in the process of development. In the early stages of industrialization physical capital is scarce, the rate of return to human capital is lower than the rate of return to physical capital and the process of development is fuelled by capital accumulation. The positive effect of inequality on aggregate saving dominates, therefore, the negative effect on investment in human capital, and inequality raises aggregate savings and capital accumulation and enhances the process of development. In later stages of development, as physical capital accumulates, the complementarity between capital and skills increases the rate of return to human capital. Investment in human capital accumulation increases and the accumulation of human capital as well as physical capital fuels the process of development. Since human capital is embodied in individuals and individuals' investment in human capital is subjected to diminishing marginal returns, the aggregate return to investment in human capital is maximized if investment in human capital is widely spread among individuals in society. Equality alleviates the adverse effect of credit constraints, and therefore has a positive effect on the aggregate level of human capital and economic growth. Moreover, the differences in the marginal propensities to save across individuals narrow as wages increase, and the negative effect of equality on aggregate saving subsides. In later stages of development, therefore, as long as credit constraints are sufficiently binding, the positive effect of inequality on aggregate saving is dominated by the negative effect on investment in human capital, and equality stimulates economic growth. As wages further increase, however, credit constraints become less binding, differences in the marginal propensity to save further decline, and the aggregate effect of income distribution on the growth process becomes less significant.[27]

The proposed unified theory generates an unexplored testable implication about the effect of inequality on economic growth. In contrast to the credit market imperfection approach that suggests that the effect of inequality on economic growth depends on the country's level of income (that is, inequality is beneficial for poor economies and harmful for rich ones), the theory suggests that the effect of inequality on growth depends on the relative return to physical and human capital. As long as credit constraints are largely binding, the higher is the relative return to human capital and the more adverse (or the less beneficial) is the effect of inequality on economic growth.

Although the replacement of physical capital accumulation by human capital accumulation as a prime engine of economic growth in the currently developed economies is instrumental for the understanding of the role of inequality in their process of development, the main insight of the chapter is relevant for the currently less developed economies that may have evolved differently. In contrast to the historical growth path of the currently developed economies, human capital accumulation may be the prime engine of economic growth in some LDCs, even in early stages of development, due to the importation of capital and skilled-biased technologies. In some of the current LDCs, the presence of international capital inflow diminishes the role of inequality in stimulating physical capital accumulation. Moreover, the adoption of skilled-biased technologies increases the return to human capital and thus, given credit constraints, strengthens the positive effect of equality on human capital accumulation and economic growth.

NOTES

1. The ratio of GDP per capita between the richest region and the poorest region in the world was only 1.1:1 in the year 1000, 2:1 in the year 1500, and 3:1 in the year 1820. In the course of the 'Great Divergence' the ratio of GDP per capita between the richest region (Western offshoots) and the poorest region (Africa) has widened considerably from a modest 3:1 ratio in 1820, to a 5:1 ratio in 1870, a 9:1 ratio in 1913, a 15:1 ratio in 1950, and a huge 18:1 ratio in 2001.
2. For instance, Classical Thermodynamics that lacked micro-foundations was ultimately superseded by the micro-based Statistical Mechanics.
3. Unified Field Theory, for instance, proposes to unify by a set of general laws the four distinct forces that are known to control all the observed interactions in matter: electromagnetism, gravitation, the weak force and the strong force. The term 'unified field theory' was coined by Einstein, whose research on relativity led him to the hypothesis that it should be possible to find a unifying theory for the electromagnetic and gravitational forces.
4. The transition from Malthusian stagnation to sustained economic growth was explored by Galor and Weil (1999, 2000), Lucas (2002), Galor and Moav (2002), Hansen and Prescott (2002), as well as others, and the association of Great Divergence with this transition was analysed by Galor and Mountford (2003).
5. As established in Section 1.2, and consistently with the revisionist view of the Industrial Revolution, neither the 19th century's take-off of the currently developed world, nor the recent take-off of less developed economies provide evidence for an unprecedented shock that generated a quantum leap in income per-capita. In particular, technological progress could not be viewed as a shock to the system. As argued by Mokyr (2002) technological progress during the Industrial Revolution was an outcome of a gradual endogenous process that took place over this time period. Similarly, technological progress in less developed economies was an outcome of a deliberate decision by entrepreneurs to adopt existing advanced technologies.
6. The socio-political economy approach provides an alternative mechanism: equality diminishes the tendency for socio-political instability, or distortionary redistribution, and hence it stimulates investment and economic growth. See the comprehensive survey of Benabou (1996b).
7. Benabou (1996a, 2000), Durlauf (1996), Fernandez and Rogerson (1996), and Mookherjee and Ray (2003) provide additional theoretical contributions and Perotti (1996) and Easterly (2001) provide evidence in support of this link between equality, human capital and growth. Banerjee and Newman (1993) and Aghion and Bolton (1997), among others, suggest that equality positively affects individuals' investment opportunities that could be in physical capital rather than human capital.
8. One may argue that the accumulation of physical capital at the individual level is also subjected to diminishing returns due to agency problems. However, the proposed hypothesis remains valid as long as the return to human capital accumulation at the individual level

diminishes significantly faster than the return on physical capital and the adverse effect of equality on saving is larger than its positive effect on the aggregate productivity of physical capital.

9. See Tomes (1981), Menchik and David (1983), and Dynan et. al. (2000).
10. For example, Cook (1995). Some studies do not find any significant effect of inequality on aggregate savings (for example, Schmidt-Hebbel and Serven, 2000).
11. For example, Flug et al. (1998) and Checchi (2001).
12. Evidence provided by Galor and Moav (2006) suggests that in the second phase of the Industrial Revolution, education reforms in Europe were designed primarily to satisfy the increasing skill requirements in the process of industrialization. It should be noted that although physical capital accumulation increased the demand for human capital, investment in education had the opposite effect on the return to human capital. For instance, the decline in the reward for education in the United States in the period 1910–40, despite a rapid skill-biased technological change, is due to the growth of the relative supply of more educated labour that accelerated during the high-school movement (Goldin and Katz, 1998, 1999).
13. They measure educational productivity by the contribution of education to the educational wage differentials.
14. This argument is supported indirectly by contemporary evidence about a higher rate of returns to human capital in less developed economies than in developed economies (Psacharopoulos and Patrinos, 2002). The greater prevalence of credit markets imperfections and other barriers for the acquisition of skills in less developed economies enabled only a partial supply response to industrial demand for human capital, contributing to this differential in the skill premium.
15. This section describes the model of Galor and Moav (2004).
16. If, alternatively, the time investment in education (foregone earnings) is the prime factor in the production of human capital, the qualitative results would not be affected, as long as physical capital would be needed in order to finance consumption over the education period. Both formulations assure that in the presence of capital markets imperfections investment in human capital depends upon family wealth.
17. This formulation of the saving function is consistent with the view that bequest as a saving motive is perhaps more important than life cycle considerations (for example, Deaton, 1992).
18. Moav (2002) shows that long-run inequality could persists in Galor and Zeira's (1993) framework, if this type of a 'Keynesian saving function' replaces the assumption of non-convexities in the production of human capital. Fishman and Simhon (2002) analyse the effect of income distribution on the division of labour and thereby on economic growth in a setting that integrates the classical and the credit market imperfections approaches. They argue that equality contributes to specialization and long-run growth if capital markets are imperfect and individuals' saving rates increase with income.
19. A more realistic formulation would link the cost of education to (teacher's) wages, which may vary in the process of development. For instance, $h_{t+1}^{i} = h(e_t^i / w_t)$ implies that the cost of education is a function of the number of efficiency units of teachers that are used in the education of individual *l*. As will become apparent from (1.10) and (1.11), under both formulations the optimal capital expenditure on education, e_t^i, is an increasing function of the capital–labour ratio in the economy, and the qualitative results are therefore identical under both formulations.
20. The Inada conditions are typically designed to simplify the exposition by avoiding corner solutions, but surely they are not realistic assumptions.

21. As follows from (1.25), there exists a feasible set of parameters $A, \alpha, \beta, k_0, \theta,$ and λ that satisfy Assumptions 1.A1–1.A3 such that $b_0^R \in (\bar{b}^u, \tilde{b})$. In particular, given the initial level of capital, if the number of the Rich in the initial period is sufficiently small $b_0^R > \bar{b}^u$.
22. In all stages of development members of group R are not credit constrained. That is, $e_t < b_t^R$, and the level of investment in human capital, e_t, permits therefore a strictly positive investment in physical capital, $b_t^R - e_t$, by the members of group R. If $e_t \geq b_t^R$ and hence, as follows from Lemma 1, $e_t > b_t^P$ there would be no investment in physical capital, the return to investment in human capital would be zero and $e_t = 0 < b_t^R$ in contradiction to $e_t > b_t^R$.
23. $b_{t+1}^R > 0$ in this interval since $b_t^R > 0$, and $\partial \psi^R(b_t^R, \ 0)/\partial b_t^R > 0$.
24. For any given $b > \tilde{b}$ (where $\tilde{b}$ is independent of A) since $\beta[w(\lambda b) + bR(\lambda b) - \theta]$ is strictly increasing in A, there exists a sufficiently large A such that $\beta[w(\lambda b) + bR(\lambda b) - \theta] > b$. Note that $\hat{b}$ decreases with A, however a sufficiently large θ assures that $\hat{k} > \tilde{k}$.
25. Note that the condition $\beta[w(\lambda \hat{b}) + \hat{b}R(\lambda \hat{b}) - \theta] > \hat{b}$ that follows from Assumption 1.A3 and assures that there is no steady-state in Stage I of Regime II, implies that $\beta R(\hat{k}) \geq 1$.
26. Note that k_t in Stage II of Regime II may decline below $\hat{k}$. In this case, conditional dynamics are described by (1.38). However, b_t^P is non-decreasing in Stage II of Regime II, that is, b_t^P is above the threshold level $b{=}\phi(b,k)$ of (1.38).
27. Inequality may widen once again due to skilled or ability-biased technological change induced by human capital accumulation. This line of research was explored theoretically by Galor and Tsiddon (1997), Caselli (1999), Galor and Moav (2000), Gould et al. (2001), and Acemoglu (2002), among others. It is supported empirically by Goldin and Katz (1998), among others.

REFERENCES

Abramowitz, M. and P.A. David (2000), 'American macroeconomic growth in the era of knowledge-based progress: the long-run perspective', in S.L. Engerman and R.E. Gallman (eds), *The Cambridge Economic History of the United States*, Cambridge, NY: Cambridge University Press.

Acemoglu, D. (2002), 'Technical change, inequality, and the labor market', *Journal of Economic Literature*, **40**, 7–72.

Aghion, P. and P. Bolton (1997), 'A theory of trickle-down growth and development', *Review of Economic Studies*, **64**, 151–72.

Alesina, A. and D. Rodrik (1994), 'Distributive politics and economic growth', *Quarterly Journal of Economics*, **109**, 465–90.

Banerjee, A. and A. Newman (1993), 'Occupational choice and the process of development', *Journal of Political Economy*, **101**, 274–98.

Benabou, R. (1996a), 'Equity and efficiency in human capital investment: the local connection', *Review of Economic Studies*, **63**, 237–64.

Benabou, R. (1996b), 'Inequality and growth', *NBER Macroeconomics Annual*, **11**, 11–74.

Benabou, R. (2000), 'Unequal societies: income distribution and the social contract', *American Economic Review*, **90**, 96–129.

Bourguignon, F. (1981), 'Pareto superiority of unegalitarian equilibria in Stiglitz model of wealth distribution with convex saving function', *Econometrica*, **49**, 1469–75.

Caselli, F. (1999), 'Technological revolutions', *American Economic Review*, **89**, 78–102.

Checchi, D. (2001), 'Inequality in incomes and access to education. A cross-country analysis (1960–95)', Mimeo, University of Milan.

Clark, G. (2003), 'The condition of the working-class in England, 1200–2000: Magna Carta to Tony Blair', UC Davis.

Cook, C. (1995), 'Saving rates and income distribution: further evidence from LDCs', *Applied Economics*, **27**, 71–82.

Crafts, N.F.R (1985), *British Economic Growth During the Industrial Revolution*, Oxford: Oxford University Press.

Deaton, A. (1992) *Understanding Consumption*, Oxford: Clarendon Press.

Dollar, D. and A. Kraay (2002), 'Growth is good for the poor', *Journal of Economic Growth*, **7**, 195–225.

Durlauf, S.N. (1996), 'A theory of persistent income inequality', *Journal of Economic Growth*, **1**, 75–94.

Dynan, K.E., J. Skinner and S. Zeldes (2000), 'Do the rich save more?', *NBER Working Paper* 7906.

Easterly, W. (2001), 'The middle class consensus and economic development', *Journal of Economic Growth*, **6**, 317–35.

Fernandez, R. and R. Rogerson (1996), 'Income distribution, communities, and the quality of public education', *Quarterly Journal of Economics*, **111**, 135–64.

Fishman, A. and A. Simhon (2002), 'The division of labor, inequality and growth', *Journal of Economic Growth*, **7**, 117–36.

Flora, P., F. Kraus and W. Pfenning (1983), *State, Economy and Society in Western Europe 1815–1975*, vol. 1, Chicago: St. James Press.

Flug, K., A. Spilimbergo and E. Wachtenheim (1998), 'Investment in education: do economic volatility and credit constraints mater?', *Journal of Development Economics*, **55**, 465–81.

Forbes, K. (2000), 'A reassessment of the relationship between inequality and growth', *American Economic Review*, **90**, 869–87.

Galor, O. (2005), 'From stagnation to growth: unified growth theory', in P. Aghion and S. Duraluf (eds), *Handbook of Economic Growth*, Amsterdam: North-Holland.

Galor, O. (2006), *Unified Growth Theory*, forthcoming.

Galor, O. and O. Moav (2000), 'Ability biased technological transition, wage inequality and growth', *Quarterly Journal of Economics*, **115**, 469–98.

Galor, O. and O. Moav (2002), 'Natural selection and the origin of economic growth', *Quarterly Journal of Economics*, **117**:1133–92.

Galor, O. and O. Moav (2004), 'From physical to human capital accumulation: inequality and the proces of development', *Review of Economic Studies*, 71, 1001–26.

Galor, O. and O. Moav (2006), 'Das human kapital: a theory of the demise of the class structure', *Review of Economic Studies*, **73**.

Galor, O. and A. Mountford, (2006), ‘Trade and the great divergence: the family connection’, *American Economic Review*. **96**.
Galor, O. and D. Tsiddon (1997), ‘Technological progress, mobility, and growth’, *American Economic Review*, **87**, 363–82.
Galor, O. and D.N. Weil (1999), ‘From Malthusian stagnation to modern growth’, *American Economic Review*, **89**:150–54.
Galor, O. and D.N. Weil (2000), ‘Population, technology and growth: from the Malthusian regime to the demographic transition", *American Economic Review*, **90**:806–28.
Galor, O. and J. Zeira (1993), ‘Income distribution and macroeconomics’, *Review of Economic Studies,* **60**, 35–52.
Goldin, C. and L.F. Katz (1998), ‘The origins of technology–skill complementarity’, *Quarterly Journal of Economics*, **113**, 693–732.
Goldin, C. and L.F. Katz (1999), ‘The return to skill across the Twentieth Century United States’, *NBER Working Paper* 7126.
Goldin, C. and L.F. Katz (2001), ‘The legacy of U.S. educational leadership: notes on distribution and economic growth in the 20th century’, *American Economic Review*, **91**, 18–23.
Gould, E.D., O. Moav and B.A. Weinberg (2001), ‘Precautionary demand for education, inequality, and technological progress’, *Journal of Economic Growth*, **6**, 285–315.
Green, A. (1990), *Education and State Formation*, New York: St. Martin’s Press.
Hansen, G. and E. Prescott (2002), ‘Malthus to Solow’, *American Economic Review*, **92**:1205–17.
Kaldor, N. (1957), ‘A model of economic growth’, *Economic Journal*, **67**, 591–624.
Keynes, J.M. (1920), *The Economic Consequences of the Peace*, Macmillan and Co. Limited.
Lewis, W.A. (1954), ‘Economic development with unlimited supply of labor’, *The Manchester School*, **22**, 139–91.
Landes, D.S. (1969), *The Unbound Prometheus. Technological Change and Industrial Development in Western Europe from 1750 to the Present*, Cambridge, MA: Cambridge University Press.
Lucas, R.E. Jr. (2002), *The Industrial Revolution: Past and Future*, Cambridge, MA: Harvard University Press.
Matthews, R.C., C.H. Feinstein and J.C. Odling-Smee (1982), in *British Economic Growth 1856–1973*, Stanford: Stanford University Press.
Menchik, P. and M. David (1983), ‘Income distribution, lifetime savings, and bequests’, *American Economic Review*, **73**, 672–90.
Moav, O. (2002), ‘Income distribution and macroeconomics: the persistence of inequality in a convex technology framework’, *Economics Letters,* **75**, 187–92.
Mokyr, J. (1990) *The Lever of Riches*, New York: Oxford University Press.
Mokyr, J. (1993), ‘The new economic history and the industrial revolution’, in J. Mokyr (ed.), *The British Industrial Revolution: an Economic Perspective*, Boulder CO: Westview Press, pp. 1–131.
Mokyr, J. (2002), *The Gifts of Athena: Historical Origins of the Knowledge Economy*, Princeton: Princeton University Press.

Mookherjee, D. and D. Ray (2003), 'Persistent inequality', *Review of Economic Studies*, **70**, 369–93.
Perotti, R. (1996), 'Growth, income distribution, and democracy: what the data say', *Journal of Economic Growth,* **1**, 149–87.
Psacharopoulos, G. and H.A. Patrinos (2002), *Returns to Investment in Education: A Further Update September 1, 2002*, Washington D.C.: World Bank.
Sanderson, M. (1995), *Education, Economic Change and Society in England 1780–1870*, Cambridge: Cambridge University Press,
Schmidt-Hebbel, K. and L. Serven (2000), 'Does income inequality raise aggregate saving?', *Journal of Development Economics*, **61**, 417–46.
Smith, A. (1776) [1976], *An Inquiry into the Nature and Causes of the Wealth of Nations*, in the Glasgow edition of the works and correspondence of Adam Smith (1976) in two volumes, R.H. Cambell, A.S. Skinner and W.B. Todd (eds), Oxford: Oxford University Press.
Tomes, N. (1981), 'The family, inheritance and the intergenerational transmission of inequality', *Journal of Political Economy*, **89**, 928–58.
Voth, H.J. (2004), 'Living standards and the urban environment', in P. Johnson and R. Floud (eds), *The Cambridge Economic History of England*, Cambridge: Cambridge University Press.

2. The role of income distribution in long run endogenous growth

Amit Bhaduri

2.1. INTRODUCTION

Two related questions have been the central concern of almost all the theories of economic growth, namely, what propels growth and what limits it? Posed this way neither question is very precise, and unsurprisingly, answers would vary considerably. However, if we look back at the great masters of classical political economy, we might at least begin to appreciate better some of the defining characteristics of the problem of economic growth. Adam Smith (1776), and nearly a century before him William Petty (1662), identified the process of increasing technical and social division of labour as the main force propelling economic growth, and raising continuously the productivity of labour. Since increasing technical division of labour in Smith's celebrated example of the pin factory leads naturally to the idea of increasing returns in the form of rising labour productivity in each successive round of production from the supply side, his observation that the division of labour is limited by the extent of the market assumes special significance. It suggests that he assigned to demand the role of setting the limit to productivity growth, while in this scheme the constraint of natural resources as non-produced primary factors of production has only a limited role to play.

David Ricardo (1817) looked at the problem from almost the opposite angle. For him, land as a symbol for all primary factors not produced within the system, sets the limit to growth. As the margin of cultivation extends, land of diminishing fertility is brought into cultivation, making land at the extensive margin subject to diminishing returns. Rent being determined by the difference between the average and the marginal productivity of land, the share of rent in national income increases at the expense of profit due to diminishing return to land as a factor of production, as the margin of cultivation increases. By making the assumption – sociologically plausible for his time – that all profits of the capitalists, but no rent of the landlords and wage of the workers are saved, Ricardo could argue that this process of redistribution of income from profit to rent would lead to a gradual drying up of profits as the source of the wage fund to limit growth, driving the economy towards its ultimate stationary state.

The link between functional or class distribution of income and saving was taken up later as an important feature in Keynesian and post-Keynesian growth theory (Kahn, 1959; Kaldor, 1957; Pasinetti, 1962; Robinson, 1956). At the same time, however, there is a crucial difference. Ricardo's analysis, in contrast to that of Smith, left no place for aggregate demand or 'the extent of the market' as a possible limiting factor to growth. From this point of view, it was the very antithesis of the later Keynesian tradition in growth theory. Ricardo ignored the demand problem by assuming implicitly that enough demand would always exist, and all the saving of the economy would be automatically invested back within the same period for extending the margin of cultivation, that is, saving equals investment by assumption. This very assumption, which results in ignoring all problems of aggregate demand, came to characterise the entire post-war neo-classical tradition in growth theory.

The Ricardian idea of inter-class distribution as an important variable driving saving features prominently in post-Keynesian growth theory, and the Ricardian circumvention of all problems of aggregate demand is a defining characteristic of neo-classical growth theory. However, neither tradition is able to deal satisfactorily with the Smithian view of intra-class competition as the driving force behind the division of labour. In the absence of increasing division of labour and without any demand constraint, Ricardo plausibly suggested that economic growth would hit ultimately the constraint of some natural resource.

It is worth pointing out in this context that the central result of post-war neo-classical growth theory is basically a reinvention of this vision of Ricardo (Solow, 1956; Swan, 1956). In models in this tradition, the produced means of production, 'capital', is considered as a factor of production in an aggregate production function on a par with labour. However, the former is producible and augmentable through saving, while the latter is non-producible within the system. In contrast to diminishing returns on the extensive margin of land postulated by Ricardo, both the factors land and labour are assumed to have diminishing returns on their intensive margins through substitution between them. Assuming the only primary and non-produced factor in the system, labour grows at some exogenous rate, the model shows that the economy also converges to that exogenously given growth rate of labour. The analogy is clear; just as the Ricardian economy converges to a stationary state with a zero growth rate of land, so does the

Solow–Swan economy to a steady state with a constant exponential growth rate of labour.

However, this result derived within the neo-classical analytical framework is logically insecure, when extended beyond a one-commodity world. In his search for an 'invariant measure of value', Ricardo was already aware of this problem. Beyond the one commodity world, the resulting capital theoretic problems might indeed render the central mechanism of factor substitution and the associated marginal productivity theory of distribution logically invalid (Sraffa, 1960; Samuelson, 1966; Pasinetti, 2000). In effect the depiction of the supply side through an aggregate production function reduces it to a one-commodity model, while the problem of Keynesian effective demand ruled out through the assumption that saving and investment are one and the same decision, reduces it to a single agent economy. In short, we have a misleadingly reductionist one individual, one-commodity model for analysing the problem of modern capitalistic growth! And, it is within this over-simplified framework that the Ricardian answer to the question of what limits growth has been reinvented in neo-classical growth theory.

Karl Marx (1867, 1891), whose unique distinction was to emphasise the importance of historical categories in economic analysis, discussed capitalistic growth in particular, rather than economic growth in general. With commodity production for the market as the most basic feature of capitalism, Marx encountered the problem of the size of the market for absorbing the surplus product or 'surplus value' generated through the exploitation of the workers. That this problem of the realization of surplus into profit is a forerunner of the analysis of effective demand becomes transparent in Kalecki's formulation preceding Keynes (Kalecki, 1971). In his analysis, Marx tried to deal simultaneously with the two related sets of issues, namely, how the surplus is generated through exploitation and how it is realised into monetary profits. And, in dealing with the latter question, he came up with an under-consumptionist explanation of the crisis of capitalism. While in his scheme of expanded reproduction the entire surplus requires to be reinvested in each period to avoid the crisis of realisation, he argued that the exploitation of labour, facilitated by a large reserve army of labour, would not allow real wages to rise. With the limited size of the market caused by the stagnant purchasing power of the workers, capitalists would find it increasingly difficult to invest profitably the rising surplus per worker.

From a contemporary perspective, a most valuable insight of Marx was to recognise the two-sided role played by intra-class competition among the capitalists as well as inter-class competition between capital and labour. Both these aspects of competition might be driving labour productivity higher through increasing division of labour and labour–saving devices, but they would also impact adversely on the real wage rate through the reserve army of labour. This would tend to widen the gap between rising labour productivity and a more or less constant real wage, raising the potential surplus per worker at the micro-level, while the realisation of this surplus on a macro-level would become more difficult due to the limited size of the market. A more precise analysis of these conditions under technical progress and different market structures was developed later, particularly in the 'stagnationist' model of Steindl (1952).

Marx's discussion of the exploitation of labour at the micro-level of the factory, juxtaposed against his macro-view of the failure of the surplus to be realised due to under-consumption, highlights how microeconomic arguments may run into various macroeconomic fallacies of composition. Thus, more exploitation of the worker or a lower real wage rate per worker would not necessarily mean that more total profits can be realised due to the insufficiency of demand. This line of argument, used later by Keynes in formulating his 'paradox of thrift', or in the 'wage-cut controversy', also has a contemporary parallel. Higher labour productivity through downsizing the labour force, when carried out by a single corporation, increases its market share and profit. However, when carried out by many, it might reduce aggregate demand, and the realized profit of all corporations.

The presence of such fallacies of composition, which defines the border between micro- and macroeconomics, seems to be a forgotten lesson in cotemporary neo-classical macroeconomics. Thus, it finds it good enough for growth theory to proceed on the assumption of an all-seeing optimising agent. His inter-temporally optimal saving is presumed to be invested automatically in each period, avoiding all problems of deficiency in aggregate demand in this single agent economy (see, for example, Romer, 1996 for an exposition). The literature on optimum saving had originally been developed in the context of normative planning theory (Ramsey, 1928; Koopmans, 1965; Cass, 1965), and understandably abstracted from all problems of effective demand or profit realisation in an idealised centrally planned economy. In a parallel vein the models of overlapping generations, despite their more plausible assumptions regarding saving, miss the same central point that the savings plans of households or the profit expectations of the firms cannot macro-economically be realised without adequate demand (Samuelson, 1958; Diamond, 1965). In short, we keep returning to some version of Say's law by leaving out all problems of aggregate demand. As a result, models in the neo-classical tradition become purely supply side growth models, while their supply side depiction through an aggregate production function tends to be logically flawed due to capital theoretic problems outside a one-commodity world.

The aim of this chapter in mapping out an alternative approach to the problem of capitalistic growth can be appreciated better against this background of the unsatisfactory state of mainstream neo-classical growth theory. The model presented here attempts to blend together diverse elements of the classical tradition with the Keynesian theory of effective demand, and the Schumpeterian emphasis on the influence of market structures on technological change. In the present model the neo-classical construct of the aggregate production function to depict the supply side is avoided. It is not only logically flawed outside a one-commodity world, but even within that framework a higher capital–labour ratio does not seem to provide an empirically satisfactory explanation of higher labour productivity (Solow, 1957; Shaikh, 1980; Mankiw et al., 1992). Without relying on a well-behaved production function with capital labour substitution and the associated marginal productivity theory of distribution, the present model takes a different approach to depicting the supply side, and the class distribution of income.

On the supply side, the main propelling force for growth is identified as the Smithian vision of generation and diffusion of technology through increasing division of labour (Young, 1928; Kaldor, 1989). It is postulated to be driven by intra-class competition among the capitalists as Smith had emphasised, and by inter-class competition between capital and labour as Marx had highlighted. Thus, labour productivity growth is viewed as the outcome of competition at various levels, shaped by different structures of the product and the labour market. An important outcome of this wide-ranging competition in the product and in the labour market is the observed tendency of the share of wages to remain roughly stable over the longer run. While this attributes 'neutrality' to technical progress in the sense of Harrod (1942), it is not incorporated in our model as a technological property by assuming a Cobb–Douglas technology, or purely labour-augmenting technical progress (Uzawa, 1961), also typical of the 'human capital' approach. In contrast, we postulate a continuous race between rising labour productivity and the real wage rate as the outcome of intra- and inter-class competition, capturing the observed phenomenon of endogenously generated neutral technical progress and the distribution of income as interwoven processes. However, while technical progress gets diffused in the economy through intra- and inter-class competition as visualised by the classical economists, we share the neo-classical view that productive knowledge has the public good character of non-excludability. Instead of theorising about how this knowledge is generated by postulating various implausible production functions for knowledge generation which is typical of the current mode in neo-classical theory (for example, Barro and Sala-i-Martin, 1995; Lucas 1988; Romer, 1986), we focus on the economic process of diffusion of productive knowledge as a public good, resulting in positive externalities, and increasing returns as the driving force on the supply side of the growth process.

The problem of effective demand is introduced along the usual route by separating investment from saving decision. While very many investment functions are plausible, and none of them particularly satisfactory, our primary motivation in this model is to highlight how effective demand plays a central role in explaining growth without necessarily guaranteeing full employment. While the saving function can also take various plausible forms, none fully satisfactory, the assumption of neither an inter-temporally optimal saving plan by an immortal agent, nor the indefinitely repetitive arrangement of an inter-temporal contract among overlapping generations adds much to a macro-model of growth with effective demand. The problem is not merely lack of realism. According to the principle of effective demand of Keynes and Kalecki, investment is treated as a variable largely independent of saving, and saving tends to adjust through changes in income, in case of a discrepancy between investment and saving. Consequently, the realisation of the saving plans of the households depends on the investments carried out by the firms. Since elaborate saving plans of households through inter-temporal optimisation or arrangement among overlapping generations are unrealisable without matching investment plans of the firms, the analysis gets rather unnecessarily complicated by specifying only elaborate saving plans. We focus instead on the central issue of the interaction between investment and saving as two separate decisions, as this interaction influences the process of economic growth from the demand side.

2.2. THE MODEL

Almost inverting the image of the Ricardian theory of differential land rent, we postulate that our industrial economy is subject to increasing returns due to the public good character of productive knowledge which gets diffused to rival firms, as the scale of output and employment expands. This means that the labour employed by the firm at the frontier or ‘margin’ of new technological knowledge is more productive than the average productivity of labour in the rest of the economy. If that marginal firm at the frontier of knowledge has a sufficiently low weight in the total output produced in the economy, then its higher productivity raises only negligibly the average productivity of the economy. Note that this assumption of ‘atomistic’ firm as innovator, considered typically by the classical economists, rules out the problem of technology diffusion in oligopolistic market structures (Steindl, 1952).

Assuming intra-class competition among the firms along the lines suggested by the classical economists, the marginal firm at the technological frontier sets its price lower in accordance with its lower production cost compared to the ruling price in the economy, and attempts to capture a higher share of the market. Thus, the cost differentials among the firms are assumed not to affect their respective profit margin per unit of sale, but get converted into a competition among them for increasing the relative volume of sales through higher market share.

Under classical competition this price-setting behaviour puts a general downward pressure on the price level, as the technologically advanced firm emerges as the price leader by lowering its price. The numerous remaining firms follow the price leader in order to survive in a competitive market by lowering their costs and prices through adopting the new technology gradually over time. This technological competition results exclusively in price competition aimed at increasing relative market shares with all firms maintaining the same mark-up on their costs. And in this fight for market shares, the benefit of cost-reducing technology is passed on to the consumers without any higher profit margin even for the innovating firm. This process of reduction in costs and prices with increasing returns through the diffusion of technology is captured more easily as a discrete process.

Let

p_t^f = price set by the firm at the frontier of technology at time t
p_t = the average ruling price of the numerous other firms in the economy.

Assuming a uniform mark-up (k), and money wage (w), the mark-up prices with labour as the only variable cost are,

$$p_t^f = kw\frac{dL}{dY} \quad \text{and} \quad p_t = kw\frac{L}{Y} \tag{2.1}$$

Where, $p_t^f < p_t$ because $(dL/dY) < (L/Y)$ due to increasing returns and w = money wage rate, L = employment, Y = output, and $(Y/L) = x$, labour productivity. If technology diffuses at a uniform rate, and it takes t periods for each innovation to be diffused completely, then the adjustment in the price level in each period is $(p_t^f - p_t)/T$. Therefore, the average price in the next period becomes,

$$p_{t+1} = p_t + \frac{p_t^f - p_t}{T},$$

or

$$(p_{t+1} - p_t) = \Delta p_t \cong \frac{dp}{dt} = \lambda(p^f - p), \ \ \lambda = \frac{1}{T} > 0 \tag{2.2}$$

The speed of adjustment λ in (2.2) is related to the ease with which technological diffusion takes place in the economy. For example, a tighter regime of intellectual property rights would make the value of λ smaller, reducing also the speed of price reduction in the economy.

With the aid of the definitional relation among the growth rates in labour productivity (*x*), output (*Y*) and employment (*L*) as,

$$g_l = g_y - g_x \tag{2.3}$$

and, using (2.1) in (2.2), we obtain on simplification the decline in the rate of average price level in percentage terms as,

$$g_p = -\lambda \frac{g_x}{g_y}, \ \ g_y \neq 0, \tag{2.4}$$

the variable *g* representing the proportional growth rate of the relevant variable denoted by the subscript.

Note that the elasticity of employment with respect to output, that is $[(dL/L)/(dY/Y)]$ is rewritten in (2.4) in terms of the growth rates by the chain rule of differentiation to capture the dynamic nature of increasing returns.

Equation (2.4) captures the consequence of technological competition among rival firms in terms of a long run downward pressure on the price level. It is generated as labour productivity rise in the economy characterised by dynamic increasing returns, and the benefit is passed on to the consumers through price reduction, as firms contest over market shares. The strength of dynamic increasing returns in a growing economy is measured in the present context by the elasticity of employment with respect to output over time, that is (g_l / g_y). The smaller is this ratio, the larger is the growth of labour productivity with respect to output growth from (2.3) and hence the larger is the reduction in unit cost and the price level.

However, while price changes according to equation (2.4), the behaviour of the real wage rate depends also on how the money wage rate changes in the labour market through inter-class competition. We assume that the percent change in the money wage rate is systematically related to the change in the unemployment rate (du/dt), as specified by the 'wage curve'. This relationship is used often in the recent literature in place of the more traditional Phillips' curve as more stable over time (for example, Blanchflower and Oswald, 1994, 1995; Card, 1995). Some estimates of the elasticity b of the wage curve places it in the range of –0.2 and –0.8 (Blanchflower and Oswald, 1995). However, by definition the elasticity of the wage curve presumes a changing rate of unemployment (u) to yield,

$$-b = \left[\frac{dw}{dt}\frac{1}{w}\right]\left[\frac{du}{dt}\frac{1}{u}\right]^{-1}, \tag{2.5}$$

where b is the absolute value of the elasticity. It may be rewritten from definition as,

$$-b = \frac{dw}{dt}\frac{1}{w}\frac{(n-g_l)(1-u)}{u},$$

where n and g_l are the growth rates of the labour force and of employment respectively. Note that unlike in the Phillips curve, the rise in money wage is not defined for any constant rate of unemployment at $n = g_l$, and the magnitude of the elasticity changes at different rates of unemployment.

We assume for expositional convenience a particular form of the wage curve given in terms of the employment rate $(1-u)$, as

$$w = a(1-u)^m, \quad m > 0, \tag{2.6}$$

where a represents the full employment wage at $u = 0$.

From the logarithmic differentiation of (2.6), the growth rate of money wage becomes,

$$g_w = g_a + m(g_l - n) \tag{2.7}$$

Consequently the rate of growth of the real wage rate (g_v) is given from equations (2.4) and (2.7) as,

$$g_v = g_w - g_p = g_a + m(g_l - n) + \lambda\frac{g_x}{g_y} \tag{2.8}$$

If labour productivity (g_x) continues to rise at a rate different from that of the real wage rate in the longer run, the share of wages in income would also continue to rise or fall, becoming ultimately incompatible with steady state growth. Therefore labour productivity growth may be postulated to adjust in such a manner as to keep the wage share constant in the long run. This is captured by the adjustment equation,

$$\frac{dg_x}{dt} = \beta[g_v - g_x] = \beta\left[g_a + m(g_y - g_x - n) + \lambda\frac{g_x}{g_y} - g_x\right] \tag{2.9}$$

where $\beta > 0$ is the speed of adjustment.

In contrast to most neo-classical models of endogenous growth, the rough constancy of the share of wages over the long run is not viewed in (2.9) as a technological datum through ad hoc assumptions like Cobb–Douglas technology or strictly human capital approach. The co-evolution of technology and wage share are instead viewed in (2.9) as the outcome of an economic process that involves intra- as well as inter-class competition, leading to an observed tendency towards constancy of the wage share over the longer run. This view of technological progress, especially its adoption and diffusion as an endogenous process, comes close to the Smith–Marx vision. But it differs from the classical, especially Marxian, view in postulating that technological development tends to keep the wage share, but not the real wage rate constant over the long run.

For introducing effective demand into the model, saving is distinguished from investment, by postulating an investment function which is different from the saving function. However, since neither function has a commonly agreed form, for expositional simplicity relatively simple specifications are used. Our purpose is to focus on the interaction between investment and saving in influencing the growth path through effective demand and the role that labour productivity plays in this process. Investment is assumed to depend positively on the current level of output as a predictor of the future state of demand, as well as on average labour productivity. Thus, expectation is assumed to be static or quasi-static, or what Keynes more helpfully described as expectations ruled by ‘conventions’ (Keynes, 1937). Note that in the context of intra-class competition among the firms specified by equation (2.2), each round of higher productivity, initiated by the innovator to reap transient benefit (for example, Schumpeter, 1961), also forces other firms in the economy to update their technologies, and thus stimulates the overall level of investment. On these assumptions the investment function is postulated as $I = I\,(Y, x)$, which on simple manipulation reduces to

$$g_I = \eta_y g_y + \eta_x g_x, \tag{2.10}$$

where η_y, η_x are positive partial elasticities of investment with respect to output and productivity respectively.

Saving is treated simply as an increasing function of income, so that,

$$g_s = \varepsilon_Y g_y, \tag{2.11}$$

where ε_y is the positive elasticity of saving with respect to income. For simplicity of exposition in the present model, all the relevant elasticities in (2.7), (2.10) and (2.11) are assumed constant, rendering the system either linear or the analysis valid for linear approximation around equilibrium. Note that a constant average and marginal propensity to save (2.11) would imply $\varepsilon_Y = 1$, while changing distribution among the classes with different saving propensities would affect the value of this elasticity (Kaldor, 1956). We refrain from introducing these complications, as well as those caused by retained and distributed profits of the corporations or institutional saving relevant for modern corporate capitalism (Pitelis, 1997).

Assuming the economy starts from an initial condition of investment saving equality, we need to capture how the growth rate of output would adjust to the growth in excess demand, resulting from a discrepancy between the growth rates of investment and saving. Therefore we consider the possibility of disequilibrium in the growth rates, rather than in the levels of output, investment and saving. Suppose the economy is initially in a state of equilibrium growth, so that the ratio $(I^*/S^*) = 1$, with both I* and S* growing at the same rate. A higher rate of growth of investment above this equilibrium path, that is $(I/I^*) > 1$, would require the corresponding saving rate (S/S^*) to increase through a higher growth rate in output for restoring back equilibrium. Formally, this may be represented by the adjustment process,

$$\left(g_Y - g_Y^*\right) \cong \frac{dg_Y}{dt} = \Psi\left(\frac{I^*}{S^*} + \frac{I}{I^*} - \frac{S}{S^*}\right),$$

$$\Psi\left(\frac{I^*}{S^*}\right) = \Psi(1) = 0, \text{ and } \Psi'\left(\frac{I^*}{S^*}\right) = \Psi'(1) = \alpha > 0 . \qquad (2.12)$$

Small deviations from the equilibrium are represented by,

$$\left(\frac{I}{I^*} - \frac{S}{S^*}\right) = \left[(1 + g_I) - (1 + g_S)\right],$$

so that the first-order approximation of (2.12) by the Taylor's series yields,

$$\frac{dg_Y}{dt} = \alpha[g_I - g_S], \quad \alpha > 0 . \qquad (2.13)$$

To interpret equation (2.13) from the opposite side, it may be integrated to yield, $Y = Y^*(I/S)^\alpha$, from which its economic rationale is easily seen. Y^* is the particular set of constants of integration representing real income at

$(I^*/S^*)=1$, with corresponding growth rate, g_y^* at $(I/S)=1$. Consequently, it has the economic interpretation that an excess of investment over saving ratio, that is $[(I/I^*)/(S/S^*)]>1$ would also raise the output growth path by the ratio (Y/Y^*) according to some power law with an exponent α.

Using (2.10) and (2.11) in (2.13), we obtain,

$$\frac{dg_Y}{dt}=\alpha(g_I-g_S)=\alpha\left[-(\varepsilon_Y-\eta_Y)g_Y+\eta_x g_x\right],\quad \alpha>0, \tag{2.14}$$

the speed of adjustment.

Equations (2.14) and (2.9) form a coupled dynamical system in the two variables, g_y, g_x.

The growth rate in output g_y reaches its stationary value at,

$$\frac{dg_y}{dt}=0\text{, implying } g_y=zg_x\text{, where } z=\frac{\eta_x}{\varepsilon_y-\eta_y}. \tag{2.15}$$

Note $z>0$, provided the usual one-variable Keynesian output adjustment stability condition, $(\varepsilon_Y-\eta_Y)>0$ holds, given $\eta_x>0$. Moreover, unless $z>1$, any positive growth rate in output would be out-stepped by labour productivity growth, so that the employment growth rate can not be positive from (2.3), implying that the economy would be experiencing jobless or even negative growth in employment.

We emphasise that this restriction on the value of z need not necessarily be satisfied, as it depends on the investment behaviour of the firms and on the saving behaviour of the households. Nevertheless, in what follows, we restrict our analysis of the economy to the case where investment and saving behaviour make positive growth rates in output, productivity and employment possible. This means from (2.15) that,

$$z>1\text{, implying } \eta_x+\eta_Y>\varepsilon_y>0,\text{ for } \eta_x>0. \tag{2.16}$$

The rest point or equilibrium of the coupled dynamical system (2.14) and (2.9) is given by,

$$g_y^*=\frac{z(g_a-mn)+\lambda}{1+m-mz}\text{, and } g_x^*=\frac{g_y^*}{z}. \tag{2.17}$$

The stability of the of the steady state equilibrium described by (2.17) requires the relevant Jacobian matrix evaluated at equilibrium to have a negative trace (T), and positive determinant (D), that is

$$T < 0\text{, implying, } \alpha(\varepsilon_y - \eta_y) + \beta\left[(1+m) - \frac{\lambda}{g_y^*}\right] > 0 \tag{2.18}$$

$$D > 0\text{, implying, } \alpha\beta(\varepsilon_y - \eta_y)(1 + m - mz) > 0\,, \tag{2.19}$$

and (2.19) is satisfied in view of (2.16) if,

$$1 + \frac{1}{m} > z > 1. \tag{2.20}$$

Note from (2.6), (2.7) and (2.20) that the weaker is the inter-class conflict between labour and capital resulting in a weaker wage response to employment growth, measured by the elasticity *m*, the less restrictive becomes condition (2.20). Consequently the system is more likely to be stable for weak inter-class conflict at relatively low values of *m*, provided the trace condition (2.18) is satisfied simultaneously. Substituting the value of g_Y^* from (2.17), the trace condition (2.18) is seen to be sufficiently satisfied in view of (2.16) and (2.20) provided,

$$(g_a - mn) > \lambda. \tag{2.21}$$

At the same time, from (2.17) and (2.19) and (2.20),

$$g_Y^* > 0\text{, further requires, } (g_a - mn) > \frac{\lambda}{z}\,. \tag{2.22}$$

Given condition (2.16), inequality (2.21) is stricter than (2.22), and becomes the binding inequality.

An interesting economic implication of (2.20) and (2.21) follows. Given the state of intra-class competition among rival capitalists by some value of the parameter λ, inequality (2.20) and (2.21) imply that weaker inter-class competition between capital and labour, that is a low value of *m* would tend to make the system stable. And, given the value of *m*, stronger intra-class competition, that is a higher value of λ, would tend to be destabilising for the system. However, while less acute inter- as well as intra-class competition tend to be stabilizing for the steady state growth path, it can also be seen from (2.17) that the growth rates of both output g_y^* and productivity g_x^* tend to be stimulated by more fierce intra-class competition among rival capitalists. Thus, the competition among rival capitalist firms operates like a double-edged weapon in this system; while stimulating growth, it tends to destabilise it.

Although the economic system depicted above seems capable of a stable configuration of steady, positive growth rates in output, productivity and employment under certain conditions, it is in fact only a quasi-steady state, because the unemployment rate (u) changes through time. And, with the unemployment rate increasing or decreasing continuously, the quasi-stationary state would be unsustainable for long. Thus, contrary to the comfortable neo-classical story of full employment growth, this model intends emphasizing that even in the long run, there is no automatic tendency in the market economy to keep the labour market either at full employment or at some constant, so-called 'natural' rate of unemployment, despite equilibrium in the commodity market, and constancy of the share of wages (see Bhaduri, 2002).

However, if we assume that the economy somehow maintains either full employment or a constant rate of unemployment, then by assumption,

$$g_L^* = n = (g_y^* - g_x^*) \ . \qquad (2.23)$$

The corresponding steady state at $dg_y / dt = 0, dg_x / dt = 0$ can be computed from (2.9) and (2.14) as,

$$g_x^* = \left(g_a + \frac{\lambda}{z} \right), \ g_y^* = z \left(g_a + \frac{\lambda}{z} \right) . \qquad (2.24)$$

Our previous conclusion can be seen to hold in (2.24) also. In the steady state, a more acute intra-class competition among rival firms raises both output and productivity growth through a higher value of λ. However, since it has to satisfy simultaneously condition (2.23) yielding,

$$\lambda = z \left[\left(\frac{n}{z} - 1 \right) - g_a \right], \qquad (2.25)$$

intra-class competition among the firms cannot go beyond a point without upsetting the exogenous labour supply constraint.

For a comparison with the standard neo-classical model, the value of λ in (2.25) is inserted into (2.24) to yield the steady state output growth rate coupled with full (or a constant rate of) employment to yield,

$$g_y^* = \frac{nz}{z-1} \ . \qquad (2.26)$$

In contrast to the standard neo-classical result made famous by the Solow–Swan model, (2.26) shows that the steady state growth rate of output is not strictly constrained by the exogenous growth rate of the labour supply. Influenced by both saving and investment behaviour through the value of the parameter $z > 1$, it exceeds the growth rate of labour supply.

2.3. A SUMMING UP

One of the main objectives of recent neo-classical endogenous growth theory has been to free the long run output growth rate from the exogenous constraint of labour supply growth (see Arrow, 1962; Frankel, 1962; Romer, 1986, Lucas, 1988; Barro and Sala i Martin, 1995). However, as the present model shows this objective is achieved more easily and plausibly through the influence of effective demand operating through saving and investment on the growth rate of output. It also shows that recourse to the concept of a non-decreasing marginal product of 'composite' capital, consisting of physical and human capital, is both unnecessary and misleading for freeing the output growth rate from being strictly constrained by the exogenous supply of labour.

However, once the problem of aggregate demand is recognised, neither full employment nor the stability of the long run growth path can be generally presumed. While Harrod (1939) might have overstated the case for instability in this respect, models in the neo-classical tradition of Solow and Swan err on the other side by overstating the case for stability by relying on a one-agent, one-commodity framework. The former assumption amounts to ignoring the problem of effective demand by postulating a single 'rational' agent whose saving is identical with his investment. The latter assumption of one commodity is necessary, because the view of 'capital' as a factor of production on a par with labour is logically indefensible outside a one-commodity world. By jettisoning both the aggregate production function with capital–labour substitution, and the single agent framework, we show why steady state growth is unlikely, and even if achieved in some limited sense, it need not guarantee full employment, or even a constant (natural) rate of unemployment.

The growth process becomes endogenous in this model through continuous diffusion of technology driven by intra-class rivalry among rival firms, as well as inter-class competition between capital and labour. In this model intra-class competition among the firms is seen to act like a double-edged weapon, while it raises the growth rate, it also tends to destabilise it. Our specification of the dynamics of inter-class competition on the other hand suggests how the growth of real wages plays a dual role in the capitalist economy in the short as well as in the long run. This is not simply because higher wage adds to demand on the one hand, but raises the costs of production on the other, leading to the possibilities of wage- and profit-led regimes of growth (see Bhaduri and Marglin, 1990; Marglin and Bhaduri, 1990). In the longer run, the growth in real wage has an added function in the capitalist economy. It poses a challenge to the firms continuously by threatening to squeeze their profit, and has to be countered by the successful firms through raising their labour productivity by updating technology. The technological dynamism of a capitalist economy is, to a significant extent, the result of this race between the growth rates in real wages and in labour productivity. It follows that the policy solution of imposing greater restraint on wages may turn out to be counter-productive in the longer run, in so far as it weakens this race contributing to the endogenously generated technological dynamism of successful capitalism. At the same time it would also rob the system of its most attractive economic feature of making high growth compatible with a rising standard of living for the working people.

REFERENCES

Arrow, K. (1962), ‘The economic implications of learning by doing’, *Review of Economic Studies*, **29**:55–73.

Barro, R. and X. Sala-i-Martin (1995), *Economic Growth*, New York: McGraw-Hill.

Bhaduri, A. (2002), ‘Chaotic implications of the natural rate of unemployment’, *Structural Change and Economic Dynamics*, **3**(3):357–65.

Bhaduri, A. and S. Marglin (1990), ‘Unemployment and the real wage: the economic basis for contesting political ideologies’, *Cambridge Journal of Economics*, **14**:375–93.

Blanchflower, D.G and A.J. Oswald (1994), *The Wage Curve*, Cambridge, MA: MIT Press.

Blanchflower, D.G and A.J. Oswald (1995), ‘An introduction to the wage curve’, *Journal of Economic Perspectives*, **9**:153–67.

Card, D. (1995), ‘The wage curve’, *Journal of Economic Literature*, **33**:785–99.

Cass, D. (1965), ‘Optimal growth in an aggregate model of capital accumulation’, *Review of Economic Studies*, **32**:233–40.

Diamond, P. (1965), ‘National debt in a neoclassical growth model’, *American Economic Review*, **55**:1126–50.

Frankel, M. (1962), ‘The production function in allocation and growth. a synthesis’, *American Economic Review*, **52**:995–1002.

Harrod, R.F. (1939), ‘An essay in dynamic theory’, *Economic Journal*, **49**:14–33.

Harrod, R.F. (1942), *Towards a Dynamic Economics*, London: Macmillan.

Kahn, R.F. (1959), ‘Exercises in the analysis of growth’, *Oxford Economic Papers*, **11**:143–56.

Kaldor, N. (1956), 'Alternative theories of distribution', *Review of Economic Studies*, **23**:83–100.

Kaldor, N. (1957), 'A model of economic growth', *Economic Journal*, **57**:591–624.

Kaldor, N. 1989), 'The irrelevance of equilibrium economics', in F. Targetti and A.P. Thirlwall (eds), *The Essential Kaldor*, London: Duckworth, pp. 373–98.

Kalecki, M. (1971), *Selected Essays on the Dynamics of the Capitalist Economy*, Cambridge, UK: Cambridge University Press.

Keynes, J.M. (1937), 'The General Theory: fundamental concepts and ideas', *Quarterly Journal of Economics*, **51**:1–15.

Koopmans, T.C. (1965), 'On the concept of optimal economic growth', in *The Econometric Approach to Development Planning*, Amsterdam: North Holland (for Pontificia Academy).

Lucas, R. (1988), 'On the mechanics of economic development', *Journal of Monetary Economics*, **22**:3–42.

Mankiw, N.G., D. Romer and D.N. Weil (1992), 'A contribution to the empirics of economic growth', *Quarterly Journal of Economics*, **107**:407–37.

Marglin, S.A. and A. Bhaduri (1990), 'Profit squeeze and Keynesian theory', in S.A. Marglin and J. Schor (eds), *The Golden Age of Capitalism*, Oxford: Clarendon Press.

Marx, K. (1867) [1977], *Capital*, Volume 1, New York: International Publishers.

Marx, K (1891), *Wage Labour and Capital*, Moscow: Progress Publishers.

Pasinetti, L. (1962), 'The rate of profit and income distribution in relation to the rate of economic growth', *Review of Economic Studies*, **29**:267–79.

Pasinetti, L. (2000), 'Critique of the neoclassical theory of growth and distribution', *Banca Nazionale del Lavoro Quarterly Review*, **53**:383–432.

Petty, W. (1662) [1963], 'A treatise on taxes and contributions', in C.H. Feinstein (ed.), *The Economic Writings of Sir William Petty*, Cambridge: Cambridge University Press, 1899 reprinted by Augustins M. Kelley, New York.

Pitelis, C. (1997), 'On Kaldor and pensions', *Cambridge Journal of Economics*, **21**:469–82.

Ramsey, F. (1928), 'A mathematical theory of saving', *Economic Journal*, **38**:543–99.

Ricardo, D. (1817) [1951], 'Principles of political economy and taxation', in P. Sraffa (ed.), *Works and Correspondence of David Ricardo*, Vol. I, Cambridge UK: Cambridge University Press.

Robinson, J. (1956), *The Accumulation of Capital*, London: Macmillan.

Romer, D. (1996), *Advanced Macroeconomics*, New York: McGraw-Hill.

Romer P. (1986), 'Increasing returns and long run growth', *Journal of Political Economy*, **94**:1002–37.

Samuelson, P.A. (1958), 'An exact consumption-loan model of interest, with or without the social contrivance of money', *Journal of Political Economy*, **66**:467–82.

Samuelson, P.A. (1966), 'A summing up', *Quarterly Journal of Economics*, **80**:568–83.

Schumpeter, J.A. (1961), *Theory of Economic Development*, New York: Oxford University Press.

Shaikh, A. (1980), ‘Laws of production and laws of algebra’, in E.J. Nell (ed.), *Growth, Profits and Property*, Cambridge, UK: Cambridge University Press, pp. 80–96.
Smith, A. (1776) [1976], *An Inquiry into the Nature and Causes of the Wealth of Nations*, in the Glasgow edition of the works and correspondence of Adam Smith (1976) in two volumes, R.H. Cambell, A.S. Skinner and W.B. Todd (eds), Oxford: Oxford University Press.
Solow, R. (1956), ‘A contribution to the theory of economic growth’, *Quarterly Journal of Economics*, **70**:65–94.
Solow, R. (1957), ‘Technical change and the aggregate production function’, *Review of Economics and Statistics*, **39**:312–20.
Sraffa, P. (1960), *Production of Commodities by Means of Commodities. Prelude to a Critique of Economic Theory*, Cambridge, UK: Cambridge University Press.
Steindl, J. (1952), *Maturity and Stagnation in American Capitalism*, 2nd edition, New York: Monthly Review Press.
Swan, T. (1956), ‘Economic growth and capital accumulation’, *Economic Record*, **32**:343–61.
Uzawa, H. (1961), ‘Neutral inventions and the stability of growth equilibrium’, *Review of Economic Studies*, **28**.
Young, A. (1928), ‘Increasing returns and economic progress’, *Economic Journal*, **38**:527–44.

3. The competition-of-capitals doctrine and the wage–profit relationship

Ferdinando Meacci*

3.1. INTRODUCTION

It is widely recognized that the development of Ricardo's theory of profit stems from Ricardo's 'dissatisfaction' with Smith's alternative theory running in terms of the 'competition of capitals'.[1] This theory is generally known as the 'competition-of-capitals doctrine'. Much research has been done in recent years both on Ricardo's 'struggle of escape' from this doctrine and on the consistency of the analytical results of this escape. The focus of attention, however, has been mostly centred on Ricardo's alternative theory. This was developed first in his *Essay on Profits* (1815), where it took the elementary form of a 'corn-ratio theory of profits',[2] and later on in the *Principles* (1821), where it took the more advanced form of the 'labour-embodied theory of profits'.[3]

The purpose of this chapter is to reconstruct Smith's (1776 [1976]) competition-of-capitals doctrine. This reconstruction, however, is not intended to provide a faithful assembly of what Smith actually wrote or a 'rational' view of what he must have thought in this connection. Rather, it is to extract from his faulty exposition and with the benefit of hindsight what is necessary to make Smith's doctrine consistent with his system of thought and vision of the future in order to determine whether, or to what extent, Ricardo's dissatisfaction is justified. This reconstruction will be based on the fragmentary statements by which the doctrine is presented in the *Wealth of Nations* and will try to highlight not only some of the ambiguities incorporated in these statements but also the links between these statements and other crucial parts of Smith's system of thought. These links, it will be argued, involve the wage–profit relationship.

This relationship is commonly traced to Ricardo's *Principles* where it is used in support of Ricardo's theory of distribution. But the wage–profit relationship was first recognized and is extensively used in the *Wealth of Nations*. The role it plays in Smith's work, however, differs greatly from that in Ricardo's. Not only is Smith's relationship put forward in the context of that competition-of-capitals doctrine from which Ricardo endeavoured to escape, but this doctrine is also coherent with that part of Smith's theory of value which was rejected by Ricardo. Smith's theory of value, however, is not the only framework in which the competition-of-capitals doctrine finds its proper place. Another framework is Smith's theory of capital. This theory was never rejected and, indeed, was instead defended by Ricardo on many controversial points.

Our reconstruction of the doctrine will run as follows. Section 3.2 presents the main fragments of the doctrine to be found in the *Wealth of Nations*. Section 3.3 examines the main ambiguities incorporated in these fragments. Section 3.4 locates the analytical foundations of the doctrine in Smith's theory of value and, particularly, in the principle of demand and supply in so far as it is part of this theory. These foundations are discussed in Sections 3.5 and 3.6 in the light of some clarifications introduced by Malthus and Senior. Section 3.7 shifts the focus of attention to the theory of capital and to the role played by the accumulation of capital in making the wage–profit relationship work in practice. This relationship, it will be argued in this section, is the link by which the competition-of-capitals doctrine interacts with Smith's theory of value as labour commanded, on the one hand, and with Smith's theory of capital as command of productive labour, on the other. Section 3.8 is concerned with the sustainability of this link as the accumulation of capital proceeds in time. Some conclusions are drawn in Section 3.9.

3.2. THREE FRAGMENTS OF THE DOCTRINE

Of the two theories which support Smith's competition-of-capitals doctrine one (the theory of value) eventually deals with the question concerning the commodity in which a variation in the 'exchangeable value' (from now onwards: e-value) originates while the other (the theory of capital) culminates in the analysis of the forces that account for a rise in the natural e-value of labour (natural wages). We will see below how these questions are instrumental to the conclusion of the doctrine. For now it is enough to note that Smith, while failing to argue how these questions relate to the doctrine, presents this doctrine in some passages so unconnected with, or so distant from, each other that it is either hard to take it as a doctrine or it is nearly impossible to discern the crucial role it plays in Smith's system of thought.

One of these fragments is found in Smith's chapter on wages and is focused on the 'natural collusion' of masters to prevent a rise of wages:

> When in any country the demand for those who live by wages, labourers, journeymen, servants of every kind, is continually increasing; when every year furnishes employment for a greater number than had been employed the year before, the workmen have no occasion to combine in order to raise their wages. The scarcity of hands occasions a competition among masters, who bid against one another, in order to get work and thus voluntarily break through the natural combination of masters not to raise wages. (*Wealth*, Book I, Chapter VIII, p. 86)

Another fragment is found right at the beginning of Smith's chapter on profits and focuses on the 'increasing or declining state of the wealth of the society':

> The rise and fall in the profits of stock depend upon the same causes with the rise and fall in the wages of labour, the increasing or declining state of the wealth of the society; but those causes affect the one and the other very differently.
> The increase of stock, which raises wages, tends to lower profit. When the stocks of many rich merchants are turned into the same trade, their mutual competition naturally tends to lower its profit; and when there is a like increase of stock in all the different trades carried on in the same society, the same competition must produce the same effect in them all. (*Wealth*, Book I, Chapter IX, p. 105)

Finally, a third fragment qualifies the fall of profits in particular sectors ('into the same trade') and in the economy as a whole ('in any country') by focusing on what happens 'at both ends' of the subtraction by which profits are calculated:

> As capitals increase in any country, the profits which can be made by employing them necessarily diminish. It becomes gradually more and more difficult to find within the country a profitable method of employing any new capital. There arises in consequence a competition between different capitals, the owner of one endeavouring to get possession of that employment which is occupied by another. But upon most occasions he can hope to justle that other out of this employment, by no other means but by dealing upon more reasonable terms. He must not only sell what he deals in somewhat cheaper, but in order to get it to sell, he must sometimes too buy it dearer. The demand for productive labour, by the increase of the funds which are destined for maintaining it, grows every day greater and greater. Labourers easily find employment, but the owners of capitals find it difficult to get labourers to employ. Their competition raises the wages of labour, and sinks the profits of stock. (*Wealth*, Book II, Chapter IV, pp. 352–3)

3.3. SOME INITIAL CLARIFICATIONS

The passages quoted above are worded in such a manner that some clarifications are needed before moving on to the systematic structure of the doctrine.

First, the terms 'profits' and 'wages' are used by Smith in the twofold sense of classical economics, that is both as the *amounts* of profits and wages and as their *rates*. It is understood that, when it comes to Smith's wage–profit relationship as implied in the passages above, what is liable to change in the opposite direction is neither the *amounts* nor the *shares* of profits and wages. Rather, as will be argued below with regard to Ricardo's different version of the same relation, it is the *rates* of profits and wages or, to put it in Cannan's terms (1917), *profits per cent* and *wages per head*. The terms 'profits' and 'wages' will be used throughout this chapter in this sense.

Second, the three fragments make it clear that the rates of profits and wages that are liable to change in the opposite direction are *natural* rates. They are, that is, the 'ordinary or average' rates that prevail in a particular time and place and that are in turn liable to change with the 'increasing or declining state of the wealth of the society'. They are not, therefore, the *market* rates which oscillate around their natural levels once the 'state of the wealth of society' is given. This qualification is adopted throughout this chapter and will be further developed in the sections to come.

Third, the difference between market and natural rates of profits and, therefore, between a fall (or rise) of market rates towards their natural level and a fall (or rise) of the natural level itself is missing in Smith's treatment of the wage–profit relationship. Nonetheless, this difference is related to the other difference, which Smith does discuss though not as thoroughly as he should, between the competition of capitals within particular sectors ('into the same trade') and within the economy as a whole ('in any country' or 'within the country').

Fourth, the two differences just indicated relate to each other in the sense that changes in market rates of profit are usually confined to particular sectors while changes in natural rates are usually common to all sectors. Granted the condition of 'perfect liberty' and the 'whole of the advantages and disadvantages of the different employments of labour and stock' (*Wealth*, Book I, Chapter X), this implies that at any moment there is a single 'ordinary or average' rate of profit in the economy as a whole. This was to be called the 'uniform' or 'general' rate of profit.[4]

Some further observations, however, are needed with regard to the third clarification. Leaving aside monopoly profits (a special form of market profits), the markets to be affected by the two forms of competition mentioned in that clarification are the market for *labour*, when it comes to the whole economy, and the market for the *products* of labour, when it comes to particular sectors. On the other hand, the competition at issue is a competition between *buyers*, when it comes to the market for labour, and a competition between *sellers*, when it comes to the market for the products of labour. Thus profits fall for different *reasons* in the two cases: they fall, in the former case, because the price of the labour to be employed (in any sector) rises in terms of the wage-goods exchanged for it (in the economy as a whole) while they fall, in the latter case, because the prices of the commodities produced in some sectors fall in terms of the commodities produced in other sectors. Moreover, profits may diverge owing to the different *consequences* of competition in the two cases: these consequences are, in the case of the market for the products of labour, a fall in the profits earned from selling some products *and* an increase in the profits earned from selling the products given in exchange for them; by contrast, in the case of the market for labour, the rise in the price of labour being in terms of its products, competition 'must produce the same effect' in all sectors and must accordingly cut the uniform or general rate of profit in the economy as a whole (that is the profit earned by turning labour into any of its products).

Finally, it should be noted that, however different these forms of competition may be, their outcome is the same when it comes to the standpoint of an individual capitalist. For they equally bring about a reduction in the difference between the two extremes within which this individual is used to calculating *his* profit: the extreme of the e-values advanced (costs) and the extreme of the e-values returned (revenues). When it comes to the standpoint of society, however, the two extremes have a different relevance. For, labour being the only commodity that the 'friends of humanity' (Ricardo, 1821, p. 100) wish to see rising in price, the competition of capitals in which they (the economists) are most interested is the competition between *buyers* in the market for labour. This is the market where labour is exchanged in view of the production of any of its products.

3.4. THE COMPETITION-OF-CAPITALS DOCTRINE AND THE PRINCIPLE OF DEMAND AND SUPPLY

Whether the market focused upon is the market for labour or the market for the products of labour, any variation in the e-value of labour or of any of its products is determined in Smith's system by the principle of demand and supply. This principle is the key for linking the two questions that lie at the roots of the competition-of-capitals doctrine. As anticipated above, these are the question concerning the commodity in which the variation of e-value originates and the question concerning the forces that account for a rise in the e-value of labour (and consequently for a fall in the rate of profit). This key is developed here in two steps. One is tackled in this section and is focused on Smith's contributions; the other is dealt with in the section to follow and focuses on Malthus's and Senior's clarifications.

The phrase 'principle of demand and supply' is not part of Smith's language. But it was introduced by Malthus in order to unveil an essential part of Smith's reasoning. This part is centred on the distinction between the natural/market prices of commodities and the natural/market compensations of the persons who contribute to their production as outlined in Chapter VII, Book I, of the *Wealth of Nations*. In spite of Ricardo's saying that this chapter is 'very well written' (1821, p. 91) and of Schumpeter's assertion that this is 'the best piece of economic theory turned out by A. Smith' (1954, p. 189), it remains nonetheless one of the three chapters (V–VII) that Smith himself says are 'in some degree obscure' (*Wealth*, p. 46). One aspect of this obscurity is that the treatment of the natural/market price of commodities as *products* of labour is mixed up with the treatment of the natural/market price of *labour* as the special commodity owned by *labourers*.[5] The obscurity, however, diminishes if the three chapters in question are considered in conjunction with the four chapters that follow (chapters VIII–XI). Chapters VIII–XI deal with the forces that determine the 'natural rates' of wages, profits and rents (that is their 'ordinary or average' rates at a given place and time) and chapters V–VII with how these natural rates determine the 'natural prices' of the products of labour. The two groups of chapters provide two different applications of the principle of demand and supply: while one application is concerned with the determination of the prices of the products of labour, the other is concerned with the determination of the incomes (wages, profits and rents) received by the individuals (workers, capitalists, landlords) involved in the production of these products.

Concerning the first application of the principle of demand and supply. The principle of demand and supply lies behind Smith's notions of 'effectual demand' (the quantity of a commodity demanded by those who are willing to pay its natural price) and 'quantity brought to market' (the quantity supplied to satisfy this demand) to the extent that natural prices are determined, like market prices, according to this principle (Malthus, 1836 [1986], Book I, Chapter II, Section III). Natural prices, however, differ from market prices in that, if the price of a commodity is at its 'natural' level, the individuals who have contributed to its production desire to *reproduce* it in the following period. So natural prices are a 'centre of repose and continuance' not only in the static sense of equalizing quantity demanded and quantity supplied ('repose') but also, given the principle of self-interest that governs the exchange of commodities, in the dynamic sense of guaranteeing the reproduction of commodities in the course of time ('continuance'). Hence the importance of distinguishing not only variations in market prices (above or below their natural levels) from variations in natural prices but also the different consequences of the latter variations on the quantities to be 'brought to market' in the periods to come.

Concerning the second application of the principle of demand and supply. It should be noted that the phrase 'to bring to market' means, in the case of the products of labour, 'to reproduce' as well as 'to supply' whereas, in the case of the individuals who own the means necessary for their production (land, labour and capital), it signifies 'to supply' rather than 'to reproduce'. This is especially true for the owners of land (landlords) since land can indeed be supplied or re-supplied but cannot be produced or re-produced. And this also holds for the owners of labour (labourers) in so far as the laws of reproduction of *labourers* are different from the laws of reproduction of commodities as *products* of their labour. Finally, concerning the owners of capital (capitalists), it is true that the object of their property is re-produced (unlike land and labour) and re-supplied (like land and labour) according to the income (profit) earned by these individuals. But the size of this income is determined according to the inverse wage–profit relationship: this works in practice according to the rule, stated by De Quincey (1844, p. 205) and shared by Marx (1969–72, Part II, Chapter XV, §B4), that 'any change that can disturb the existing relations between wages and profits must originate in wages'. Hence, leaving rent aside, everything boils down to understanding what determines the natural e-value of labour (natural wage) in a particular period and in a particular country (*Wealth*, Book I, Chapter VIII); and to identifying the forces that increase this rate from period to period (in the same country) and from country to country (in the same period) (ibid., Book II, Chapter III).

3.5. CAUSES, MAGNITUDES AND VARIATIONS OF E-VALUES

Now let us move on to the link between the principle of demand and supply and the competition-of-capitals doctrine. This link can be brought to light by means of the distinctions between the *causes* and the *magnitudes* of e-values and between the *intrinsic* and *extrinsic* causes of variations in these magnitudes (Malthus, 1836 [1986], Book I, Chapter II; Senior, 1836 [1965], pp. 116–20).

Concerning the causes and magnitudes of e-values. This distinction casts new light on the difference between the question of 'why commodities have value' and the question of 'what determines the magnitude of this value'. According to Smith and Malthus, these questions deserve two diverging answers and a common clarification depending on whether they refer to the 'early and rude state of society' or to the capitalist state. These answers and clarification may be summarized as follows: 1) labour embodied is the *cause* of e-values in the early as well as in the capitalist state although it is not *sufficient* to determine their magnitudes in the latter state; 2) labour commanded is *necessary* for measuring these magnitudes in the capitalist state for profits must be *added* in determining these magnitudes in the latter state; 3) the principle of demand and supply is meant to explain not so much the *cause* of e-values but only the determination of their (natural or market) *magnitudes* in the early as well as in the capitalist state.

Concerning the intrinsic and extrinsic causes of variations in the magnitudes of e-values. This distinction relates to, and casts new light on, the question concerning the commodity in which the variations take place. While the *intrinsic* causes affect 'demand' (Senior) or the 'desire to possess' (Malthus) and 'supply' (Senior) or 'the difficulty to obtain possession' (Malthus) of a particular commodity, the *extrinsic* causes affect 'demand' or the 'desire to possess' and 'supply' or 'the difficulty to obtain possession' of any other commodity for which the former is exchanged.[6] However, the cause of e-values must be distinguished from the forces that determine the variations in their magnitudes if only because the former affects *both* commodities exchanged while the latter may affect only *one* of them.[7]

The importance of this distinction is best noticed if it is applied to labour as a commodity substantially different from any of its products. While, concerning the determination of the (magnitude of the) e-value of labour, the demand for it requires a corresponding supply of wage goods (demand being 'the will combined with the power to purchase' as argued in note 8 below), concerning the variations in the magnitude of this e-value one should first determine whether these variations stem from changes in the demand for labour (that is from the will) or from *autonomous* changes in the supply of wage-goods (that is from the power): changes of the first kind are the 'intrinsic' while those of the second are the 'extrinsic' causes of the variations. Thus extrinsic causes, such as a bumper crop or a fall in the coefficients of production of wage-goods, may indeed result in an increase in wages. But, lacking a rise in the will to purchase labour, such an increase can only be *temporary*. Hence the importance of the Malthus–Senior clarification of Smith's doctrine: if the 'friends of humanity' want to trace the origin of (permanent) increases in the (natural) e-value of labour, they must first distinguish between the intrinsic and the extrinsic causes of its variations and, once Smith's view of the accumulation of capital is accepted, they must accordingly regard it as the only intrinsic cause of these increases.

3.6. INTENSITY AND EXTENT OF THE DEMAND FOR LABOUR

The distinction between intrinsic and extrinsic causes is best understood if it is combined with the distinction between the *intensity* and the *extent* of demand; and in particular if the variations in the e-value of commodities (labour and the products of labour) are traced to the variations in the intensity, rather than in the extent, of the demand for them: while the intensity of demand reflects the sacrifice that buyers are willing to make (the price they are willing to pay) to procure the commodity, the extent of demand refers to the quantity purchased by the buyers who are able to pay the price for it (Malthus, 1836 [1986], Book I, Chapter II, Section II). Thus any rise in price is due to an increase in the intensity of demand, it being understood that this increase is always *in relation* to the state of supply and that its long-period impact on the price (at which the commodity is exchanged) and the quantity (which is exchanged at this price) is determined by the conditions of reproduction. If goods are *not reproducible* (that is if they are not commodities), the impact is a rise in price but not in quantity; if goods are reproducible *without limits*, the impact is a rise in quantity but not in price; if goods are reproducible *with some limits*, the impact is a rise both in price and in quantity. The result of these clarifications is that labour should be consistently understood by Smith as a commodity reproducible *with* some limits (*Wealth*, Book I, Chapter VIII, and *passim*) and should be accordingly contrasted with how it is actually understood by Ricardo, that is as a commodity reproducible *without* limits (*Principles*, Chapters XXI, XXXII and *passim*). Likewise, a long-period increase in the demand for labour should be consistently understood by Smith as an increase in its *intensity* while it is actually understood by Ricardo as an increase in its *extent*.[8]

3.7. POSITIVE PROFIT, RELATIVE PROFIT AND THE WAGE–PROFIT RELATIONSHIP

We have examined above the links between the competition-of-capitals doctrine and Smith's theory of value. Let us now turn to the links between this doctrine and Smith's theory of capital. This theory provides the foundations for a theory of *profit*, on the one hand, and for a theory of the *rate of profit*, on the other. Although the links between these two theories are left in the dark by Smith, they are implicit in his unconfessed use of James Steuart's distinction between 'positive' and 'relative' profit[9] (Meacci, 2003).

The point of departure for tracing these links is Smith's discussion of dwelling-houses as distinct from profitable buildings.[10] If we focus on the chapter that contains this discussion (*Wealth*, Book II Chapter I, *Of the Division of Stock*) rather than on Smith's chapter on profits, we are more likely to identify these links. For the chapter on the division of stock deals with the differences and similarities between capital from the point of view of an *individual* (that is that part of the 'stock which a man possesses' which is to yield a profit to this individual) and capital from the point of view of *society* (that is that part of the 'general stock of any country or society' which is to yield a profit to the whole society). As is well known, these similarities and differences are developed in this chapter along two lines: one is concerned with the two different 'ways' in which the capital of an individual may be employed; the other with the two 'portions' in which the capital of society is divided once that employment has been determined. The issue, however, as to how (relative) profit accrues to (the capital of) an individual and how (positive) profit accrues to (the capital of) society is never tackled explicitly by Smith. He comes closest to this issue when he contrasts the manufacturer's 'consideration of his own *private* profit' with the fact that 'the different quantities of productive labour which it may put into motion and the different values which it may *add* to the annual produce of the land and labour of the society … never enter into his thoughts' (*Wealth*, Book II, Chapter V, p. 374, italics added). What is here called 'private profit' is Steuart's 'relative profit' while the increase in the 'annual produce', which is clumsily identified by Smith as an addition of 'different values', is Steuart's 'positive profit' and coincides with what is otherwise called 'surplus', 'surplus produce' or 'social surplus'. Leaving aside the obscurities incorporated in Smith's notion of 'annual produce',[11] it can be concluded that the profit of the wage–profit relationship in Smith's sense is 'private' or 'relative' profit while the increase in the 'annual produce' from which the opposite variations of wages and (relative) profits are drawn is the (positive) profit accruing to the whole society from the employment of capital and (productive) labour.

There is more, however, to the connection between these two forms of profit and the wage–profit relationship. For not only is the expectation of relative profit necessary for the realization of positive profit, but the former is also a partial or total appropriation of the latter. Furthermore, one thing is the *fact*, another is the *extent*, of this appropriation: while it is the task of the theory of profit to explain this fact, it is the task of the theory of the rate of profit to explain its extent. The latter theory was developed by the classics in two directions: one results in Ricardo's rent–profit relationship, the other in Smith's and Ricardo's wage–profit relationship. It should be noted, however, that this relationship represents two different sets of variations and assumes two different meanings depending on whether the variation in wages is intended in Smith's sense (that is as a variation in the *quantity* of 'necessaries, conveniences and amusements' given in exchange for labour) or in Ricardo's (that is as a variation in the *proportion* of the value of total product appropriated by labour). But whether the variation in wages be understood in one sense or in the other, both views regard this variation as temporary unless it results from a continuous process of accumulation. In Smith's view, however, the accumulation of capital is not only the main source of positive profit; it is also, via the principle of demand and supply, the intrinsic cause of variations in the (natural) e-value of labour (natural wages).

3.8. ACCUMULATION, TECHNICAL PROGRESS AND CAPITAL DEEPENING

After tackling the two main issues of Smith's theory of value and theory of capital (that is in which commodity the variation in e-value originates and which forces account for the variation in the natural e-value of labour), we turn to the problem that comes at the end of the latter theory. This problem concerns the *sustainability* of the process of accumulation in the presence of a continuous rise in the e-value of labour (and consequent fall in the rate of profit). This problem can be put in the following manner: how can the accumulation of capital continue if its outcome is the rise of (natural) wages and consequent fall of (natural) profits? Concerning the rise of wages, capitalists would be pleased if only they agreed with Smith's argument on the beneficial effects of the 'liberal reward of labour' (*Wealth*, Book I, Chapter VIII, p. 91ff). But what about the resulting fall of profits?[12]

Smith does not raise this question either in his chapter on profits or anywhere in Book II and even less in the brief passage on the invisible hand in Book IV of the *Wealth*. Where he comes closest to an adequate answer is at the end of Chapter VIII of Book I when he regards the rise of wages as compatible with a decrease in the cost of labour per unit of output. Unfortunately, when he reaches this point, he has not yet developed the notion of the wage–profit relationship although he has already begun to argue (while finishing the argument that a rise of wages need not be harmful to society) that this rise is a consequence of the competition of capitals. So the answer Smith begins to give to that question is unconnected with the notion of the wage–profit relationship and, accordingly, with the competition-of-capitals doctrine. This answer focuses first on the ambiguous assertion that the rise in wages increases the ‘price of many commodities’ and soon after on the clear-cut recognition of a further effect of this rise:

> The same cause, however, which raises the wages of labour, the increase of stock, tends to increase its productive powers, and to make a smaller quantity of labour produce a greater quantity of work … There are many commodities, therefore, which, in consequence of these improvements, come to be produced by so much less labour than before, that the increase of its price is more than compensated by the diminution of its quantity. (*Wealth*, Book I, Chapter VIII, p. 104)

This clear-cut recognition helps to solve the issue of the sustainability of the process of accumulation raised above. This issue can be addressed by noting that the accumulation of capital (‘the increase of stock’) may be intended in two senses and brings about two consequences.

Concerning the two senses. The accumulation of capital may be intended as an increase in *free* capital and/or as an increase in *invested* capital (Jevons, 1879): if intended in the first sense, it presents itself (immediately) as an increase in the demand for labour (that is as an increase in the competition between the buyers of labour); if intended in the second sense, it presents itself (with lags) either as an increase in output with constant coefficients (*capital widening*) or as an increase in labour productivity (*capital deepening*) with or without an upgrading of the products of labour (*product deepening*). This upgrading, it should be noted in passing, is what is needed for the e-value of labour in Smith’s sense to increase in the course of time. For the continuous rise of wages calls for an increase in the number and quality of the wage-goods produced by, and given in exchange for, labour. This increase is the outward form of increasing natural wages in Smith’s sense. This increase should be contrasted with an increase both in Ricardo’s ‘proportional wages’ and in the natural e-value of labour intended as the amount of labour embodied in wage-goods.

Concerning the two consequences. While an increase in free capital entails (the supply of labour remaining constant) an increase in wages in Smith’s sense, the resulting increase in invested capital, in so far as it results in an increase in labour productivity, leads to a decrease in the cost of labour per unit of output (wages remaining constant) (*Wealth*, Book I, Chapter VIII, p. 104–57, partly quoted above). The role of this process is to shift out the wage–profit frontier, that is, to bring the rate of profit back to the level from which it had fallen owing to the increasing demand for labour and the resulting increase in wages. Hence Smith’s notion of the ‘progressive state’ as ‘the cheerful and the hearty state to all the different orders of the society’; that is, not only to labourers but also, in spite of the wage–profit trade-off, to their counterparts in the market for labour. Hence the importance, for the sustainability of the process of accumulation, that inventions be made and new techniques and products be periodically introduced. This phenomenon is inevitable as the accumulation of capital advances and is indeed the result of the associated advance in the *vertical* division of labour.[13]

3.9. CONCLUDING REMARKS: THE LAW OF INCREASING WAGES

If cleared of the ambiguity by which the variations in the e-value of commodities in terms of each other are mixed with the variations in the e-value of labour in terms of wage-goods, and if strengthened by the distinction between intrinsic and extrinsic causes of variations in the magnitudes of e-values, the competition-of-capitals doctrine presents itself as a link between Smith’s theories of value and capital, on the one hand, and his views of the wage–profit relationship in relation to the sustainability of the process of accumulation (the main source of technical progress and further advances in the division of labour: *Wealth*, Book II and, particularly, Chapter III, pp. 343–32), on the other. Smith’s explicit and implicit argument may be developed and summarized by the following sequence of connections: accumulation of free capital (increasing funds for the maintenance of productive labour) → increasing demand for labour as an intrinsic cause of increases in its e-value → competition between capitalists in the market for labour → increasing natural wages in Smith’s sense→ decreasing relative profits across sectors→ labour-saving technical progress → increasing relative profits at constant wages in Smith’s sense → resumption of the process of accumulation (free capital → invested capital → technical progress → division of labour).

This argument underlies Smith’s most general but poorly highlighted conclusion: if capital continues to be accumulated in conditions of ‘perfect liberty’, the demand for labour (the competition between the *buyers* of labour) is destined to surpass the supply (the competition between the *sellers* of labour) – however stimulated the latter may be by the former – so that the (natural) e-value of labour in Smith’s sense grows over time through consecutive appropriations of positive profit. This e-value goes up more easily in an economy (such as ‘our American colonies’ in Smith’s time) in which accumulation is intense (so that the supply of labour does increase but not as much as the demand for it) than in an economy (such as China in Smith’s time) where, the accumulation of capital being weak or non-existent, the supply of labour tends to exceed the annual demand. All this is reflected in Smith's famous aphorism that it is not the actual greatness of the wealth of a country but its continual increase 'which occasions a rise in the wages of labour'. This rise, it should be noted, is made effective by a multiplication in the number, and an improvement in the quality, of the wage-goods produced by, and given in exchange for, labour. This thesis brings the competition-of-capitals doctrine to a close and might be called the 'law of increasing wages'. This is prepared in Book I, is brought to conclusion in Book II, and permeates the whole system of thought of the *Wealth of Nations*.

NOTES

* I thank the participants at the Conference and two anonymous referees for their helpful criticisms and suggestions.

1. See for instance Hollander (1973a, 1983), Eatwell (1975), Garegnani (1982) and Peach (1993, Chapters 2 and 3).
2. The 'rational foundation' of this theory is that 'in agriculture the same commodity, namely corn, forms both the capital (conceived as composed of the subsistence necessary for workers) and the product; so that the determination of profit by the difference between total product and capital advanced, and also the determination of the ratio of this profit to the capital, is done directly between quantities of corn without any question of valuation' (Sraffa, 1951, p. xxxi).
3. The 'rational foundation' (to use Sraffa's expression again) of this later theory is that 'the rate of profits was no longer determined by the ratio of the corn produced to the corn used up in production, but, instead, by the ratio of the total labour of the country to the labour required to produce the necessaries for that labour' (Sraffa, 1951, p. xxxii).
4. The central argument of Chapter X, *Of Wages and Profit in the Different Employments of Labour and Stock*, Book I of the *Wealth* seems to be that 'in a society where things were left to follow their natural course' the differences between the 'ordinary or average' rates of wages and profits across sectors are not only compatible with the idea of a single natural rate in the economy as a whole; they also have nothing to do with *changes* in natural rates, let alone with *differences* between market rates and natural rates in particular sectors. As for a society where things are *not* left 'to follow their natural course', see the part of the same chapter devoted to the 'policy of Europe' and the equally thoughtful Chapter VII, *Of Colonies*, Book IV of the *Wealth*.
5. The main obscurity, however, lies in the argument of this chapter known today as the 'adding-up theorem'. This is no place to go into the problems of analysis and interpretation raised by this argument if only because we are here concerned with the (natural) *rates* of wages and profits rather than with the (natural) *prices* of the commodities produced by means of capital and (productive) labour.
6. 'The causes which affect the desire to possess, and the difficulty of obtaining possession of, any one commodity may with propriety be denominated the *intrinsic* causes of its power of purchasing; because the more these causes increase, the greater power will the commodity possess of purchasing all those objects which continue to be obtained with the same facility. The causes which affect the desire to possess, and the difficulty of obtaining possession of, all the different commodities with which the first commodity might be exchanged may with

propriety be denominated the *extrinsic* causes of its power of purchasing' (Malthus, 1836 [1986, p. 48]).

7. This is explained by Bailey as follows: 'The value of *A* and *B* is the effect of causes acting on *both*, but a change in their mutual value may arise from causes acting on *either*: as the distance of two objects is to be referred to the circumstances which have fixed both of them in their particular situation, while an alteration of the distance between them might originate in circumstances acting on one alone' (Bailey, 1825 [1931, p. 184]; italics added).
8. The lag between changes in demand and changes in supply and the different consequences that this lag exerts on the supply of different commodities have been examined by Ricardo in connection with his criticisms of the principle of demand and supply (1821, Chapters XIII and XXX). Here it is impossible to go into these criticisms and the related disputes between Ricardo and Malthus. But it must at least be noted that the principle of demand and supply underlying the competition-of-capitals doctrine has nothing to do (*contra* Hollander, 1973b) with the 'curves of demand and supply' of the neoclassical theory. The most that can be said when one 'has in mind' these curves is that Malthus's demand is nothing but 'total purchasing power directed towards a commodity' (O'Brien, 1975, p. 105) or, more briefly, the quantity demanded at a particular price (Garegnani, 1983, 2003, §§14–15), a change in the intensity of demand depending on changes in the relation between the quantity supplied and the quantity demanded at this very price. It should however be noted that the modern habit of collapsing the neoclassical theory into the so-called 'demand-and-supply approach' and of contrasting this with the 'surplus approach' is dangerous in that, by obfuscating the role played by demand and supply in classical theory, it prevents a better understanding of the theory being defended or challenged. It should indeed be noted that Malthus's principle of demand and supply belongs so fully to the classical theory that it was introduced and developed by Malthus in order 1) to reject Ricardo's *version* of the classical theory of *value* and, within this version, Ricardo's *doctrine* of the variations in the natural e-value of labour (and consequently in the natural rate of profit); and 2) to defend Smith's *different* version of this theory along with Smith's *different* doctrine of these variations. It should also be noted that Ricardo did share Malthus's notion of demand as 'the will combined with the power to purchase' (to the extent that 'the greater is the degree of this will and power with regard to any particular commodity, the greater or the more intense may be fairly said to be the demand for it') (see, for instance, Ricardo, 1820 in *Works*, II, pp. 38–9; see also *Works*, VII, pp. 56–8) even when he takes issue with him on the possibility of gluts (see, for instance, *Works*, VI, pp. 130–35).
9. 'Positive profit implies no loss to anybody; it results from an augmentation of labour, industry, or ingenuity, and has the effect of swelling or augmenting the public good. Relative profit is what implies a loss to somebody; it marks a vibration of the balance of wealth between parties, but implies no addition to the general stock' (J. Steuart, 1767 [1966], Book II, Chapter VIII).
10. While a dwelling house may yield a 'revenue or profit' to its proprietor (the tenant paying 'the rent out of some other revenue which he derives either from labour, or stock, or land' so that 'the revenue of the whole body of the people can never be in the smallest degree increased by it'), profitable buildings are to procure a 'revenue or profit' not only 'to their proprietor who lets them for rent' but also 'to the person who possesses them and pays that rent' (*Wealth*, Book II, Chapter I, p.281).
11. See, for instance, O'Donnell (1990, Chapter 3) and Vianello (1999).
12. The problem of the sustainability of the process of accumulation is addressed by Smith only implicitly and indirectly. It was brought to the fore at a later stage by authors such as Ricardo, J.S. Mill and Marx who made use of different terminology, developed diverging

arguments and eventually reached conclusions in disagreement either with Smith or with one another. See, to begin with, Malthus (1836 [1986], Book II, *On the Progress of Wealth*).

13. The notion of vertical division of labour comes from the Austrians but is implied in Smith's treatment of the accumulation of capital (*Wealth*, Book II) if not of the division of labour as such (*Wealth*, Book I, Chapter I). As for the ambiguity of the phrase 'accumulation of capital' (which makes it unclear whether the capital accumulated is *free* or *invested* and whether accumulation itself is of the *deepening* or the *widening* kind), it should be noted not only that the wage–profit relationship implies a constant productivity of labour (cf. Bailey, 1825 [1931], Chapter IV; McCulloch, 1864 [1965]; Marx, 1969–72, Part II, p. 187) but also that it is only through 'invested' capital and through the 'deepening' of capital that the productivity of labour normally rises (and the wage–profit frontier shifts out). This ambiguity, however, is justified by the fact that both forms of capital and both forms of accumulation are needed for (natural) wages to increase in time with undiminishing (natural) profits. This interlacing of forms and consequences is needed, along with Smith's principle of 'perfect liberty' and call for changes in (China's) 'laws and institutions', to understand how an economy can avoid that 'full complement of riches' where, besides being 'fully peopled' (*Wealth*, Book I, Chapter VIII, pp. 89–90), it is also 'fully stocked' (ibid., Chapter IX, pp. 111–12). This is also what lies behind Malthus's argument on the superior strength of the 'regulating' (Smith's) principle of profits over the 'limiting' (Ricardo's) principle (1836 [1986], Book I, Chapter V).

REFERENCES

Bailey, S. (1825) [1931], *A Critical Dissertation on the Nature, Measures, and Causes of Value*, London: London School of Economics Reprints.

Cannan, E. (1917) [1967], *A History of the Theories of Production and Distribution in English Political Economy from 1776 to 1848*, New York: Kelley Reprints of economic classics.

De Quincey, T. (1844), *The Logic of Political Economy*, Edinburgh: Blackwood.

Eatwell, J. (1975), 'The interpretation of Ricardo's Essay on Profits', *Economica*, **42**:182–7.

Garegnani, P. (1982), 'On Hollander's interpretation of Ricardo's early theory of profits', *Cambridge Journal of Economics*, **6**:65–77.

Garegnani, P. (1983), 'The classical theory of wages and the role of demand schedules in the determination of relative prices', *American Economic Review*, Papers and Proceedings, **73**:309–13.

Garegnani, P. (2003), 'Professor Blaug on understanding classical economics', Centro di Ricerche e Documentazione Piero Sraffa, Quaderno di Ricerca no. 3

Hollander, S. (1973a), 'Ricardo's analysis of the profit rate 1813–15', *Economica*, **40**:260–82.

Hollander, S. (1973b), *The Economics of Adam Smith*, London: Heinemann.

Hollander, S. (1983), 'Professor Garegnani's defence of Sraffa on the material rate of profit', *Cambridge Journal of Economics*, **7**:167–74.

Jevons, S. (1879), *The Theory of Political Economy*, London: Macmillan, 2nd edn.

Malthus, T.R. (1836) [1986], *Principles of Political Economy*, in E.A. Wrigley and D. Souden (eds), *The Works of Thomas Robert Malthus*, London: Pickering, 1986.

Marx, K. (1969–72), *Theories of Surplus-Value*, London: Lawrence Wishart.
Meacci, F. (2003), ‘Positive profit, relative profit and capital as command of productive labour’, Paper presented at the Annual Meeting of the European Society for the History of Economic Thought, Paris, January 30–February 2, 2003, mimeo.
McCulloch, J.R. (1864) [1965], *The Principles of Political Economy*, New York: Kelley Reprints of economic classics.
O’Brien, D.P. (1975), *The Classical Economists*, Oxford: Clarendon Press
O’Donnell, R. (1990), *Adam Smith’s Theory of Value and Distribution. A Reappraisal*, London: Macmillan.
Ricardo, D. (1820) [1951], *Notes on Malthus's Principles of Political Economy*, in P. Sraffa (ed.), (1951–73), Vol. II.
Ricardo, D. (1821), *On the Principles of Political Economy and Taxation*, in P. Sraffa (ed.), (1951–73), Vol. I.
Peach, T. (1993), *Interpreting Ricardo*, Cambridge: Cambridge University Press.
Schumpeter, J. (1954), *History of Economic Analysis*, London: Allen & Unwin.
Senior, N.W. (1836) [1965], *An Outline of the Science of Political Economy*, New York: Kelley Reprints of economic classics.
Smith, A. (1776) [1976], *An Inquiry into the Nature and Causes of the Wealth of Nations*, Campbell and Skinner edition, Oxford: Clarendon Press.
Sraffa, P. (1951), ‘Introduction’, in Sraffa (1951–73), Vol. I, pp. XIII–LXII.
Sraffa, P. (ed.) (1951–73), *The Works and Correspondence of David Ricardo*, Cambridge: Cambridge University Press, 11 Vols.
Steuart J. (1767) [1966], *An Inquiry into the Principles of Political Economy*, London: Oliver & Boyd.
Vianello, F. (1999), ‘Social Accounting with Adam Smith’, in G. Mongiovi and F. Petri (eds), *Value, Distribution and Capital. Essays in Honour of Pierangelo Garegnani*, London: Routledge, pp. 165–80.

4. A heterodox growth and distribution model

Duncan K. Foley and Lance Taylor

4.1. INTRODUCTION

A major intellectual fault line in contemporary economics separates the 'orthodox' representative–agent rational-expectations based school of mainstream macroeconomics from the broad range of Keynesian, post-Keynesian, structuralist and Marxist models of growth and distribution, which we will refer to as 'heterodox'. Our aim in this chapter is to describe a synthetic, canonical heterodox macroeconomic model with two aims. The first is to establish a benchmark for a methodological discussion of the orthodox and heterodox approaches. The second is to emphasize that the diverse heterodox approaches share a common core of modelling presumptions, a fact sometimes lost sight of in the vigorous debate among the heterodox school over specific modelling strategies.

In our view the core insights that unify heterodox perspectives are: a focus on the functional distribution of income (the division of national income between wages and profits); the avoidance of model closures that imply full employment of a given labour force; differential modelling of the consumption and savings decisions of workers and capitalists; the adoption of an investment demand function independent of savings decisions; and a separate treatment of the firm as an economic agent independent of its owner households. These insights contrast sharply with the insistence of the orthodox approach on attained equilibrium models with full employment of labour, continuously fulfilled expectations, and a representative household, which imply a savings-constrained growth process.

The model we study here is eclectic in that it has features taken from a number of heterodox contributions, including notably the work of Michal Kalecki, Nicholas Kaldor, Joan Robinson, Donald Harris, Stephen Marglin and Amit Bhaduri, and Gérard Duménil and Dominique Lévy. We draw freely on our own earlier work, particularly Lance Taylor (2004) and Duncan Foley and Thomas Michl (1999). In the exposition we will call attention to the key points of disagreement among the heterodox schools as well as the important common elements. Some of the key innovations of this model are intended to shed light on macroeconomic issues that have become more important in recent years, particularly the interplay between the financial markets and the real economy, the impact of government borrowing, and the role of international capital movements in influencing macroeconomic outcomes. The model is also designed to distinguish variables such as wages that vary over the business cycle from variables such as capitalist consumption, which are determined by long-run considerations.

4.2. THE MODEL

Our model studies a six-sector, four-asset, two-class, one-commodity open capitalist economy.

The sectors are: firms, worker households, capitalist households, government, financial institutions including the central bank, and the rest of the world. Variables representing claims on the rest of the world are indicated by a bar. The assets are physical capital, *K*, domestic short-term debt, *B*, domestic equity, *Q*, and foreign assets, *F*. The sector issuing an asset is indicated by subscripts, and the sector holding the asset by superscripts: *f*, *w*, *c*, *g* and *b* for the domestic sectors. Holdings of assets by sectors are measured in net terms, and thus allow for negative values when appropriate. Thus, for example, B_f is the debt issued by the firm sector, and B^c is the debt held by capitalist households.

Firms produce output *Y*, measured as real Gross Domestic Product, which is the numéraire, using a single capital good *K*, interchangeable with output, which depreciates at the rate δ, and labour *N*.[1] The ratio of output to the accumulated real capital stock is capacity utilization, $u = Y/K$. The (real) wage is *w*, so that in any period the wage bill is $W = wN$, and the wage share is $\omega = W/Y$. Before-tax profits are $P = Y - W - \delta K$. The government taxes the value of output at the rate t_i, wage income at the rate t_w, property income at the rate t_c, and firm profits at the rate t_f. The gross (before-tax) domestic profit rate is $r = P/K = (1-\omega)u - \delta$, while the after-tax net profit rate is $\tilde{r} = (1-t_f)[u(1-t_i-\omega)-\delta]$.

The domestic price level is *p*. The value of the world money in terms of domestic money is the exchange rate *e*.[2] For simplicity we assume the foreign price of real output in foreign currency is 1, so that the real terms of trade are $\bar{e} = e/p$. The interest rate on domestic debt is *i*. The rate of return to foreign assets, assumed to be a generic balanced portfolio of securities, is $\bar{r}$.

In the text that follows, we set out the model's behavioural relationships and accounting in analytical terms. As an aid to understanding, we also present the flow accounting in Table 4.1 in the form of a social accounting matrix or SAM. The matrix incorporates a few conventions which make it straightforward to read. Corresponding rows and columns should have equal sums. The first row gives the demand breakdown of GDP into private and public consumption, net exports, and investment. The first column gives its decomposition in terms of market prices into wages, profits and indirect taxes. The upper rows labelled '*w*' through '*r*' give sources of income for worker households, capitalist households, firms, government, the financial sector, and the rest of the world, that is factor payments, dividends, interest incomes, taxes (for the government) and payments to nationals from the rest of the world.[3] The corresponding columns show uses of those incomes, basically for current spending on output, interest payments in and out, taxes, and flows of savings.

Table 4.1 Social accounting matrix of the model

Sector		w	c	f	g	b	r						Sum
		C^w	C^c		G		X	I					Y
w	W												Y^w
c													Y^c
f	rK												Y^f
g	T^i	T^w	T^c	T^f									Y^g
b				iB_f/p	iB_g/p								Y^b
r						$i\bar{B}/p$							$\bar{Y}$
w		S^w								$-\Delta B^w/p$	$-p_Q\Delta Q_f^w$		0
c			S^c							$-\Delta B^c/p$	$-p_Q\Delta Q_f^c$	$-\bar{e}\Delta\bar{F}^c$	0
f				S^f				$-I$	$\Delta B_f/p$		$p_Q\Delta Q$		0
g					S^g				$\Delta B_g/p$				0
b						S^b			$-\Delta B^b/p$	$\Delta B_b/p$			0
r							$\bar{S}$			$-\Delta\bar{B}/p$	$-p_Q\Delta\bar{Q}$	$\bar{e}\Delta\bar{F}^c$	0
Sum	Y	Y^w	Y^c	Y^f	Y^g	Y^b	$\bar{Y}$	0	0	0	0	0	

The second set of '*w*' through '*r*' rows summarize flows of funds for the different groups of actors. The accounting convention is that 'sources' of funds (saving and increases in liabilities) are given a positive sign and 'uses' (investment and increases in financial assets) carry a negative sign. The columns show how flow changes of assets balance out. Thus, investment I adds to aggregate demand in the first row and represents a use of funds for firms in row '*f*'. The columns further to the right show flow balances for domestic bonds and equity, and foreign equity.

As discussed below, the change in net worth for each group of actors is the sum of its savings from the SAM and capital gains on financial assets. The flows of funds in the SAM thereby cumulate smoothly into changes in balance sheets.

4.2.1. Firms

The firm sector holds real domestic capital, K, and issues equity, Q, and real net domestic debt B_f / p. The financial markets value firm equity at the real price p_Q explained below. The firm's balance sheet can be written:

$$J^f = K - \frac{B_f}{p} - p_Q Q \tag{4.1}$$

where J^f is the net worth of the firm sector, valued at market prices.[4]

The firm sector's net profit after interest payments is $\tilde{r}K - iB_f / p$, which we assume is used to pay dividends $D = \sigma K$, finance investment, or retire debt and equity. Firms' investment in new capital is I.

Firm sector saving, S^f, profit income less transfers, is equal to investment minus the change in its liabilities:

$$\begin{aligned} S^f &= \tilde{r}K - i\frac{B_f}{p} - \sigma K \\ &= I - \delta K - \frac{\Delta B_f}{p} - p_Q \Delta Q \end{aligned} \tag{4.2}$$

(Δ is the time difference operator. To reduce notation we denote current period variables without a time subscript, and next period variables with the subscript $_{+1}$. We write, for example, $\hat{K} = \Delta K / K$.)

The savings equality can be re-arranged to show the equality of the firm sector's sources and uses of funds:

$$\tilde{r}K + \frac{\Delta B_f}{p} + p_Q \Delta Q = I - \delta K + \sigma K + i\frac{B_f}{p} \tag{4.3}$$

The time-difference of the firm sector's net worth includes capital gains or losses due to changes in asset prices over the period:

$$\begin{aligned} \Delta J^f &= S^f - \Delta\left[\frac{1}{p}\right] B_{f+1} - \Delta p_Q Q_{+1} \\ &= \Delta K - \frac{\Delta B_f}{p} - p_Q \Delta Q - \Delta\left[\frac{1}{p}\right] B_{f+1} - \Delta p_Q Q_{+1} \end{aligned} \tag{4.4}$$

We assume that capital markets value the equity of the firm at a real price $p_Q Q = qK$ by capitalizing the current after-tax, after-interest profits at a discount rate, ρ, so that:

$$q = \frac{p_Q Q}{K} = \frac{\tilde{r} - i\frac{B_f}{pK}}{\rho} \tag{4.5}$$

Firm investment demand is:

$$I = \Delta K + \delta K = g^K[i,\ q]K \tag{4.6}$$

Investment demand is constrained by high domestic interest rates, $g_i^K < 0$, and stimulated by a high profit rate relative to the financial target rate of return, $g_q^K > 0$.

Equation (4.3) determines the firm sector's total issue of new liabilities given profits, interest payments, and investment. We assume that firms issue or retire equity in proportion to after-tax net profit:

$$p_Q \Delta Q = qK\frac{\Delta Q}{Q} = qK\hat{Q} = -\alpha^f \tilde{r}K \tag{4.7}$$

4.2.2. Worker Households

The worker household sector receives a share of total domestic wage income ϕW, and net foreign wage income, $\overline{eW}$, and saves for life-cycle reasons, holding domestic equity, $Q^w = \theta^w Q$, and domestic debt (issued by the financial sector), B^w, as assets, thus receiving interest and dividends as well. The worker household sector's balance sheet is:

$$J^w = p_Q Q^w + \frac{B^w}{p} = qK\theta^w + \frac{B^w}{p} \tag{4.8}$$

The actual number of outstanding shares of equity, Q, plays no real economic role and is indeterminate. We focus instead on the proportion of equity held by worker households, $\theta^w = Q^w / Q$. Taking time differences, we see that:

$$\frac{\Delta Q^w}{Q} = \Delta\theta^w(1+\hat{Q}) + \theta^w\hat{Q} \tag{4.9}$$

Worker household income is:

$$Y^w = \phi W + \overline{eW} + \frac{iB^w}{p} + \sigma\theta^w K \tag{4.10}$$

Worker income including wages, both foreign and domestic, interest and dividends are taxed at the rate t_w, so that worker household taxes are $T^w = t_w Y^w$.

Worker-households have a target ratio of wealth to after-tax wage income, $\nu = J^{w*}/(1-t_w)(\phi W + \bar{e}\bar{W})$. Workers are assumed to adjust their wealth–income ratio to their target level at the rate γ.[5] They also must allow for the change in the number of worker households due to growth in employment, and the change in the ratio of the absolute number of equity shares to capital.

Workers want to hold a fraction $\alpha^w = p_Q Q^w / J^w$ of their wealth as equity and $(1-\alpha^w)$ as domestic debt. Worker household target holding of equity is $p_Q Q^{w*} = \alpha^w \nu(1-t_w)(\phi W + \bar{e}\bar{W})$ and of debt is $B^{w*}/p = (1-\alpha^w)\nu(1-t_w)(\phi W + \bar{e}\bar{W})$. We assume that worker households take account of the growth of employment, which is proportional to the growth of the capital stock, in making this stock adjustment. In addition to whatever net acquisition of debt is required to adjust toward their desired wealth-wage income ratio, worker households also increase net debt by $(gK[i,\ q]-\delta)(B^w/p)$ and net equity by $qK\theta^w\hat{Q}$ to allow for growth.

Putting together these assumptions, we see that worker-household acquisition of equity (after purchases or sales of stock by the firm sector) and domestic debt in each period satisfy:

$$\begin{aligned} p_Q \Delta Q^w &= qK[\Delta\theta^w(1+\hat{Q}) + \theta^w\hat{Q}] = \\ &= \gamma\left[\alpha^w(1-tw)\nu K(\bar{e}\bar{W} + \phi u\omega[u]) - qK\theta^w\right] + qK\theta^w\hat{Q} \end{aligned} \tag{4.11}$$

$$\frac{\Delta B^w}{p} = \gamma\left[(1-\alpha^w)(1-tw)\nu K(\bar{e}\bar{W} + \phi u\omega[u]) - \frac{B^w}{p}\right] + (gK[i,\ q]-\delta)\frac{B^w}{p} \tag{4.12}$$

The consumption of the worker household sector is income less taxes and saving. Worker-household consumption is thus:

$$\begin{aligned} C^w &= (1-t_w)\left(\phi W + \bar{e}\bar{W} + i\frac{B^w}{p} + \sigma\theta^w K\right) + \\ &- \left[\gamma(J^{w*} - J^w) - (g^K[i,\ q]-\delta)\frac{B^w}{p}\right] - qK\theta^w\hat{Q} \end{aligned} \tag{4.13}$$

The savings of the worker household sector is equal to its net acquisition of debt and equity:

$$S^w = \frac{\Delta B^w}{p} + qK\left[\Delta\theta^w(1+\hat{Q}) + \theta^w\hat{Q}\right] \tag{4.14}$$

4.2.3. Capitalist Households

Capitalists hold their real wealth as domestic debt issued by the financial sector, B^c / p, domestic equity, $p_Q Q^c = qK\theta^c$ and foreign assets, $\bar{F}^c$. The capitalist household sector balance sheet is thus:

$$J^c = qK\theta^c + \bar{e}\bar{F}^c + \frac{B^c}{p} \tag{4.15}$$

Capitalist households receive a share $1-\phi$ of wage income. The savings of the capitalist household sector is income less taxes and consumption spending, and is equal to the change in assets after purchases and sales of stock by firms:

$$\begin{aligned} S^c &= (1-\phi)W + \overline{re}\bar{F}^c + i\frac{B^c}{p} + \sigma K\theta^c - T^c - C^c \\ &= \frac{\Delta B^c}{p} + p_Q\Delta Q^c + \bar{e}\Delta\bar{F}^c \\ &= \frac{\Delta B^c}{p} + qK\left[\Delta\theta^c(1+\hat{Q}) + q\theta^c\hat{Q}\right] + \bar{e}\Delta\bar{F}^c \end{aligned} \tag{4.16}$$

The savings equality can be re-arranged to show the equality of the capitalist household sector's sources and uses of funds:

$$(1-\phi)W + \overline{re}\bar{F}^c + i\frac{B^c}{p} + \theta^c\sigma K = C^c + p_Q\Delta Q^c + \bar{e}\Delta\bar{F}^c + \frac{1}{p}\Delta B^c \tag{4.17}$$

The time difference of the capitalist household sector's net worth includes capital gains or losses due to changes in asset prices over the period:

$$\Delta J^c = S^c + \Delta p_Q Q^c_{+1} + \Delta\bar{e}\bar{F}^c_{+1} + \Delta\left[\frac{1}{p}\right]B^c_{+1} \tag{4.18}$$

We assume that capital income, including interest and capital gains from sales of stock to the firm sector is taxed at the rate $t_{c,}$ so that capitalist-household sector taxes are $T^c = t_c[(1-\phi)W + \overline{re}\bar{F}^c + i(B^c / p) + \sigma\theta^c K]$.

Capitalists consume a proportion $1-\beta$ of their beginning-of-period after-tax wealth in each period, so that capitalist consumption is:

$$C^c = (1-\beta)J^c \tag{4.19}$$

Capitalist households face a portfolio decision in dividing their total wealth between domestic equity, domestic debt and foreign assets. We assume that equity and foreign assets are a proportion α^c of capitalist portfolios. (This proportion may be a function of the domestic interest rate and the rate of return on foreign assets.) The long-run proportion of domestic debt in capitalist portfolios is $1-\alpha^c$. The division of capitalist household wealth between domestic equity and foreign assets is determined by market clearing, given firms' issuance of domestic equity. To keep the proportion of debt in capitalist portfolios equal to $1-\alpha^c$, purchases of debt must be adjusted to allow for the differential rate of growth of capital and equity.

Putting together these assumptions we see that:

$$\begin{aligned}\frac{\Delta B^c}{p} &= (1-\alpha^c)(1-t_c)\left[(1-\phi)W + \overline{re}\overline{F}^c + i\frac{B^c}{p} + \sigma\theta^c K\right] + \\ &-(1-\alpha^c)\left[(1-\beta)\left(qK\theta^c + \overline{e}\overline{F}^c + \frac{B^c}{p}\right) - (\hat{K}-\hat{Q})qK\theta^c\right]\end{aligned} \tag{4.20}$$

$$\begin{aligned}&\frac{\Delta B^c}{p} + qK[\Delta\theta^c(1+\hat{Q})] + qK\theta^c\hat{Q} + \overline{e}\Delta\overline{F}^c = \\ &= \left\{(1-t_c)\left[(1-\phi)W + \overline{re}\overline{F}^c + i\frac{B^c}{p} + \sigma\theta^c K\right] - (1-\beta)\left(qK\theta^c + \overline{e}\overline{F}^c + \frac{B^c}{p}\right)\right\}\end{aligned} \tag{4.21}$$

4.2.4. Government

The government issues debt. The government net worth is thus:

$$J^g = -\frac{B_g}{p} \tag{4.22}$$

Government saving is the difference between tax income and expenditures plus interest on the outstanding government debt:

$$S^g = T^w + T^c + T^i + T^f - G - i\frac{B_g}{p} = -\frac{\Delta B_g}{p} \tag{4.23}$$

The savings equality can be re-arranged to show the equality of the government sector's sources and uses of funds:

$$T^w + T^c + T^i + T^f + \frac{\Delta B_g}{p} = G + i\frac{B_g}{p} \tag{4.24}$$

The time difference of the government sector net worth includes capital gains and losses due to changes in the price level over the period:

$$\Delta J^g = -\frac{\Delta B_g}{p} - \Delta\left[\frac{1}{p}\right] B_{g+1} \tag{4.25}$$

Putting these assumptions together, we see the law of evolution of the government debt:

$$\begin{aligned}\frac{\Delta B_g}{p} &= G + i\frac{B_g}{p} - t_w\left(\phi W + \overline{e}\overline{W} + i\frac{B^w}{p} + \sigma\theta^w K\right)\\ &-t_c\left[(1-\phi)W + \overline{re}\overline{F}^c + i\frac{B^c}{p} + \sigma\theta^c K\right] - t_f\tilde{r}K - t_i uK\end{aligned} \tag{4.26}$$

4.2.5. Financial Sector

In order to make the model as compatible as possible with available flow-of-funds data, we consolidate the central bank and all other banks and financial intermediaries into a financial sector. The financial sector holds the domestic debt of firms, B_f, and the government, B_g. It issues debt which is held by households, B^w and B^c, and the rest of the world, $\overline{B}_b$.

The reserve position of the central bank is included in net financial sector borrowing from the rest of the world, so that we will not need to model reserve policy separately from exchange rate and interest rate policy.

In order to take the domestic interest rate on debt as exogenous, at least in the short run, we assume that the central bank adjusts the composition of the supply of debt through open market operations in order to enforce the domestic interest rate i. Behind the scenes, as it were, the composition of the liabilities of the financial sector may also be changing as the interest rate changes (for example, between 'money' and 'bonds'). We avoid detailed modelling of the institutional structure of capital markets and financial intermediation in order to make the model applicable to as wide a range of economies as possible. The domestic interest rate enforceable by the central bank may be constrained by the premium the international bond market charges domestic borrowers over the rate of return to foreign assets $\bar{r}$.[6]

The financial sector's net worth is thus:

$$J^b = \frac{(B_f^b + B_g^b) - (B_b^w + B_b^c + \bar{B}_b)}{p} = \frac{B^b - B_b}{p} \tag{4.27}$$

Even if we assume on average that the interest rates on financial sector assets and liabilities are the same, the financial sector will have non-zero net income if assets and liabilities are not equal. We assume that the financial sector saves all of this income:[7]

$$S^b = i\frac{B^b - B_b}{p} = \frac{\Delta B^b - \Delta B_b}{p} \tag{4.28}$$

The time difference of the financial sector net worth includes capital gains and losses due to changes in the price level over the period:

$$\Delta J^b = \frac{\Delta B^b - \Delta B_b}{p} + \Delta\left[\frac{1}{p}\right](B_{+1}^b - B_{b+1}) \tag{4.29}$$

We treat the balance sheet of the financial sector as a residual, at least in the short run. The financial sector absorbs the debt issued by firms and government, and issues the domestic debt demanded by households, letting borrowing from the rest of the world adjust to make up the difference.

4.2.6. Rest of the World

The rest of the world's net worth, writing $\bar{\theta} = \bar{Q}/Q$ is:

$$\bar{J} = \frac{\bar{B}_b}{p} + qK\bar{\theta} - \bar{e}\bar{F}^c \tag{4.30}$$

The rest of the world has interest income from its lending to the financial sector, and dividends on domestic equity, while its spending is net exports from the domestic economy, foreign wages of working households, and interest and dividends on capitalist household foreign assets. Thus the saving of the rest of the world is the negative of the domestic current account in the balance of payments, and equal to the capital account surplus in the balance of payments:

$$\bar{S} = i\frac{\bar{B}_b}{p} - \bar{e}(\bar{W} + r\bar{F}^c) + \sigma\bar{\theta}K - X = \frac{\Delta\bar{B}_b}{p} + p_Q\Delta\bar{Q} - \bar{e}\Delta\bar{F}^c$$
$$= \frac{\Delta\bar{B}_b}{p} + qK[\Delta\bar{\theta}(1+\hat{Q}) + \bar{\theta}\hat{Q}] - \bar{e}\Delta\bar{F}^c \tag{4.31}$$

We assume that net exports, measured in domestic currency, as a fraction of the domestic capital stock are a function of the terms of trade and the level of capacity utilization, $x[\bar{e},\ u] = X/K$, with $x_{\bar{e}} > 0,\ x_u < 0$.

Thus there is a relation between the terms of trade and net capital outflow:

$$X = x[\bar{e},\ u]K =$$
$$= i\frac{\bar{B}_b}{p} + \sigma\bar{\theta}K - \bar{e}(\bar{W} - r\bar{F}^c) - \left(\frac{\Delta\bar{B}_b}{p} - p_Q\Delta\bar{Q} + \bar{e}\Delta\bar{F}^c\right) \tag{4.32}$$

The sectoral flows of funds described here are conveniently and transparently summarized in the Social Accounting Matrix (SAM) in Table 4.1.

4.2.7. Distribution

The heterodox tradition eschews the assumption of continuous clearing of the labour market, and substitutes a distribution schedule relating the wage share to the level of capacity utilization:

$$\omega = \omega[u] \tag{4.33}$$

with $\omega' > 0$.[8] Thus the higher is capacity utilization and the tighter the labour market (or in Marxian terms, the smaller reserve armies of labour) the higher will be wages and the wage share.

4.2.8. Aggregate Demand and Saving

Output can be expressed in terms of expenditure or domestic income:

$$Y = I + C^w + C^c + G + X = W + rK + T^i \tag{4.34}$$

where I is investment (abstracting from depreciation), C^w is consumption of worker households, C^c is consumption of capitalist households, G is government expenditure on goods and services, X is the value of net exports in domestic currency.

Domestic equity at the end of the period must be held by worker households, capitalist households, and the rest of the world, so that:

$$1 = \theta^w + \Delta\theta^w + \theta^c + \Delta\theta^c + \overline{\theta} + \Delta\overline{\theta} \tag{4.35}$$

We take $\overline{\theta}$ and $\Delta\overline{\theta}$ as parameters in each period. On a steady-state growth path, $\Delta\overline{\theta} = \Delta\theta^w = \Delta\theta^c = 0.$.

The domestic economy's aggregate net worth is thus:

$$J = J^f + J^w + J^c + J^g + J^b = K + \overline{e}\overline{F}^c - \frac{\overline{B}_b}{p} - qK\overline{\theta} = K - \overline{J} \tag{4.36}$$

Aggregate domestic saving is:

$$\begin{aligned} S &= S^f + S^w + S^c + S^g + S^b = \\ &= rK + W + \overline{e}\overline{W} + \overline{re}\overline{F}^c - \sigma\overline{\theta}K - i\frac{\overline{B}}{p} - C^w - C^c - G = \\ &= I + X - i\frac{\overline{B}}{p} - \sigma\overline{\theta}K + \overline{e}(\overline{r}\overline{F}^c + \overline{W}) + q\overline{\theta}\hat{Q} = \\ &= I - \overline{S} = I + \overline{e}\Delta\overline{F}^c - \frac{\Delta\overline{B}_b}{p} - qK\overline{\theta}\hat{Q} \end{aligned} \tag{4.37}$$

4.2.9. Aggregate Demand Equilibrium

We can divide equation (4.34) by the domestic capital stock K, to get an expression for the equilibrium level of capacity utilization, given the wage share, the terms of trade, and the real financial valuation of domestic equity. Here we write $\overline{w} = \overline{W}/K$, $b^w = B^w/pK$, $b^c = B^c/pK$, $\overline{b} = \overline{B}/pK$, $\overline{f} = \overline{F}/K$ and $z = G/K$:

$$\begin{aligned} u &= g^K[i,\, q] + z + x[\overline{e},\, u] + (1 - t_w)(\phi\omega u + \overline{e}\,\overline{w} + ib^w + \sigma\theta^w) + \\ &\quad - \left\{\gamma\left[\nu(1 - t_w)(\phi\omega u + \overline{e}\overline{w}) - (b^w + q\theta^w)\right] + (g^K[i,\, q] - \delta)b^w + q\theta^w\hat{Q}\right\} + \\ &\quad + (1 - \beta)(q\theta^c + b^c + \overline{e}\,\overline{f}) \end{aligned} \tag{4.38}$$

Differentiating with respect to u, q and ω (representing an autonomous shift of the distribution schedule), taking $\hat{Q} = 0$ and $t_w = 0$ to simplify, we see that:

$$du = \frac{u(1-\gamma\nu)\phi d\omega + \left[\gamma\theta^w + (1-\beta)\theta^c + (1-b^w)g_q^K\right]}{1 - x_u - (1-\gamma\nu)\left[\phi(\omega + \omega' u)\right]} \tag{4.39}$$

$$dq = -\frac{(1-t_f)\left[ud\omega + (1-t_i-\omega-u\omega')du\right]}{\rho} \tag{4.40}$$

The derivative $du/d\omega$ can be positive or negative depending on whether an autonomous rise in the wage share stimulates demand more by raising wages than it reduces demand by lowering the profit rate. In the first case the economy is wage-led, and in the second case profit-led.

4.2.10. The Complete Model

We can write the complete model as twelve equations in the eleven variables, u, q, Q, $\Delta\theta^w$, $\Delta\theta^c$, Δb_f, Δb_g, Δb^w, Δb^c, $\Delta\bar{b}$ and $\Delta\bar{f}$. One of the equations is redundant by the accounting (or budget) constraints. The model has as parameters the functions $g_K[.]$, $\omega[.]$, $x[.]$, the behavioural constants α^f, σ, α^w, γ, ν, $\bar{w}$, α^c, β, δ, $\bar{\theta}$ and $\Delta\bar{\theta}$, the policy variables t_w, t_c, t_i, t_f, z and i, and the price levels p and $\bar{e}$. In each period the state variables θ^w, θ^c, b_f, b_g, b^w, b^c, $\bar{b}$ and $\bar{f}$ are given by the history of the system.

For convenience we summarize the equations of the model here, in a form in which they can be solved hierarchically. To begin with, equations (4.38), (4.5) and (4.7) can be solved implicitly for u, q, and Q, since $\tilde{r} = (1-t_f)[(1-t_i-\omega[u])u-\delta]$ is a function of u and the parameters:

$$\begin{aligned} u = {} & g^K[i,q] + z + x[\bar{e},u] + (1-t_w)(1-\gamma\nu)(\phi\omega[u]u + \bar{e}\,\bar{w}) + \\ & + \left[(1-t_w)i + 1 + g^K[i,q] - \delta\right]b^w + \left[(1-t_w)\sigma + q(1+\hat{Q})\theta^w\right] \\ & + (1-\beta)(q\theta^c + b^c + \bar{e}\bar{f}) \end{aligned} \tag{4.41}$$

$$q = \frac{\tilde{r} - ib^f}{\rho} \tag{4.42}$$

$$q\hat{Q} = -\alpha^f\tilde{r} \tag{4.43}$$

Given the values of u, q and $\hat{Q}$ (which determine the after-tax net profit rate $\tilde{r}$ and $\hat{K} = g^K[i,\ q]-\delta$), and $\theta^c = 1-\theta^w-\bar{\theta}$, six of the remaining variables can be solved for directly:

$$\Delta b_f = \frac{(1-b_f)\hat{K} - \tilde{r} + \sigma + ib^f - q\hat{Q}}{1+\hat{K}} \tag{4.44}$$

$$\Delta\theta^w = \frac{\gamma\left[\alpha^w \nu(1-t_w)(\phi u\omega[u] + \overline{ew}) - q\theta^w\right]}{q(1+\hat{Q})} \tag{4.45}$$

$$\Delta b^w = \frac{\gamma\left[(1-\alpha^w)\nu(1-t_w)(\phi u\omega[u] + \overline{ew}) - b^w\right]}{1+\hat{K}} \tag{4.46}$$

$$\Delta\theta^c = -(\Delta\theta^w + \overline{\theta}) \tag{4.47}$$

$$\begin{aligned}(1+\hat{K})\Delta b^c &= (1-\alpha^c)(1-t_c)\left[(1-\phi)u\omega[u] + ib^c + \overline{r\overline{ef}} + \sigma\theta^c\right] + \\ &-(1-\alpha^c)\left[(1-\beta)(b^c + q\theta^c + \overline{\overline{e}f}) - (\hat{K}-\hat{Q})q\theta^c\right] - \hat{K}b^c\end{aligned} \tag{4.48}$$

$$\begin{aligned}(1+\hat{K})\overline{e}\Delta\overline{f} &= (1-t_c)\left[(1-\phi)u\omega[u] + ib^c + \overline{r\overline{ef}} + \sigma\theta^c\right] + \\ &-(1-t_c)\left[(1-\beta)(b^c + q\theta^c + \overline{\overline{e}f}) + \hat{K}(b^c + \overline{\overline{e}f}) + q\hat{Q}\theta^c\right] + \\ &-(1-t_c)\left\{(1+\hat{K})\Delta b^c\left[1+\hat{K} + q\Delta\theta^c(1+\hat{Q})\right]\right\}\end{aligned} \tag{4.49}$$

The dynamic equations for the other two state variables, $b_g = B_g / pK$, and $\overline{b} = \overline{B} / pK$, are:

$$\begin{aligned}(1+\hat{K})\Delta b_g &= (i-\hat{K})b^g - t_w(\phi u\omega[u] + \overline{ew} + ib^w + \sigma\theta^w) + \\ &-t_c\left[(1-\phi)u\omega[u] + ib^c + \overline{r\overline{f}} + \sigma\theta^c\right] - \frac{t_f}{1-t_f}\tilde{r} - t_i u\end{aligned} \tag{4.50}$$

$$\begin{aligned}(1+\hat{K})\Delta\overline{b} &= (i-\hat{K})(b_f + b_g - b^w - b^c) + \\ &+(\Delta b_f + \Delta b_g - \Delta b^w - \Delta b^c)(1+\hat{K}) - \hat{K}\overline{b}\end{aligned} \tag{4.51}$$

The balance of payments equation (4.31) then follows as an identity by the accounting constraints.

4.3. SHORT-RUN COMPARATIVE STATICS

The short-run equilibrium of the model can be conceptualized as the intersection of two loci in (u, q) space, the first showing the (u, q) pairs that satisfy the aggregate demand equation (4.41), and the second showing the (u, q) pairs that satisfy the asset price equation (4.42). These are both upward sloping near the equilibrium. To assure short-run stability, the first locus must cross the second from below, as illustrated in Figure 4.1.

The asset price curve depends on the tax parameters t_f, t_i, the depreciation rate δ, and i, $\overline{r}$ and B_f / pK. An increase in any of these shifts the asset price curve downward, leading to a lower short-run equilibrium u and q.

The aggregate demand curve depends on the parameters and state variables. In general any upward shift in the components of demand shifts the aggregate demand curve outward, leading to higher short-run equilibrium u and q.

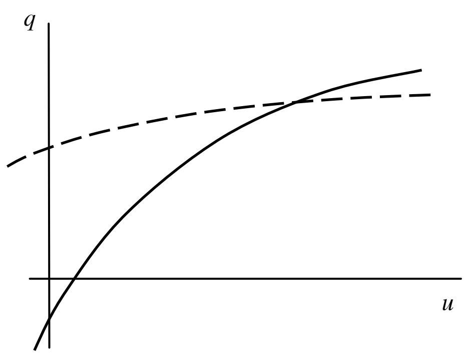


Figure 4.1 The solid line represents the solutions of equation (4.41), and the dashed line the solutions of equation (4.42). The intersection is the short-run equilibrium of the economy

To explore comparative statics of the core model, it makes sense to work with simplified versions of the foregoing equations. Expressed in terms of excess demand for output, a compact version of (4.38) becomes

$$g^K[i,\ q] + x[\overline{e},\ u] + m_w \phi \omega[u]u + \zeta[q] - u = 0 \tag{4.52}$$

in which m_w is the marginal propensity to consume of worker households (boiled down from the saving and tax parameters in (4.38), with foreign wage income suppressed) and $\zeta[q]$ summarizes the effects on aggregate demand of an increase in q via changes in households' levels of wealth.

Omitting tax rates and depreciation and recalling that $b_f = B_f / pK$, (4.42) takes the form

$$(1-t_f)\left[(1-t_i-\omega[u])u-\delta\right]-ib_f-\bar{r}q=0 \tag{4.53}$$

in which q in the financial market is assumed to respond to shifts in distribution as represented by $\omega[u]$.

After taking total differentials and rearranging terms, the system (4.52)–(4.54) can be restated in matrix notation as

$$\begin{bmatrix} -\left[1-X_u-m_w\phi(\omega+\omega')\right] & g_q^K+\zeta_q \\ 1-\phi(\omega+\omega') & -\bar{r} \end{bmatrix}\begin{bmatrix} du \\ dq \end{bmatrix}=\begin{bmatrix} -m_w u\phi d\omega-g_i^K di-X_{\bar{e}}d\bar{e} \\ \phi u d\omega+b_f di \end{bmatrix} \tag{4.54}$$

The new subscripts denote derivatives and $d\omega$ is an exogenous shift in the labour share. The usual stability conditions for adjustment of u and q to shocks to (4.52) and (4.53) in temporary equilibrium are that the trace of the matrix on the left-hand side should be negative and the determinant positive. Typically one would assume that $X_u<0$ and $1>m_w\geq 0$. The implication is that unless a positive response of the wage share to an increase in u (the term $\phi(\omega+\omega')$) in the northwest entry in the matrix is 'very strong' the trace condition will be satisfied. If $\phi(\omega+\omega')<1$, the determinant condition will be satisfied as well. It is easy to verify that it implies the configuration of the solid and dashed lines shown in Figure 4.2, with the former now corresponding to (4.52) and the latter to (4.53).

Using the diagram, we can get immediate results in comparative statics. In Figure 4.2 an increase $d\omega>0$ in the wage share shifts the demand curve outward for a given level of q. On the other hand, it makes q decline for a given level of capacity utilization. The effects on both variables as the equilibrium is displaced from point A to point B are ambiguous. As it is drawn, the diagram shows an increase in u, so that effective demand is wage-led.

A variation on this theme would be an increase in labour productivity with a constant real wage. The wage share is the ratio of the real wage to the output–labour ratio. Higher productivity means more output per unit labour input. Unless it is matched by an equivalent increase in the real wage, therefore, a productivity increase makes ω go down. The outcome in Figure 4.2 would be a movement from B to A, or a fall in output accompanied by an increase in q. It is easy to verify that output would tend to rise in a profit-led economy, which is more receptive to productivity increases than a 'Luddite' wage-led system.

Taken by itself, real devaluation or $d\bar{e} > 0$ would shift the solid line outward in Figure 4.2, leading to higher output. However, in practice devaluation may also affect the wage share by driving up local prices of traded goods. If nominal wages are not fully indexed to commodity price increases, ω will decline.[9] In other words $d\omega < 0$ is a consequence of $d\bar{e} > 0$ and the overall effect could be a leftward shift of the solid curve in Figure 4.2. The dashed curve would tend to rise but if investment demand and wealth effects in consumption are not strongly responsive to a higher *q*, the final outcome could be a reduction in *u*. This is an example of 'contractionary devaluation', which often seems to occur in developing economies.

Finally, we can consider an increase in the interest rate $di > 0$. As shown in Figure 4.3, the demand curve shifts inward and *q* is driven down by an increased corporate debt burden. The outcome unambiguously combines output reduction and lower profitability.

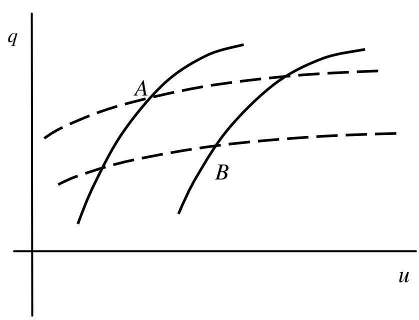


Figure 4.2 *An increase* $d\omega > 0$ *in the labour share in a wage-led economy*

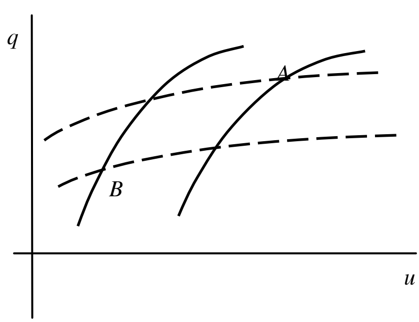


Figure 4.3 *Effects of an increase in the interest rate,* i. *Both* u *and* q *decline*

These results are familiar generalizations of aggregate-demand based macroeconomic models. The chief novelty here is the mediation of asset market prices on investment demand.

4.4. BALANCED GROWTH PATHS

In the steady state the proportions of domestic equity held by worker households, capitalist households, and the rest of the world are constant, as are the ratios of the real stocks of debt to the value of the capital stock, $\Delta\theta^w = 0$, $\Delta\bar{\theta} = 0$, $\Delta b_f = 0$, $\Delta b^w = 0$, $\Delta b^c = 0$, $\Delta\bar{f} = 0$, $\Delta b_g = 0$ and $\Delta\bar{b} = 0$. Plugging these values into the dynamic equations we get the steady-state equations setting $\bar{w} = 0$ to reduce clutter:

$$\tilde{r}^* = (1-t_f)(1-t_i-\omega[u^*])u^* \tag{4.55}$$

$$b^{f*} = \frac{g^* + \sigma - (1-\alpha^f)\tilde{r}^*}{g^* - i} \tag{4.56}$$

$$q^* = \frac{\tilde{r}^* - ib^{f*}}{\rho} = \frac{g^*(\tilde{r}^* - i) + \alpha^f i\tilde{r}^*}{\rho(g^* - i) + \sigma i} \tag{4.57}$$

$$q^*\hat{Q}^* = -\alpha^f\tilde{r}^* \tag{4.58}$$

In order for these equations to have a consistent solution given a steady-state rate of growth g^*, the investment function must satisfy:

$$g^* = g_K[i,\ q^*] - \delta \tag{4.59}$$

We can solve for the steady state values in terms of the steady-state growth rate of capital, g^* and the other parameters:

$$q^*\theta^{w*} = \alpha^w(1-t_w)v\phi\omega[u^*]u^* \tag{4.60}$$

$$b^{w*} = (1-\alpha^w)(1-t_w)v\phi\omega[u^*]u^* \tag{4.61}$$

$$\theta^{c*} = 1 - \theta^{w*} - \bar{\theta} \tag{4.62}$$

Steady-state capitalist wealth, $j^c = J^c / K$ is:

$$j^{c*} = b^{c*} + q^* \theta^{c*} + \overline{ef}^*$$
$$= \theta^{c*} \frac{(1-t_c)\left[(1-\phi)u^* \omega[u^*] + \sigma - q^* \overline{r}\right] + q^*(g^* - \hat{Q}^*)}{g^* + (1-\beta) - (1-t_c)\left[\alpha^c \overline{r} + (1-\alpha^c) i\right]} \tag{4.63}$$

$$b^{c*} = (1-\alpha^c)\, j^{c*} \tag{4.64}$$

$$\overline{ef}^* = \alpha^c j^{c*} - q^* \theta^{c*} \tag{4.65}$$

To make the steady state consistent, some element of aggregate demand, such as z or $x[\overline{e},\ u^*]$ must be chosen consistently with the steady-state level of capacity utilization:

$$u^* = g^K[i, q^*] + z + x[\overline{e}, u^*] + (1-t_w)(\phi\omega[u^*]u^* + ib^{w*} + \sigma\theta^{w*}) +$$
$$-\left\{\gamma\left[\nu(1-t_w)(\phi\omega[u^*]u^* + \overline{e}\,\overline{w}) - (b^{w*} + q^*\theta^{w*})\right] + g^* b^{w*}\right\} + \tag{4.66}$$
$$+ (1-\beta)(q^* \theta^{c^*} + b^{c*} + \overline{ef}^*)$$

These steady-state conditions offer some insights into the structure of this type of economy. An economically meaningful steady state, for example, clearly must have $q^* = (\tilde{r}^* - ib^{f*}) / \rho > 0,$ which requires the after-tax profits of the firm sector to exceed its debt service, and rules out Minsky's Ponzi regime. In the steady state the firm sector as a whole has to be in a speculative or hedged state. The steady state is speculative when $b^{f*} > 0,$ so that the firm sector has to borrow in order to finance its net investment. The steady state is hedged when $b^{f*} < 0,$ in which case the firm sector generates financial surpluses which are transferred to other sectors.[10]

Another condition for an economically meaningful steady state is $\overline{f}^* > 0,$ since the capitalist household sector can hold foreign assets, but cannot issue them.

4.5. CALIBRATION

We have begun to try calibrate the model to represent the structure of the US economy in the late 1990s. While available accounting data describe the balance sheets of the firm, consolidated household, government, financial and rest-of-the-world sectors, we have a more difficult time in separating out the capitalist- and worker-household sectors. We should also keep in mind that the US economy was not necessarily close to a steady-state growth path in any particular year in this period.[11]

Our stylized facts for the U.S. in the late 1990s put $g=0.0313$ and $\delta=0.0394$. National accounting data suggest that $t_i=0.1536$ and $t_f=0.14$, with household taxes net of transfers at about $t_w=t_c=0.066$ times household income. In this period the wage share in GDP was $\omega=0.68$, but in order to make the steady state more realistic, we take $\omega=0.6$ and the ratio of GDP to the value of total assets $u=0.37$. These figures imply $\tilde{r}=0.045$. Firms paid a large proportion of their after-tax net profits in dividends and stock buy-backs. We estimate $\sigma=0.0125$ and $\alpha^f=0.1$, yielding $q\hat{Q}=-0.0035$. We have no direct observation of the rate at which the stock market discounted earnings, but we can estimate $q=1.26$, which is consistent with $\rho=0.03$. With a real interest rate $i=0.02$, these figures imply b^f of 0.33. This is in the observed range of b^f, which ranges from $0.27-0.5$.

We have very little information about worker and capitalist households separately. In the simulations below, we assume that worker households have a target wealth equal to one year's wage income. The main asset in many US households is residential real estate, which we would model as holding equity, financed by borrowing. We set $\alpha^w=1.25$ to reflect this. We have no way of estimating capitalist households' propensity to consume out of wealth, but in the simulations we set $1-\beta=0.09$, which seems to give somewhat reasonable results, with $\alpha^c=0.8$. Assuming that foreigners hold a share $\bar{\theta}=0.1$ of domestic equity, this results in steady-state levels (as a proportion of the capital stock) $\theta^{w*}=0.15$, $b^{w*}=-0.04$, $\theta^{c*}=0.74$, $b^{c*}=0.25$, $\bar{f}^*=0.07$, with worker-household consumption equal to 0.21, capitalist household consumption equal to 0.107, and capitalist household consumption equal to 0.114.

When $z=0.08$, implying a government-expenditure to GDP ratio of 0.23, $b_g^*=-0.004$, implying a steady-state government debt to GDP ratio close to zero.

Foreign borrowing is the residual in this model, and for these parameter values has a steady-state value of $\bar{b}=0.11$.

4.5. DYNAMIC SIMULATION

In this section we present a tentative example of the use of the model for dynamic simulation, in part to demonstrate the consistency of the specification, and in part to suggest possibilities for further investigation.

For the purposes of the simulation we take the derivative of the wage share with respect to capacity utilitization at the steady state, $\omega'^{*} = 0.1$, the derivative of the rate of gross investment to q at the steady state $g_q^{K*} = 0.1$, and the derivative of net exports to capacity utilization, $\xi_u^{*} = -0.01$.

Figures 4.4–4.9 report the results of simulating 50 years in the model after a fall in the investment rate from $\bar{g} = 0.033$ to $\bar{g} = 0.0313$, starting in the steady state corresponding to the higher investment. The immediate effect of this change is to create a sharp recession in the economy. The path shows the transient in capitalist and worker household wealth that results and the consequent return to steady-state levels.

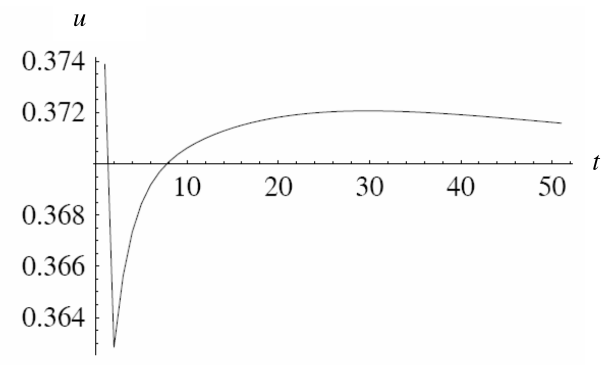


Figure 4.4 The response of u *to a fall in the investment rate*

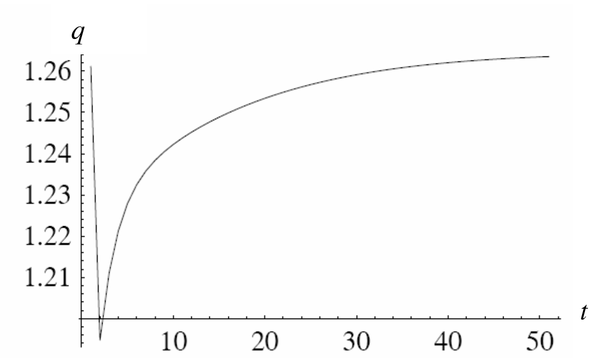


Figure 4.5 The response of q *to a fall in the investment rate*

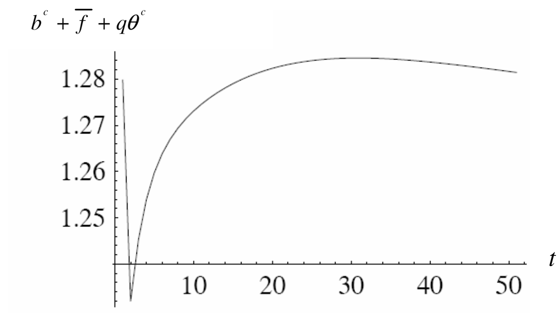


Figure 4.6 The response of capitalist wealth to a fall in the investment rate

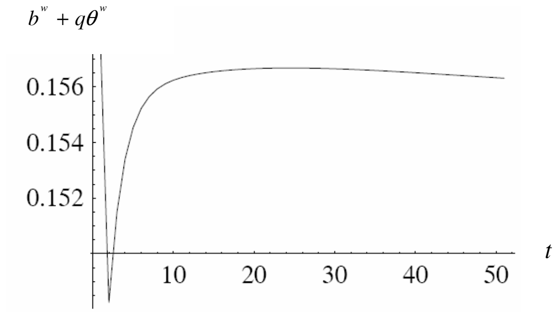


Figure 4.7 The response of worker wealth to a fall in the investment rate

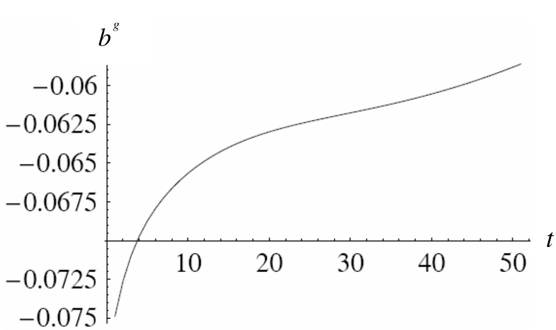


Figure 4.8 The response of government debt to a fall in the investment rate

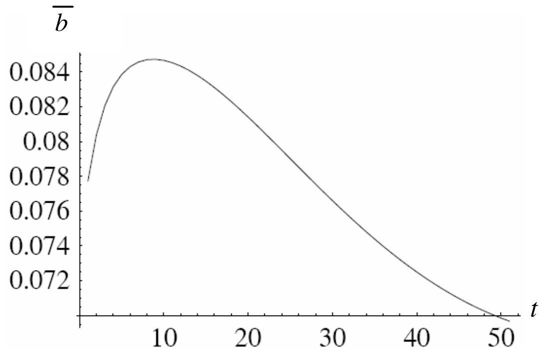


Figure 4.9 The response of foreign borrowing to a fall in the investment rate

4.7. CONCLUSION – HETERODOX METHODS

Heterodox macroeconomic theory combines classical, Marxist, structuralist, Kaleckian, and Keynesian approaches to output determination, growth, and distribution. Three key principles – determination of total income by effective demand, determination of the functional income distribution by processes of bargaining and social conflict, and recognition of the sensitivity of the macroeconomic system to potentially destabilizing interactions of its real and financial components – undergird theory and data analysis which have made great strides over the past two decades.

The heterodox model we put forward here reflects these key principles. Income is determined by aggregate demand through the interaction of an independent firm investment function, government spending and taxation, and household consumption decisions. Income distribution reflects class bargaining positions that depend on the level of capacity utilization and hence on employment opportunities. The financial valuation of firms can differ from the reproduction cost of the capital stock, and this financial valuation feeds back on the economy through wealth impacts in household consumption decisions.

Complete accounting through the Social Accounting Matrix, the representative tool of contemporary heterodoxy, lets us centralize information that can be used to discern the main features of the economy, and guarantees the stock-flow consistency of the model. The model can answer such questions as: Is effective demand wage- or profit-led? Do distributive conflicts as observed from production cost decompositions point to one or another type of inflation? Will financial market fluctuations lead to a crisis?

We hope that this model can serve as both a benchmark and a foundation for future heterodox macroeconomic analysis in the Keynesian–structuralist tradition. It supplies a basic analytical framework which can be modified and extended to deal with specific issues in specific economies. For example, it might be desirable to account for residential housing and its valuation separately from firm investment in studying the macroeconomic dynamics of economies where housing represents a major component of household net worth. The analysis of economies with complex and dualistic financial institutions might require an elaboration of our single financial sector. The distinction between capitalist and worker households might have to be abandoned in situations where data simply does not allow the estimation of the different behavioural parameters of the two sectors. We have tried to provide a flexible, adaptive tool in this model, and invite others to make their own use of it.

NOTES

1. In assuming the existence of a single capital good we abstract from the important issues raised by the Cambridge critique of capital theory based on the work of Piero Sraffa.
2. If pesos are the domestic currency and dollars are the world money, e has the dimensions pesos per dollar.
3. One major payment flow, government transfers to households (around 10 per cent of GDP in the US) is omitted for simplicity. In the numerical calibrations discussed below, transfers are netted out of household direct taxes.
4. This is one point where our approach diverges from the 'mainstream' macroeconomic tradition, which, following Modigliani and Miller (1958), assumes that the composition of firm liabilities has no impact on the valuation of the firm, which depends only on the real value of its assets.
5. Wynne Godley emphasizes the importance of this kind of stock-adjustment process in macroeconomic modelling.
6. This treatment of finance is compatible with the long tradition in heterodox macroeconomics of treating money and credit as 'endogenous', and assuming that in the short run financial institutions accommodate the demands of firms for finance at the going interest rate.
7. Thus we abstract from the real costs of financial intermediation.
8. This follows the tradition of Richard Goodwin. Nicholas Kaldor emphasized the possibility that ω' might be negative due to the slow adjustment of money wages to rising prices leading to forced saving of workers.
9. A similar impact from oil price increases was in part responsible for stagflation in the US economy in the 1970s.
10. For a more complete discussion of Minsky's regimes as applied to national economies and their sectors, see Foley (2003).

11. There are several 'discrepancies' and inconsistencies on the order of 0.5 per cent or more of capital in the US NIPA and Flow-of-Funds statistics, which add to the uncertainty in this exercise.

REFERENCES

Bhaduri, A. and S.A. Marglin (1990), 'Unemployment and the real wage: the economic basis for contesting political ideologies', *Cambridge Journal of Economics*, **14**:375–393.

Duménil, G. and D. Lévy (1994), *The Economics of the Profit Rate*, Aldershot, UK and Brookfield, USA: Edward Elgar.

Foley, D.K. (2003), 'Financial fragility in developing economies', in A. Dutt and J. Ros (eds), *Development Economics and Structuralist Macroeconomics: Essays in Honor of Lance Taylor*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Foley, D.K. and T.R. Michl (1999), *Growth and Distribution*, Cambridge, MA: Harvard University Press.

Godley, W. and T.F. Cripps (1983), *Macroeconomics*, London: Fontana.

Goodwin, R.M. (1967), 'A growth cycle', in C.H. Feinstein (ed.), *Socialism, Capitalism, and Growth*, Cambridge: Cambridge University Press, pp. 54–58.

Harris, D.J. (1978), *Capital Accumulation and Income Distribution*, Stanford: Stanford University Press.

Kaldor, N. (1956), 'Alternative theories of distribution', *Review of Economic Studies*, 23:83–100.

Kaldor, N. (1957), 'A model of economic growth', *Economic Journal*, **67**:591–624.

Kaldor, N. (1961), 'Capital accumulation and economic growth', in F.A. Lutz and D.C. Hague (eds), *The Theory of Capital Accumulation*, London: Macmillan,

Kalecki, M. (1971), *Selected Essays on the Dynamics of the Capitalist Economy: 1933–1970*, Cambridge: Cambridge University Press.

Keynes, J.M. (1936), *The General Theory of Employment, Interest, and Money*, London: Macmillan.

Marglin, S.A. (1984), *Growth, Distribution, and Prices*, Cambridge, MA: Harvard University Press.

Minsky, H.P. (1975), *John Maynard Keynes*, New York: Columbia University Press.

Modigliani, F. and M.H. Miller (1958), 'The cost of capital, corporation finance, and the theory of investment', *American Economic Review*, **48**:261–97.

Robinson, J. (1956), *The Accumulation of Capital*, London: Macmillan.

Robinson, J. (1962), 'A model of accumulation', in *Essays in the Theory of Economic Growth*, London: Macmillan.

Sraffa, P. (1960), *Production of Commodities by Means of Commodities*, Cambridge: Cambridge University Press.

Taylor, L. (2004), *Reconstructing Macroeconomics: Structuralist Proposals and Critiques of the Mainstream*, Cambridge, MA: Harvard University Press.

5. Distribution and growth in a post-Keynesian stock-flow consistent model

Gennaro Zezza and Claudio H. Dos Santos

5.1. INTRODUCTION

Despite the emphasis laid by many post-Keynesian (PK) authors on real financial interactions (Davidson, 1972; Minsky, 1986), most PK growth models either neglect financial issues altogether or assume economies with over-simplified, unrealistic financial structures (Dos Santos, 2004a). In this chapter, we present a stock-flow consistent (SFC) PK growth model of an economy with well-developed financial markets.

Godley (1996, 1999) and others (for example, Taylor, 2004; Foley and Taylor, 2006) have analysed SFC models extensively. In fact, the model presented here is meant as an extension of Lavoie and Godley's (2001–2002, L&G from now on) ground-breaking contribution. In particular, we have incorporated a government sector and a central bank to L&G's story, which assumes an economy with only households, firms and private banks. This extended L&G economy, we argue, allows realistic and integrated analyses of a broad range of inter-related (and, to some extent, obscure) issues in Keynesian economics, such as the role played by the stocks of financial wealth/debt in flow behaviour (of firms, families and banks), the functioning of financial markets (and, therefore, the 'transmission mechanisms' of monetary policy), and the (dynamically) optimum monetary/fiscal mix to be adopted by policy-makers.

We have, indeed, used the model to analyse some of these issues elsewhere, such as in Zezza and Dos Santos (2004), and Dos Santos and Zezza (2004, DSZ from now on). In this chapter we use it to address the relationship between growth and the distribution of income among households, firms and the government, and among three groups in the household sector, namely wage earners, owners of equities issued by non-financial firms and recipients of bank profits. We assume, in the tradition of Kaldor–Pasinetti models (Kaldor, 1966), that the expenditure behaviour of such groups differs, and analyse the implications that this assumption may have for growth.

The rest of this chapter is divided into three parts. The model is briefly described in Section 5.2. In Section 5.3 we explore the model's properties through deterministic simulations over a broad spectrum of parameters, and address the interrelations between growth and the distribution of income. We are interested to check, in particular, how feedbacks from financial markets to capital and wealth accumulation modify some standard results in post-Keynesian growth models presented in the literature. Section 5.4 summarizes and concludes.

5.2. THE MODEL

We follow here Taylor's (2004) 'structuralist' methodology of phrasing macroeconomic models explicitly as 'closures' of a given accounting framework (see also Barbosa Filho, 2004). We have already described our own SFC accounting framework elsewhere (for example, Zezza and Dos Santos, 2004; Dos Santos, 2004a and 2004b), so we will go relatively quickly in Section 5.2.1 below: the interested reader will find a more comprehensive analysis in the aforementioned texts. The behavioural assumptions are presented in Section 5.2.2.

5.2.1. Structural Hypotheses (and their Logical Implications)

Virtually no one disagrees that households, firms, banks and the government each play crucial roles in actual capitalist economies. Interestingly enough, only a small number of macroeconomists (for example Backus et al., 1980; Franke and Semmler, 1989; Godley, 1996; Zezza and Dos Santos, 2004) have sought to fully explore the logical implications of such an institutional structure. This section aims to do just that.

Perhaps the most obvious consequence of working with such a rich institutional framework is the multiplication of the financial stocks assumed in one's model. Indeed, if one wants to be fully consistent, modelling four institutional sectors means having to model four (inter-related) balance sheets. We believe our assumptions about the latter (summarized in Table 5.1) are very intuitive. As the reader will notice, the following, standard, simplifying assumptions were adopted: (i) households neither invest (that is buy 'capital'[1]) nor get bank loans; (ii) firms do not hold government bills or money; (iii) the government neither invests nor holds equity or money (in any form); and (iv) banks (including the central bank) neither issue nor hold

Table 5.1 Balance sheets

	Households	Firms	Banks	Central Bank	Government	Total
High powered money	$+Hh$		$+Hb$	$-H$		0
Central bank advances			$-A$	$+A$		0
Bank deposits	$+M$		$-M$			0
Loans		-L	$+L$			0
Bills	$+Bh$		$+Bb$	$+Bc$	$-B$	0
Capital		$+pK$				$+pK$
Equities	$+E\,pe$	$-E\,pe$				0
Total (net worth)	$+V$	$+Vf$	0	0	$-B$	$+pK$

equities, do not invest, and distribute all their profits (so their net worth is zero).

The hypotheses summarized in Table 5.1 above have important logical implications. First, accounting consistency alone allows us to write the following identities, which determine, respectively, the stock of bank deposits (M) as a residual component of households' wealth (V); the stock of firms' wealth (Vf); the stock of bonds held by banks (Bb); the stock of advances (A) from the central bank; the total stock of cash (H); and the stock of government bonds held by the central bank (Bc), which is assumed to clear the market for government bonds.

$$M_t \equiv V_t - (Bh_t + E_t pe_t + Hh_t) \tag{5.1}$$

$$Vf_t \equiv p_t K_t - L_t - E_t pe_t \tag{5.2}$$

$$Bb_t \equiv A_t + M_t - Hb_t - L_t \tag{5.3}$$

$$A_t \equiv H_t - Bc_t \tag{5.4}$$

$$H_t \equiv Hh_t + Hb_t \tag{5.5}$$

$$Bc_t \equiv B_t - Bh_t - Bb_t \tag{5.6}$$

Second, the balance sheets above have 'cash flow' implications. As Minsky (1975, p. 118) reminds us, 'cash flows are the result of (i) the income-producing system, which includes wages, taxes and non-financial corporate gross profits after taxes, (ii) the financial structure, which is composed of interest, dividends, rents and repayments on loans, and (iii) the dealing or trading in capital assets and financial instruments'. Tables 5.2 and 5.3 formalize these 'flows of funds'.

Flow identities are summarized in Table 5.2 where, as usual,[2] monetary payments are recorded in columns, while rows register receipts. The second row of Table 5.2, for example, means that households receive income in the form of wages (*WB*), dividends from firms (*FD*) and banks (*FB*) and interest payments (from their holdings of government bills, *Bh*, and bank deposits, *M*). The second column of Table 5.2, on the other hand, says that households spend part of this income in consumption (*pC*) and taxes (*Td*) and save the rest (*Sh*) in the form of four financial assets (first column of Table 5.3), that is cash (*Hh*), bank deposits, government bills and firms' equities (*peE*). Total Production (*pY*), in turn, is assumed to be done all by firms and can be decomposed in wage bill (*WB*), indirect taxes (*TI*) and gross profits (*FT*) (Table 5.2, first column). Firms do not retain all profits, though. They only get what is left (*FU*) after payments of indirect and profit taxes (*TI* and *TF*), interest on loans (*iL*), and dividends to equity holders are made (Table 5.2, third row and column), using it (along with bank loans and equity emissions) to finance their investment ($p\Delta K$) in physical capital (*K*) (Table 5.2, second column).

Turning our attention now to the government sector, we note that its receipts come from the aforementioned taxes and the dividends paid by the central bank (*Fc*, see Table 5.2, sixth row) and are spent in public goods (*pG*) acquired from firms and in interest payments to the holders of government bills (Table 5.2, sixth column). As one would expect, the public deficit (that is a negative *Sg*) often needs to be financed by the central bank (when the private sector does not do it itself) through the issuing of high-powered money to buy government bills (see Table 5.2, columns 5 and 6). The central bank may also receive interest on its loans to banks (if they need them), and distributes all its profits (that is interest payments received) to the government. Finally, banks make money from the interest payments they receive on loans to firms and on their holdings of government bills and spend it on interest payments to households and (if necessary) the central bank. For simplicity, banks are assumed not to pay taxes and to distribute all their profits to households.

The identities arising from Table 5.2 may be summarized as follow:

$$Y_t \equiv C_t + G_t + K_t - K_{t\text{-}1} \tag{5.7}$$

$$FT_t \equiv p_t Y_t - WB_t - TI_t \tag{5.8}$$

$$Yhd_t \equiv WB_t + FD_t + i_{Mt-1}M_{t-1} + FB_t + i_{Bt-1}Bh_{t-1} - TD_t \tag{5.9}$$

Table 5.2 Social accounting matrix

	Prod.	House-holds	Firms	Banks	Central Bank	Govern-ment	Capital Account	Total
1. Production		$+pC$				pG	$p\Delta K$	pY
2. Households	$+WB$		$+FD$	$+iM$ $+Fb$		$+iBh$		$+Yh$
3. Firms	$+FT$							$+FT$
4. Banks			$+iL$			$+iBb$		$+Yb$
5. Central Bank				$+iA$		$+iBc$		$+Yc$
6. Government	$+Ti$	$+Td$	$+Tf$		$+Fc$			$+Yg$
7. Capital Account		$+Sh$	$+FU$	0	0	$+Sg$		$+SAV$
Total	$+pY$	$+Yh$	$+FT$	$+Yb$	$+Yc$	$+Yg$	$p\Delta K$	

Table 5.3 Sources and uses of funds(*)

Changes in	Households	Firms	Banks	Central Bank	Govern-ment	Total
Cash	$+\Delta Hh$		$+ \Delta Hb$	$- \Delta H$		0
Central Bank advances			$- \Delta A$	$+ \Delta A$		0
Bank deposits	$+ \Delta M$		$- \Delta M$			0
Loans		$-\Delta L$	$+ \Delta L$			0
Treasury Bills	$+ \Delta B$h		$+ \Delta Bb$	$+ \Delta B$c	$-\Delta B$	0
Capital		$+p\ \Delta K$				$+p\ I$
Equities	$+ \Delta E\ pe$	$-\Delta E\ pe$				0
Total	Sh	Fu	0	0	Sg	SAV

(*) The Δ operator signifies a discrete change, so for example, $\Delta E = E_t - E_{t-1}$

$$Sh_t \equiv Yhd_t - p_t C_t \tag{5.10}$$

$$FD_t \equiv FT_t - iL_{t-1} L_{t-1} - TF_t - FU_t \tag{5.11}$$

$$FB_t \equiv iL_{t-1} L_{t-1} + iB_{t-1} Bb_{t-1} - iM_{t-1} M_{t-1} - iA_{t-1} A_{t-1} \tag{5.12}$$

$$FC_t \equiv i_{At-1} A_{t-1} + i_{Bt-1} Bc_{t-1} \tag{5.13}$$

$$Sg_t \equiv TI_t + TD_t + TF_t + Fc_t - i_{Bt-1} B_{t-1} - p_t G_t \tag{5.14}$$

Our detailed modelling of the cash flows in our artificial economy has a simple rationale. Without it, we cannot know for sure how the balance sheets of the institutional sectors (which are crucially affected by sectoral saving flows, portfolio shifts, and capital gains) will evolve over time, for an accurate calculation of sectoral saving flows presupposes accurate accounting of the income redistributions among sectors. Knowledge of the balance sheets, in turn, is crucial in Keynesian models, for these assume that asset prices are determined in markets for stocks (see Section 5.2.2 below). Miscalculated balance sheets, therefore, would imply wrong conclusions about financial market behaviour.

From Table 5.3 we can determine the stock of households' wealth (V) from households' savings (Sh) – taking properly into account capital gains on equities (CG); the amount of loans (L), considered to be a residual source of finance for firms; the amount of government bonds (B) needed to finance government negative saving (Sg).

$$V_t \equiv V_{t-1} + Sh_t + CG_t \tag{5.15}$$

$$L_t - L_{t-1} \equiv p(K_t - K_{t-1}) - FU - pe_t (E_t - E_{t-1}) \tag{5.16}$$

$$B_t - B_{t-1} \equiv -Sg_t \tag{5.17}$$

Capital gains can be obtained on the stock of equity, since equities are, by assumption, the only financial asset with a market price.

$$CG_t \equiv (pe_t - pe_{t-1}) E_{t-1} \tag{5.18}$$

5.2.2. Behavioural Hypotheses

Institutions matter in the PK world depicted above. This section details how exactly the institutional sectors are assumed to behave in this model, beginning with households.

5.2.2.1. Households

We assume that households' consumption depends on real disposable income, the opening stock of wealth and real capital gains, where we allow for different propensities to consume out of income for the three groups in the economy (assuming that the propensity to save for owners of equities and recipients of bank profits is higher than that for wage earners, that is $0 < \alpha_1', \alpha_2'' < \alpha_1$).

$$C_t = \left[\alpha_1 \frac{Yhd_t}{p_t} - (\alpha_1' FD_t + \alpha_2'' FB_t) \frac{1 - \tau d}{p_t} + \alpha_2 \frac{V_{t-1}}{p_t} + \alpha_3 \frac{CG_t^e}{p_t} \right] \quad (5.19)$$

Our consumption function is the first source of 'real' feedbacks from financial stock accumulation to expenditure flows: a higher value of real wealth will decrease savings, and the same effect is given by capital gains on equities.

Focusing now on the portfolio choice of households, we assume that households' demand for cash is in a stable relation to consumption, and wealth is allocated to financial assets according to a Tobinesque set of equations, where each asset share on wealth is determined by real expected returns on each asset. Since expected values will generally differ from realized values, one asset must act as a buffer to provide ex-post equilibrium, and we assume that bank deposits play such a role (see equation (5.1)). Formally

$$Hh_t = v_1 \, p_t C_t \quad (5.20)$$

$$E_t pe_t = \left(\lambda_{00} - \lambda_{01} r_{M\,t} + \lambda_{02} r_{E\,t} - \lambda_{03} \frac{Yhd_t^e}{V_t^e} - \lambda_{04} r_{B\,t} \right) (V_t^e - Hh_t) \quad (5.21)$$

$$Bh_t = \left(\lambda_{10} - \lambda_{11} r_{M\,t} - \lambda_{12} r_{E\,t} - \lambda_{13} \frac{Yhd_t^e}{V_t^e} + \lambda_{14} r_{B\,t} \right) (V_t^e - Hh_t) \quad (5.22)$$

Where the real rates of interest are obtained based on nominal interest rates (i_{Mt}, i_{Bt} and i_{Et}) and expected inflation ($\dot{p}^e$):

$$(1+r_{Mt})(1+\dot{p}_t^e)=1+i_{Mt} \tag{5.23}$$

$$(1+r_{Bt})(1+\dot{p}_t^e)=1+i_{Bt} \tag{5.24}$$

$$(1+r_{Et})(1+\dot{p}_t^e)=1+i_{Et} \tag{5.25}$$

$$i_{Et}=\frac{FD_t+CG_t^e}{pe_{t-1}E_{t-1}} \tag{5.26}$$

with the nominal return on equities i_{Et}, in equation (5.26), being given by distributed profits plus expected capital gains.

Finally, we assume (in equation (5.28)) that households' unit wage demand (w) depends on expected inflation ($\dot{p}_t^e$) and on expected productivity growth ($\dot{\pi}_t^e$).[3] Moreover, the aggregated wage bill of firms is given by the nominal wage rate times the number of workers employed (equation (5.27)).

$$WB_t=w_tN_t \tag{5.27}$$

$$\dot{w}_t=\dot{p}_t^e+\chi_t\dot{\pi}_t^e \tag{5.28}$$

5.2.2.2. Firms

Firms are assumed to always get the point of effective demand right[4] (hiring as many workers as necessary to produce that amount, see equation (5.29)). Along L&G lines, investment is driven by four key variables: the cash-flow rate (*rfc*, equation (5.31)), which provides a measure for the ability of firms to self-finance their expenditure decisions; a measure of leverage (*lev*) given by the ratio of interest payments on bank loans to the stock of capital (equation (5.32)), which implies that, as firms' debt with banks rises, some investment decisions will be abandoned; Tobin's q (equation (5.33)) and finally the utilization rate u (equation (5.34)).[5] As depicted in Table 5.2, investment is financed with retained earnings (*FU*, which are assumed to be a fixed proportion ϕ of after tax profits net of the interest bill, see equation (5.36) below). Firms also issue new equities, assumed to be a fixed share of the amount of 'external funds' required to finance investment (see equation (5.37)) and, if necessary, firms demand bank loans (see equation (5.16)). In other words, as stated earlier, firms demand bank loans as a residual source of finance. Firms also fix their mark-up,[6] thereby determining the price level p (equation (5.35)). Formally:

$$N_t = \frac{Y_t}{\pi_t} \tag{5.29}$$

$$\dot{K}_t = \gamma_0 + \gamma_1 rfc_{t-1} - \gamma_2 rrl_{t-1} lev_{t-1} + \gamma_3 q_{t-1} + \gamma_4 u_{t-1} \tag{5.30}$$

$$rfc_t = \frac{FU_t}{K_{t-1} pt} \tag{5.31}$$

$$lev_t = \frac{L_t}{K_{t-1} pt} \tag{5.32}$$

$$q_t = \frac{L_t + pe_t E_t}{K_{t-1} pt} \tag{5.33}$$

$$u_t = \frac{Y_t}{K_{t-1}} \tag{5.34}$$

$$p_t = \frac{1+\rho_t}{1-\tau i} \cdot \frac{w_t}{\pi_t} \tag{5.35}$$

$$FU_t = \phi(FT_t - r_{Lt-1} L_{t-1} - TF_t) \tag{5.36}$$

$$pe_t \Delta E_t = x(p_t \Delta K_t - FU_t) \tag{5.37}$$

The last equation can be solved together with equation (5.21) to yield the equilibrium value of equities, *pe*.

We are fully aware that investment functions are controversial. In fact, we choose to adopt the formulation in L&G, specifically because it allows competing theories of investment as special cases (by setting some parameters of the initial, 'general' investment function to zero). Alternative closures – implying, say, convergence of the actual utilization rate to a 'normal', exogenous utilization rate (as in Shaikh, 1989; or Moudud, 1998); and by endogenizing the mark-up (Wood, 1975; Eichner, 1976), could also have been assumed, though lack of space prevents us from discussing them here.

5.2.2.3. Government and banks

Beginning with banks, we assume (*à la* Godley, 1999) that they are price makers (and, therefore, quantity takers) in both the markets for loans (to firms) and deposits (from households).[7] Specifically, banks are assumed to fix the interest rate on deposits and bank loans, as mark-ups over the central bank's interest rate on advances (equations (5.38)–(5.39)), accept any amount of deposits from households (a fact that shows up in equation (5.1)), and supply whatever amount of loans demanded by firms (and hence equation (5.16)). Banks also need to meet a reserve requirement (imposed by the central bank on its deposits, see equation (5.40)), and, as stated above, distribute all their profits to households (equation (5.12)).

$$i_{Lt} = i_{At} + \mu_1 \tag{5.38}$$

$$i_{Mt} = i_{At} + \mu_2 \tag{5.39}$$

$$Hb_t = v_2 M_t \tag{5.40}$$

$$i_{At} = i_{Bt} + \mu_3 \tag{5.41}$$

Our hypotheses about the government sector are also very simple. The rate of growth in government expenditure is linked to the growth rate of capital (equation (5.42), where we start with $\eta_0 = 0$, $\eta_1 = 1$) and taxes are fixed proportions of the relevant flows (equations (5.43)–(5.45)). As depicted in equation (5.14), government receipts include also the profits of the central bank (to be explained below), while government expenditures include also the service of government debt. The stock-flow implications of these assumptions were depicted in equations (5.14) and (5.17)

$$\dot{G}_t = \eta_0 + \eta_1 \dot{K}_t \tag{5.42}$$

$$TD = \tau d \frac{Yhd}{1 - \tau d} \tag{5.43}$$

$$TI = \tau i \, pY \tag{5.44}$$

$$TF = \tau f \, FT \tag{5.45}$$

Last, but not least, central bank behaviour is also extremely simplified: the central bank (i) fixes the interest rate on its loans to banks; (ii) provides as much in loans as banks want at this chosen rate; (iii) fixes the interest rate on government bills by purchasing as many Treasury bills as are needed to achieve its goal (equation (5.6)); and (iv) pays back to the government any 'profits' arising from its revenues being larger than its payments (equations (5.13) and (5.14)).

Using the above assumptions on banks' behaviour, however, the amount of central bank advances, or the purchases of government bills from banks, remain undetermined. We may close the model with alternative assumptions,[8] and here we simply assume that banks do not purchase government bills, so that the banks' balance sheet constraint, in equation (5.3), determines the amount of central bank advances A, while the assumption that the central bank clears the market for government bills, in equation (5.6), determines the amount of government bills purchased by the central bank Bc. It will always be true, given balance sheet identities, that the amount of cash created by the central bank equals the sum of central bank's loans to banks and the government, so identity 4 should be dropped from the model. Formally:

$$A_t \equiv Hb_t + L_t - M_t \tag{5.3'}$$

$$c_t \equiv B_t - Bh_t \tag{5.6'}$$

To conclude this section, we note that the assumptions above imply the existence of four explicit financial markets, that is (i) a stock market (where firms are suppliers and households demanders); (ii) a market for government bills (in which the treasury is the supplier and households, and the central bankers are the demanders); (iii) a market for bank deposits (supplied by banks and demanded by households); and (iv) a market for bank loans (demanded by firms and supplied by banks). The amounts of central bank advances to banks, and high-powered money held by banks and households, are both determined endogenously.

The model has 44 equations with 44 endogenous variables, namely A, B, Bc, Bh, C, CG, E, FB, FD, FG, FT, FU, G, H, Hb, Hh, i_A, i_E, i_L, i_M, K, L, lev, M, N, p, pe, q, r_B, r_E, r_M, rfc, Sg, Sh, TD, TF, TI, u, V, Vf, w, WB, Y, Ydh, depending on the various parameters in the model – more prominently the exogenous growth rates in the stock of capital, γ_0, and government expenditure, η_0; the exogenous interest rate on government bills, i_B, and productivity, π. We should add expected values for wealth, income, capital gains, inflation and productivity growth to the list of endogenous variables above: we choose to model all expected values according to an adaptive model, such as

$$\dot{p}_t^e = \dot{p}_{t-1} + \psi(\dot{p}_{t-1}^e - \dot{p}_{t-1})$$

5.2.2. Comments on our Modelling Strategy

We attach great importance to two features of the model described above, that is the fact that it is stock-flow consistent (SFC) and the fact that the 'artificial economy' above encompasses those underlying virtually all post-

Keynesian (one sector, closed economy) growth models as special cases (Dos Santos, 2004a). Indeed, the use of our general framework appears indispensable to any rigorous attempt to discuss the nature and impact of the simplifying and/or implicit assumptions usually adopted in the literature in question. Such systematization, we believe, is a pre-requisite to the development of a future consensual 'formal post-Keynesian growth model' that is both rigorous and flexible enough to be applied to the analysis of macroeconomic policies in actual economies.

Beginning with the importance of SFC requirements, we note that dynamic analyses of non-SFC models are necessarily biased. Of course, there are often very good reasons to focus on specific mechanisms operating within a given economy. But this should not obscure the fact that any rigorous dynamic analysis of the 'relative importance' of these specific mechanisms presupposes a SFC model of the 'economy as a whole'. Indeed, there is no *a priori* reason to believe that the insights obtained in partial models will remain valid in models in which all system system-wide implications of their hypotheses are taken into consideration. Moreover, discrete time SFC models seem a natural 'weapon of choice' for PKs who believe that the 'long run' can only be fruitfully analysed if understood as a sequence of short-period equilibria governed by effective demand. Specifically, the '(short) period-by-(short)period' dynamics of the system we propose here depend – as in the PK tradition – on expectations (which are assumed here to be formed adaptively), and on the accumulation of stocks (which will affect expenditure decisions in the following periods).

As regards the financial architecture assumed, we believe the artificial economy described above is a rigorous formalization of those assumed in most PK literary writings (Dos Santos, 2004b). We could easily have simplified it, of course, but in doing so we would like to make sure that these simplifications were robust enough. This, in turn, can only be ascertained through rigorous knowledge of both 'complex' and 'simplified' versions of these same ideas. The reader is welcome to think about this chapter as a step in the exploration of the former.

We are well aware of the costs of our approach. Specifically, the model we will be working with is such that analytical solutions are uninformative (given its size), so its properties must be explored through simulations (as exemplified in Section 5.3). However, this should not discourage the interested reader from carefully analysing the equations of the model and figuring out why is that so.

5.3. GROWTH AND INCOME DISTRIBUTION

To investigate the properties of the model by virtue of dynamic, deterministic simulation, we obtain a baseline solution by 'calibrating' an initial, plausible[9] set of parameters. We use Eviews for dynamic simulation.[10]

The main exogenous variables determining the growth rate of the economy will be autonomous investment, fiscal parameters and the interest rate set by the central bank, with consumption and government expenditure adjusting to investment growth. As stated above, the utilization rate is endogenous.[11] In our baseline steady growth, all (endogenous) determinants of investment, namely the utilization rate, Tobin's q, the cash-flow rate and the leverage, quickly converge to their steady values.

To illustrate the major properties of the model under our chosen set of parameters, we have shocked the exogenous growth rate in private investment, namely the constant in equation (5.30). Changes to growth rates of output, capital and consumption are reported in Figure 5.1.

The acceleration in investment generates the standard Keynesian multiplier effects on consumption. However, as the stock of capital starts growing faster than sales, the utilization rate drops (see Figure 5.2), and this effect counters the initial shock. Additional investment, moreover, requires new finance: investment does generate an increase in profits, but only with a lag, and initially an increase in bank loans to firms is required, and this increases the flow of interest payment from firms to banks in the subsequent simulation periods. The increase in the leverage is such as to more than offset the increase in gross profits, so that retained profits, as measured by the cash-flow ratio, eventually decline below the baseline level, again countering the initial, positive shock to growth. Finally, the increase in profits will make equities more attractive, even though the shock to investment is accompanied by a proportional increase in the supply of equities, which tend to lower their market price. In the new steady growth path, equities will be a larger share of household wealth, with a corresponding decrease in bonds and bank deposits. However, since demand for equities does not grow as fast as the increase in the stock of capital, Tobin's q will decrease, countering the initial positive shock to growth.

Interestingly, the share of income accruing to firms – retained profits – is lower in the new steady growth path, as is the rate of profit, while the share of the government receipts on income increases. As we have seen, increases in bank loans mean higher interest payments, and therefore lower retained profits, even though gross profits increase. Recipients of both firms and banks profits will thus experience an increase in income. Since in our model the government sector sets public expenditure at the expected growth rate of

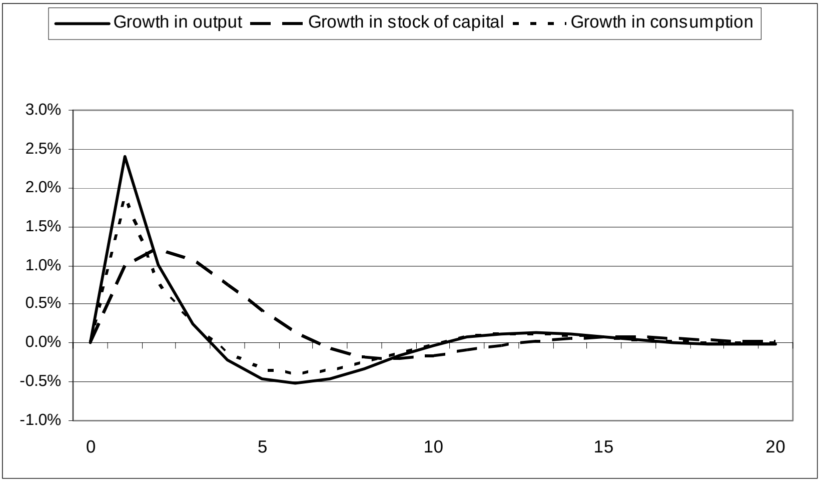


Figure 5.1 Shock to private investment. Difference with baseline

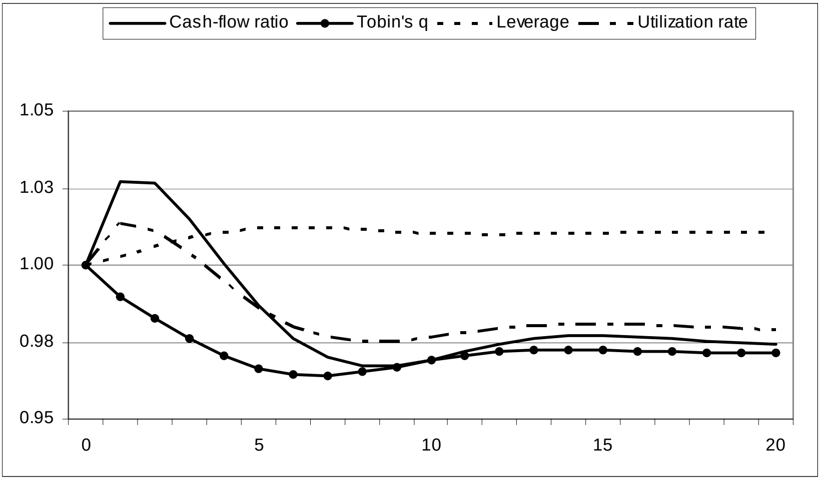


Figure 5.2 Shock to private investment. Ratio to baseline values

the economy, any acceleration in income will make government receipts grow faster than expenditure, and lower the deficit. This in turn implies a smaller public debt, compared to the baseline, and thus lower interest payments on Treasury bills, which will contribute to increase the share of government income on total output, while decreasing the share of households' income.

5.3.1. The 'Paradox of Thrift'

Our next experiment verifies how our model compares with respect to the Keynesian 'paradox of thrift', according to which an increase in the propensity to save reduces capital accumulation and profits. We have thus simulated our model by assuming a positive increase in the propensity to save out of non-profit income.

An increase in the propensity to save implies an immediate drop in consumption. This in turn generates a reduction in the utilization rate, as the stock of capital is now growing faster than output, and a fall in profits (and the profit rate), which translates into a lower cash-flow ratio. Both these variables will make investment fall, and as investment decelerates the utilization rate stabilizes towards its new, lower, long-run value.

Since we assume that the propensity to save out of income will be lower for owners of firms' equities receiving distributed profits, and for recipients of bank profits, the validity of the paradox of thrift will imply that any change to the distribution of income in favour of those groups will result in the economy stabilizing on a lower growth path. We will return to this issue below.

5.3.2. The 'Paradox of Costs'

It is interesting to check whether our model also exhibits the post-Keynesian 'paradox of costs', namely that increases in the real wage lead to increases in the rate of accumulation and in the profit rate. The relationship between the rate of accumulation and the profit rate is given, in simpler post-Keynesian models than the one we adopt here, by the Cambridge formula $r = g/s$; where r is the rate of profit, g the rate of accumulation and s the propensity to save out of profits. Therefore, for a given propensity to save, a higher rate of accumulation implies a higher profit rate.[12]

We tested for such effects by shocking the real wage:[13] results are reported in Figures 5.3 and 5.4. In our model an increase in the real wage implies a higher accumulation rate, but a lower profit rate. This is due to the fact that a higher real wage implies a higher utilization rate and a higher accumulation rate. The profit margin falls, lowering the profit rate and the rate of accumulation, and the net effect on the profit rate is negative, under a wide range of parameters in our investment function.

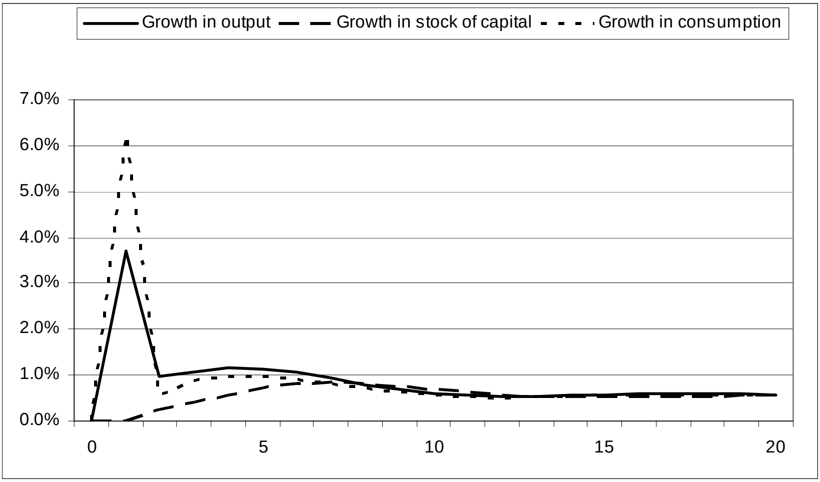


Figure 5.3 *Shock to real wage. Difference with baseline*

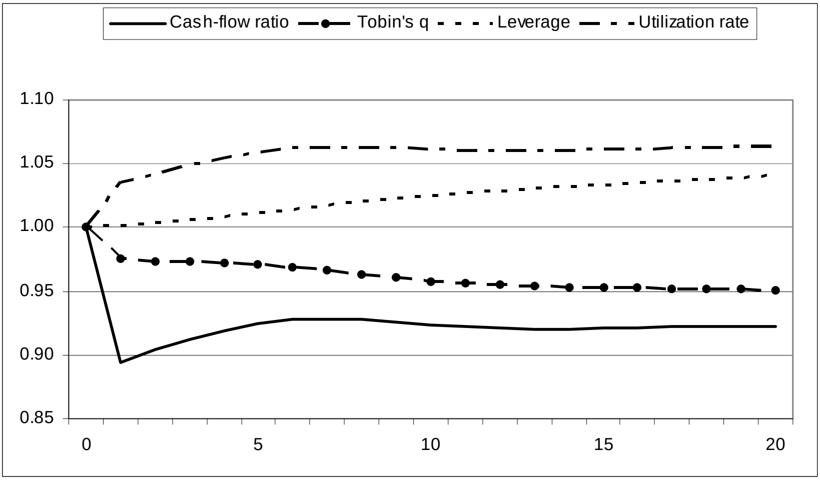


Figure 5.4 *Shock to real wage. Ratio to baseline values*

5.3.3. The Distribution of Income: Banks

In our model, banks can appropriate a larger share of income by increasing the interest rate on loans, which is assumed to be set according to a mark-up on the interest rate banks pay to the central bank to obtain advances. As discussed earlier, in the present simple version of the model banks distribute all of their profits to households, and do not buy either equities or Treasury bills. We wish to make it clear, however, that we chose a set of parameters such that the structure of returns in our model is entirely plausible: interest rates on loans are lower than realized return on equities, to reflect a risk differential, and interest rates on Treasury bills are lower than interest rates on loans, again to consider the risk differential in the probability of defaults of the different debtors. With these sets of values, the assumption that banks do not diversify their portfolio – into other assets rather than loans – should not be too implausible, even though we realize further work is needed to model banks' finance decisions more appropriately.

In our model, banks can raise their share on income either by lowering the interest rate on deposits, or by increasing the interest rate on loans, or both. In both cases, the net increase in profits is transferred to households. The long-run effect of this redistribution in income on growth turns out to depend on the parameter relating investment to the utilization rate.

When the interest rate on bank deposits is lowered, income is transferred from wage earners to recipients of bank profits, who have a higher propensity to save. Consumption drops, and so does the utilization rate, and the cash-flow ratio. However, since the relative return on deposits has fallen, demand for equities will rise, implying a rise in Tobin's q that stimulates investment. If the negative effect of the drop in the utilization rate on investment is sufficiently large, it will outweigh the positive effect arising from the increase in the market value of equities, and the economy will move to a lower growth path. If the effect of the utilization rate is small, on the contrary, the net effect of the redistribution of income on growth may be positive. In both cases, however, the impact on the growth rate will be small (about 0.1 per cent for a 1 per cent change in the real interest rate).

When banks increase the interest rate on loans, the probability of ending on a lower growth path increases. Now, part of the redistribution of income is from owners of equities to recipients of bank profits and, since the two categories have the same propensity to save, there is no direct effect from the redistribution of income on consumption. The increase in the cost of borrowing to firms will make investment fall, and the increase in savings will translate into higher demand for equities: since the supply of new equities is falling with investment, the price of equities will rise, generating capital gains and a rise in Tobin's q. However, for a wide range of plausible parameters, the net effect on growth will be negative, and even substantial (about 1 per cent lower growth for a 1 per cent increase in the interest rate on loans) for a sufficiently high reaction of investment to the fall in the utilization rate.

5.3.4. The Distribution of Income: Distributed Profits

What happens if firms increase the share of distributed profits (decrease the share of retained profits)? The impact on growth depends, again, on the values of parameters in the investment function and the propensity to save of both profit earners and recipients of bank profits. The drop in retained profits will reduce the cash-flow ratio, and increase the need for bank loans: both these variables will drive investment decisions downwards. The drop in investment relative to sales will eventually drive up the utilization rate, and slowly mitigate, or reverse, the drop in capital accumulation. Any attempt at increasing distributed profits, at the expenses of firms' retained profits, will provide only temporary benefits to owners of equities, since the drop in gross profit generated by slower growth and the higher cost of borrowing will slowly offset the initial gain. The only beneficiary of this redistribution of income will be the banking sector, which will increase its stock of loans to firms and thus obtain a growing stream of interest payments.

5.3.6. The Distribution of Income: the Government

We finally examine how the model reacts when the government increases its share of output by increasing a tax rate. We choose to shock the direct tax rate for this exercise, increasing it by 1 per cent. For our choice of parameters, the economy stabilizes on a lower growth path, with a drop of about 0.3 per cent in income growth, with respect to the baseline.

5.4. FINAL REMARKS

We have presented a consistent stock-flow model for a closed, one-sector economy, describing the main relations among households, firms, banks, the government and a central bank, and assuming that the propensity to save of households who are recipients of bank or firms' profits is higher than that of wage earners.

The method we adopt requires tracking all monetary flows among sectors, the allocation of savings to real and financial assets, and feedbacks from stock-flow ratios to expenditure decisions. We believe that the consistency of this approach is preferable as a 'foundation' for macroeconomics to the widely adopted (though theoretically unacceptable) neoclassical 'representative agent' hypothesis.[14, 15]

The model can also be interpreted as the structural form of other simpler demand-led models, allowing us to evaluate the internal consistency of such models, and to study how results from such models change when allowing for more complex interactions, such as the introduction of financial markets.

In this contribution, we have used the model to study – in preliminary fashion – the relationship between growth and the distribution of income, where the latter depends on exogenous parameters (some of which may be endogenized) reflecting conflict among different institutions or social groups. In our economy the paradox of thrift holds, and therefore attempts to shift the distribution of income out of low-saving wage earners to recipients of profits will slow down accumulation. We have also analysed the effects of attempts of banks, firms and the government to increase their share in the distribution of income: while increases in the tax rate univocally determine slower growth, the results of the other experiment are mixed, and ultimately depend on the choice of parameters for the investment function.

We believe our approach to be a useful starting point for integrating real and financial markets in a post-Keynesian model of a growing economy. The importance of such integration cannot be underestimated. A research programme that only consists of partial models cannot possibly hope to have an impact in the profession. Our approach emphasizes the importance of testing for the robustness of partial hypotheses in economy-wide contexts and, as such, should be seen as a contribution to the ongoing systematization of the PK research programme.

APPENDIX

Model Variables

A	Central bank advances
B	Stock of Treasury bills – total
Bb	Stock of Treasury bills held by banks
Bc	Stock of Treasury bills held by the central bank
Bh	Stock of Treasury bills held by households
C	Real consumption of goods
CG	Capital gains on equities
E	Stock of equities
Fb	Bank profits
Fc	Central bank: difference between receipts and payments
Fd	Distributed profits
Ft	Gross profits
Fu	Retained profits
G	Real public expenditure
gr	Growth rate of the real stock of capital

gry	Growth rate of real sales
H	Total stock of cash
Hb	Bank reserves
Hh	Cash held by the public
IA	Interest paid on central bank advances
i_A	Interest rate on central bank advances
i_B	Interest rate on Treasury bills
iBb	Interest paid on Treasury bills to banks
iBc	Interest paid on Treasury bills to central bank
iBh	Interest paid on Treasury bills to households
iL	Interest paid on loans
i_L	Interest rate on loans
IM	Interest paid on bank deposits
i_M	Interest rate on deposits
K	Stock of capital (real)
L	Stock of loans
lev	Leverage ratio
M	Stock of bank deposits
N	Employment
p	Price level
π	Productivity
pe	Market price of equities
q	Tobin's *q*
r_A	Real interest rate on central bank advances
r_B	Real interest rate on Treasury bills
r_E	Real rate of return on equities
r_L	Real interest rate on loans
r_M	Real interest rate on deposits
rfc	Cash flow ratio
SAV	Savings of the economy
Sg	Savings of the government sector
Sh	Savings of the household sector
Td	Taxes on income
Tf	Taxes on profits
Ti	Indirect taxes
u	Capacity ratio
V	Total net worth of the household sector
Vf	Total net worth of firms
W	Unit wages
WB	Wage bill
Y	Real sales
Yb	Total receipts of banks

Yc	Total receipts of central bank
Yhd	Households disposable income
Yg	Total receipts of the Government sector
Yh	Total receipts of the household sector

A superscript (^{e}) denotes an expected value. A dot ($\dot{}$) above the variable name denotes a rate of change.

Model Parameters

α_i	Propensities to consume
λ_{ij}	Parameters in asset demand functions
γ_i	Parameters in the investment function
η_i	Parameters determining growth in government expenditure
ρ	Mark-up
τ_d, τ_f, τ_d	Tax rates on income, profits, sales respectively
υ_1, υ_2	Parameters in the demand for cash from households and banks, respectively
χ	Parameter to measure workers' strength in the labour market
ϕ	Retained profit ratio
x	Ratio of investment financed by issuing equities

NOTES

1. Firms' capital is valued 'at replacement cost' in Table 5.1. Tobin (1982) and Taylor and O'Connell (1985), for example, value firms' capital 'at market value' (forcing *Vf* to zero), but, in our view, this procedure obscures the analysis unnecessarily. Note also that, since any financial asset must have a counterpart financial liability, rows 1–5 and 7 must add up to zero.
2. See Taylor (1983) among others. Note that accounting consistency requires that the total for each row equals the total for the corresponding column, and therefore from Table 5.2 we can derive a system of 7 accounting identities, one of which is a linear combination of the others.
3. The parameter χ can be endogenized in more sophisticated versions of the model, reflecting different hypotheses about the bargaining power of workers.
4. So as to avoid complications related to inventory cycles.
5. We do not introduce a long-run 'normal capacity' constraint to the model. See DSZ for further results on alternative assumptions.
6. We assume that firms fix their mark-up ρ so that $FT = \rho WB$. Using the definition of profits from equation (5.8), and our derivation for indirect taxes, then $pY - WB - \tau d\ p\ Y = \rho\, WB$. Using equation (5.27) for *WB*, and equation (5.29) for employment *N*, we get equation (5.35) in the text for the price level *p*.
7. Fair (1984, 1994) reports no significant evidence of bank credit constraints in the US economy, so the passive behaviour above may not be a bad empirical approximation. In any

case, the modelling of 'credit crunch' regimes – such as those discussed, for example, by Davidson (1972, p. 280) or more recently by Stiglitz and Greenwald (2003) – would only require the specification of a loan supply demand. In this particular case, however, either the interest rate on bank loans is assumed to fluctuate to clear the market (to make this interest-elastic supply of loans to equal equation (5.16)), or *à la* Stiglitz and Greenwald (ibid.) it does not and aggregate investment has to adjust (that is equation (5.30) has to be dropped) to whatever (inelastic) amount of 'external funds' can be raised by firms to finance it (that is equation (5.16) has to be solved for $p(K_t - K_{t-1})$).

8. See Dos Santos and Zezza (2004) for an extended discussion on the implications of such closures on money endogeneity.
9. We started from parameter values in our behavioural functions that are close to those usually obtained in econometric estimates for the US economy, whenever available. We then proceeded to adjust some of them to obtain 'correct' paths for all endogenous variables: positive profits for all sectors, rates of return on assets that respected relative risk, meaningful values for the stocks of assets and stock-flow ratios, and so on. We investigated the model for stability by simulation, changing parameters one at a time. For instance, model stability depends crucially by the speed of adjustment in expectations: as we move towards 'rational' expectations the model quickly becomes unstable.
10. The macro programs are available from the authors on request.
11. See Lavoie (2003) for a discussion of the endogeneity of the utilization rate in post-Keynesian models.
12. See Lavoie (2003) among others.
13. The shock is administered by keeping the same baseline path for prices, and thus an increase in the real wage corresponds to a drop in the mark-up.
14. Essentially the same points were noted before – with different terminology – by Tobin (1980, 1982), Godley and Cripps (1983), and Godley (1996, 1999), among others. See Kirman (1992).
15. It has been pointed out that the adoption of adaptive expectations may introduce an inconsistency in our approach. However, we tested the model under model-consistent expectations, and we could not determine any stable solution: more work should be done in this direction.

REFERENCES

Backus, D., W.C. Brainard, G. Smith and J. Tobin (1980), 'A model of U.S. financial and non-financial economic behavior', *Journal of Money, Credit, and Banking*, 12(2).

Barbosa Filho, N.H. (2004), 'A simple model of demand-led growth and income distribution', Paper presented at the Conference and available in the Conference web site, 2004.

Davidson, P. (1972), *Money and the Real World*, Armonk, NY: M.E. Sharpe.

Dos Santos, C. (2004a), 'A stock-flow consistent general framework for formal Minskyan analyses of closed economies', Working Paper No. 403, Levy Economics Institute of Bard College, Annandale-on-Hudson, New York. Forthcoming in the *Journal of Post-Keynesian Economics*.

Dos Santos, C. (2004b), 'Keynesian theorizing during hard times', Working Paper No. 408, Levy Economics Institute of Bard College, Annandale-on-Hudson, New York.

Dos Santos, C. and G. Zezza (2004), 'A post-Keynesian stock-flow consistent macroeconomic growth model: preliminary results', Working Paper No. 402, Levy Economics Institute of Bard College, Annandale-on-Hudson, New York.

Eichner, A.S. (1976), *The Megacorp and Oligopoly*, Cambridge: Cambridge University Press.

Fair, R.C. (1984), *Specification, Estimation, and Analysis of Macroeconometric Models*, Cambridge, MA: Harvard University Press.

Fair, R.C. (1994), *Testing Macroeconometric Models*, Cambridge, MA: Harvard University Press.

Foley, D. and L. Taylor (2006), 'A heterodox growth and distribution model', this volume, pp. 75–99.

Franke, R. and W. Semmler (1989), 'Debt financing of firms, stability and cycles in a dynamical macroeconomic growth model', in W. Semmler (ed.), *Financial Dynamics and Business Cycles*, New York: M.E. Sharpe, pp. 38–64.

Godley, W. (1996), 'Money, finance and national income determination: an integrated approach', Working Paper No. 167, Levy Economics Institute of Bard College, Annandale-on-Hudson, New York.

Godley, W. (1999), 'Money and credit in a Keynesian model of income determination', *Cambridge Journal of Economics,* **23**(2):393–411.

Godley, W. and F. Cripps (1983), *Macroeconomics,* Oxford: Oxford University Press.

Kaldor, N. (1966), 'Marginal productivity and the macro-economic theories of growth and distribution', *Review of Economic Studies*, **33**:309–19.

Kirman, A. (1992), 'Whom or what does the representative individual represent?' *Journal of Economic Perspectives*, **6**(2):117–36.

Lavoie, M. (2003), 'Kaleckian effective demand and Sraffian normal prices: towards a reconciliation', *Review of Political Economy*, **15**(1):53–74.

Lavoie, M. and W. Godley (2001–2002), 'Kaleckian growth models in a stock and flow monetary framework: a Kaldorian view', *Journal of Post Keynesian Economics*, **24**(2):277–311.

Minsky, H. (1975), *John Maynard Keynes*, New York: Columbia University Press.

Minsky, H. (1986), *Stabilizing an Unstable Economy,* New Haven: Yale University Press.

Moudud, J.K. (1998), 'Endogenous growth cycles and money in an open economy: a social accounting matrix approach', Ph.D. dissertation, New School for Social Research.

Shaikh, A. (1989), 'Accumulation, finance, and effective demand in Marx, Keynes, and Kalecki', in W. Semmler (ed.), *Financial Dynamics and Business Cycles: New Prospects*, New York: M.E. Sharpe.

Solow, R. (1956), 'A contribution to the theory of economic growth', *Quarterly Journal of Economics*, **70**(1):65–94.

Stiglitz, J.E. and B. Greenwald (2003), *Towards a New Paradigm in Monetary Economics*, Cambridge: Cambridge University Press.

Taylor, L. (1983), *Structuralist Macroeconomics: Applicable Models for the Third World*, New York: Basic Books.

Taylor, L. (2004), *Reconstructing Macroeconomics: Structuralist Proposals and Critiques of the Mainstream*, Cambridge, MA: Harvard University Press.

Taylor, L., and S.A. O'Connell (1985), 'A Minsky crisis', *Quarterly Journal of Economics*, **100**, Supplement: 871–85.

Tobin, J. (1965), 'Money and economic growth', *Econometrica*, **33**(4):671–84.

Tobin, J. (1980), *Asset Accumulation and Economic Activity*, Chicago, IL: University of Chicago Press.

Tobin, J. (1982), 'Money and the macroeconomic process', *Journal of Money, Credit and Banking*, **14**(2):171–204.

Wood, A. (1975), *A Theory of Profits*, Cambridge: Cambridge University Press.

Zezza, G. and C. Dos Santos (2004), 'The role of monetary policy in post-Keynesian stock-flow consistent macroeconomic growth models: preliminary results', in M. Lavoie and M. Seccareccia (eds), *Central Banking in the Modern World: Alternative Perspectives*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

6. Technological progress, income distribution and capacity utilization

Fabio Hideki Ono and José Luís Oreiro

6.1. INTRODUCTION

The long-term stability of capitalist economies has been an exhaustively discussed topic in the history of economic thought. Authors such as Marx (1867), Keynes (1936) and Schumpeter (1934) pointed out the inherently unstable nature of capitalist economies. Extensive interest in the theory of economic growth in the 20th century, starting from the seminal articles from Harrod (1939) and Domar (1946), was followed by a recrudescence of scepticism with regard to the supposedly self-controlled characteristic of these economies. In fact, one of the fundamental findings of the Harrod–Domar model is the so-called 'Harrod principle of instability', which states that any slip from a stable growth-path tends to amplify the economic disequilibrium, resulting in explosive growth trajectories or chronic depressions.

In the post-Keynesian tradition, following on from original formulations from Kalecki (1954), Robinson (1962) and Rowthorn (1981), the matter of stability remains a fairly debated topic. As a result, recent developments in this tradition, such as Dutt (1994), You (1994) and Lima (1999), focus on the analysis of economic stability assuming alternative hypotheses about the relation between: a) growth and income distribution (wage-led accumulation regimes versus profit-led ones); b) technological progress and market concentration; and c) capital accumulation and productivity expansion. These models have demonstrated that instability *à la* Harrod is not an essential attribute of modern capitalist economies. Thereby, the existence of non-linearities in these models creates a bounded instability, and limits the likelihood of explosive growth trajectories or massive decrease in production levels just as predicted in Harrod's original model.

A common assumption adopted in models that follow a post-Keynesian tradition is the constancy of the capital–output ratio: these models explicitly suppose neutral technological progress *à la* Harrod, that is, the type of progress that does not alter the amount of capital that is technically required to produce an additional unit of output. This hypothesis is sustained by two main arguments, a theoretical and an empirical one. From a theoretical standpoint, the assumption of a neutral technological progress *à la* Harrod seems to be the only way of reconciling technological progress with the construction of balanced-growth models (Bresser-Pereira, 1988, p. 49; Solow, 2000, p. 4). From an empirical point of view, the long-term stability of the capital–output ratio was presented by Kaldor (1957) as one of the 'stylized facts' of capitalist economies' growth. Thus, the assumption of a constant capital–output ratio would be justified not only for its theoretical convenience, but also its presumed realism.

Nevertheless, empirical studies indicate that the capital–output ratio is unlikely to be constant in the long term. Table 6.1, reproduced from Maddison (1991, p. 54), clearly shows an upward trend of the capital–output ratio in a group of six developed countries during the period of 1890–1987.

Therefore, the central objective of this chapter is to analyse implications of different assumptions with regard to the behaviour of capital–output ratio and the dynamics of investment towards accumulation regimes and the long-term stability of capitalist economies. In particular, we analyse the effects of distinct hypotheses about the nature of technological progress, according to Harrod's terminology (neutral, capital saving or capital intensive), over conditions for long-term economic stability. To that end we develop a post-Keynesian growth model, in the tradition of Kalecki (1954), Robinson (1962) and Rowthorn (1981), in which (i) the mark-up rate varies in the long term due to a misalignment between the actual profit rate and the 'desired' profit rate;[1] and (ii) the capital–output ratio is not necessarily constant. On the contrary, it may shift as a result of the technological progress that the economy experiences in the long term.

In this context, we demonstrate that the conditions for economic stability can only be fulfilled if technological progress is 'neutral' or 'capital intensive' and the investment is sensitive to fluctuations in mark-up rate.

Table 6.1 Capital–output ratio in selected countries (1890–1987)

	1890	1913	1950	1973	1987
France	n.a	1.64	1.68	1.75	2.41
Germany	2.29	2.25	2.07	2.39	2.99
Japan	0.91	1.01	1.80	1.73	2.77
Netherlands	n.a.	n.a.	1.75	2.22	2.74
United Kingdom	0.95	1.48	1.68	1.96	2.59
United States	2.09	2.91	2.26	2.07	2.30

Source: Maddison (1991, p.54)

However, in the case of 'capital-saving' technological progress or if investment is insensitive to variations in profit margins, then the conditions for stability may never be fulfilled; in other words, these economies would be essentially unstable. The corollary for these assumptions is the attested behaviour of the capital–output ratio during the last 120 years which complies with the long-term stability of capitalist economies. Hence, if these economies are truly unstable, then the reasons for this instability must be pursued elsewhere, except for the type of technological progress.

The chapter is divided into four sections, including this introduction. In Section 6.2 we present the model's framework. In Section 6.3 we examine the long-term dynamics of this model assuming different hypotheses concerning the types of technological progress and the intensity of the accelerator and profitability effects on investment decisions. In Section 6.4 we conduct computer simulations on the model and proceed to perform stress tests on it. Finally, in Section 6.5 we provide a summary of our findings.

6.2. THE MODEL'S STRUCTURE

We suppose an economy where firms produce a homogeneous output and have market power, that is to say, the firms set prices by adding a mark-up rate to unit costs, which is constant in the short term. The mark-up price equation is then written as:

$$p = (1+z)wq \tag{6.1}$$

where, p is the price level, z is the *actual* mark-up rate[2] $(z > 0)$, w is the money wage rate and q is the inverse of labour productivity, a technical coefficient that represents the amount of labour needed per unit of final output. In the following analysis, we will suppose that labour productivity is growing at an exogenous rate α, so that q is decreasing at a rate $1/\alpha$.[3]

Let R be the profit rate generated by the capital stock, assumed given in the short term, u is the degree of capacity utilisation (defined as $u = X/\overline{X}$, where X is the actual output/income and $\overline{X}$ the full-capacity potential output/income), m is the share of profits in income, which is equal to $a/(1+z)$ and σ is the output–capital ratio (defined as $\sigma = \overline{X}/K$, where K is the capital stock value in the economy), which is the reciprocal of the capital–output ratio (Bresser-Pereira, 1988, p. 196). It is straightforward to show that:

$$R = um\sigma \tag{6.2}$$

Therefore, assuming the output–capital ratio to be constant in the short term, then actual profit rate is the product of the actual degree of capacity utilisation and the actual mark-up rate.

By following a post-Keynesian tradition of Kaldor (1956), Robinson (1962) and Pasinetti (1962) we assume the separation between two social classes: capitalists and workers. Capitalists save a constant fraction (s_c) of their income which only consists of profits, while workers 'spend what they earn', in other words, they consume all their wages $(s_w = 0)$. Thus, these assumptions permit us to show that the aggregate saving per capital unit is given as:

$$\frac{S}{K} = s_c R \tag{6.3}$$

The desired growth rate of the capital stock in this economy is denoted by:

$$\frac{I}{K} = \alpha_0 + \alpha_1 m + \alpha_2 u\sigma \quad \text{where: } \alpha_0 > 0,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0 \tag{6.4}$$

Equation (6.4) is nothing less than an investment function. This specification follows Bhaduri and Marglin's (1990) function, by which the desired growth rate of the capital stock is a separable function of m and u. Perhaps the only difference with respect to this previous work consists in including the output–capital ratio explicitly as an independent argument in the investment function.[4]

This formulation is justified by the fact that the accelerator effect of output growth on investment decisions depends upon the rate of capacity utilisation and 'capital productivity' as well. In fact, given the rate of capacity utilisation, the higher *sigma* is, the higher would be the output level associated with the current capital stock in economy, and so, the higher would be the total level of sales. Thus, an increase in output–capital ratio, *ceteris paribus*, would induce a higher investment level due to the accelerator effect.

Finally, if we suppose a closed economy without government, then the goods market equilibrium requires an equivalence between savings per unit of capital and the desired growth rate of capital stock:

$$\frac{S}{K} = \frac{I}{K} \tag{6.5}$$

Substituting (6.2) into (6.3) and the resulting equation into (6.5) we have:

$$\frac{I}{K} = s_c mu\sigma \tag{6.6}$$

Substituting (6.4) into (6.6) and solving the resulting equation for *u*, we have:

$$u^* = \frac{\alpha_0 + \alpha_1 m}{(s_c m - \alpha_2)\sigma} \tag{6.7}$$

Equation (6.7) presents the short-term equilibrium of degree of capacity utilisation, that is to say, the degree of capacity utilisation that equalises planned investment to savings out of profits. To ensure that $u^* > 0$, it is necessary that $s_c m - \alpha_2 > 0$, which implies that the share of profits in income must exceed a critical value m^* denoted by α_2 / s_c.

The effect of a change in the share of profits in income over the short-term equilibrium of the degree of capacity utilisation is given by the derivative below:

$$u_m = \frac{\partial u^*}{\partial m} = -\frac{\alpha_1 \alpha_2 + \alpha_0 s_c}{\sigma (s_c m - \alpha_2)^2} \tag{6.8}$$

We observe from equation (6.8) that an increase in the share of profits in income will result in a decrease in the degree of capacity utilisation. This is a surprising outcome since, in principle, there are two forces exerting pressure against the degree of capacity utilisation. On the one hand, an increase in the share of profits in income gives rise to an increase in aggregate saving because capitalists save a higher share of their income than workers do. This effect tends to reduce the volume of effective demand and thus, the degree of capacity utilisation.

On the other hand, investment depends on the share of profits in income which is a proxy for 'profitability'. Thus investment would rise as a consequence of an income redistribution towards profits, and this would engender a higher volume of effective demand and a higher degree of capacity utilisation. However, according to our specification of the investment function, the first effect is stronger than the second, so that the degree of capacity utilisation will reduce in response to an increase in *m.* Hence, we conclude that a wage-led accumulation regime prevails in the economy under consideration.

The effect of a change in output–capital ratio on the short-term equilibrium of the degree of capacity utilisation is denoted by the partial derivative below:

$$u_{\sigma} = \frac{\partial u^*}{\partial \sigma} = -\frac{u^*}{\sigma} < 0 \tag{6.9}$$

We observe from equation (6.9) that an increase in the output–capital ratio (in other words, a decrease in the capital–output ratio) will result in a fall in the short-term equilibrium of the degree of capacity utilisation.

From (6.8) and (6.9) it is easy to rewrite u^* as an implicit function of m and σ:

$$u^* = u^*(m, \sigma);\ u_m < 0, u_{\sigma} < 0 \tag{6.10}$$

6.3. LONG-TERM DYNAMICS AND STABILITY

In the long term, the share of profits in income and output–capital ratio can no longer be regarded as constant. With reference to income distribution between capitalists and workers, the devising of a 'desired' profit rate by capitalists – which is a 'social convention' that prevails at a particular time (see Bresser-Pereira, 1988, p. 125) – makes the mark-up rate, and also the share of profits in income, *endogenous* variables in the long term. And what if the actual profit rate is different from the 'desired' rate? This may result from three different situations.

First, the degree of capacity utilisation may be below its desired level so that, under competitive conditions, it would be appropriate to reduce the mark-up as an attempt to increase sales. Second, it is also possible that a firm could sell all it expected – when the actual degree of capacity utilisation equals the desired level – but the profit rate is below the desired rate, which means that the firm is selling well but the mark-up rate is too small. In this situation, a natural reaction would be to increase the actual mark-up rate (given the desired mark-up rate) in order to enhance profitability. Third, both the actual mark-up rate and the degree of capacity utilisation could be below the desired level. So the firm sold less and got a profit margin below what was desired. In this case, the firm would have to choose whether to increase the mark-up rate to recover profitability (by losing market share) or to reduce the mark-up rate in order to gain market share. This choice depends on the market structure.

In our model, where prices are set based on a mark-up on unit costs, we assume that markets are oligopolised, and so firms have a high degree of monopoly. Therefore, if the actual profit rate is *lower* than its 'desired' level, then the capitalists would increase the mark-up rate as a device to raise the actual profit rate, since demand is price-inelastic.[5] If the actual profit rate were *higher* than the 'desired' rate, then capitalists would reduce their mark-up in order to lower the actual profit rate down to the level given by this 'desired' rate.[6] Thus, we have the following differential equation:

$$\dot{m} = -\theta\left(R - \bar{R}\right) \qquad \theta > 0 \tag{6.11}$$

where $\bar{R}$ is the 'desired' profit rate and θ determines the speed of adjustment of the profit rate to the 'desired' level.

Lavoie (2002) proposes another adjustment mechanism for profit rate misalignments. When firms do not manage to achieve expected profits, they reduce the mark-up rate and the price level. This behaviour suggests a competitive market structure, where demand is price-elastic and the negative effect of mark-up rate reductions on the profit rate is more than offset by an increase in sales. In this framework, however, capitalists would not be capable of inducing actual profit rate adjustments toward its 'desired' level, because the actual profit rate is ultimately determined by the market and not by firms. Thus capitalists could only adjust their 'desired' profit rate, according to the following equation: $\dot{\bar{R}} = -\theta(\bar{R} - R)$.

If the actual rate of profit is lower than the capitalists' desired level then firms, which have moderate market power, could only affect the actual profit rate indirectly. According to Lavoie's framework, firms' price decisions affect, together with several other factors (for instance the given saving propensities of households), the level of aggregate demand, which in turn determines the degree of capacity utilisation and the actual profit rate. For Lavoie, income distribution has a very strong effect on aggregate demand such that an increase in mark-up rate would be offset by a reduction in the level of capacity utilisation, holding the actual profit rate constant. Lavoie implicitly considers that the actual mark-up rate equals a long-term 'desired' rate. Hence the actual profit rate is lower than 'desired' profit rate only if the degree of capacity utilization is low and firms are selling less than expected. In this case, it is logical to admit that the mark-up rate is too high and to bring it down. The problem is that within Kalecki's (1954) price formation framework, the actual mark-up rate may be different from the desired rate.[7]

As we know, the profit rate depends upon the mark-up rate, which is controlled by firms, and on the degree of capacity utilisation, determined by demand. In our model, demand's price-elasticity depends on the sensitivity of the degree of capacity utilisation to changes in the share of profits in income. If we assume that income distribution has a weak effect on demand – in other words, if the elasticity of u with regard to m is less than 1 – then capitalists can effectively change their profit rate. On these terms, Lavoie's framework would be reasonable only if the elasticity of u with regard to m was more than 1. In equation (6.11), we assume that capitalists, who have a high degree of monopoly, change their mark-up rate in order to adjust the profit rate to an exogenously determined 'desired' level, and then they implicitly change the share of profits in income.[8]

6.3.1. The Dynamics of Different Technological Progress

The type of technological progress that occurs in the economy determines its output–capital ratio dynamics. The output–capital ratio can only be regarded as constant if technological progress is neutral in Harrod's sense. In this configuration we have:

$$\frac{\dot{\sigma}}{\sigma} = 0 \tag{6.12}$$

If technological progress is 'capital saving', then the output–capital ratio will increase over time, indicating that the production of one unit of output requires less and less capital. In this case, we suppose that the output–capital ratio increases at a constant and exogenous rate $h > 0$, in such a manner that:

$$\frac{\dot{\sigma}}{\sigma} = h > 0 \tag{6.13}$$

Eventually, if technological progress is 'capital intensive', then the output–capital ratio will diminish over time, indicating that more and more capital is necessary to produce an additional unit of output. Again, we suppose that the output–capital ratio decreases at a constant and exogenous rate $h < 0$, that is

$$\frac{\dot{\sigma}}{\sigma} = h < 0 \tag{6.14}$$

One could question whether the labour–output ratio would also change due to variations in output–capital ratio. In fact, the model embraces the relationship between these two ratios. The output–capital ratio can be rewritten as

$$\sigma = \frac{X}{K} = \frac{X/L}{K/L} = \frac{\tilde{y}}{\tilde{k}} \tag{6.15}$$

where $\tilde{y}$ is the labour–output ratio and $\tilde{k}$ is the capital–labour ratio. Log-linearising this expression we have $\ln \sigma = \ln \tilde{y} - \ln \tilde{k}$. If we apply time derivates then $\dot{\sigma}/\sigma = (\dot{\tilde{y}}/\tilde{y}) - (\dot{\tilde{k}}/\tilde{k})$. Since $\dot{\sigma}/\sigma = h$ and $\dot{\tilde{y}}/\tilde{y} = \alpha$ then

$$\hat{k} = \alpha - h \tag{6.16}$$

where a hat above a variable denotes the variation rate. Since $\alpha > 0$ then the model allows a negative relation between q (the amount labour needed per unit of final output) and σ which indicates a 'mechanisation' of productive processes.

6.3.1.1. Long-term dynamics when technological progress is neutral

When technological progress is neutral *à la* Harrod, the output–capital ratio is constant and the long-term dynamics of this economy is completely outlined by equation (6.11).

Substituting (6.10) into (6.2) and the resulting equation into (6.11) we obtain the following differential equation that presents the dynamic behaviour of the share of profits in income:

$$\dot{m} = -\theta\left[m\sigma u^*(m,\sigma) - \bar{R}\right] \tag{6.17}$$

Differentiating (6.17) with respect to $\dot{m}$ and m, we have:

$$\frac{\partial \dot{m}}{\partial m} = -\theta\sigma u^*\left(1 - \varepsilon_{u,m}\right) \tag{6.18}$$

where, $\varepsilon_{u,m} \equiv -(m/u)u_m$ denotes the elasticity of the degree of capacity utilisation with respect to the share of profits in income.

This elasticity is an important determinant of economic stability and it differs from one economy to another because of structural factors, namely the propensity to save out of profits (s_c) and wages (s_w). If these two propensities were the same and investment were not sensitive to the profit share then an increase in the share of profits in income would have 'zero sum' effect in aggregate demand and on the degree of capacity utilisation since, although capitalists consume more, workers reduce their spending by the same amount. This situation is discarded, since we assume that $s_w = 0$. If investment decisions are not very sensitive to the profit share (u_m is zero or very low) and/or s_c is also close to zero, then $\varepsilon_{u,m}$ is small. On the other hand, if investment decisions are sensitive to the profit share and/or the propensity to save out of profits is a high value, then the degree of capacity utilisation is elastic with regard to variations in the share of profits in income. Certainly, 'close to' and 'high' are imprecise definitions of intensities. Thus numerical simulations are employed to avoid these ambiguities.

Let m^* be a fixed value from (6.17), that is, the value of m that holds the share of profits in income constant over time. This value is a *stable* *equilibrium* if and only if $\partial\dot{m}/\partial m < 0$ (see Takayama, 1993, p. 336). But this requires the fulfilment of one condition, namely:

$$\varepsilon_{u,m} < 1 \tag{6.19}$$

In words: the long-term equilibrium will be *stable* if and only if the elasticity of the degree of capacity utilisation with respect to the share of profits in income is *less* than unity.

In order to better understand the economic logic beneath this result let us suppose an initial situation where the actual profit rate is below the 'desired' profit rate. In these circumstances, capitalists would increase the mark-up rate to raise the actual profit rate up to the 'desired' level. However, an increase in mark-up rate will result in income redistribution toward profits, which will reduce the volume of effective demand and hence the degree of capacity utilisation. If the fall in the degree of capacity utilisation is very steep, then it will outweigh the effects of an increase in the mark-up rate on the profit rate. Thus, the actual profit rate will decrease instead of increase, which would persuade the capitalists to raise the mark-up rate even further and so evidently the profit rate dynamics will be non-convergent. In order to avoid this instability, the degree of capacity utilisation must present a lower sensitivity to changes in the share of profits in income, in other words, condition (6.19) must be satisfied.

6.3.1.2. Long-term dynamics when technological progress is 'capital saving'

When technological progress is 'capital saving', the dynamic behaviour of the economy is characterised by a system of differential equations in σ (6.13) and m (6.17).

Linearising the system around its steady state and expressing the resulting equations in a matrix form we have:

$$\begin{bmatrix} \dot{m} \\ \dot{\sigma} \end{bmatrix} = \begin{bmatrix} -\theta\sigma u^*(1-\varepsilon_{u,m}) & -m\theta(u^* + \sigma u_\sigma) \\ 0 & h \end{bmatrix} \begin{bmatrix} m - m^* \\ \sigma - \sigma^* \end{bmatrix} \tag{6.20}$$

From (6.9) we have $-m\theta(u^* + u_\sigma \sigma) = -m\theta[u^* - (u^*/\sigma)\sigma] = 0$.

The system of equations represented in (6.20) will be stable only if the *determinant* and the *trace* of the Jacobian matrix are respectively positive and negative (see Takayama, 1993, pp. 407–8). From these relations the conditions for system's stability follow:

$$DET = -h\theta\sigma u^*(1-\varepsilon_{u,m}) \tag{6.21}$$

$$TR = -\theta\sigma u^*(1-\varepsilon_{u,m}) + h \qquad (6.22)$$

From equation (6.21) we may notice that the determinant of the Jacobian matrix will be positive if and only if $\varepsilon_{u,m} > 1$, namely, if the degree of capacity utilisation is very sensitive to changes in the share of profits in income. Notwithstanding, if this condition holds then the trace of the Jacobian matrix will be necessarily positive, making the system unstable. On the other hand, if the previously mentioned elasticity is less than one, then the trace of the Jacobian matrix can be negative but its determinant will also be negative and the system is also inevitably unstable. Therefore, we conclude that in the case of 'capital saving' technological progress the conditions for stability will never be met and the system is intrinsically unstable.

6.3.1.3 Long-term dynamics when technological progress is 'capital intensive'

When technological progress is 'capital intensive', the dynamic behaviour of the economy is characterized by a system of differential equations analogous to that represented in (6.20), with a difference that in the new system we have $h < 0$. Thus, the determinant and the trace of the Jacobian matrix remain expressed by (6.21) and (6.22).

Given that $h < 0$, the determinant of the Jacobian matrix will be positive if the elasticity of the degree of capacity utilisation with respect to the share of profits in income is less than unity. In these conditions, the trace of the Jacobian matrix is necessarily negative and the system is stable. As a consequence, we conclude that in the case of a 'capital intensive' technological progress the economic system is stable only if the degree of capacity utilisation presents low sensitivity to variations in the share of profits in income.

6.3.2. The Acceleration and Profitability Effects over Investment

The investment function specified in (6.4) complies with a post-Keynesian tradition (see Robinson, 1962, chap. 2; Rowthorn, 1981; Bhaduri and Marglin, 1990) and is influenced by an interaction between the profitability and acceleration effects,[9] respectively represented by the terms $\alpha_1 m$ and $\alpha_2 \sigma u$. Investment sensitivity to both effects is directly bound to the values of parameters α_1 and α_2. Consequently, these parameters define the model's long-term dynamics and the particular conditions for a steady-state equilibrium. In the next subsections we identify the attributes required for stability and assess the mentioned effects separately and simultaneously.

6.3.2.1. Acceleration effect

Assuming that $\alpha_1 = 0,$ we isolate the acceleration effect on investment. In this situation, the short-term equilibrium of the degree of capacity utilisation is given as

$$u^* = \frac{\alpha_0}{\sigma(s_c m - \alpha_2)} \tag{6.7a}$$

This will be a positive value if and only if $s_c > \alpha_2 / m = s_c^*$, in other words, the savings out of profits ratio must exceed a critical value s_c^*, which depends inversely upon the share of profits in income.

With regard to the elasticity of the degree of capacity utilisation with respect to the share of profits in income we conclude, after algebraic transformations, that $\varepsilon_{u,m} \equiv -(m/u^*)u_m = (s_c m / s_c m - \alpha_2) > 1,$ given that $\alpha_2 > 0$. Hence, from (6.20) we have $\partial \dot{m} / \partial m = -\theta\sigma u^*(1 - \varepsilon_{u,m}) > 0$, so we deduce that an increase in the share of profits in income would lead capitalists to raise their mark-up rate even further. The explosive trajectory of the share of profits in income suggests the occurrence of a 'worker's euthanasia' in the long term.

6.3.2.2. Profitability effect

Supposing that $\alpha_2 = 0,$ we separately consider the profitability effect over investment. In this case, the equilibrium level of the degree of capacity utilisation is given by:

$$u^* = \frac{\alpha_0 + \alpha_1 m}{s_c m \sigma} \tag{6.7b}$$

and the dynamics of *m* is given by equation

$$\dot{m} = -\theta\left(\frac{\alpha_0 + \alpha_1 m}{s_c} - \overline{R}\right) \tag{6.11b}$$

The steady-state value of *m* is:

$$m^* = \frac{s_c \overline{R} - \alpha_0}{\alpha_1} \tag{6.23}$$

Two conditions result from this outcome: firstly, m^* is a positive value provided that $s_c > \alpha_0 / \overline{R};$ secondly, it is economically inconceivable that $m^* > 1.$ Therefore, to ensure that $1 > m^* > 0$ a coherent combination of parameters is needed, so that $\alpha_0 / \overline{R} > s_c > (\alpha_0 + \alpha_1) / \overline{R}$.

6.3.2.3. Simultaneous operation of acceleration and profitability effects

In this situation we assume that $\alpha_1 > 0$ and $\alpha_2 > 0$. Substituting (6.7) into (6.17), the dynamics of the share of profits in income is written as the equation below:

$$\dot{m} = -\theta\left[\frac{m(\alpha_0 + \alpha_1 m)}{s_c m - \alpha_2} - \bar{R}\right] \tag{6.11c}$$

Deriving (6.11c) with respect to *m* we obtain

$$\frac{\partial \dot{m}}{\partial m} = \frac{\left[\alpha_0\alpha_2 - \alpha_1 m(s_c m - 2\alpha_2)\right]}{(s_c m - \alpha_2)^2}.$$

As we verified in Section 6.3.1.1, the system's stability depends upon the condition $\partial\dot{m}/\partial m < 0$ holding, that is to say, an increase in the share of profits in income must not induce capitalists to raise the mark-up rate further and further. Thus, the savings out of profits ratio must be sufficiently high as to ascertain that:

$$s_c > \frac{\alpha_2(\alpha_0 + 2\alpha_1 m)}{\alpha_1 m^2} \tag{6.24}$$

This condition must hold along the entire trajectory of *m*, from its initial condition until its steady-state value. Additionally, from (6.7), in order that the degree of capacity utilisation is a positive number, another condition must be satisfied:

$$s_c > \frac{\alpha_2}{m} \tag{6.25}$$

Nevertheless, once condition (6.24) holds, so does condition (6.25) since

$$\frac{\alpha_2(\alpha_0 + 2\alpha_1 m)}{\alpha_1 m^2} > \frac{\alpha_2}{m}.$$

6.3.3. Configurations for the Model's Stability

Table 6.2 shows how the above analysis applies to different combinations among technological progress types and specifications of the investment function, and we present the long-term behaviour of the output–capital ratio

Table 6.2 Stability conditions

Technological Progress	Neutral $h = 0$	Capital Saving $h > 0$	Capital Intensive $h < 0$
Investment			
Acceleration Effect $\alpha_1 = 0$	m: Explosive σ: Constant	m: Explosive σ: Explosive	m: Explosive $\sigma \to 0$
Profitability Effect $\alpha_2 = 0$	m: Stable in the long term σ: Constant	m: Stable in the long term σ: Explosive	m: Stable in the long term $\sigma \to 0$
Combined Effects $\alpha_1 > 0$ and $\alpha_2 > 0$	m: Stable in the long term σ: Constant	m: Stable in the long term σ: Explosive	m: Stable in the long term $\sigma \to 0$

and the share of profits in income. We verify that the steady-state equilibrium is a feasible result in the long term only if technological progress is neutral or capital intensive and, besides, if the investment is susceptible to variations in the share of profits in income.

6.4. COMPUTER SIMULATIONS

To evaluate the dynamics and interactions between the variables of the model for different types of technological progress we carried out computer simulations,[10] by adopting top-down methodology.[11] So, we assigned economically plausible[12] values to the model parameters, according to Table 6.3 below.

As initial values for the economy under consideration, we used a share of profits in income of 0.3[13] and an output–capital ratio of 0.5, which is tantamount to a capital–output ratio of 2 and is within the range of Maddison's (1991) estimates. When technological progress is capital saving or capital intensive we suppose respectively a variation in the output–capital ratio of between 0.01 and –0.01, that is to say, a variation of 1 per cent in each period. We arbitrarily stipulate 70 periods as the range for the analysis.

In simulations we examined the dynamics of the share of profits in income (m), the degree of capacity utilisation (u^*), the output–capital ratio (σ) and the profit rate over time and under different assumptions with regard to technological progress (neutral, capital saving and intensive) and to investment sensitivity to changes in the profit margins and the degree of capacity utilisation. With reference to the latter we focused on the trajectory

Table 6.3 Parameters

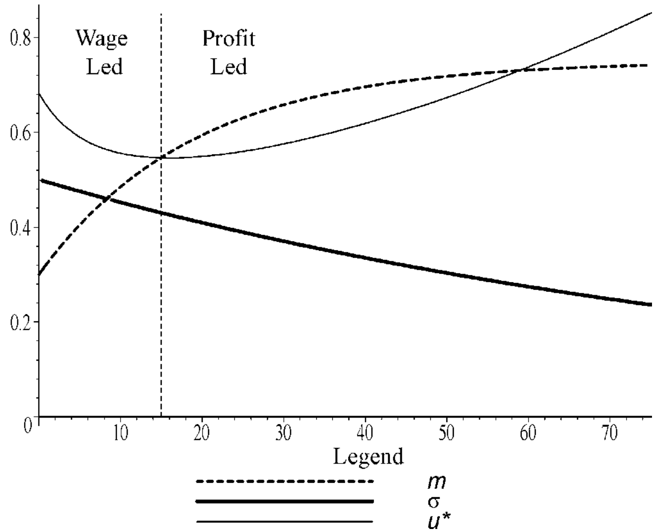
Parameter	Value	Description
θ	0.5	Adjustment factor of the profit rate
α_0	0.2	Autonomous term of the investment function which represents the animal spirits of capitalists
$m^* = \dfrac{s_c\bar{R} - \alpha_0}{\alpha_1}$	0.3	Coefficient that expresses investment sensitivity to profit fluctuations
α_2	0.2	Coefficient that measures the accelerator effect, namely the influence of changes in the degree of capacity utilisation and the output–capital ratio on investment
$\bar{R}$	0.15	Desired profit rate
s_c	0.28	Savings out of profits

over time of the variables under consideration in three distinct scenarios, which are: $\alpha_1 = 0$ (investment is not sensitive to changes in profit margins); $\alpha_2 = 0$ (investment is not sensitive to changes in the degree of capacity utilisation) and $\alpha_1 > 0$, $\alpha_2 > 0$ (the investment is sensitive to both changes in m and u).

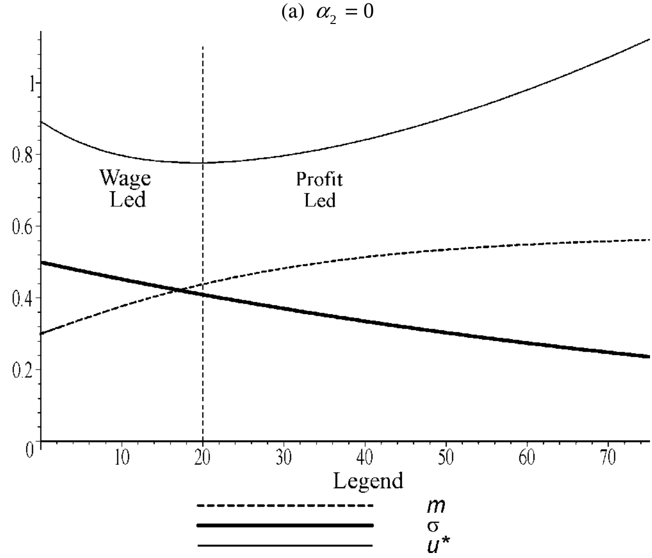
Simulations did not bring any surprising result when compared with those displayed in Table 6.2, except when technological progress was capital intensive ($h < 0$) and the investment was sensitive to changes in the share of profits in income $(\alpha_1 \neq 0)$. In these cases, an endogenous transition of accumulation regimes arose. Thus we concentrate the analysis only on this situation.

We observe economic dynamics when we disregard the acceleration effect on investment and consider only the profitability effect $(\alpha_2 = 0)$ and when profitability and accelerator effects on investment decisions are considered all together ($\alpha_1 > 0$ and $\alpha_2 > 0$). In both cases, the share of profits in income and the output–capital ratio converged toward their own steady-state equilibrium and σ fell to zero in the long term.

At the beginning of the analytical period, m rose sharply which caused a decrease in the degree of capacity utilisation (see equation (6.23)). However, as σ declined (tending to zero) and m increased slightly, then the degree of capacity utilisation reversed its downward trend and displayed increasing variation rates. In this situation, output grew more slowly than the capital stock growth rate (because technological progress was capital intensive: $\sigma \to 0$), hence it imposed an additional effort on the installed productive capacity in order to maintain the equilibrium level of the share of profits in income. Interestingly, there was a wage-led accumulation regime in the initial stage, when an increase in the share of profits in income induced a reduction of u^*. However, the behaviour of u^* endogenously changed[14] over time and the model exhibited a profit-led accumulation regime. In other words, as m increased it raised the degree of capacity utilisation.



(a) $\alpha_2 = 0$



(b) $\alpha_1 > 0$ and $\alpha_2 > 0$

Figure 6.1 Endogenous transition of accumulation regimes

In Figure 6.1(b), the degree of capacity utilisation starts at 0.892 (u_0^*) and falls to 0.777 $(u_{\min}^*)$ in the 19th period (t_{ch}), when the accumulation regime changes. In the long term, the share of profits in income tends to become stable at 0.576 (m_∞). These facts are gathered in the first row of Table 6.4 and will be used as a standard to test parameter sensitivity.

Even though u^* and σ show opposite trends in the long term, both forces are offset resulting in the profit rate convergence (see equation (6.2)) toward the 'desired level': $\lim_{t\to\infty} R = \bar{R} = 0.15$. As a consequence, we conclude, from equation (6.6), that the investment rate converges to 0.042.

Two interesting outcomes derive from numerical simulations:

1. In all cases listed in Table 6.2, m started at 0.3 but rose over time which implied a higher income concentration on profits (possessed by capitalists) whereas the degree of capacity utilisation fell in almost every case, except for $\alpha_2 = 0$ and $h < 0$ and for $\alpha_1 > 0$, $\alpha_2 > 0$ and $h < 0$. We then deduce that investment sensitivity to changes in $u\sigma$ and m is only relevant to the dynamics of the degree of capacity utilisation when technological progress is capital intensive.
2. The profit and investment rate remained relatively stable in all situations, indicating that capitalists are well able to sustain these rates. Simulation for all different combinations included in Table 6.2 also revealed that neither profit rate nor investment rate are determined by the type of technological progress and hence, by changes in the output–capital ratio.

6.4.1. The Robustness of Results

In order to verify the robustness of the endogenous transition, we submitted the parameters to a stress test which demonstrated their sensitivity. We considered the investment function to be dependent on both accelerator and profitability effects $(\alpha_1 > 0$ and $\alpha_2 > 0)$ and then we changed some parameters in Table 6.3, *ceteris paribus*, in 'capital intensive' technological progress.

Table 6.4 describes the behaviour of the degree of capacity utilisation, the share of profits in income, the elasticity of the degree of capacity utilisation with respect to the share of profits in income and the period when the transition occurs (t_{ch}). Conditions denoted by equations (6.24) and (6.25) impose some restrictions on parameter values that fulfil some notional limits on variables (for example, $0 < m < 1$).

The tests disclosed the sensitivity of the share of profits in income to the 'desired' profit rate, and to the determinants of the investment decision

Table 6.4 Robustness test – endogenous transition

Parameter	Value	Values u_0^*	$u_{\min}^*$	m_∞	$\varepsilon_{u,m}$	t_{ch}
Original values	See Table 6.3	0.892	0.777	0.576	0.997	19
α_0	0.021	0.923	0.856	0.527	1.007	17
	0.015	0.738	0.542	0.790	0.932	19
α_1	0.035	0.938	0.916	0.454	0.963	10
	0.020	0.800	0.524	0.970	1.076	28
α_2	0.021	0.906	0.804	0.564	1.017	19
	0.015	0.828	0.683	0.631	0.903	18
$\overline{R}$	0.155	0.892	0.785	0.634	0.997	20
	0.145	0.892	0.837	0.515	0.997	16
s_c	0.035	0.682	0.458	0.981	0.925	22
	0.027	0.950	0.938	0.471	1.017	10
θ	0.750	0.892	0.720	0.576	0.997	18
	0.250	0.892	0.874	0.576	0.997	12
h	–0.020	0.892	0.874	0.576	0.997	6
	–0.005	0.892	0.685	0.576	0.997	33

(α_0, α_1 and α_2). For example, a subtle increment of 0.005 in the 'desired' profit rate raised the long-term value of the share of profits in income from 0.576 to 0.634. Besides, an increase of 0.001 in α_0 and α_2 changed the initial degree of capacity utilisation from 0.892 to 0.923 and 0.906, respectively. For values of $\overline{R}$ higher than 0.16 some variables simply became inconsistent (m_∞ was beyond 1). These results indicate that the standard values of Table 6.3 define a system which is close to its upper threshold, since significantly lower values of α_0 and α_1 produced no inconsistent outcomes.

Changes in θ and h had no effects over the long-term values of the share of profits in income; they only changed the speed of transition. For example, the endogenous transition occurred in period 6, when h was equal to –0.02, or in period 33, when h was equal to –0.005.

6.5. CONCLUDING REMARKS

In the previous sections we analysed the dynamics of some economic variables: the share of profits in income, the output–capital ratio, the profit and investment rate under different assumptions about the type of technological progress and the specification of the investment function. All these factors enable us to examine the conditions for the long-term stability of capitalist economies.

We verified that the type of technological progress is essential to determine the requirements for economic stability. Thus, if technological progress is capital saving then the conditions for stability will never be met, which implies that such a capitalist economy is necessarily unstable. Stability is a probable result only if technological progress is neutral or capital intensive. Under these situations, it is a necessary condition for stability that the elasticity of the degree of capacity utilisation with respect to the share of profits in income is *less* than unity, that is to say, changes in functional income distribution must have a slight influence on effective demand, and consequently, on the degree of capacity utilisation. Another determinant of long-term stability is the prevalence of a high propensity to save out of profits. If this ratio is not comparatively low, it could not ensure the equilibrium (at a positive value) of the share of profits in income, in other words, in face of an increase in the share of profits in income, the entrepreneurs would be impelled to increase the profit margin even further, which would cause the profit's trajectory to be explosive. Besides, if investment is quite dependent on the accelerator effect, which means that profit changes do not affect investment, then the share of profits in income will not achieve its long-term equilibrium.

Computer simulations have revealed that, despite the types of technological progress and the specifications of the investment function, the profit rate has kept stable during its long-term trajectory. Variation range in the profit rate, from its initial level until equilibrium, was relatively small. This outcome denies Marxist predictions that the profit rate would inexorably decline over time.

Moreover, when technological progress is capital intensive and investment is sensitive to changes in the share of profits in income, the surprising phenomenon of an endogenous transition from a wage-led to a profit-led accumulation regime arises, the reason being the simultaneous effect of the decrease in output–capital ratio, and increase in the share of profits in income up to an equilibrium level. In all other situations, as income distribution becomes more concentrated on profits we may observe a decrease in the degree of capacity utilisation (in some situations it falls to zero), which indicates a wage-led accumulation regime during the whole economy's trajectory.

Recent empirical studies about the long-term dynamics of capitalist economies have shown an apparent upward trend in the capital–output ratio in the last 120 years. Hence, we conclude that technological progress has been, up to the present, 'capital intensive' *à la* Harrod. Unless there are strong motives to believe that the growth rate of the capital stock is independent of the share of profits in income, the requirements for the long-term stability of capitalist economies would have been satisfied. In other words, technological progress, at least in its present state, is not the source of economic instability.

Nevertheless, instability is actually a distinctive attribute of capitalist economies. According to Maddison (1991), the average growth rate of developed capitalist countries has varied substantially over time. Long periods of rapid growth – such as the 'golden age' of capitalism during the period of 1950–73 – alternate with periods of semi-stagnation or moderate growth.

The post-Keynesian literature has pointed out two alternative sources for economic instability. The first, with a Marxist inspiration, would involve a class struggle between capitalists and workers. Goodwin (1967) demonstrated that a class conflict may breed a limit cycle dynamics for the share of wages (and profits) in income and for the unemployment rate. The second, inspired in Minsky's (1975, 1982) approach, is based on the interaction between Harrod's principle of acceleration and the endogenous money supply assumptions. This approach is formally developed by Calvet (1999), among other authors. It follows from these analyses that the causes of economic instability might be investigated by taking into account these alternative sources except for technological progress.

NOTES

1. This assumption is inspired by Bresser-Pereira (1988). He states that: 'the hypothesis, therefore, is that firms, specially in oligopolistic sectors, would establish some sort of "desired" profit-rate, that would be historically determined, according to what managers and share-holders consider to be a reasonable profit rate. This rate would probably be estimated around 10 per cent to 15 per cent of the firm's total capital' (ibid., p. 125).
2. According to Kalecki (1954), the *actual* mark-up rate may be different from the *desired* mark-up rate. The desired mark-up rate results from a firm's long-term strategic decisions; it is also 'a tacit agreement among firms of an industry to protect 'profits', and consequently to increase prices in relation to unit prime costs' (Kalecki, 1954, p. 17). This desired mark-up rate is expected to lead the profit rate to its long-term value when the degree of capacity utilisation reaches its desired level. On the other hand, with regard to the Kaleckian mark-up theory, Possas (2005) agrees that the actual mark-up rate is a 'sort of compromise between

the desired mark-up by a firm (or its long run strategic mark-up) and current competitive conditions. While low cost firms enjoy the advantage of making additional profits in the short run, in excess of what would result from applying the strategic mark-up, high cost firms sacrifice their desired mark-up for keeping their market share' (Possas, 2005, p. 12 *apud* Silverberg, 1987, p. 130). In other words, to define their actual mark-up rate firms take into account not only the desired profit rate in the long term, but the actual demand for their goods.

3. Thus, equation (6.1) could be rewritten as: $w/p=[1/(1+\overline{z})](1/q)$. Since the mark-up $(\overline{z})$ is constant in the short term, any change in labour productivity (q) is entirely incorporated into real wages (w/p). This effect implies that unions and workers succeed well in bargaining with capitalists, which also suggests the prevalence of the second stage of capitalism as stated by Kaldor (1957).
4. Bhaduri and Marglin (1990, p. 380) define their investment function in an implicit form: $I=I(m,u)$, given that $I_m>0$ and $I_u>0$. But, unlike equation (5.4), the authors implicitly consider the influence of the output–capital ratio on investment decisions, since σ affects the average rate of profit and it is assumed given in the short period (Bhaduri and Marglin, 1990, p. 379).
5. We must emphasize that actual profit rate will only rise, in response to an increase in the mark-up rate, if the degree of capacity utilisation has little sensitivity to mark-up rate variations. The reason for this is the prevalence of a wage-led accumulation regime in the economy. Under these circumstances, an increase in the mark-up rate will raise the share of profits in income that would ultimately result in a lower degree of capacity utilisation. If the decline in degree of capacity utilisation is very steep, then the actual profit rate may decline by virtue of an increase in the mark-up rate.
6. Why should capitalists deliberately undertake measures to reduce their profit rate? A possible answer for this question is given by the theory of 'barriers to new competitors' from Bain (1956) and Sylos-Labini (1962). According to this theory, the desired profit rate can be considered the rate that is compatible with the long-term stability of an industry, that is to say, the maximum profit rate that would not induce the entry of new competitors in the industry. Hence, if the actual profit rate is above its 'desired' level, then this would invite new competitors to enter and consequently, in the long term, it would cause a decrease in profits and reduce the market power of established firms. A post-Keynesian macroeconomic model consistent with the theory of barriers to new competitors is presented in Oreiro (2004).
7. Kalecki (1954, p. 21) states that 'changes in the ratio of proceeds to prime costs for a single industry which, according to the above, is determined by changes in the degree of monopoly, reflect changes in conditions particular to that industry. For instance, a change in the price policy of one big firm may cause a fundamental change in the degree of monopoly in that industry'. Moreover, the actual mark-up as a 'factor of "protection" of profits, is especially apt to appear during periods of depression … As a result there is a tendency for the degree of monopoly to rise in the slump, a tendency which is reversed in the boom' (ibid., pp. 17–18).
8. In Lavoie's specification, the 'desired' profit rate is endogenous in the long term, and in our terms *mutatis mutandis*, he would have $\theta<0$. Nevertheless, this very idea is not free from criticism. For example, what if the actual profit rate were so low that the 'desired' profit rate were likely to be a negative value, in Lavoie's rule? In this situation raising the mark-up rate would be a legitimate expedient to avoid bankruptcy. Lavoie assumes that firm market power is not very strong. Moreover, he does not seem to acknowledge that the 'desired' profit rate may be restrained within some institutional limits, as proposed by Bresser-Pereira (1988, p. 125).

9. The acceleration effect occurs when investment, which is a decision to enlarge productive capacity, reacts to current or foreseen changes in aggregate demand (that in our analysis is represented by the degree of capacity utilisation). When profit rates increase, capitalists are impelled to expand productive capacity, that is, to invest and this is called profitability effect.
10. To process the numerical simulations we used Maple 7.0 from Waterloo Maple Inc.
11. This procedure involves adjustment of a pre-existing economic model to a numerical simulation environment. An opposite method would involve the creation of a model designed to perform simulations. The first method entails some difficulties in setting up real and/or plausible values for the variables and parameters. However it enriches the assessment of interactions among the selected variables, which is hard to accomplish with ordinary quantitative analysis.
12. For example, the savings out of profits ratio was assigned with the purpose of fulfilling conditions (6.23) and (6.24). It is not 'manna from heaven' since a savings out of profits ratio of 28 per cent seems to be a reasonable value. For example, in 1998, gross domestic savings in proportion to GDP reached 23 per cent in Italy (World Bank, 2001). Besides, these numbers are also consistent with Samuelson's 'principle of correspondence' that corroborates the reasoning that values might be chosen in order that the model's dynamics would prove to be minimally realistic.
13. In 1998, according to the World Bank (2001) 'World Development Indicators', household final consumption expenditure, which is a proxy for the share of wages in income, totalled 59 per cent of GDP in Italy and 67 per cent in the United States. Considering the functional income distribution between wages and profits this would correspond to a share of profits in income of about 30 per cent.
14. Bhaduri and Marglin (1990) identified the requisites for profit-led and wage-led accumulation regimes. Given that their model involved comparative statics analyses, the transition from one accumulation regime to another does not take place over time but it would depend on the degree of capacity utilisation level. By using our specification for variables and parameters we obtain the result, in accordance with Bhaduri and Marglin (1990), that: a) a wage-led accumulation regime occurs if $u^* > (\alpha_1 / s_c)$; b) a profit-led accumulation regime occurs if $u^* < (\alpha_1 / s_c)$. This relation is not supported by Figure 6.1, because we observe situations where distinct accumulation regimes occur for identical values of u^* . This feature is attributed to different investment function specification, which now considers output–capital ratio variations, and to the different type of mathematical analysis (comparative statics versus dynamic analysis).

REFERENCES

Bain, J. (1956), *Barriers to New Competition*, Cambridge, MA: Harvard University Press.

Bhaduri, A. and S. Marglin (1990), 'Unemployment and the real wage: the economic basis for contesting political ideologies', *Cambridge Journal of Economics*, **14**(4): 375–95.

Bresser-Pereira, L.C. (1988), *Lucro, Acumulação e Crise*, São Paulo, Brasiliense, 2nd edn.

Calvet, J.G. (1999), 'Los ciclos: aspectos reales y financieros', in J. Bricall and O. Juan (eds), *Economia Política del Crecimiento, Fluctuaciones y Crisis*, Barcelona: Ariel, pp. 139–72.

Domar, E. (1946), 'Capital expansion, rate of growth and employment', *Econometrica*, **14**:137–47.

Dutt, A.K. (1994), 'On the long-run stability of capitalist economies: implications of a model of growth and distribution', in A.K. Dutt (ed.), *New Directions in Analytical Political Economy*, Aldershot, UK and Brookfield, USA: Edward Elgar, pp. 93–120.

Goodwin, R.M. (1967), 'A growth cycle', in C.H. Feinstein (ed.), *Socialism, Capitalism and Economic Growth*, Cambridge: Cambridge University Press.

Harrod, R. (1939), 'An essay in dynamic theory', *Economic Journal*, **49**:14–33.

Kaldor, N. (1956), 'Alternative theories of distribution', *Review of Economic Studies*, **23**(2):83–100.

Kaldor, N. (1957), 'A model of economic growth', *Economic Journal*, **67**:591–624.

Kalecki, M. (1954), *The Theory of Economic Dynamics*, London: Allen & Unwin.

Keynes, J.M. (1936), *The General Theory of Employment, Interest and Money*, London: Macmillan Press.

Lavoie, M. (2002), 'The Kaleckian growth model with target return pricing and conflict inflation', in M. Setterfield (ed.), *The Economics of Demand-Led Growth: Challenging the Supply-Side Vision of the Long Run*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 172–88.

Lima, G.T. (1999), 'Progresso tecnológico endógeno, crescimento econômico e distribuição de renda', in G.T. Lima, J. Sicsú and L.F. Paula (eds), *Macroeconomia Moderna: Keynes e a Economia Contemporânea*, Rio de Janeiro: Editora Campus, pp. 190–207.

Maddison, A. (1991), *Historia del Desarollo Capitalista*, Barcelona: Ariel.

Marx, K. (1867), *Capital*, Harmondsworth: Penguin.

Minsky, H.P. (1975), *John Maynard Keynes*, New York: Columbia University Press.

Minsky, H.P. (1982), *Can 'It' Happen Again?*, New York: M.E. Sharpe.

Oreiro, J.L. (2004), 'Accumulation regimes, endogenous desired rate of capacity utilization and income distribution', *Investigación Económica*, **63**(248):41–70.

Pasinetti, L. (1962), 'Rate of profit and income distribution in relation to the rate of economic growth', *Review of Economic Studies*, **29**(4): 267–79.

Possas, M. (2005), 'A multisectoral micro-macrodynamic model', UFPR Working Papers.

Robinson, J. (1962), *Essays in the Theory of Economic Growth*, London: Macmillan.

Rowthorn, R. (1981), 'Demand, real wages and economic growth', *Thames Papers in Political Economy*, Autumn.

Schumpeter, J.A. (1934), *The Theory of Economic Development*, Cambridge, MA: Harvard University Press.

Silverberg, G. (1987), 'Technical progress, capital accumulation and effective demand: a self-organization model', in D. Batten, J. Casti and B. Johansson (eds), *Economic Evolution and Structural Adjustment*, Berlin: Springer Verlag.

Solow, R. (2000), *Growth Theory: An exposition*, Cambridge, MA: MIT Press.

Sylos-Labini, P. (1962), *Oligopoly and Technical Progress*, Cambridge, MA: Harvard University Press.

Takayama, A. (1993), *Analytical Methods in Economics*, Michigan: The University of Michigan Press.

World Bank (2001), *World Development Indicators*, CD-ROM.

You, J.I. (1994), 'Macroeconomic structure, endogenous technical change and growth', *Cambridge Journal of Economics*, **18**:213–34.

7. Demand-led growth and the classical approach to value and distribution: are they compatible?*

Graham White

7.1. INTRODUCTION

The Sraffian objection (Garegnani, 1992; Kurz, 1994) to the neo-Keynesian approach to growth, at least in terms of its embodiment in and interpretation of the so-called Cambridge equation, stressed the seeming inconsistency between this interpretation and the notion of a Keynesian long-run, to the extent that the latter centres on the idea of investment being determined independently of saving and aggregate demand determined independently of the level of output and output capacity. Yet the counter-claim has been advanced more recently that, given technology, in so far as the rate of profit is determined along Sraffian lines (either via an exogenous real wage or an exogenous rate of interest) the long run, interpreting this as the steady state, is not Keynesian. Interestingly, this latter view is not limited to those adopting a non-Sraffian position – it can be found amongst some adopting a Sraffian or 'classical' approach to value and distribution (for example, Dumenil and Levy, 1999). This result has also been noted by Commendatore et al. (2003, p. 123).[1] Such inferences are particularly interesting in light of the role played by a Sraffian approach in the critique of marginalist theory and therefore of the supposed foundations for the thesis of supply-driven growth in capitalist economies.

To add to the complexity in the debate, the claim of inconsistency between steady state analysis and a Keynesian long run has seen some, notably among the Sraffian camp, point to the need to reject the usefulness of the former; in particular, growth analysis based on comparisons of 'fully-adjusted' situations, where the latter refer to capacity being aligned with demand in such a way that utilization is at the desired rate, and where the normal or general rate of profit corresponds with the actual rate of profit (Trezzini, 1995, 1998; Palumbo and Trezzini, 2003).

Theis chapter is an attempt to clarify some of the issues in this debate about the relation between growth and income distribution, with particular reference to the Sraffian analysis of value and distribution. It is particularly concerned with inferences in recent literature that the Sraffian approach to value and distribution has implications which are at odds with a Keynesian or demand-constrained view of long-run growth.

Section 7.2 opens the discussion with a consideration of different interpretations of the Cambridge equation and the suggestion of inconsistency between Sraffian approaches and a Keynesian long run. Section 7.3 focuses on a recent attempt by Lavoie to reconcile Sraffian and Kaleckian approaches in the context of long-run growth. This discussion provides an opportunity to consider whether a Keynesian view of growth necessarily requires a growth rate exogenous with respect to distribution. Section 7.4 deals with what has been suggested as a separate channel through which growth and accumulation could supposedly influence normal distribution, namely, through the adaptation of normal utilization over time to the actual rate of capacity utilization. It is suggested that this particular channel is not as clear cut as has been supposed in the literature. Section 7.5 outlines an alternative way of approaching (at least steady state) growth which treats the independence of demand in terms of the existence of non-capacity-creating autonomous demands. It is relatively straightforward in this case to demonstrate analytically the consistency of a Sraffian explanation of distribution and a Keynesian long run, interpreted in terms of an exogenous growth rate. Section 7.6 considers this alternative further and in particular the notion of the warranted growth rate being governed by the growth of autonomous, non-capacity-creating demand; including recent doubts from both post-Keynesian and Sraffian standpoints. Section 7.7 considers in terms of a simple model the possible complexities which arise, where autonomous demand is treated in terms of public sector and export expenditures. Section 7.8 provides some brief concluding notes.

7.2. THE CAMBRIDGE EQUATION: EXOGENOUS GROWTH OR EXOGENOUS DISTRIBUTION?

At a formal level the problem which presents itself in the interpretation of the Cambridge equation is as follows: looking at the Cambridge growth equation

$$g = s_c r \tag{7.1}$$

from a Sraffian standpoint one would take r as being determined exogenously, to the extent that r represents the 'normal' rate of profit.

Supposing, as is usually the case, that s_c is given, then g becomes endogenous. The 'problem' with this result is how to reconcile it with g being exogenous. Certainly, if one supposes that normal capacity utilization is maintained, then even with investment being exogenous, g is still endogenous as the capital stock adjusts to maintain normal capacity utilization.[2] A problem appears to exist to the extent that an exogenous (at least with respect to s_c and r) rate of growth is seen as essential to a demand-led perspective on growth.

In effect, this feature of the Cambridge equation and its more common neo-Keynesian interpretation has given rise to two interesting positions, symptomatic in a sense of differences between Sraffian and Kaleckian perspectives on growth. The first or Sraffian position, most notably put by Garegnani (1992) and Kurz (1994), is that the Cambridge equation could not be seen as a way of determining the normal rate of profit consistent with the acceptance of the Keynesian principle of effective demand in the long run. If r refers to the normal rate of profit and thus to a utilization rate equal to normal, the rate of accumulation could not refer to the actual level. If it did, this would imply an actual rate of utilization equal to normal and thus presuppose that investment was governed by the flow of saving at normal utilization. Looked at this way, the neo-Keynesian interpretation of the Cambridge equation, as an explanation of the normal rate of profit, is itself distinctly un-Keynesian.[3]

The second position referred to above, put recently by Lavoie in an attempt at reconciling Sraffian and Kaleckian approaches, stresses that with r being exogenous – at least determined independently of s_c and g – in the Sraffian interpretation, g is thereby determined in a non-Keynesian way. Thus, having identified a similarity in the treatment of distribution between classical and 'new growth theory', Lavoie goes on to note that '[f]rom the point of view of effective demand, these analyses are supply-led in the sense that higher savings generate higher growth rates' (Lavoie, 2003 p. 54). As Commendatore et al. (2003) note explicitly, 'along the equilibrium path, effective demand does not affect growth' (p. 123);[4] moreover, a rise in s_c would be associated with a rise in the steady state rate of growth. Support for this interpretation of what is implied by a classical/Sraffian explanation of income distribution is also to be found in the work of Dumenil and Levy who make use of the unfortunate (for those who see the relevance of both effective demand for the long run while adopting a Sraffian approach) phrase 'being Keynesian in the short-term and classical in the long-term'.[5]

In view of this second position, together with the significance of the Sraffian critique of the orthodox belief in Say's Law, it is not surprising that at least for some of those who accept both a Sraffian view of distribution and a desire for an explanation of growth which gives prominence to effective demand there exists serious doubts about the degree of insight provided by the Cambridge equation, when taken to refer to normal distribution. Indeed, criticism along these lines has gone beyond criticism of the usefulness of the Cambridge equation and has questioned the usefulness of normal utilization and steady state analysis for growth theory (for example, Ciccone, 1986; Garegnani, 1992; Trezzini, 1998; Garegnani and Palumbo, 1998; Palumbo and Trezzini, 2003).[6]

For the present, the significant point to note is that what has emerged as a key difference in the Sraffian and Kaleckian positions may be seen as manifesting itself in two radically different opinions about the Cambridge equation. On the one hand, Kaleckians would appear to have little problem with the usefulness of the concept for growth theory, but with the interpretation that distribution is endogenous to the growth process. On the other hand, some (if not all) of those adopting a Sraffian approach to distribution see any inconsistency between it and a Keynesian view of growth in the context of the Cambridge equation[7] as indicative not of the need for an alternative approach on distribution or growth but rather of the need to reconsider the worth of steady state analysis.

In the following two sections we consider two different ways in which theorists have sought to maintain the view that the (normal) rate of profit will be endogenous to the growth process, that is influenced by changes in the rate of accumulation and/or the saving propensity: through permanent effects on the profit share; and through effects on the normal rate of capacity utilization.

7.3. LAVOIE ON SRAFFIAN AND KALECKIAN APPROACHES

We turn first to Lavoie's recent attempt at reconciliation of Kaleckian and Sraffian views on growth, since a consideration of his views may help clarify some of the issues at stake. Lavoie's argument aims at dealing with two matters: first, the Sraffian criticism of the conventional interpretation of the Cambridge equation; and, second, the need to make consistent the Kaleckian explanation of the steady state rate of capacity utilization and the correspondence between actual and normal rates in equilibrium.[8] The latter correspondence along with the equality of actual and normal rates of profit are required for the economy to be in a 'fully-adjusted' position. Thus, a key aim of Lavoie's analysis is to explain growth in a manner consistent with 'full adjustment'[9] as a feature of the steady state and consistent with the neo-Keynesian interpretation of the Cambridge equation. Embodied in the latter is not just the possible role of the rate of accumulation in determining the normal rate of profit, but what is also referred to as the 'paradox of thrift': in terms of Kaleckian growth models this amounts to the result that changes in the saving propensity of capitalists have effects of the opposite sign on both the steady state rate of accumulation and the normal rate of profit.

To ensure a steady state with equality between actual and normal utilization rates and as a result between actual and normal profit rates, Lavoie supposes that divergences between actual and normal utilization rates lead to divergences between the actual and target (that is normal) rates of return. The latter divergence feeds back on the system in a way which brings the actual utilization rate into line with its normal level. In other words, the normal or 'target' rate of return eventually responds to the actual rate of return: prices will change to reflect the new adjusted target rate, this also impacting on the real wage (the money wage being taken as given). This adjustment in the real wage in turn affects expenditures in a way which moves the actual utilization in the direction of the normal rate. Thus, for example, if the actual rate of return exceeds the target rate, this will reflect an actual utilization higher than the normal rate. The target rate of return starts to rise, the real wage falls, actual utilization is reduced as a consequence, so that the actual and target rates both move towards each other. As Lavoie notes, in the long run, actual utilization adapts to an exogenous normal utilization, while the target rate of return becomes endogenous.

What is significant about this argument is that it also apparently represents Lavoie's response to the Sraffian criticism of conventional interpretations of the Cambridge equation. Effectively this response is to allow for changes in the actual ratio of profits to capital to generate changes in the normal profit rate, via permanent changes in the profit share. Thus changes in distribution, both actual and normal, would occur from autonomous changes in the rate of accumulation (that is changes in the constant term of the 'accumulation function' used – see next paragraph) in Lavoie's analysis.[10]

Particularly interesting for the purposes of the present discussion are two modifications to the model, both of which are intended to allow for a classical/Sraffian account of the rate of profit. These modifications appear at first sight to confirm the possibility that a model with a classical/Sraffian approach to distribution may sit alongside what at first sight appear to be non-Keynesian outcomes. Both modifications involve not only the requirement of 'full adjustment' in the steady state, but seemingly also a Sraffian/classical determination of the rate of profit. In both cases, the normal rate of profit exhibits no tendency to move in line with the actual rate of profit, via changes in the profit share, price and real wage levels. In the first of these modifications, based on an interpretation of Dumenil and Levy (D–L) (1999), a divergence between actual and normal utilization triggers real interest rate changes in such a manner as to bring actual utilization into line with the normal rate and thus to correct any divergence between actual and normal profit rates. In his interpretation of D–L, Lavoie uses an accumulation function in which, apart from an autonomous component, the rate of accumulation is positively related to capacity utilization and to the excess of the rate of profit over the real rate of interest (p. 61). Hence, for example, a rise in the propensity to save would generate a short-run fall in utilization below normal, and in the actual rate of profit below normal, triggering a fall in the real rate of interest. The latter would bring utilization and the profit rate back into line with their respective normal levels. With the return of the rate of profit to its normal value, but a lower real rate of interest, the rate of accumulation would be higher in the new 'fully-adjusted' position. The paradox of thrift would therefore not apply, at least in terms of a comparison of fully adjusted situations: the higher saving rate is matched by a higher rate of accumulation, with the unchanged normal rate of profit.

The second of the two modifications with a 'classical/Sraffian' approach to distribution has the rate of profit being determined by the real rate of interest. In this approach, Lavoie shows how a rise in the rate of accumulation generates a rise in the rate of profit. In particular, an autonomous rise in the rate of accumulation (through 'an increase in the animal spirits of entrepreneurs', p. 64), leads to a rise in utilization above normal and the actual rate of profit above the normal rate; real interest rates are raised, in turn raising the normal rate of profit. Utilization returns to normal, with a higher long-run rate of profit corresponding to the higher rate of accumulation. In the case of a rise in the saving propensity, this leads to a fall in utilization below normal, a fall in the actual profit rate relative to normal and a short-run fall in the rate of accumulation. Real interest rates fall which increases actual utilization back in line with the normal rate. The normal rate of profit, however, will be lower, although the rate of accumulation returns to its original level. In this case, a rise in the saving propensity is matched by a long-run fall in the rate of profit, but no long-run change in the rate of accumulation.

In both of the above cases the paradox of thrift (as defined above) disappears, which seems to lend itself to the perception that a classical/Sraffian view of distribution comes at some cost at least in terms of considerations of effective demand. On the other hand, as Lavoie notes, the second modification holds out the possibility of 'some reconciliation' between Sraffian and post-Keynesian positions in relation to long-run growth, in so far as the rate of profit will adapt to changes in the long-run rate of accumulation – a Kaleckian/neo-Keynesian feature; but with allowance for the rate of profit to be 'determined' by interest rates, interpreted by Lavoie as a concession to the Sraffian approach.

But what of the more fundamental inference running through Lavoie's analysis, namely, that an exogenous rate of profit, specifically one determined in a classical/Sraffian way, would entail a non-Keynesian or 'supply-led' explanation of growth? At the very least, the two cases of an exogenous rate of profit examined by Lavoie suggest that a key demand-side effect is lost, namely, the depressing effect on the steady state rate of accumulation of a rise in the saving propensity.[11]

In order to shed some further light on this issue one can consider whether an exogenous rate of profit, determined along classical/Sraffian lines and a steady state rate of accumulation falling into line with $s_c r$ is inconsistent either with the adaptation of saving to investment through changes in the scale of productive capacity, and thus with the long-run working of the principle of effective demand; or with a rise in the saving propensity, s_c, having a negative impact on effective demand, including investment demand.

As to whether an exogenous r and endogenous g are inconsistent with the adaptation of saving to investment, it seems sufficient to point out that an exogenous level or rate of growth of investment is quite consistent with an endogenous rate of accumulation, depending on the behaviour of the capital stock. In fact, with the rate of profit exogenous, along with the saving propensity, the capital stock *must* take the burden of the adjustment of saving in line with investment.

Now, as noted above, the Cambridge equation implies that with r given, a rise in s_c will lead to an increase in g. What this implies is of course that some part of the adjustment will involve a rate of growth of the aggregate capital stock different from the growth rate of investment. And this brings one to the question above, as to whether this is inconsistent with the rise in s_c having a negative impact on effective demand. The short answer to this is no, it is not.

A simple two-commodity (capital good, consumption good) model such as that of Vianello (1985), but with $0 < s_c < 1$ is instructive in this regard. Assuming for simplicity everlasting fixed capital, then in equilibrium, with demand equal to output in each sector, and thus aggregate investment demand equal to gross output of the capital good sector, the rate of accumulation could be expressed as

$$g = \frac{Y_i}{K_c + K_i} = \frac{Y_i}{K_i} \frac{K_i}{K_c + K_i} = u_{ni} \, \beta_i \, \frac{K_i}{K_c + K_i} \tag{7.2}$$

where the c and i subscripts refer to consumption and capital goods respectively, Y is gross output, L labour employment, u_n the normal rate of capacity utilization, β the output capacity of a unit of fixed capital and K the capital stock.

The significance of expression (7.2) for the present discussion is simply to assist in pointing out that a rise in the steady state rate of accumulation associated with an increase in the propensity to save is consistent with a negative impact on effective demand, taking the form of a fall in the levels or growth rates of the capital stock in both sectors and, as would seem to be implied by that (at least temporarily), a fall either in the level or rate of growth of aggregate investment demand. What *is* implied in this case is that any slowdown in effective demand is more pronounced in the consumption good sector so that any negative impact on the aggregate capital stock is proportionately larger than the negative impact on investment demand, and thus on output and the capital stock in the investment good sector. In other words, the rate of accumulation will rise consistent with negative impacts on effective demand of the rise in the saving propensity, to the extent that the impact of effective demand is proportionately greater in the consumption good sector.

Hence, the Cambridge equation with an exogenous rate of profit does not automatically rule out the adjustment of saving to investment through changes in the scale of productive capacity or, indeed, a negative impact of a rising saving propensity on effective demand.[12]

Taking the argument a step further, were one to argue that an exogenous rate of profit and exogenous saving propensity in the context of the Cambridge equation implies a non-Keynesian view of growth, then one should also regard Harrod's warranted rate of growth as representing a non-Keynesian result; to the extent that the warranted rate of growth will be higher for a higher aggregate saving propensity (given the desired capital–output ratio). But the legitimate response to such an interpretation would be that the positive relation between the warranted growth rate and the saving rate does not indicate a non-Keynesian hypothesis: rather it indicates that for continuous normal utilization of productive capacity a higher saving rate requires investment to grow at a faster rate. Moreover, as is well known, Harrod clearly distinguishes this requirement for steady growth from the question of whether investment will *actually* grow at this faster rate. He thus clearly provides for investment to grow in a manner independent of the requirements of desired saving rate and steady growth. What matters from a Keynesian perspective, is that nothing guarantees that this faster growth rate will materialize.

Yet to acknowledge this is to acknowledge in turn that the Cambridge equation just like Harrod's 'Fundamental Equation', tells us about the requirements of the steady growth but very little about what actually governs the growth rate of investment. A more fundamental question then arises about the significance of steady state analysis; specifically whether the assumption that producers are able to adapt capacity in such a way in each sector so as to restore normal utilization rates is itself consistent with the presumption that investment is independent of saving. This issue is taken up further in Section 7.6 below.

7.4. UTILIZATION – ACTUAL, NORMAL – AND LONG-RUN GROWTH

The claim that the Cambridge equation with a classical/Sraffian approach to distribution does not rule out the Keynesian adjustment of saving to investment through changes in the scale of productive capacity interestingly (and seemingly paradoxically) draws added support from the consideration of the second of the two above-mentioned arguments (p. 4) which have been advanced to support the alternative idea that it is income distribution which takes the burden of long-run adjustment between investment and saving.

The argument in question makes use of the notion, advanced by a number of writers (for example, Lavoie, 1996; Park, 1997; Dutt, 1997; Commendatore, 2006), that the normal utilization rate could be subject to influence by the average utilization rate realized over a sufficiently long period of time. The significance of such a claim is that, were such an adaptation to take place, it may be feasible to argue that 'normal' capacity utilization was 'demand-determined' (Barbosa-Filho, 2000). Further, to the extent that the normal rate of profit is reckoned at normal utilization, then the rate of profit might then be regarded as demand-determined. The fuller significance of the possibility of a 'demand-determined' normal utilization in the long run is therefore to allow for a change in distribution to provide the means by which saving adapts to investment in the long run, in turn lending support to the neo-Keynesian interpretation of the Cambridge equation.[13]

While it cannot be denied that a lengthy deviation of the actual average utilization rate relative to the normal rate may lead to changes in the normal rate itself, it is important to consider carefully the nature of this influence. In particular, such consideration would lead one to conclude at the very least that in general the normal rate could not take on all of the adjustment between the two rates, assuming the normal rate is not a purely historical magnitude (for example, an average rate realized over the last *n* years).

This point can be explained in terms of a simple exercise. For the purposes of the exercise we consider that the normal rate of utilization is chosen as the anticipated profit-maximizing average rate of capacity utilization over a specified period of time into the future, given technology and relative prices, and assuming that capacity is allowed to grow over this period of time so as to allow supply to meet expected demand levels. The normal rate in this case will depend *inter alia* on the expected pattern of demand fluctuations, for example, the ratio of peak to trough demand as well as the frequency of fluctuations (see White, 1996).

Suppose that looking back in time from the present – period t, over a span of time long enough to include a number of cycles, producers discover that the actual average rate of capacity utilization was higher than the intended 'normal' rate at the beginning of that span of time – period $t - n$ – and that demand had grown over this time at a faster rate on average over the cycle than had been anticipated. As Garegnani (1992) has noted, this higher rate of growth may reasonably be supposed to have manifested itself in terms of longer booms and shorter slumps compared with the pattern of demand expected at $t - n$. For argument's sake we assume that the realized path of demand over time $t - n$ to t shows a higher ratio of trough to peak demand compared with what was anticipated by producers at $t - n$. With given relative prices (including shift premiums for labour, and the costs associated with holding of inventories) and technology, it can be shown that a rise in the anticipated ratio of trough to peak demand will push up the cost-minimizing utilization rate. Hence, the higher than anticipated trend rate of growth of demand could conceivably lead to an upward revision of the normal utilization rate.[14]

However, the important question is how much the normal rate adjusts upwards towards the actual average rate? On this the critical point to note is that the actual rate of utilization over the time period $t - n$ through t reflects not only the realized trend rate of growth of demand but also the actual rate of growth of capacity. And, importantly, *the latter would have been higher had the realized growth of demand been correctly anticipated at $t - n$*. In other words, the normal utilization rate may rise, but in response to a revision in expectations about the likely trend rate of growth of demand in the future, the planned rate of growth of capacity is also likely to be adjusted upwards. Hence, although the anticipated change in the characteristics of the cycle – a higher ratio of trough to peak demand – might act to push up normal utilization, the planned higher rate of growth of capacity over the subsequent n periods would act to limit the extent of this adjustment in normal utilization, certainly preventing the normal rate from rising to the level of the actual rate over timespan $t - n$ to t.

This result would be sufficient to ensure that any exogenous change in the rate of growth of investment, taking the form of a faster trend rate of growth over a number of cycles, need not (and most likely would not) entail an equivalent increase in the trend rate of accumulation.[15] In other words, an increase in the rate of accumulation even over a number of cycles need not be sustainable to the extent that this increase reflected a growth in capacity built on incorrect expectations about demand. The increased realized rate of accumulation over $t-n$ to t, through its effect on the cycle, may well lead to a revision of normal utilization and a change in distribution in turn providing some change in the aggregate saving to match part of the expansion in investment. Yet to the extent that the growth in investment is not matched by a commensurate rise in normal utilization, then, with the profit share in output given, and assuming an unchanged sectoral composition of the capital stock, part of the saving expansion to match the expansion in investment demand must come about by means of changes in the scale of productive capacity, thus reducing the extent of the long-run change in the rate of accumulation.

7.5. THE ALTERNATIVE TREATMENT OF AUTONOMOUS DEMAND – EXOGENOUS GROWTH AND EXOGENOUS DISTRIBUTION

At this point it is useful to recap on two results so far. First, an exogenous rate of profit does not appear to rule out the adaptation of saving to investment, at least partly through a change in the scale of productive capacity, or, to put the same point another way, a lasting impact of effective demand on output. Second, this result seems to hold even where the normal rate of capacity utilization is allowed to vary (and with it the normal rate of profit) as a result of adjustments to producers' expectations about future demand movements.

Yet there remains a one troubling aspect to these results which goes back to a problem raised in Section 7.2. Supposing that producers are able to restore normal utilization through the adaptation of capacity to an exogenous rise in the growth rate of investment (which is implied by the case of an exogenous s_c and r), no room is left for exogenous influences on the rate of accumulation. To the extent that a demand-led view of growth refers to an exogenous influences on the rate of accumulation, there remains some tension between a Sraffian view of distribution and a demand-led view of growth, at least in terms of the Cambridge equation.

An alternative approach – at least in terms of modelling the steady state – which appears to offer some way around this tension has made its appearance in recent literature originating from the Sraffian camp (Serrano, 1995; Trezzini, 1995; Palumbo and Trezzini, 2003; Cesaratto et al., 2003), although its origins go back to Harrod (1939). This alternative involves explicit recognition of the independence of effective demand in terms of elements of demand which are non-capacity-creating and which are independent of expected rate of growth of demand.[16] Harrod's warranted growth rate in the case of non-capacity-creating autonomous demand is given as

$$g_w = \frac{s}{v} - \frac{I^A}{vY} \tag{7.3}$$[17]

where s is the aggregate saving propensity, Y is income, I^A is autonomous demand and v is the desired capital–output ratio. Assuming only profits are saved and expressing the aggregate saving propensity in terms of the propensity of capitalists to save out of profit then

$$g_w = s_c r - \frac{I^A}{Y}\frac{1}{v} \tag{7.4}$$

In effect, making explicit the distinction between autonomous demand and expenditure directly tied to income or to the anticipated growth in income seems to be one means of reconciling an exogenous rate of profit and an exogenous rate of accumulation. Indeed, if one were to suppose that the warranted rate of growth is determined exogenously by the rate of growth of autonomous demand then an exogenous rate of profit requires that the ratio of autonomous demand to income take the burden of adjustment to changes either in the rate of growth of autonomous demand or to changes in the rate of profit, at least in the absence of technical change and changes in the normal rate of capacity utilization.

At this point we consider a simple example based on relation (7.4) above. We rewrite expression (7.4) by substituting $1/(u_n\beta)$ for the desired capital–output ratio, where u_n and β refer to the normal utilization rate and the output capacity of a unit of fixed capital, so that

$$g_w = s_c r - \frac{I^A u_n \beta}{Y} \tag{7.5}$$

Consider the case of a rise in the propensity to save s_c in a situation where the economy has been growing at the warranted rate. We suppose that the effect at least temporarily is a fall in actual utilization and a fall in the actual growth rate of output. Suppose also that the rate of growth of autonomous demand remains unchanged. As such the ratio I^A/Y begins to rise. One could also assume here for the sake of argument that the fall in the actual rate of capacity utilization would reduce the actual rate of profit below the normal rate.

How does the system adjust, assuming that equality between the warranted and actual rates is restored after some time? If prices are given along with technology and characteristics of the cycle remain unchanged except for the trend rate of growth, we might reasonably suppose that producers maintain their notion of normal utilization and hence slow investment to bring the growth of capacity to a rate consistent with a return to normal utilization.

Presumably the key to any subsequent adjustment of the actual growth rate in line with the rate of growth of autonomous demand is the persistence of growth in autonomous demand and in particular how this affects expectations about demand growth in general. To the extent that the fall in the actual growth rate below that of autonomous demand is deemed not to be permanent, then the required slow-down in the growth of capacity will be seen as being necessary only temporarily. Thus, the nature of the adjustment would seem to depend on how producers anticipate demand in their own sectors, how that relates to demand in the economy as a whole and how the latter relates to the growth of autonomous demand.

The point of this example is that the rise in s_c could conceivably be matched by a rise in the ratio of autonomous demand to income, consistent with an unchanged equilibrium growth rate and normal rate of profit.

7.6. THE WARRANTED GROWTH RATE DETERMINED BY THE RATE OF GROWTH OF AUTONOMOUS DEMAND

In the discussion above it is assumed that the warranted growth rate is determined by the rate of growth of non-capacity-creating autonomous demand. Within this simplified framework, our immediate concern is with two types of criticism advanced against the representation of growth in the previous section – interestingly, coming partly from Sraffian circles. The first criticism focuses on the stability of the growth rate equal to the rate of autonomous demand growth – which we will refer to as the 'exogenous equilibrium growth rate' compared with the stability of the warranted path given by the ratio of the saving propensity and capital–output ratio – which we refer to here as the 'endogenous equilibrium growth rate'. The argument has been put by Park (2000) and by Barbosa-Filho (2000), in terms of simplified one-sector models that the latter is a stable growth path, while the former, determined as in expression (7.3) above is unstable. In particular, the 'endogenous' equilibrium growth rate will set the maximum rate compatible with continuous normal utilization, so that with a positive rate of growth of autonomous, non-capacity-creating demand, an actual growth rate initially greater than the exogenous equilibrium growth rate but smaller than the endogenous equilibrium growth rate will converge over time to the latter. Thus over time the growth rate converges to the Harrodian warranted rate $g =$

s/v. This process will entail that the ratio I^A/Y tends to zero, as income grows faster than autonomous demand.

A key question raised by these results is whether they are indicative of a more general result. In particular, it is interesting to consider whether the instability of the 'exogenous equilibrium growth rate' is a function, as some might suspect, of the simplicity of the models examined. Both Park (2000) and Barbosa-Filho (2000), use what is effectively a multiplier–accelerator interaction. It is well known that the dynamic behaviour of such interaction is dependent on the values of key parameters – saving propensities, the desired capital–output ratio – as well as on the nature of lags in the income–expenditure process and the nature of expectations formation. To this end, the Appendix to this chapter provides some simulation results for a multiplier–accelerator model, which generates different results to those of Park and Barbosa-Filho. These simulation results demonstrate the possibility of the actual growth rate of demand fluctuating around the rate of growth of autonomous demand rather than converging to the Harrodian warranted rate equal to s/v.

Arguably stability results in relation to the two equilibrium growth rates are also conditional on the technology – including the normal utilization rate and the output capacity of fixed capital (which together determine the desired capital–output ratio); and on the assumptions made regarding the calculation of the expected growth rate of demand. In the model used the assumption is made that the anticipated growth rate is a weighted average of a measure reflecting recently observed growth rates and the growth rate of autonomous demand. In other words, it is assumed that producers factor into their forecasts of demand growth the growth rate of autonomous demand.[18]

Moreover, the latter assumption can be justified on the following intuition: producers believe that the rate of growth of demand in their own sector is partly dependent on the growth rate of the economy as a whole, and through this, on the rate of growth of components of autonomous demand, for example, export demand, public sector expenditure.[19] For the purposes of the exercise detailed in the Appendix we however also assume that producers are ignorant of the precise relation between growth in their own sector and growth in the economy as a whole.

At the very least these simulation results suggest that any claim that the steady state growth rate is ultimately governed by the saving propensity and the capital–output ratio, regardless of the existence of a growing autonomous, non-capacity-creating demand, cannot be assumed to be general.[20] These results demonstrate the possibility of growth rates fluctuating around a level determined by the growth rate of autonomous demand.

The second type of criticism of the approach of the previous section has emerged in the work of Trezzini (1995, 1998) and Palumbo and Trezzini (2003) (hereafter P–T). Here the focus has been in part on the ability of the system to move from one autonomous demand governed steady state growth path to another as a result of a change in the rate of growth of autonomous demand and on the long-run changes in capacity utilization involved. The deeper question raised in this work concerns the usefulness of steady state analysis for Keynesian growth theory. P–T are clearly doubtful about the usefulness of the notion of 'full-adjustment' between capacity and demand and about the usefulness of steady state analysis. Significantly, at least part of this doubt has to do with what they perceive as difficulties with the adjustment of the system's actual growth path to changes in the warranted rate due to a change in the rate of growth of autonomous demand and in particular the restoration of a desired ratio of capacity to demand.

Consider the case of a system having grown until the current period at the warranted rate equal to the rate of growth of autonomous demand and suppose in the current period a rise in the latter and that this higher growth rate is then persistent. For steady state growth at the new higher rate, with a given saving propensity and capital–output ratio, a smaller proportion of output growth must be absorbed by the growth of autonomous demand. In other words, the ratio of autonomous demand to total demand I^A/Y must fall.

P–T note that this fall can only come about if output grows faster than autonomous demand at the latter's new higher rate and that the system's adjustment must involve a period of over-utilization relative to normal. More precisely there are two aspects to this over-utilization of capacity. The first aspect is the initial effect: where capacity is initially growing at the old, lower rate of growth of autonomous demand (that is the old warranted rate and thus consistent with normal utilization), a rise in the rate of growth of autonomous demand must force utilization above normal. The second aspect is more complex. Supposing for the sake of argument that output *does* grow through the adjustment phase faster than the new higher rate of growth of autonomous demand, this is likely to involve at some point an over-utilization of capacity, to the extent that producers would have adjusted their investment so that capacity grows in line with the new rate of growth of autonomous demand.[21]

Over-utilization together with a consideration of the demand effects of investment leads P–T to the conclusion that the long-run adjustment of capacity to demand must entail considerable variation in relation to the normal rate even where the latter is defined as an average rate of utilization.[22] It also leads them to doubt that it is unlikely that full adjustment could be achieved except with a very lengthy period of adjustment in which the rate of growth of autonomous demand was unchanged. As such, they regard ‘as incorrect, the claim that, given th[e] tendency [of productive capacity to adjust to demand] we can represent actual processes of accumulation and the relations between actual quantity variables by means of theoretical positions characterized by the full adjustment of capacity to demand’ (p. 123).

It is worth reflecting a little further on P–T’s exercise in order to clarify the strength of their doubts about the possibility of full adjustment. In particular, taking the case discussed above of a rise in the rate of growth of autonomous demand, what is not immediately clear are the precise mechanics by which the necessary fall in I^A/Y would come about. If expectations of growing demand are the key force driving investment and output, for Y to grow faster than I^A would require a reasonably firmly held expectation that demand was going to grow faster than the new higher rate of growth of autonomous demand. And if the expected growth rate of demand was based on a weighted average of recently observed growth rates of autonomous demand and recently observed growth rates of sectoral demand, then it is difficult to see how an expected rate of demand growth could exceed the new higher growth rate of autonomous demand.

Similarly, in the case of a fall in the rate of growth of autonomous demand the adaptation of the warranted growth rate to this new lower rate would require a rise in the ratio I^A/Y and thus a fall in the rate of growth of output below the newer lower rate of growth of autonomous demand. In this case, producers would have to be anticipating a growth in demand at a rate lower than the new rate of growth of autonomous demand.

It is thus not easy to envisage the adjustment of the economy to a path associated with normal utilization and steady growth at a rate corresponding to the new rate of growth of autonomous demand, at the very least without significant variations in capacity utilization relative to the normal rate,[23] as is maintained by P–T, and/or variations in the normal rate itself. As noted above, what is required in the case of a rise in g_a is a fall in I^A/Y, in order for the system to return to a steady growth path. The question is whether there could exist conditions which would generate a growth rate of income, even for a relatively small amount of time, which exceeded the new higher rate of growth of autonomous demand? The answer to this rests with how the system responds to the initial surge in the growth rate of autonomous demand.

It seems possible to conceive of a situation where a surge in the rate of growth of autonomous demand triggers a cyclical or cumulative response in the system, although this response may be intrinsically damped. All that is required for a fall in the ratio I^A/Y is an upswing in demand which is faster than the new higher rate of growth of autonomous demand.

The question is then what would generate a cumulative expansion for a period involving a rate of growth of income in excess of the new higher rate of growth of autonomous demand. Supposing such a period of growth, which gives way to a cyclical downturn in the growth rate of income, this having the opposite effect, namely, raising for a period I^A/Y and reducing capacity utilization. What would be required over the cycle as a whole, in order that the system move towards 'full adjustment', is that the cyclical upturn in growth and the fall in I^A/Y exceed the cyclical downturn in growth and the rise in I^A/Y; and thus a net fall in I^A/Y. It is difficult to conceive of how this could not generate an average utilization rate over the cycle which differs from the average rate anticipated by producers on the basis of previous growth, in the aggregate, and of autonomous elements of demand. Had the system grown at a constant rate before the rise in the rate of growth of autonomous demand, even with cycles, the pattern of demand fluctuations and the trend rate of growth anticipated on that basis would differ from what would have resulted from the unanticipated rise in the rate of growth of autonomous demand.

Now such cyclical responses to a rise in the rate of growth of autonomous demand should certainly not be ruled out in any model incorporating multiplier–accelerator interaction, as does expression (7.3) above. Indeed, White (2003) demonstrates by means of simulation experiments that the initial reaction in a multiplier–accelerator model could give rise to a sufficiently large enough surge in growth rates of demand that I^A/Y falls – thus the adjustment to a new rate of growth of autonomous demand involves a cycle in growth rates around a higher trend rate of growth. These experiments show that in these cases there *is* gravitation of the actual utilization rate around the normal rate over time.[24] However, these results should be treated with caution, since in the model examined the normal utilization rate is taken as given; the rate of growth of autonomous demand does not fluctuate over time; expectations of demand growth are partly based on anticipations about autonomous growth in demand [25] and no allowance is made in decisions about fixed capital investment for fluctuations in demand.

More generally, the ability of multiplier–accelerator interaction to generate a cyclical rise in the growth rate in excess of the rise in the growth rate of autonomous demand would depend most importantly (though of course not wholly) on the responsiveness of investment to any initial surge in the growth rate due to the higher growth rate of autonomous demand. If for example producers routinely judge investment prospects with the view that demand behaves cyclically, it becomes more difficult to account for the above cyclical responses on the basis of the interaction between the multiplier and an accelerator applied to investment in fixed capital. In other words, where producers anticipate fluctuations in demand, the reaction of producers to a higher rate of growth may well be to interpret it as a non-trend or cyclical increase in the growth rate, thus reducing the chances that the increase in the rate of growth of autonomous demand triggers a larger increase in the growth rate of income; and thus a fall in I^A/Y. Interestingly, this reasoning suggests that the less volatile the system is in relation to shocks – in the form of changes in the growth of autonomous demand – the less likely it may be that the system is able to fully adjust its capacity to demand so as to bring actual and warranted growth rates in line at the new rate of growth of autonomous demand, without considerable long-run variation in the degree of capacity utilization.[26]

Recapping, it seems possible to support the view of Trezzini and P–T that the process of adjustment to a new warranted growth path following a change in the rate of growth of autonomous demand will likely involve considerable non-normal utilization of capacity. The question arises as to what would lead producers to expand capacity faster or slower than the new rate of growth of autonomous demand: to the extent that producers do not, then periods of over- or underutilization relative to normal are likely to be longer. However, in contrast to Trezzini and P–T, the fact that producers would likely take into account demand fluctuations in planning investment would act to limit the extent to which attempts to adjust capacity to demand are frustrated by the feedback effects of those investments on demand.

Whether the difficulty of full adjustment within a time period during which the growth rate of autonomous demand remains unchanged points to the need to abandon steady state analysis like that of Section 7.5 remains unclear. In this regard it might be useful to consider the statement of Committeri (in an earlier version of the debate over Sraffian distribution theory and effective demand) as to

> whether there are forces capable of keeping the average actual utilization rate in line with its normal level … we would like to suggest that while actual average [utilization] rates may diverge from their normal levels, these deviations will be contained within certain bounds, which may vary according to prevailing conditions. (1986, p. 180)[27]

7.7. AUTONOMOUS DEMAND – *HOW* AUTONOMOUS IN THE LONG RUN?

There remains one other issue pertinent to the approach adopted in Section 7.5, namely, the argument that autonomous, non-capacity-creating expenditures can mostly be tied to income or are constrained by income in some manner in the long run and hence the approach to growth adopted in the previous two sections does not tell the complete story (for example, Park, 2000).

In this section we consider two types of autonomous demand: public sector expenditure and expenditure on exports. With regard to the former the relevant constraint is here taken to be concern by policymakers about the ratio of public debt to income. With regard to the latter, the issue is more complex; one of these complexities being whether one is discussing small or large open economies. Quite clearly, as numerous writers have noted, identifying exports as the key element of autonomous demand is not logically consistent when applied to all economies simultaneously. But two points are worth making in relation to this fact. First, clearly, this type of expenditure is not 'independent of income' in a strict sense, even for small open economies, since that 'smallness' can be interpreted in terms of the income of that economy in relation to the incomes of other economies.[28] However, and this is the second point, it is not clear that either the logical inconsistency referred to above or the lack of a strict independence of income undermines the usefulness of the approach of Section 7.5, even where some part of that autonomous demand is associated with exports, at least in relation to small open economies.[29]

As a means of alluding to some of the issues involved, we consider a very simple one-sector case with two types of autonomous demand: public sector expenditure and export expenditure. In relation to each of these we take note first of the connection between the behaviour of expenditure and the stock of debt (public sector debt in the former case, foreign debt in the latter) as a proportion of income; and second, that policymakers would not be unconcerned about either ratio. Our concern is with the implications of different rates of growth of the two types of expenditure. These implications are traced through mainly in terms of the relation between actual and warranted growth rates.

We write the open economy saving–investment equality – namely, that the sum of the excess of private sector investment over saving and the public sector deficit is equal to the current account deficit – can be written for any given period as:

$$S+T+rEL=s_c(1-t)\,r\,(B+K-EL)+T+rEL=I+G+rB+X-mY \qquad (7.6)$$

where S and T refer to aggregate saving and taxation respectively; t the marginal (and average) tax rate on income, r the rate of profit, B the stock of government bonds outstanding, K the capital stock, I induced investment, G government expenditure, X exports, m the marginal propensity to import and EL the stock of external liabilities (foreign debt). We assume the interest rate on public debt, the interest rate paid on foreign debt (assumed to be held solely in the hands of the private sector) and the rate of profit on capital are identical.

Denoting the expected level of income (demand) in $t+1$ as Y_{t+1}^{e} induced investment along the steady state path is given by

$$I_t = v\left(Y_{t+1}^{e} - Y_t\right) = vY_t g_w \tag{7.7}$$

with v, the desired capital–output ratio. Substituting this expression for I in equation (7.6), expressing variables as ratios to income, and solving for the warranted rate of growth g_w yields

$$g_w = r\, s_c\left(1-t\right) - \frac{\left(B^Y - E^Y\right) r\left[1 - s_c\left(1-t\right)\right] + \left(X^Y + G^Y - m - t\right)}{v} \tag{7.8}$$

where B^Y represents the stock of public debt, E^Y external liabilities, X^Y exports and G^Y public sector expenditure all as proportions of income.

What needs to be added to the analysis is a treatment of the dynamics of both public sector debt and foreign debt as proportions of income, as well as the preferences of policymakers regarding these two ratios. We take a very simplistic approach in regard to both of these issues and assume that policymakers regard the initial ratios of both types of debt to income as the maximum desirable and would prefer for these ratios to fall over time. Regarding the modelling of the dynamics of these ratios, for the purposes of the simple discussion, we assume that in order for these ratios not to rise it is necessary for the following two relations to hold (see Pasinetti, 1998):

$$D_t^{PY} \leq (g_t - r)B_t^Y \tag{7.9}$$

where D_t^{PY} is the ratio of the primary public sector deficit (G – T) to income at time t and g_t represents the actual growth rate of income (nominal and real), with the inflation rate assumed to be zero; and

$$T_t^Y \leq (g_t - r)E_t^Y \tag{7.10}$$

where T_t^Y is the trade deficit as a proportion of income at time t.[30]

We consider only two cases, which raise some interesting questions at least in terms of expression (7.8) given the constraints represented by expressions (7.9) and (7.10). In particular, expression (7.8) relates the warranted growth rate positively to the foreign debt ratio and negatively to the public debt ratio, the public sector expenditure ratio and the export ratio (and thus given m, positively to the trade deficit as a proportion of income).[31] On the other hand all four ratios will be affected by movements in the actual growth rate.

Initially, we suppose a situation where actual and warranted growth rates are equal, export and public sector expenditures are growing at the same rates, equal to the warranted rate and, given the profit rate (and relevant interest rates) equal to r, expressions (7.9) and (7.10) are satisfied as equalities, so that debt to income ratios are constant. Now suppose a rise in the growth rate of export expenditure with the growth rate of public sector expenditure remaining unchanged. X^Y will rise so long as the actual growth rate of income does not rise up in line with the higher growth rate of exports. If the actual growth rate nonetheless rises above that of public sector expenditure, G^Y will fall, and thus, in combination with the higher growth rate of income, B^Y will fall. In addition the rise in X^Y and in the actual growth rate of income will lead to a fall in E^Y.

The net effect of these changes on the warranted growth rate is unclear and turns largely on the implications of the rise in the growth rate of exports on the actual growth rate of income. If one supposes that the actual growth rate of income tends towards the highest growing component of autonomous demand – in this case, the new higher growth rate of exports – the rise in X^Y will eventually be overtaken by falling G^Y and B^Y. If the latter is sufficiently large relative to a falling foreign debt ratio, E^Y, the warranted rate of growth must rise. And this seems one possible path by which the warranted rate may adjust in line with a higher actual rate, the latter reflecting the higher rate of growth of exports.

However, even in this latter case, to the extent that the rate of growth of public expenditure remains at its initial lower level, the ratio of public expenditure to income G^Y and the ratio of public debt to income B^Y will continue to fall. If not balanced by matching falls in E^Y the warranted rate could rise above the actual growth rate, generating, seemingly perversely, a situation of underutilization of capacity. In this sense at least, there may be pressure to bring the rate of growth of public expenditure into line with the growth rate of exports.

In the opposite case, where the growth rate of export expenditure falls, again the implications turn very much on the impact on the actual growth rate of income. Suppose that after a temporary fall, the actual growth rate of income tends to move back in line with the unchanged growth rate of public expenditure, which now exceeds the growth rate of exports. X^Y will be falling. This, together with any temporary fall in the actual growth rate of income, will worsen the foreign debt ratio so that E^Y will be rising. The fall in X^Y together with the rise in E^Y pushes the warranted above the actual growth rate, resulting in underutilization of capacity. One might suspect, in light of a rising foreign debt to income ratio, governments may be tempted to constrain public sector expenditure.[32] However in order to bring the warranted growth rate in line with the actual growth, a reduction in the growth rate of government expenditure would need to be sufficiently large to bring the actual growth below that of the rate of growth of exports thus increasing X^Y and by a sufficient amount to reduce E^Y and to offset any fall in G^Y. At the very least, in this second case, the possibility of the warranted rate coming back into line with the actual growth rate seems more problematic compared with the case of a rise in the growth rate of exports.

Though undoubtedly extremely simplified, the discussion above at the very least points to the complexity of the analysis when allowance is made for different types of autonomous demand subject to different growth rates.

7.7. CONCLUSION

The analysis of this chapter began with the question of the compatibility of the Sraffian approach to value and distribution and a Keynesian explanation of economic growth. This chapter suggests that these two approaches are indeed compatible.

What the chapter leaves open is the precise way in which these two aspects of economic theory may be synthesized. In particular, we can choose to ignore autonomous elements of demand which are non-capacity-creating, so that, with an exogenous rate of profit, the steady state rate of accumulation is governed by the former and the saving ratio. Yet the endogeneity of the rate of accumulation in this sense does not preclude the long-run adaptation of saving to investment via changes in productive capacity. Indeed, the endogeneity of the rate of accumulation is a manifestation of this adjustment process. Yet this approach leaves little room for a truly exogenous rate of accumulation. The alternative involves distinguishing autonomous, non-capacity-creating components of demand. This in turn opens up the possibility that an exogenous rate of accumulation – governed by the growth rate of autonomous demand – coexists with the exogenous (Sraffian) rate of profit; the adjustment of investment to saving involving variation in the proportion of income absorbed by autonomous demand. In fact, depending again on the time-frame of the analysis, it is possible to distinguish non-capacity-creating demand, but to simultaneously distinguish some of this demand as subject to income constraints and some not. The interesting complexity in this latter case concerns the presence of differential growth rates for different components of autonomous demand. It is here in particular that the most interesting questions relating to growth and economic policy are likely to arise.

An unresolved issue, which this chapter does not provide a definitive answer to, is the appropriateness of a model based on comparisons of warranted growth paths, at least for the analysis of demand-led growth. To the extent that any light on this emerges from the preceding discussion it would be that the answer to this issue should await further analysis of the factors which are likely to determine the extent of deviation between realized average utilization rates and normal rates.

APPENDIX

Consider the following simple discrete-time, one-commodity model. With no government or foreign sector, demand at the end of period t is the sum of consumption and investment, so that

$$D_t = Y_{t-1}wl + (1 - s_c w)(D_{t-1} - wlY_{t-1}) + \frac{1}{u_n \beta} D_{t-1}\left(1 + g_{t-1}^{ed}\right)^2 - K_t + I_t^A \quad (7.A.I)$$

where D_t is demand at the end of t, Y_{t-1} is output at the end of $t-1$; g_{t-1}^{ed} is the expectation of demand growth between end of $t+1$ and the end of t, with $D_{t-1}(1+g_{t-1}^{ed})^2$ the expected level of demand at the end of $t+1$; β is the output capacity of a unit of capital and u_n the desired or 'normal' utilization rate; K_t is the capital stock in place during period t and I_t^A is autonomous non-capacity-creating demand forthcoming at the end of t. The capital stock is assumed to last indefinitely, workers do not save (in the sense that they consume at the end of t all wages earned in $t-1$[33]) and capitalist consumption at the end of t is based on the realized flow of profit at the end of $t-1$. It is also assumed that the most recent demand on which investment at the end of t can be based is the level at the end of $t-1$. Demand is thus driven by a combination of an income–expenditure multiplier with a lag of one period and investment governed by anticipated growth in demand (that is a simple accelerator).

We assume further that output is based on demand at the end of the preceding period as well as on the anticipated growth rate g^{ed}. Thus

$$Y_t = D_{t-1}\left(1 + g_{t-1}^{ed}\right) \quad (7.A.II)$$

Since investment at the end of $t-1$ is equal to the change in capital stocks between t and $t-1$, we can also write

$$K_t = \frac{1}{u_n \beta} D_{t-2} \left(1 + g_{t-2}^{ed}\right)^2 \tag{7.A.III}$$

With the actual growth rate of demand, g_t^d given by

$$g_t^d = \frac{D_t}{D_{t-1}} - 1 \tag{7.A.IV}$$

expressions (7.A.II) and (7.A.III) allow one to eliminate Y_{t-1} and K_t from expression (7.A.I) and to express D_t as

$$D_t = \frac{\left(1+g_{t-1}^d\right)\left[D_{t-1}\left(1+g_{t-1}^{ed}\right)^2 + \left(D_{t-1}+I_t^A - D_{t-1}s_c\right)u_n\beta\right] - D_{t-1}\left(1+g_{t-2}^{ed}\right)\left(1+g_{t-2}^{ed} - ls_c u_n w\beta\right)}{\left(1+g_{t-1}^d\right)u_n\beta} \tag{7.A.V}$$

Writing for the ratio of autonomous to total demand I_t^{AD}, then

$$I_t^{AD} = \frac{I_t^A}{D_t} = I_{t-1}^{AD}\left(\frac{1+a}{1+g_t^d}\right) \tag{7.A.VI}$$

where a is the growth rate of autonomous demand (assumed constant for the purpose of the present exercise). Bearing in mind expression (7.A.IV), expression (7.A.V) can be transformed into an expression for the rate of growth of demand as a function of past actual and expected growth rates and the growth rate of autonomous demand. Thus

$$g_t^d = \frac{G_{t-1}\left\{\left(1+g_{t-1}^{ed}\right)^2\left[I_{t-1}^{AD}(1+a) - s_c\right]u_n\beta\right\} - g_{t-2}^{ed}\left(g_{t-2}^{ed} + 2 - l.s_c u_n w\beta\right) + l.s_c u_n w\beta - 1}{G_{t-1}u_n\beta} \tag{7.A.VII}$$

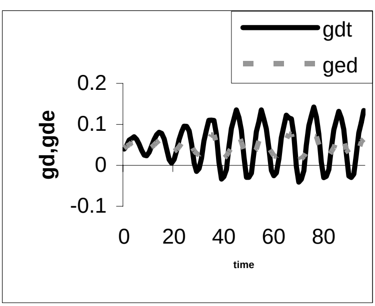
with $G_{t-1} = (1 + g_{t-1}^d)$.

There remains the anticipated growth rate of demand. Following White (2003) this is assumed to be a weighted average of two components: one of these components is the most recent growth rate of autonomous demand; the other component is an average of the growth rate of the previous two periods, with a discount for the dispersion between those growth rates. Hence

$$g_t^{ed} = \varepsilon \left\{ \frac{g_t^d + g_{t-1}^d}{2} \right\} X_t + (1-\varepsilon) a \quad \text{where} \quad X_t = \frac{1}{1+\sigma\left(g_{it}^d - g_{it-1}^d\right)^2}$$

(7.A.VIII)[34]

The figures below show the results of simulations based on expressions (7.A.VII) and (7.A.VIII). For each simulation the model is assumed to initially be growing steadily at the rate of growth of autonomous demand equal to 4 per cent per period and is then subject to a shock in the form of a rise in the rate of growth of autonomous demand to 5 per cent, whereafter it remains constant. In each case the 'endogenous' equilibrium growth rate in the absence of autonomous demand is 50 per cent per period. The first case, with σ = 500 and ε = 0.6 shows cycles of gradually increasing amplitude with time. By constrast, a higher σ (1000) and a lower ε (0.4) and hence a larger weighting on autonomous demand growth in expectations appears to have (as expected) a stabilizing effect on the time path of the growth rate.



$sc = 0.6$
$\beta = 1.2$
$l = 1$
$r = 0.41$
$w = 0.6$

Figure 7.A.1 Growth rates of demand: actual and expected. б = 500, ε = 0.6

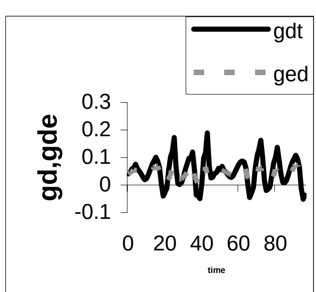


Figure 7.A.2 Growth rates of demand: actual and expected. б = 1000, ε = 0.4

NOTES

* I am indebted to Tony Aspromourgos, Fabio Petri, Marc Lavoie, Pierangelo Garegnani, Attilio Trezzini and to two anonymous referees for useful comments on earlier drafts of this chapter, as well as to participants at the Conference. Any remaining errors and omissions are of course my own responsibility.

1. Commendatore et al. (2003) cites the results Park (2000) and Barbosa-Filho (2000), both of whom suggest that equilibrium growth paths determined by the rate of growth of an autonomous demand are unstable.
2. For the purpose of discussion in this and the following two sections, it is assumed that the rate of growth g represents an asymptotically stable equilibrium growth path. Hence g also represents the level towards which the actual growth rate tends at least in the long-run. In the absence of such an assumption (for example, Harrod, 1939) controversy over the direction of causation between g and r in equation (7.1) may be thought of as somewhat redundant.
3. This argument also involves a rejection of the notion that changes in the rate of capital accumulation necessarily require changes in income distribution. Of course, this is not the only criticism of the neo-Keynesian position put by Garegnani and others: there is also the point that outside of continuous steady growth, and outside of a one-commodity model the rate of accumulation cannot be defined independently of relative prices and therefore of the rate of profit. We focus however on the criticism above because it goes to the heart of the issue examined in this chapter, namely, the compatibility of different approaches with a Keynesian view of the long run.
4. The authors note as a consequence the emergence of 'neo-Ricardian' responses designed 'to reassign a role to demand'. We consider one such response in Section 7.5.
5. The suggestion that the Sraffian approach may entail an orthodox view of growth is compounded further with Lavoie's reference to 'loanable funds theory' in his discussion of the work of Dumenil and Levy (Lavoie, 2003b, p. 61).
6. It is important to note that these same authors nonetheless hold a role for 'normal utilization' in the analysis of relative prices; specifically, the set of relative prices consistent with a uniform rate of profit across sectors. In other words, for these authors, the analytical significance of normal prices and the associated normal rate of profit should not be construed as an argument in support of analysis in terms of comparisons of situations where capacity is fully adjusted to demand and yielding on average a utilization rate equal to that anticipated when current capacities were installed.
7. It should be noted here that Keynesian and Sraffian views are not seen by Sraffians as being inconsistent in the context of the Cambridge equation to the extent that the latter refers to the *actual* rate of accumulation and the *actual* or *realized* ratio of profits to capital (see Garegnani, 1992, pp. 60–62). It is important also to mention the position advanced by Panico who suggests that a Sraffian determination of the profit rate by reference to the rate of interest, and a Kaleckian/post-Keynesian view of the steady state rate of growth as exogenous are not necessarily at odds, at least when the government sector is introduced to the model and when financial equilibrium is also considered (see Panico, 1997). Suffice it to say that it is a result not unlike one of the results achieved by Lavoie.
8. Lavoie has in mind here the result of some earlier Kaleckian models whereby actual and normal rates could diverge in equilibrium. In private correspondence, Marc Lavoie has reminded the author that the inconsistency between steady states and non-normal utilization rates had been pointed out in Sraffian criticisms of earlier Kaleckian models.
9. It appears to be the case that bringing together 'full adjustment' and the Kaleckian growth model is for Lavoie at least one way of reconciling Sraffian and Kaleckian approaches – to

the extent that 'Sraffa's price equations are implicitly based on fully adjusted positions' (Lavoie, 2003, p. 61), though, as Lavoie recognizes, the identification of full adjustment and normal prices is seen as questionable by some Sraffians. In fact, whether or not growth theory should make use of the notion of full adjustment seems however to be a matter of some dispute amongst Sraffians (see for example, Serrano, 1995; Trezzini, 1998 and Palumbo and Trezzini, 2003).

10. One problem with Lavoie's argument sketched above arises when one considers that the utilization rate returns to normal in his narrative following the adjustment in the profit share, but the change in the profit share is permanent. One might reasonably argue that, to the extent that the change in capacity utilization resulting from a change in the saving propensity is temporary, why wouldn't the rise in the actual profit rate above normal (or the original level of the target rate of return) be seen as similarly temporary. If the upward adjustment in the target rate is not permanent, the real wage will eventually move back to its initial level and utilization will diverge from normal. Hence the profit rate would move away from normal. The question then arises as to how the system gets back into equilibrium?
11. Strictly speaking there are two points here: whether a rise in the saving propensity has a depressing effect on aggregate demand – the issue associated with the traditional understanding of the paradox of thrift; and whether one should expect this effect to necessarily lead to a fall in the rate of accumulation. As emphasized by an anonymous referee, what happens to the latter depends in part on the effect of a rise in the saving propensity on the capital stock. As this referee has also noted, Lavoie's notion of the paradox of thrift is not an altogether legitimate extension of the concept as understood from traditional short-period analysis, particularly because of the uncertainty concerning subsequent changes in the capital stock and thus in the rate of accumulation.
12. This sort of interpretation was in fact that provided by Vianello in his 1985 paper. It should be noted that in communication with the author, Vianello has distanced himself from his 1985 argument to the extent that it involves the assumption that the economy returns to a position of normal utilization and therefore suggests a gravitation of actual around normal utilization analogous to the gravitation of prices around normal levels. What writers such as Garegnani, Vianello, Trezzini and Palumbo appear to favour is an interpretation of the long run giving a prime place to the adjustment of saving to investment via changes in the scale of productive capacity, but one which allows for even long-run deviations between actual and normal utilization, these latter being seen as an important symptom of demand-led growth. As noted already above, this position appears to leave little room for compatibility of both demand-led growth and steady state analysis (Palumbo and Trezzini, 2003, pp. 123–6).
13. Hence, although the profit share may be given, if normal utilization responds in the long run to deviations between normal and actual utilization, created for example by a faster rate of accumulation, the normal rate of profit will rise (see the argument in Barbosa-Filho, 2000, pp. 28–9). Here again the supposed conflict between the Sraffian approach to distribution and demand-led growth referred to in earlier sections is posed; with a demand-determined long-run capacity utilization offering 'a demand theory of income distribution but [one which] does not fit so well into the "classical theory" of income distribution' (ibid.).
14. We leave aside here the complication that changes in normal utilization rates will affect relative prices and distribution; namely, that a change in relative prices may impact on the profit-maximizing average rate of capacity utilization, thus leading to further changes in the normal rate.
15. This process would conceivably entail the rise in the rate of growth of investment over $t-n \rightarrow t$ giving rise in turn to a growth in capacity from $t \rightarrow t+n$ greater than the

anticipated growth rate of demand over $t \to t+n$ as producers 'catch-up' in response to their underestimation of demand growth over $t-n \to t$.

16. This is to adopt a different way of looking at the independent nature of effective demand – at least with respect to autonomous investment demand – from that in the Kaleckian growth literature. In the latter, the demand-led nature of growth is represented in terms of an 'accumulation function' with an element which is independent of the profit rate, utilization rate or the rate of growth of demand (see Blecker, 2003 and Commendatore et al., 2003 for surveys). This element is sometimes interpreted as representing investment which is dependent on 'animal spirits'. Yet this element of expenditure is nonetheless capacity-creating.
17. This formulation is based on a very simple accelerator to explain investment (see Trezzini, 1995). Expression (7.3) also involves a number of important assumptions: if I^A includes demand effects from technical change, our discussion so far implicitly ignores the effect of this on v (on this see Cesaratto et al., 2003); to the extent that I^A involves demand originating from government expenditures, our discussion ignores the constraints imposed by the requirement of financing those expenditures in a sustainable manner. It is also obvious that the complexity associated with different rates of growth of autonomous demand for different commodities is also ignored in the present exercise. Both issues are taken up in Section 7.7.
18. That such an assumption may allow for growth to be governed by the growth rate of autonomous demand is hinted at by Barbosa-Filho himself (ibid., p. 32).
19. Clearly, this assumption does impart to producers belief in some sort of economic model – a not uncommon practice in economic modelling. What is unusual here is that the 'model' in question is one of demand-led growth.
20. Consideration of more complex models would seem to be required, including a multi-commodity approach, a more flexible treatment of the accelerator (though retaining its essential element that growth of capacity is driven by expected growth in demand) and a more general treatment of autonomous non-capacity-creating demand. Some attempt to consider growth and autonomous demand in a more complex framework is provided in White (2003).
21. These two aspects don't appear to be clearly separated in either Trezzini's or P–T's analysis.
22. A similar argument is put in earlier literature by Garegnani (1992).
23. It is worth remarking here the divergence between actual and normal utilization referred to by Trezzini and by P–T as a result of the autonomous nature of demand is not the same matter as the divergence between the two rates in Kaleckian models, a divergence with which Lavoie (2003) is partly concerned. The latter relates to the inconsistency between the desire for the appropriate capacity in relation to demand and a steady state or dynamic equilibrium with a utilization rate different from the normal or desired rate. In a sense, the problem Lavoie seeks to correct is a notion of equilibrium which allows for a failure of producers to adapt capacity to demand. For P–T, on the contrary, the difference between actual and normal utilization happens in spite of and possibly in part due to attempts at this adjustment. Above all, for P–T, the probability of long-run divergences between actual and normal utilization reflects the autonomous nature of demand. As well, as was noted earlier, P–T use the divergence in the long run as a pointer to the lack of usefulness of the notion of normal utilization in the analysis of growth.
24. However, as White also demonstrates, the convergence of the actual utilization rate to the normal rate is much less apparent when the growth rate of autonomous demand is also governed by random fluctuations about discernible trends.
25. As White notes, this aspect of the model nonetheless has an important stabilizing effect on the system.

26. This is somewhat different from the P–T argument in that the latter argue that the difficulty of achieving full-adjustment after a change in the growth rate of autonomous demand is partly due to the demand effects of investment designed to restore full adjustment (p. 118). The view here however is that, in the absence of significant change in the growth rate of income triggered by the interaction of multiplier and inventory investment, the demand effects of fixed capital investment actually need to be significant for the adjustment in I^A/I. To the extent that such effects are weak, the overutilization or underutilization caused by the change in the growth rate of autonomous demand may be more persistent.
27. An anonymous referee has pointed out that recent Kaleckian literature includes efforts (for example, Lavoie, 1996; Dutt, 1997 and Commendatore, 2006) to move away from steady growth equilibria, which may offer some alternative to the position adopted by P–T. Reasons of space however preclude any exploration of these attempts in this chapter.
28. I am indebted to Attilio Trezzini for impressing this point on me.
29. The assumption of a small open economy, which is effectively made in the discussion below, amounts to the assumption that the growth of exports is independent of the growth of income of the domestic economy – in other words, a faster growth of the domestic economy does not influence growth in the rest of the world sufficiently to affect the growth of export demand for the country in question.
30. Strictly speaking, the relevant primary deficit ratio and the trade deficit ratios are proportions of nominal income. Hence to avoid complexity we ignore here foreign inflation rates as well as changes in exchange rates.
31. Larger X^Y, G^Y in other words entails a larger share of autonomous demand in income so that for steady growth the share of demand expansion induced by output growth (that is $v(Y^e_{t+1} - Y_t)/Y_t$) must be smaller and thus g_w must be smaller.
32. As a means of reducing the fiscal deficit in order to supposedly improve the current account position.
33. We ignore any interest on saving accruing to workers between receipt of wages at the end of $t - 1$ and expenditure at the end of t.
34. The significance of a discount for the dispersion of growth rates in the calculation of *expected* growth rates for the dynamic behaviour of multiplier–accelerator interaction was demonstrated by Franke and Weghorst (1988).

REFERENCES

Barbosa-Filho, N.H. (2000), ‘A note on the theory of demand-led growth’, *Contributions to Political Economy*, **19**:19–32.

Blecker, R.A. (2002), ‘Distribution, demand and growth in neo-Kaleckian macro-models’, in M. Setterfield (ed.), *The Economics of Demand-Led Growth: Challenging the Supply-Side Vision of the Long Run*, edited by Mark, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 129–52.

Cesaratto, S., F. Serrano and A. Stirati (2003), ‘Technical change, effective demand and employment’, *Review of Political Economy,* **15**(1):33–52.

Ciccone, R. (1986), ‘Accumulation and capacity utilization: some critical comments on Joan Robinson’s *Theory of Distribution*’, *Political Economy: Studies in the Surplus Approach*, **2** (1):17–36.

Commendatore, P. (2006), ‘Are Kaleckian models of growth and distribution relevant for the long run?’, in N. Salvadori and C. Panico (eds), *Classical, Neoclassical* *and Keynesian Views on Growth and Distribution*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Commendatore P., S. D'Acunto, C. Panico and A. Pinto (2003), 'Keynesian theories of growth', in N. Salvadori, *The Theory of Economic Growth: A 'Classical' Perspective*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 114–22.

Committeri, M. (1986), 'Some comments on recent contributions on capital accumulation, income distribution and capacity utilization', *Political Economy: Studies in the Surplus Approach*, **2**(2):161–86.

Dumenil, G.and D. Lévy (1999), 'Being Keynesian in the short term and classical in the long term: the traverse to classical long term equilibrium', *Manchester School of Economics and Social Studies*, **67**(6):684–716.

Dutt, A.K. (1997), 'Equilibrium, path dependence and hysteresis in post-Keynesian models', in P. Arestis, G. Palma and M. Sawyer (eds), *Markets, Unemployment and Economic Policy: Essays in honour of Geoff Harcourt*, Volume two, London and New York; Routledge, pp. 238–53.

Franke, R. and W. Weghorst (1988), 'Complex dynamics in a simple input–output model without the full capacity utilization hypothesis', *Metroeconomica,* **39**(1):1–29.

Garegnani, P. (1992), 'Some notes for an analysis of accumulation', in J. Halevi, D. Laibman and E. Nell (eds), *Beyond the Steady State: A Revival of Growth Theory*, London: Macmillan.

Garegnani, P. and A. Palumbo (1998), 'Accumulation of capital', in H.D. Kurz and N. Salvadori, *The Elgar Companion to Classical Economics*, Aldershot, UK and Brookfield, USA: Edward Elgar.

Harrod, R.F. (1939), 'An essay in dynamic theory', *Economic Journal*, **49**(193):14-33.

Kurz, H.D. (1994), 'Growth and distribution', *Review of Political Economy*, **6**(4): 393–411.

Lavoie, M. (1996), 'Traverse, hysteresis, and normal rates of capacity utilisation in Kaleckian models of growth and distribution', *Review of Radical Political Economics*, **28**(4):113–47.

Lavoie, M. (2003)*,* 'Kaleckian effective demand and Sraffian normal prices: towards a reconciliation'*, Review of Political Economy*, **15**(1):53–74.

Palumbo, A. and A. Trezzini (2003), 'Growth without normal capacity utilization', *European Journal of the History of Economic Thought*, **10**(1):109–35.

Panico, C. (1997), 'Government deficits in post-Keynesian theories of growth and distribution', *Contributions to Political Economy*, **16**(1):61–86.

Park, M.S. (1997), 'Accumulation, capacity utilization and distribution', *Contributions to Political Economy*, **16**:87–101.

Park, M.S. (2000), 'Autonomous demand and the warranted rate of growth', *Contributions to Political Economy*, **19**:1–18.

Pasinetti, L.L. (1998), 'The myth (or folly) of the 3% deficit/GDP Maastricht "parameter"', *Cambridge Journal of Economics*, **22**(1):103–16.

Serrano, F. (1995), 'Long period effective demand and the Sraffian Supermultiplier', *Contributions to Political Economy*, **14**:67–90.

Trezzini, A. (1995), 'Capacity utilization in the long run and the autonomous components of aggregate demand', *Contributions to Political Economy*, **14**:33–66.
Trezzini, A. (1998), 'Capacity utilization in the long run: some further considerations', *Contributions to Political Economy*, **17**:53–67.
Vianello, F. (1985), 'The pace of accumulation', *Political Economy: Studies in the Surplus Approach*, **1**(1).
White, G. (1996), 'Capacity utilization, investment and normal prices: Some issues in the Sraffa–Keynes synthesis', *Metroeconomica*, **47**(3):281–304.
White, G. (2003), 'Growth, autonomous demand and a joint-product treatment of fixed capital', mimeo.

8. Macroeconomic volatility and income inequality in a stochastically growing economy*

Cecilia García-Peñalosa and Stephen J. Turnovsky

8.1. INTRODUCTION

Research on the macroeconomic determinants of income inequality has focused mainly on three aspects: growth, trade and inflation. Studies of the impact of growth on distribution range from the analyses of the impact of structural change, such as the Kuznets hypothesis, to theories of skill-biased technical change. Based on the Heckscher–Ohlin theory, international trade has been argued to be a major determinant of income distribution, and this aspect has recently acquired prominence in the debate on the increase in inequality in a number of industrialized countries. One of the most consistently supported empirical correlations is that between inflation and inequality, explained by the fact that because inflation is a regressive tax, it generates greater income inequality. Our chapter seeks to introduce a new and so far ignored factor, the degree of aggregate risk in the economy, into the analysis of inequality.

Empirical evidence suggests that macroeconomic volatility is potentially an important channel through which income inequality and growth may be mutually related. A striking difference when we compare Latin American economies with those of the OECD is that the former are associated with much greater income inequality. In 1990, the Gini coefficients of the distribution of income in Brazil, Chile, Mexico and Venezuela ranged between 49–64 per cent, while those for the OECD range between 24–44 per cent. At the same time, the former were subject to much greater fluctuations in their respective growth rates than were the latter: during the 1980s, the standard deviation of the rate of output growth was, on average, 4.9 per cent for the four Latin American economies, and 2.7 per cent for the OECD.[1] In fact, using a broader set of data, Breen and García-Peñalosa

(2005) obtain a positive relationship between a country's volatility (measured by the standard deviation of the rate of GDP growth) and income inequality.

Our objective in this chapter is to model a mechanism through which aggregate risk – which we shall attribute to production shocks – has a direct impact on distribution. We employ an extension of the stochastic growth model developed by Grinols and Turnovsky (1993, 1998), Smith (1996), Corsetti (1997) and Turnovsky (2000b). This is a one-sector growth model in which, due to the presence of an externality stemming from the aggregate capital stock, equilibrium output evolves in accordance with a stochastic AK technology. Adopting this framework, aggregate production risk jointly determines the equilibrium growth rate, its volatility, and the distribution of income.

Previous studies have been unable to analyse the impact of volatility on income distribution, as they either abstract from labour, or otherwise, assume that agents are identical in all respects. We introduce the assumption that agents are heterogeneous with respect to their initial endowments of capital, and allow for an elastic supply of labour. As a result, the labour supply responses to different degrees of risk will induce changes in factor prices and affect the distribution of income.

Our analysis proceeds in several stages. To start with, we derive the equilibrium balanced growth path in a stochastic growth model with given tax rates. We show how this equilibrium has a simple recursive structure. First, the equilibrium mean growth rate and labour supply (employment) are jointly determined to ensure that rates of return are in equilibrium and that the product market clears. These equilibrium quantities depend upon the degree of risk in the economy but are independent of the distribution of wealth, which since the economy is always on its balanced growth path, remains unchanging over time. Second, the equilibrium labour supply, together with the given initial distribution of capital among agents, is shown to determine the volatility of the growth rate, on the one hand, and the degree of income inequality, on the other. We find that an increase in production risk raises the mean growth rate, its volatility, and the degree of income inequality.

The intuition for these results is straightforward. Because agents are sufficiently risk-averse, a greater variance of output has a strong income effect that induces them to increase their labour supply, increase their savings, and thus raise the growth rate. The increase in the labour supply raises the return to capital and lowers the real wage, thereby affecting the distribution of income. Since labour is more equally distributed than is capital, the income gap between any two individuals widens, and, for a given initial distribution of wealth, income distribution becomes more unequal.

The latter part of the chapter uses this framework to analyse the effects of taxation. As is well known, the externality associated with the capital stock implies that the competitive growth rate is too low. The first-best allocation can then be attained through suitable taxes and subsidies. When agents are heterogeneous, the use of growth-enhancing policies raises the question of the impact first-best policies have on the distribution of income. Two general conclusions emerge from our analysis. First, increasing (average) welfare and the growth rate does not necessarily entail an increase in inequality, as faster growth tends to be associated with lower post-tax income inequality. Second, we find that fiscal policy has conflicting effects on the distributions of gross and net income. First-best policies result in changes in factor prices that increase pre-tax inequality, but the direct redistributive effect of taxes tends to yield a more equal post-tax distribution. As a result gross and net income inequality often move in opposite directions.

The chapter contributes to the recent literature looking at the relationship between income distribution and growth.[2] It is related to Alesina and Rodrik (1994), Persson and Tabellini (1994) and Bertola (1993), who develop (non-stochastic) AK growth models in which agents differ in their initial stocks of capital. The first two papers have, however, a very different focus as they take initial inequality as given and argue that it has a negative impact on the rate of growth. In contrast to their results, this chapter emphasizes that growth and distribution are jointly determined, and presents a possible mechanism that generates a positive relationship between these two variables in line with the evidence presented by Forbes (2000). Bertola (1993) is closer to our approach in that he emphasizes how technological parameters, specifically the productivity of capital, jointly determine distribution and growth. He also examines how policies directed at increasing the growth rate affect the distribution of consumption, although his assumption of a constant labour supply implies that the distribution of income is independent of policy choices. Our approach shares with these three papers an important limitation, namely, that the assumption that agents differ only in their initial stocks of capital coupled with an AK technology implies that there are no income dynamics.[3]

The paper closest to our work, at least in spirit, is Aghion, Banerjee and Piketty (1999), who find that greater inequality is associated with more volatility. They show how combining capital market imperfections with inequality in a two-sector model can generate endogenous fluctuations in output and investment. In their model it is unequal access to investment opportunities and the gap between the returns to investment in the modern and the traditional sectors that cause fluctuations. We reverse the focus, examining how exogenous production uncertainty determines output volatility and income distribution.

The chapter is organized as follows. Section 8.2 presents the model and derives the equilibrium growth rate, labour supply and volatility. Section 8.3 examines the determinants of the distribution of income. Section 8.4 shows, analytically and numerically, that in the absence of taxation greater risk is associated with a more unequal distribution of income. Section 8.5 starts by obtaining the first-best optimum, and shows that the competitive growth rate is too low. It is followed by an analysis of first-best taxation, and a number of second-best policies. Numerical analysis is then used to illustrate the distributional implications of the various policies. Section 8.6 concludes by reviewing the main results, as well as relating the theoretical implications of this model to a range of empirical evidence. Finally, technical details are provided in the Appendix.

8.2. THE MODEL

8.2.1. Description of the Economy

8.2.1.1. Technology and factor payments

Firms shall be indexed by *j*. We assume that the representative firm produces output in accordance with the stochastic Cobb–Douglas production function

$$dY_j = A(L_jK)^{\alpha} K_j^{1-\alpha}(dt+du) \equiv F\left(L_jK, K_j\right)(dt+du) \tag{8.1a}$$

where K_j denotes the individual firm's capital stock, L_j denotes the individual firm's employment of labour, K is the average stock of capital in the economy, so that L_jK measures the efficiency units of labour employed by the firm; see for example, Corsetti (1997). The stochastic shock du is temporally independent, with mean zero and variance $\sigma^2 dt$ over the instant dt. The stochastic production function exhibits constant returns to scale in the private factors – labour and the private capital stock.

All firms face identical production conditions and are subject to the same realization of an economy-wide stochastic shock. Hence they will all choose the same level of employment and capital stock. That is, $K_j = K$ and $L_j = L$ for all *j*, where L is the average economy-wide level of employment. The average capital stock yields an externality such that in equilibrium the aggregate (average) production function is linear in the aggregate capital stock, as in Romer (1986), namely

$$dY = AL^{\alpha} K(dt + du) \equiv \Omega(L)K(dt + du) \tag{8.1b}$$

where $\Omega(L) \equiv AL^{\alpha}$ and $\partial\Omega / \partial L > 0$.

We assume that the wage rate, z, over the period $(t, t + dt)$ is determined at the start of the period and is set equal to the expected marginal physical product of labour over that period. The total rate of return to labour over the same interval is thus specified nonstochastically by

$$dZ = zdt = \left(\frac{\partial F}{\partial L_j} \right)_{K_j = K, L_j = L} dt \ . \tag{8.2a}$$

where

$$z = \alpha\Omega L^{-1} K \equiv wK \ .$$

The private rate of return to capital, dR, over the interval $(t, t + dt)$ is thus determined residually by

$$dR = \frac{dY - LdZ}{K} \equiv rdt + du_K \tag{8.2b}$$

where

$$r \equiv \left(\frac{\partial F}{\partial K_j} \right)_{K = K, L = L} = (1 - \alpha)\Omega \text{ , and } du_K \equiv \Omega du \ .$$

These two equations assume that the wage rate, z, is fixed over the time period $(t, t + dt)$, so that the return on capital absorbs all output fluctuations. The rationale for this assumption is that in industrial economies wages are usually fixed ex ante, while the return to capital is, at least in part, determined ex post and thus absorbs most of the fluctuations in profitability.[4] Differentiating the production function and given that firms are identical, we find that the equilibrium return to capital is independent of the stock of capital while the wage rate is proportional to the average stock of capital, and therefore grows with the economy.[5] In addition, we have $\partial r / \partial L > 0$ and $\partial w / \partial L < 0$, reflecting the fact that more employment raises the productivity of capital but lowers that of labour.

8.2.1.2. Consumers

There is a mass 1 of infinitely-lived agents in the economy. Consumers are indexed by i and are identical in all respects except for their initial stock of capital, K_{i0}. Since the economy grows, we will be interested in the share of individual i in the total stock of capital, k_i, defined as $k_i \equiv K_i/K$, where K is the aggregate (or average) stock. Relative capital has a distribution function $G(k_i)$, mean $\sum_i k_i = 1$, and variance σ_k^2.

All agents are endowed with a unit of time that can be allocated either to leisure, l_i, or to work, $1-l_i \equiv L_i$. A typical consumer maximizes expected lifetime utility, assumed to be a function of both consumption and the amount of leisure time, in accordance with the isoelastic utility function

$$\max E_0 \int_0^\infty \frac{1}{\gamma}\left(C_i(t)l_i^\eta\right)^\gamma e^{-\beta t}dt, \quad \text{with } -\infty < \gamma < 1, \eta > 0, \gamma\eta < 1 \tag{8.3}$$

where $1-\gamma$ equals the coefficient of relative risk aversion. Empirical evidence suggests that this is relatively large, certainly well in excess of unity, so that we shall assume $\gamma < 0$.[6] The parameter η represents the elasticity of leisure in utility. This maximization is subject to the agent's capital accumulation constraint

$$\begin{aligned} dK_i = (1-\tau_k)rK_i dt + (1-\tau_k')K_i du_K + (1-\tau_w)w(1-l_i)Kdt \\ -(1+\tau_c)C_i dt + sE(dK_i) + s'[dK_i - E(dK_i)] \end{aligned} \tag{8.4}$$

where $du_K = \Omega du$, and capital is assumed not to depreciate. According to (8.4), the agent's deterministic component of capital income is taxed at τ_k, while the tax on the stochastic component is τ_k'. The tax rate on the (non-stochastic) labour income is τ_w and on consumption (which is also non-stochastic) is τ_c. Finally, the deterministic and the stochastic components of investment in physical capital are subsidized, at rates s and s', respectively. Taking expectations of this expression and substituting back for $E(dK_i)$, we can express this budget constraint as

$$dK_i = \frac{(1-\tau_k)rK_i + (1-\tau_w)w(1-l_i)K - (1+\tau_c)C_i}{1-s}dt + K_i\frac{1-\tau_k'}{1-s'}du_K \tag{8.5a}$$

It is important to observe that with the equilibrium wage rate being tied to the aggregate stock of capital, the rate of accumulation of the individual's capital stock depends on the aggregate stock of capital, which in turn evolves according to

$$dK = \frac{[(1-\tau_k)r + (1-\tau_w)w(1-l)]K - (1+\tau_c)C}{1-s}dt + K\frac{1-\tau_k'}{1-s'}du_K \tag{8.5b}$$

where l denotes the average (aggregate) fraction of time devoted to leisure. The agent therefore needs to take this relationship into account in performing her optimization.

8.2.1.3. Government policy

The government balances the public budget each period, implying

$$sE(dK)dt + s'\frac{1-\tau_k'}{1-s'}Kdu_K = \left[\tau_c C + \tau_k rK + \tau_w w(1-l)K\right]dt + \tau_k' Kdu_K \quad (8.6)$$

Note that both expenditures and tax receipts have a deterministic and a stochastic component. Equating them respectively yields the following constraints required to maintain a balanced budget:

$$\tau_k' = s', \quad (8.7a)$$

$$\tau_k r + \tau_w w(1-l) + \tau_c \frac{C}{K} = s\frac{E(dK)}{K} \equiv s\psi, \quad (8.7b)$$

where ψ denotes the average growth rate.

Two points should be noted. First, some of the taxes may be negative, in which case they become subsidies, in addition to the investment subsidy. However, neither the deterministic nor the stochastic component of the two income taxes can exceed unity. Second, the assumption that the deterministic and stochastic components of income are taxed at different rates requires that the agent (and the tax authority) disentangle the deterministic from the stochastic components of income, something that may not be unlikely in practice. However, this assumption is mainly made for analytical simplicity. Taxing both components at the same rate, with the government using public debt in order to compensate any surplus or deficit, would not change any of our results, since as we will see below, the tax rate on the stochastic component of capital income does not affect any of the equilibrium relationships (see Turnovsky, 2000b and García-Peñalosa and Turnovsky, 2005).

8.2.2. Consumer Optimization

The consumer's formal optimization problem is to maximize (8.3) subject to equations (8.5a) and (8.5b). The first-order conditions with respect to consumption and leisure yield

$$\frac{1}{C_i}\left(C_i l_i^{\eta}\right)^{\gamma} = \frac{1+\tau_c}{1-s} X_{K_i} \quad (8.8a)$$

$$\frac{\eta}{l}\left(C_i l_i^{\eta}\right)^{\gamma} = \frac{1-\tau_w}{1-s} wKX_{K_i} \tag{8.8b}$$

where $X(K_i, K)$ is the value function and X_{K_i} its derivative with respect to K_i (see Appendix).

In the Appendix we show that utility maximization implies that the dynamic evolution of the stock of capital of agent *i* is given by (where we employ (8.7a))

$$\frac{dK_i}{K_i} = \left[\frac{r(1-\tau_k)-\beta(1-s)}{(1-\gamma)(1-s)} - \frac{\gamma}{2}\Omega^2\sigma^2\right]dt + \Omega\left(\frac{1-\tau_k'}{1-s'}\right)du \equiv \psi dt + \Omega du\,, \tag{8.9}$$

and Ω and r are defined in equations (8.1) and (8.2). There we have expressed them as functions of equilibrium employment, L. Assuming that the aggregate labour market clears, yields

$$L \equiv \sum\nolimits_j L_j = \sum\nolimits_i (1-l_i) \equiv 1-l \tag{8.10}$$

so that we can equally well write Ω and r as functions of $(1-l)$.[7]

From (8.9) we see that the only difference between agents, namely their initial stock of capital, does not appear in this equation. Hence all individuals choose the same rate of growth of their stock of capital. This has two implications. First, the *aggregate* rate of growth of capital is identical to the *individual* rate of growth and unaffected by the initial distribution of endowments, hence

$$\frac{dK}{K} = \psi dt + \Omega du\,. \tag{8.9$'$}$$

Second, since the capital stock of all agents grows at the same rate, the distribution of capital endowments does not change over time. That is, at any point in time, the wealth share of agent i, k_i, is given by her initial share $k_{i,0}$, say. We also observe that the rate of growth of capital has a deterministic and a stochastic component, so that the average growth rate, ψ is defined by (8.9) and its standard deviation, σ_ψ, is $\sigma_\psi = \Omega\sigma$.

Dividing equation (8.8a) by (8.8b), we obtain the consumption–capital ratio of agent *i,*

$$\frac{C_i}{K_i} = \frac{w}{\eta}\frac{1-\tau_w}{1+\tau_c}\frac{l_i}{k_i}\,, \tag{8.11}$$

Aggregating over the individuals and noting that $\sum_i k_i = 1, \ \sum_i l_i = l,$ the aggregate economy-wide consumption–capital ratio is

$$\frac{C}{K} = \frac{w}{\eta} \frac{1-\tau_w}{1+\tau_c} l \tag{8.11$'$}$$

In addition, the following transversality condition must hold

$$\lim_{t \to \infty} E\left[K_i(t)^{\gamma} e^{-\beta t} \right] = 0 \tag{8.12}$$

With $K_i(t)$ evolving in accordance with the stochastic path (8.9), (8.12) can be shown to reduce to (see Turnovsky, 2000b)

$$\gamma \left[r \frac{1-\tau_k}{1-s} - \frac{\gamma}{2}(1-\gamma)\Omega^2 \sigma^2 \right] < \beta$$

which, when combined with the definition of ψ can be shown to be equivalent to the condition

$$r\left(\frac{1-\tau_k}{1-s} \right) > \psi \tag{8.13}$$

that is the equilibrium rate of return on capital must exceed the equilibrium growth rate. Dividing the aggregate accumulation equation, (8.4b), by K, this condition can also be shown to be equivalent to

$$(1+\tau_c)\frac{C}{K} > (1-\tau_w) w (1-l) \tag{8.13$'$}$$

implying that part of income from capital is consumed.[8] Combining with (8.9′), this can be further expressed as

$$l > \frac{\eta}{1+\eta}. \tag{8.13$''$}$$

Recalling the individual budget constraint, (8.4a), we can write the individual's mean rate of capital accumulation as

$$\frac{E(dK_i / K_i)}{dt} = r\frac{1-\tau_k}{1-s} + w\frac{1-\tau_w}{1-s} \frac{1-l_i}{k_i} - \frac{1+\tau_c}{1-s} \frac{C_i}{K_i} \tag{8.14}$$

Together with equation (8.11), this expression implies that agent i's supply of labour is

$$1-l_i=\frac{1}{1+\eta}\left[1-\eta\frac{1}{1-\tau_w}\frac{r(1-\tau_k)-\psi(1-s)}{w}k_i\right] \tag{8.15}$$

Noting the transversality condition, $r(1-\tau_k)/(1-s)>\psi$, (8.15) implies that an increase in the agent's capital (wealth) has a negative effect on her labour supply; wealthier individuals chose to 'buy' more leisure. In effect, they compensate for their larger capital endowment, and the higher growth rate it would support, by providing less labour and having an exactly offsetting effect on the growth rate.

Because the rate of growth is the same for all agents, individual labour supplies are linear in the wealth shares of agents. The aggregate labour supply, $1-l=1-\sum_i l_i$, is then independent of the initial distribution of capital. Summing equation (8.15) over the agents and using the fact that $\sum_i k_i=1$, we obtain the aggregate labour supply relation,

$$1-l=\frac{1}{1+\eta}\left[1-\eta\frac{1}{1-\tau_w}\frac{r(1-\tau_k)-\psi(1-s)}{w}\right], \tag{8.15$'$}$$

and combining (8.15) and (8.15′) we can derive the following expression for the 'relative labour supply'

$$l_i-l=\left(l-\frac{\eta}{1+\eta}\right)(k_i-1) \tag{8.15$''$}$$

Again we see that the transversality condition, now expressed as (8.13″), implies a positive relationship between relative wealth and leisure. This relationship provides the fundamental mechanism whereby, given the initial distribution of capital endowments across agents, policy and risk are able to influence the distribution of income.

8.2.3. Macroeconomic Equilibrium

The key equilibrium relationships can be summarized by

Equilibrium growth rate

$$\psi=\frac{r(1-\tau_k)-\beta(1-s)}{(1-\gamma)(1-s)}-\frac{\gamma}{2}\Omega^2\sigma^2 \tag{8.16a}$$

Equilibrium volatility

$$\sigma_\psi = \Omega\sigma \tag{8.16b}$$

Individual consumption–capital ratio

$$\frac{C_i}{K_i} = \frac{w}{\eta}\frac{1-\tau_w}{1+\tau_c}\frac{l_i}{k_i} \tag{8.16c}$$

Aggregate consumption–capital ratio

$$\frac{C}{K} = \frac{w}{\eta}\frac{1-\tau_w}{1+\tau_c} l \tag{8.16d}$$

Individual budget constraint

$$\psi = r\frac{1-\tau_k}{1-s} + w\frac{1-\tau_w}{1-s}\frac{1-l_i}{k_i} - \frac{1+\tau_c}{1-s}\frac{C_i}{K_i} \tag{8.16e}$$

Goods market equilibrium

$$\psi = \Omega - \frac{C}{K} \tag{8.16f}$$

Government budget constraint

$$\tau_k r + \tau_w w(1-l) + \tau_c \frac{C}{K} = s\psi \tag{8.16g}$$

Recalling the definitions of $r(l)$, $w(l)$, and $\Omega(l)$, and given k_i, these equations jointly determine the individual and aggregate consumption–capital ratios, C_i/K_i, C/K, the individual and aggregate leisure times, l_i, l, average growth rate, ψ, volatility of the growth rate, σ_ψ, and one of the fiscal instruments given the other three policy parameters. Note that the tax and subsidy on the stochastic components of investment and the return to capital, have no effect on the equilibrium variables and thus $\tau_k' = s'$ can be set arbitrarily.

Using (8.16a), (8.16d) and (8.16f), the macroeconomic equilibrium of the economy can be summarized by the following pair of equations that jointly determine the equilibrium mean growth rate, ψ, and leisure l:

$$\text{RR} \quad \psi = \frac{(1-\alpha)\Omega(l)(1-\tau_k) - \beta(1-s)}{(1-\gamma)(1-s)} - \frac{\gamma}{2}[\Omega(l)]^2\sigma^2, \tag{8.17a}$$

PP $$\psi = \Omega(l)\left[1 - \frac{\alpha}{\eta}\frac{1-\tau_w}{1+\tau_c}\frac{l}{1-l}\right]. \tag{8.17b}$$

The first equation, denoted RR, describes the relationship between ψ and l that ensures the equality between the risk-adjusted rate of return to capital and return to consumption. The second describes the combinations of the mean growth and leisure that ensure product market equilibrium holds, and will be denoted PP.

8.2.4. The Laissez-Faire Economy

It is convenient to examine the equilibrium in the absence of taxation. Setting all taxes and subsidies to zero, the equilibrium mean growth rate and leisure are determined by the following pair of equations:

RR $$\psi = \frac{(1-\alpha)\Omega(l) - \beta}{1-\gamma} - \frac{\gamma}{2}\Omega(l)^2\sigma^2,$$

PP $$\psi = \Omega(l)\left(1 - \frac{\alpha}{\eta}\frac{l}{1-l}\right),$$

The laissez-faire RR and PP locuses are depicted in Figure 8.1, and their formal properties are derived in the Appendix (see also Turnovsky, 2000b). First, note that equation PP is always decreasing in l, reflecting the fact that more leisure time reduces output, thus increasing the consumption–output ratio and having an adverse effect on the growth rate of capital. On the other hand, for RR we have

$$\frac{\partial \psi}{\partial l} = \left[\frac{1-\alpha}{1-\gamma} - \gamma\Omega(l)\sigma^2\right]\Omega'(l)$$

This expression is unambiguously negative for $\gamma < 0$, as the empirical evidence suggests, and the case that we shall assume prevails. Intuitively, a higher fraction of time devoted to leisure reduces the productivity of capital, requiring a fall in the return to consumption. This is obtained if the growth of the marginal utility of consumption rises, that is, if the balanced growth rate falls. Under plausible conditions, the two schedules are concave, and an equilibrium exists if

$$\alpha - \gamma + \frac{\beta}{A} > -\gamma\frac{1-\gamma}{2}A\sigma^2.$$

We will see in our numerical calibrations that this condition is met for reasonable parameter values.

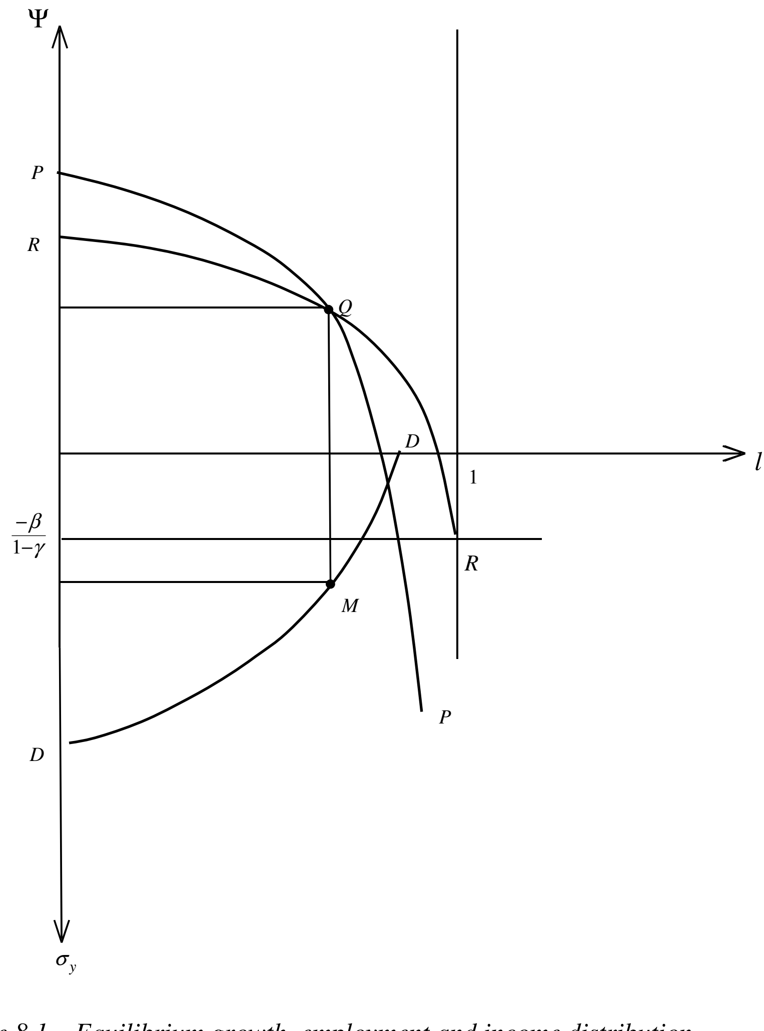


Figure 8.1 Equilibrium growth, employment and income distribution

8.3. THE DETERMINANTS OF THE DISTRIBUTION OF INCOME

In order to examine the effect of risk on income distribution, we consider the expected relative income of an individual with capital k_i. Her (expected) gross income is simply $E(dY_i) = rK_i + wK(1-l_i)$, while expected average income is $E(dY) = rK + wK(1-l)$. Using equation (8.15) to substitute for labour, we can express the relative (expected) income of individual i, $y_i \equiv E(dY_i)/E(dY)$, as

$$y_i(l, k_i) = k_i + \frac{w}{(1+\eta)\Omega}(1-k_i) = k_i + \frac{\alpha}{(1+\eta)(1-l)}(1-k_i) \qquad (8.18)$$

which we may write more compactly as:

$$y_i(l, k_i) = 1 - \rho(l)(1-k_i), \quad \text{where} \quad \rho(l) \equiv 1 - \frac{\alpha}{(1+\eta)(1-l)}, \qquad (8.18')$$

Equation (8.18') emphasizes that the distribution of income depends upon *two* factors, the initial (unchanging) distribution of capital, and the equilibrium allocation of time between labour and leisure, insofar as this determines factor rewards. The net effect of an increase in initial wealth on the relative income of agent *i* is given by $\rho(l)$. As long as the equilibrium is one of positive growth, it is straightforward to show that[9]

$$0 < \rho(l) < 1 \qquad (8.19)$$

Thus relative income is strictly increasing in k_i, indicating that although richer individuals choose a lower supply of labour, this effect is not strong enough to offset the impact of their higher capital income. As a consequence, the variability of income across the agents, σ_y, is less than their (unchanging) variability of capital, σ_k.

The second point to note is that we can rank different outcomes according to inequality without needing any information about the underlying distribution of capital. For a given distribution of capital, changes in risk or policy affect the distribution of income solely through their impact on relative prices, as captured by $\rho(l)$. Correia (1999) has shown that when agents differ only in their endowment of one good, there exists an ordering of outcomes by income inequality, as measured by second-order stochastic dominance.[10] That ordering is determined by equilibrium prices, and is independent of the distribution of endowments.

The DD locus in the lower panel of Figure 8.1 illustrates the relationship between the standard deviation of relative income, σ_y, our measure of income inequality and the standard deviation of capital endowments, σ_k, namely

$$\text{DD} \qquad \sigma_y = \rho(l)\sigma_k \qquad (8.17c)$$

Given the standard deviation of capital, σ_k, the standard deviation of income is a decreasing and concave function of aggregate leisure time. This is because as leisure increases (and labour supply declines) the wage rate rises and the return to capital falls, compressing the range of income flows between the wealthy with large endowments of capital and the less well endowed. Thus, having determined the equilibrium allocation of labour from the upper panels in Figure 8.1, (8.17c) determines the corresponding unique variability of income across agents.

Because taxes also have direct redistributive effects, we need to distinguish between the *before-tax* and *after-tax* distribution of income. We therefore define the agent's after-tax (or net) relative income as

$$y_i^{NET}(l,k_i,\tau_k,\tau_w) \equiv \frac{r(1-\tau_k)k_i + w(1-\tau_w)(1-l_i)}{r(1-\tau_k)+w(1-\tau_w)(1-l)} = 1-\rho^{NET}(l,\tau_w,\tau_k)(1-k_i) \tag{8.20a}$$

where, ρ^{NET} summarizes the distribution of after-tax income and is related to corresponding before-tax measure, $\rho(l)$, by

$$\rho^{NET}(l,\tau_w,\tau_k) = \rho(l) + \left[1-\rho(l)\right](1-\alpha)\frac{(\tau_w-\tau_k)}{\alpha(1-\tau_w)+(1-\alpha)(1-\tau_k)} \tag{8.20b}$$

with the standard deviation of after-tax income given by

$$\sigma_y^{NET} = \rho^{NET}(l,\tau_w,\tau_k)\sigma_k \tag{8.20c}$$

From (8.20a) and (8.20b) we see that fiscal policy exerts two effects on the after-tax income distribution. First, by influencing *gross* factor returns it influences the equilibrium supply of labour, l, and therefore the before-tax distribution of income, as summarized by $\rho(l)$. In addition, it has a direct redistributive effect, which is summarized by the second term on the right-hand side of (8.20b). The dispersion of pre-tax income across agents will exceed the after-tax dispersion if and only if $\tau_k > \tau_w$. As we will see below, in most cases tax increases affect the before-tax and after-tax distributions in opposite ways.

Lastly, we compute individual welfare. By definition, this equals the value function used to solve the intertemporal optimization problem evaluated along the equilibrium stochastic growth path. For the constant elasticity utility function, the optimized level of utility for an agent starting from an initial stock of capital, $K_{i,0}$, can be expressed as

$$X(K_{i0}) = \frac{1}{\gamma}\left[\frac{C_i}{K_i}l_i^{\eta}\right]^{\gamma}\left[\beta-\gamma\left(\psi+\frac{1}{2(\gamma-1)\sigma^2}\right)\right]^{-1}K_{i,0}^{\gamma} \tag{8.21}$$

The welfare of individual i relative to that of the individual with average wealth is then

$$x(k_i) = \left(\frac{C_i}{K_i}\right)^{\gamma}\left(\frac{C}{K}\right)^{-\gamma}\frac{l_i^{\eta\gamma}}{l^{\eta\gamma}}k_i^{\gamma} = \left(\frac{l_i}{l}\right)^{(1+\eta)\gamma}, \tag{8.22}$$

where the second term has been obtained by substituting for the consumption–capital ratio. Using equations (8.15), we can express relative welfare as

$$x(k_i) = \left[1+\left(1-\frac{\eta}{1+\eta}\frac{1}{l}\right)(k_i-1)\right]^{\gamma(1+\eta)}. \tag{8.22′}$$

Consider now two individuals having relative endowments $k_2 > k_1$. Individual 2 will have both a higher mean income but also higher volatility. The transversality condition (8.13″) implies that if $\gamma > 0$, then their relative welfare satisfies $x(k_2) > x(k_1) > 0$, while if $\gamma < 0$, $x(k_1) > x(k_2) > 0$. However, in the latter case absolute welfare, as expressed by (8.19) is negative. Thus in either case, the better endowed agent will have the higher absolute level of welfare.

8.4. THE RELATIONSHIP BETWEEN VOLATILITY AND INEQUALITY

We now turn to the relationship between volatility, growth, and the distribution of income, focusing on how these relationships respond to an increase the volatility of production, σ^2. In this section we examine the case of an economy without taxation. We discuss the relationship analytically and then supplement this with some numerical simulations.

8.4.1. Analytical Properties

The effect of risk operates through its impact on the incentives to accumulate capital. An increase in σ^2 shifts the RR curve only, and for $\gamma <$ 0, it shifts the RR curve upwards, as seen in Figure 8.2. Given the fraction of time devoted to leisure, the shift in RR tends to increase the growth rate. The higher ψ increases the return to consumption, which raises the labour supply, and hence the return to capital relative to that of consumption, causing a further increase in the growth rate. Thus the increase in risk raises

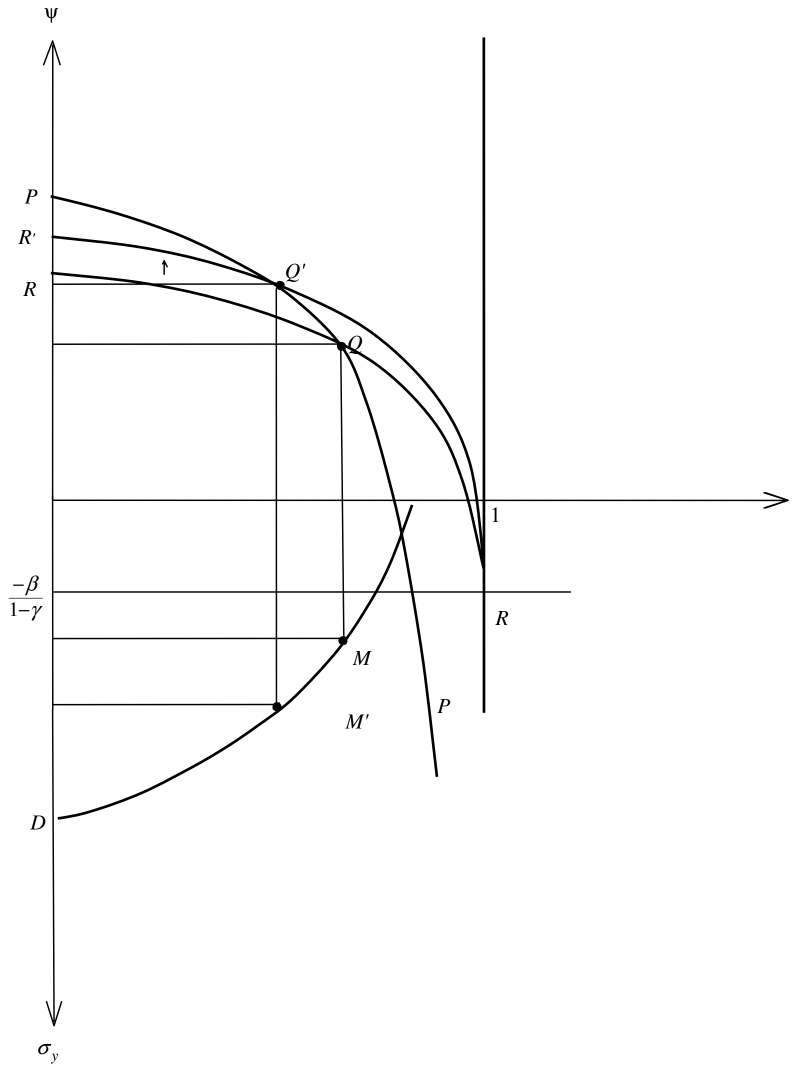


Figure 8.2 Increase in production risk

the mean growth rate and reduces leisure unambiguously, as the equilibrium moves from Q to Q' along PP. In addition greater risk increases the variance of the growth rate, $\sigma_{\psi}^{2} = \Omega^{2}\sigma^{2}$, because of both the direct effect of σ^{2} and the indirect impact of a lower l on Ω.

From equations (8.18′) and (8.20), we see that the effect of an increase in risk on the gross and net distributions of income are given by

$$\frac{d\sigma_y}{d\sigma^2} = \sigma_k \frac{\partial\rho}{\partial l}\frac{dl}{d\sigma^2} > 0,$$

$$\frac{d\sigma_y^{NET}}{d\sigma^2} = \sigma_k \left[\frac{1-\tau_w}{\alpha(1-\tau_w)+(1-\alpha)(1-\tau_k)}\right]\frac{\partial\rho}{\partial l}\frac{dl}{d\sigma^2} > 0$$

An increase in l raises both pre-tax and post-tax inequality. Greater volatility of the production shock, by reducing the amount of time devoted to leisure, increases income inequality, as measured by the standard deviation of relative incomes. Pre-tax inequality will increase more than post-tax inequality if and only $\tau_k > \tau_w$, that is, if and only if the initial pre-tax income inequality exceeds the initial post-tax inequality.

Risk will also increase measures of inequality other than the standard deviation. To see this it suffices to note that the effect of an increase in risk on the relative gross income of an agent with capital share k_i is given by

$$\mathrm{sgn}\left[\frac{dy(k_i)}{d\sigma^2}\right] = -\mathrm{sgn}\left[\frac{\partial\rho}{\partial l}\frac{dl}{d\sigma^2}(1-k_i)\right] = \mathrm{sgn}\left(k_i - 1\right)$$

An increase in risk raises the income share for those with a wealth share above the average, and reduces the income share of those with wealth below. Consequently, inequality rises.

The intuition for these results is as follows. Because agents are sufficiently risk-averse, a greater variance of output has a strong income effect that makes them increase savings. Consequently, the growth rate increases. Note from the PP locus that the allocation of labour is unaffected by σ^2 *for a given growth rate*. A higher growth rate, however, implies higher future wages, and hence higher consumption for any extra time spent at work. It therefore reduces leisure time and increases the labour supply. The change in the labour supply, in turn, affects the distribution of income. A higher labour supply increases the return to capital and lowers the wage rate. Since labour is distributed more equally than is capital, the income gap between any two individuals increases, and income inequality increases.

Note that with an inelastic supply of labour, risk would not affect relative incomes. In this case, the income of agent i would be given by $y_i = (w + rk_i)/(w + r)$. With the AK technology resulting in a constant wage and interest rate, this expression would be unaffected by risk. In our setup risk matters because it affects the growth rate, and this, in turn, impacts the labour supply and factor rewards.

8.4.2. Numerical Examples

To obtain further insights into the impact of risk on the equilibrium, and in particular the relationship between growth and income inequality, we perform some numerical analysis. In order to do so we use the following, mostly conventional, parameter values:

Production	$A = 0.75,\ \alpha = 0.60$
Taste	$\beta = 0.04,\ \gamma = -2,\ \eta = 1.75$
Risk	$\sigma = 0.05,\ 0.10,\ 0.20,\ 0.30,\ 0.40$

The choice of production elasticity of labour measured in efficiency units implies that 60 per cent of output accrues to labour. One consequence of the Romer technology being assumed, is that whereas this value is realistic in terms of the labour share of output, it implies an implausibly large externality from aggregate capital which implies extreme solutions for the first-best fiscal policy, discussed below. The choice of the scale parameter A = 0.75, is set to yield a plausible value for the equilibrium capital–output ratio.

Turning to the taste parameters, the rate of time preference of 4 per cent is standard, while the choice of the elasticity on leisure, $\eta = 1.75$, is standard in the real business cycle literature, implying that about 72 per cent of time is devoted to leisure, consistent with empirical evidence. Estimates of the coefficient of relative risk aversion are more variable throughout the literature. Values of the order of $\gamma = -18$ (and larger) have sometimes been assumed to deal with the equity premium puzzle (see Obstfeld, 1994). However, these tend to yield implausibly low values of the equilibrium growth rate. By contrast, real business cycle theorists routinely work with logarithmic utility functions ($\gamma = 0$). More recently, a consensus seems to be emerging of values between 2 and 5 (see Constantinides et al., 2002) and our choice of $\gamma = -2$ is well within that range.

Our main focus is on considering increases in exogenous production risk, which we let vary between $\sigma = 0.05$ and $\sigma = 0.40$. The value $\sigma = 0.05$ is close to the mean for OECD countries considered by Gali (1994) and Gavin and Hausmann (1995). Gavin and Hausmann present estimates for a wide range of countries and $\sigma = 0.10$ corresponds to countries subject to medium production risk. For virtually all countries they find $\sigma < 0.20$ so that the values $\sigma = 0.30$, $\sigma = 0.40$ are beyond the bounds of plausibility and are reported only to broaden the sensitivity analysis.

Choosing the distribution of wealth is less straightforward, as data on the distribution of wealth are difficult to obtain. Moreover, income distributions are reported in terms of Gini coefficients (rather than standard deviations as employed in our theoretical discussion) and Table 8.1 reports some actual distributions. The first two lines are the distributions of income in the US and Sweden in 1991 and 1992, respectively (from Deininger and Squire, 1996). The third is our hypothetical distribution of wealth. The values assumed are consistent with the data. For example, in the US in 1992, the bottom 40 per cent of the population held 0.4 per cent of total wealth, while the top 20 per cent owed 83.8 per cent of the total (see Wolff, 1998).

The last line reports the income distribution generated by the model, using the hypothetical wealth distribution for the case of low risk $\sigma = 0.05$. To obtain it we have assumed that the bottom income group has no wealth and no labour endowment (that is, zero income). Otherwise, since wages are identical for all workers, we would have a very large group with the same income at the bottom of the income distribution. Our assumption implies that the income share of the two bottom groups is 18.9 and hence of a similar magnitude to that observed in the data (15.2 and 19.1 for the US and Sweden, respectively). The resulting Gini coefficient lies between those of these two countries.

Table 8.1 The distribution of income and wealth

	Q1	Q2	Q3	Q4	Q5	Gini
US: income shares	4.6	10.6	16.6	24.6	43.6	39.1
Sweden: income shares	6.3	12.8	19.2	24.8	36.9	31.1
Assumed wealth shares	0	0	1.2	12	86.8	74.2
Assumed wealth levels	0	0	0.06	0.6	4.34	
Simulated income shares ($\sigma = 0.05$)		18.9	19.1	21.8	40.2	33.30

Table 8.2 reports the impact of increases in the volatility of the output shock on the equilibrium labour supply, the average rate of growth and its standard deviation, the Gini coefficient of income, and on overall welfare. Welfare changes reported are calculated as the percentage equivalent variations in the initial stock of capital of the average individual necessary to maintain the level of utility following the increase in risk from the benchmark level $\sigma = 0.05$ reported in the first row.

Table 8.2 Growth and distribution of income and wealth

	l	ψ	σ_ψ	ΔX	Gini(y)
$\sigma = 0.05$	72.5	3.30	1.73	–	33.30
$\sigma = 0.1$	72.4	3.40	3.46	–0.68	33.40
$\sigma = 0.2$	72.2	3.79	6.96	–3.41	33.82
$\sigma = 0.3$	71.8	4.45	10.53	–8.04	34.52
$\sigma = 0.4$	71.2	5.43	14.22	–14.70	35.51

Line 1 of the table suggests that treating $\sigma = 0.05$ as a benchmark case leads to a plausible equilibrium, having a 3.3 per cent mean growth rate and 1.73 per cent relative standard deviation, with 72.5 per cent of time allocated to leisure and a capital–output ratio (not reported) of approximately 3.[11] The implied distribution of income is also plausible, as noted.

As risk increases from $\sigma = 0.05,$ Table 8.2 indicates the following. The mean growth rate increases, as does its standard deviation. The net effect of the greater risk dominates the positive effect of the higher growth rate, so that the increase in risk reduces average welfare. It should be noted that for the plausible range of $\sigma < 0.20,$ the welfare loss is relatively modest. This is a characteristic limitation of this class of model having only aggregate risk, and has been discussed elsewhere in the literature.[12] More to the point here, we see that greater risk is associated with a substitution toward more labour (less leisure), and an increase in income inequality – as measured by the Gini coefficient – consistent with the formal analysis presented in Section 8.3.1.

In terms of magnitudes, the effect of risk on the Gini coefficient is quite modest, at least for plausible degrees of risk. It is interesting to note that income inequality in the US increased by 2.5 Gini points between 1980 and 1990, and that this has been considered a sizeable increase. From Table 8.2 it is seen that for risk alone to generate a similar increase it would have had to increase from $\sigma = 0.05$ to around $\sigma = 0.4,$ which is obviously implausible. Clearly other structural and policy changes are primarily responsible. However, small changes in risk may still play a significant role if they give rise to large policy responses.

8.5. TAXATION

A familiar feature of the Romer (1986) model is that by ignoring the externality associated capital, the decentralized economy generates a sub-optimally low growth rate. This suggests that an investment subsidy that increases the growth rate will move the equilibrium closer to the social optimum. With heterogeneous agents, two questions arise. First, how to finance this subsidy if the government is concerned about inequality as well as about average welfare. An investment subsidy raises the return to capital and will tend to favour those with large capital holdings. If the subsidy were financed by a lump-sum tax, the system would redistribute away from those with lower incomes to those with higher incomes. Are there ways in which this reverse redistribution can be avoided? Second, we want to know whether the use of first-best policies has any implication for the relationship between volatility and inequality. In this section we investigate these questions in some detail. We begin by deriving the first-best optimal rate of growth and allocation of labour.

8.5.1. The First-best Optimum

Given the externality stemming from the aggregate capital stock, finding the first-best optimums amounts to solving the following problem:

$$\max E_0 \int_0^\infty \frac{1}{\gamma}\left(C_i(t) l_i^\eta\right)^\gamma e^{-\beta t} dt, \tag{8.23a}$$

subject to

$$dK_i = \left(\Omega K_i - C_i\right) dt + \Omega K_i du \tag{8.23b}$$

In the Appendix we show that the solution to this problem is given by the equations

$$\text{R}'\text{R}' \qquad \tilde{\psi} = \frac{\Omega(\tilde{l}) - \beta}{1-\gamma} - \frac{\gamma}{2}\Omega(\tilde{l})^2 \sigma^2 \tag{8.17a$'$}$$

$$\text{PP} \qquad \tilde{\psi} = \Omega(\tilde{l})\left[1 - \frac{\alpha}{\eta}\frac{\tilde{l}}{1-\tilde{l}}\right] \tag{8.17b$'$}$$

$$\text{DD} \qquad \sigma_y = \rho(\tilde{l})\sigma_k \tag{8.17c}$$

$$\frac{\tilde{C}}{\tilde{K}} = \frac{\alpha\Omega(\tilde{l})}{\eta}\frac{\tilde{l}}{1-\tilde{l}} \tag{8.17d}$$

where the tilde denotes the first-best optimum. Note that the only difference with the solution to the competitive equilibrium in the absence of taxes is that the social rate of return to capital now takes into account the production externality and hence exceeds the private return.[13] The R′R′ schedule lies above RR. Given that the PP schedule is steeper than RR, the upward shift of RR results in a higher growth rate, lower leisure, and therefore increases inequality, as can be seen from Figure 8.2.[14]

8.5.2. First-best Taxation

Comparing the first-best optimum, described by R′R′, and PP with the decentralized equilibrium, RR, PP we can see that the tax-subsidy system can be used to attain the optimal growth rate, leisure time, and consumption–capital ratio by setting

$$\frac{1-\tau_w}{1+\tau_c} = 1, \quad \text{i.e. } \tau_w = -\tau_c, \tag{8.24a}$$

$$(1-\alpha)\frac{1-\tau_k}{1-s} = 1, \quad \text{i.e. } s = \alpha + (1-\alpha)\tau_k, \tag{8.24b}$$

$$\frac{1-\tau_k}{1-\tau_w} = \frac{1}{1-\alpha}\left(1-\eta\frac{1-\tilde{l}}{\tilde{l}}\right), \tag{8.24c}$$

where the last equation is obtained from the government's budget constraint, (8.16g). The first two equations represent intuitive optimality conditions. The first states that any wage tax should be offset with an equivalent consumption tax so as not to distort the leisure-consumption choice. Interpreting the tax on wage income as a negative tax on leisure, (8.24a) states that the two utility enhancing goods, consumption and leisure, should be taxed uniformly. The second condition simply ensures that the private rate of return on investment must equal the social return, and for this to be so the subsidy to investment must exceed the externality by an amount that reflects any tax on capital income.[15] Note from the third equation that unless consumption equals total income (in which case there is zero growth), the replication of the first-best optimum requires differential taxes on wages and capital; $\tau_w \lesseqgtr \tau_k$ according to whether there is positive or negative growth.

Equations (8.24) indicate the existence of a degree of freedom in the optimal tax-subsidy structure. One instrument can be set arbitrarily and we shall take it to be *s*. In this case (8.24a)–(8.24c) imply the following first-best optimal tax rates:

$$\hat{\tau}_k = \frac{s-\alpha}{1-\alpha} \tag{8.25a}$$

$$\hat{\tau}_w = \frac{s\tilde{l} - \eta(1-\tilde{l})}{\tilde{l} - \eta(1-\tilde{l})} = -\hat{\tau}_c \tag{8.25b}$$

from which we may conclude the following relative magnitudes, in an economy with positive growth $(1 > (\alpha/\eta)[\tilde{l}/(1-\tilde{l})]$; see (8.17b′)):

$$0 < s < \alpha: \qquad \tau_w = -\tau_c < \tau_k < 0 < s$$

$$\alpha < s < \eta\left(\frac{1-\tilde{l}}{\tilde{l}}\right): \qquad \tau_w = -\tau_c < 0 < \tau_k < s$$

$$\eta\left(\frac{1-\tilde{l}}{\tilde{l}}\right) < s < 1: \qquad 0 < \tau_w = -\tau_c < \tau_k < s$$

There are two things to note about the optimal tax structure. First, is that for a sufficiently small investment subsidy, the optimal tax structure will call for subsidies to both wage income and capital income, all financed by the consumption tax. But as the subsidy increases, both forms of income should be taxed, with the revenues financing both the initial investment subsidy and now a subsidy to consumption. This pattern will be seen to be borne out by our simulations. Second $\hat{\tau}_k, \hat{\tau}_w$ are both highly sensitive to the (arbitrary) choice of *s*.

What is the impact of the first-best taxation system on distribution? Recall that the dispersion of gross income is given by (8.17c), where $\rho(l)$ is a decreasing function of leisure time. Since the policy increases the time allocated to labour, it will increase gross income inequality. The dispersion of net income in the decentralized economy that mimics the centrally planned equilibrium is obtained by substituting the tax rates, (8.25a), (8.25b), into (8.20b) to yield

$$\rho^{NET}(l;\tau_w,\tau_k) = 1 - \frac{\alpha}{(1+\eta)(1-l)} \frac{l}{(1+\alpha)l - \eta(1-l)} \tag{8.26}$$

The striking aspect about (8.26) is that the distribution of net income is *independent* of the (arbitrary) choice of fiscal instruments employed to achieve this objective. As long as the equilibrium is one with positive growth, the optimal tax requires $\hat{\tau}_w < \hat{\tau}_k$. Then $\rho^{NET} < \rho,$ and net income is less dispersed than is gross income. In addition, in all of our simulations we find that the direct redistributive effect of taxation dominates the indirect effect of changes in factor prices so that the distribution of net income is less unequal than in the economy without taxes.

When the first-best tax system is implemented, the effect of risk on growth and leisure is equivalent to that in the laissez-faire economy, as can be easily verified from equations (8.17). Greater risk is therefore associated with a greater supply of labour and hence with more pre-tax inequality. The effect of risk on after-tax inequality is, however, ambiguous. Differentiating (8.26) with respect to l, we can see that there are two opposing effects. On the one hand, more leisure tends to reduce pre-tax inequality. On the other, and as long as the subsidy rate is less than 1, a lower labour supply implies that a higher wage tax is required in order to finance any given subsidy rate (see (8.25b)), making the fiscal system less progressive. Either effect can dominate, implying that greater risk need not result in a more unequal distribution of after-tax income.

8.5.3. Alternative Policy Responses

To attain the first-best equilibrium is likely to require the tax rates to assume extreme values, even for plausible parameter values (see Table 8.3 below). These will generate dramatic changes in the distribution of income that may render them politically infeasible. Indeed, our numerical analysis (see Table 8.3 below) implies differences between the gross and the net Gini coefficients of 14 to 20 Gini points, whereas actual differences in OECD countries range between 1.5 and 4 points. Thus, we now consider some less drastic policy responses, which nevertheless, as our simulations show, may still yield substantial welfare gains.

8.5.3.1. Subsidy to investment financed by a tax on capital income

Suppose that the fiscal authority decides to finance the subsidy to investment with a tax on capital income, alone. Setting $\tau_w = \tau_c = 0$ in the government budget constraint (8.24c) the required tax on capital income is:

$$\tau_k = \frac{s}{(1-\alpha)}\left(1-\frac{\alpha}{\eta}\frac{l}{1-l}\right) \tag{8.27}$$

From equations (8.17a) we see that this policy shifts the RR schedule upwards and leaves the PP schedule unchanged, increasing the growth rate and reducing leisure. The reduction in leisure increases the pre-tax degree of income inequality, $\rho(l)$. Recall that the net distribution of income was characterized by (8.18′). Then, taxing capital income ensures that $\rho^{NET}(l,\tau_w\tau_k) < \rho(l)$. If the redistributive effect dominates, as our simulations suggest, the after-tax inequality actually declines.

8.5.3.2. Subsidy to investment financed by a tax on wage income

Alternatively, the subsidy may be fully financed by a wage tax

$$\tau_w = \frac{s}{\alpha}\frac{\eta(1-l)-\alpha l}{\eta(1-l)-sl} \tag{8.28}$$

In this case, both the RR and PP schedules shift up, resulting in a higher growth rate and greater or lower leisure, depending on the relative shifts. The reason for the ambiguous effect on leisure is that the wage tax tends to reduce the supply of labour, while the higher growth rate tends to increase it.

The ambiguous response of labour complicates the impact on the inequality of income. First, the increase–decrease in leisure time will reduce/increase the variance of gross incomes, as seen from (8.18′). However, the required (positive) wage tax implies taxing the factor that is more equally distributed, and for any given distribution of gross incomes this raises the variability of net incomes (see (8.20b) above). If the policy reduces leisure time, it would then unambiguously increase pre-tax and post-tax income inequality. However, when leisure time increases, the two effects work in opposite directions: there will be a reduction in the variability of gross income, while net income inequality may increase or decrease as compared to the equilibrium without taxes.[16]

8.5.3.3. Subsidy to investment financed by a tax on consumption

As a third example, the subsidy may also be financed by setting the consumption tax equal to

$$\tau_c = \frac{s\left[\eta(1-l)-\alpha l\right]}{\alpha l - s\eta(1-l)} \tag{8.29}$$

in which case $\rho^{NET}(l,\tau_w,\tau_k) = \rho(l)$. Again both schedules shift upwards, increasing the growth rate. In this case it can be shown that under the weak condition $\gamma < 0$, leisure increases, so that gross income inequality declines. Since the consumption tax has no direct redistributive effect, the gross and the net distributions of income are identical and hence net income inequality declines as well.

8.5.4. Numerical Analysis

Tables 8.3–8.5 report the numerical effects of a number of different policies, using the parameter values described in Section 8.4.2.

We begin with Table 8.3, which summarizes the first-best equilibrium in the centrally planned economy. It offers a number of insights that both reinforce and complement our analytical results. First, we see that the policy involves a substantial reduction in leisure time (between 10 and 12 percentage points), raising the growth rate enormously (by a factor of 3!), and only slightly increasing volatility.[17] ΔX is the increase in the welfare of the average individual, measured as the percentage variation over that in an economy with the same level of risk in the benchmark equilibrium (that is those in Table 8.2). First-best taxation increases the welfare of the average individual in the economy by over 20 per cent.

The effects on income distribution are substantial. The large reduction in leisure time results in a large increase in pre-tax inequality. However, the redistributive effect is strong enough to offset this effect and yield an overall reduction in the Gini coefficient of net income. The reduction in post-tax inequality relative to the economy without subsidies is large, amounting to between 6 and 12 Gini points.

Table 8.3 also illustrates the analytical results that the first-best equilibrium can be replicated by a variety of tax-subsidy configurations, each of which leads to precisely the same post-tax distribution of income. The sensitivity of the tax regime to changes in the subsidy rate is also borne out. Consider for example the case of low risk, $\sigma = 0.05$. In the absence of a subsidy to investment, the first-best equilibrium can be sustained if income form capital and labour are subsidized at the rates of 150 per cent and nearly 600 per cent, respectively, while consumption is taxed at nearly 600 per cent! This is hardly a politically viable tax structure. But the first-best equilibrium can also be attained if, more reasonably, investment is subsidized at around 86 per cent, being financed by a tax on capital income of around 64 per cent, leaving consumption and labour income untaxed. Or, if investment and consumption are subsidized at 90 per cent and 30 per cent respectively, with taxes on labour income and capital income of 30 per cent and 75 per cent, respectively.

Table 8.3 also considers the effects of increasing risk. This is shown to reduce leisure, thereby increasing the gross income inequality. At the same time, the decrease in leisure increases the growth rate and reduces the

Table 8.3 First-best taxation

	s	τ_w $(=-\tau_c)$	τ_k	l	ψ	σ_ψ	Gini(y)	Gini(ny)	ΔX
	0	–597.8	–150.0						
	30.0	–388.5	–75.0						
$\sigma = 0.05$	60.0	–179.1	0	67.1	11.53	1.92	41.02	27.06	20.56
	85.67	0	64.17						
	90.0	30.22	75.0						
	0	–615.9	–150.0						
	30.0	–401.1	–75.0						
$\sigma = 0.1$	60.0	–186.4	0	67.0	11.66	3.85	41.13	26.94	20.63
	86.03	0	65.08						
	90.0	28.40	75.0						
	0	–701.7	–150.0						
	30.0	–461.2	–75.0						
$\sigma = 0.2$	60.0	–220.7	0	66.7	12.20	7.76	41.57	26.42	20.92
	87.53	0	68.82						
	90.0	19.80	75.0						
	0	–919.1	–150.0						
	30.0	–613.3	–75.0						
$\sigma = 0.3$	60.0	–307.6	0	66.0	13.13	11.78	42.30	25.39	21.45
	90.0	–1.91	75.0						
	90.19	0	75.47						
	0	–1663	–150.0						
	30.0	–1134	–75.0						
$\sigma = 0.4$	60.0	–605.1	0	65.0	14.54	15.99	43.36	23.44	22.33
	90.0	–76.30	75.0						
	94.33	0	85.82						

redistributive effect, thus reducing the inequality of net income. In all our examples, gross and net income inequality move in opposite ways, with greater volatility increasing pre-tax and reducing post-tax inequality. Our numerical results highlight the fact that the divergence between pre- and post-tax inequality is greater the more risky the economy is. The reason for this is that risk has a strong distortionary effect on the labour supply. Greater risk, by raising the labour supply and hence the wage bill, requires a lower wage tax, thus making the tax system more progressive. The effect of the increased labour supply is to raise pre-tax inequality, the impact of the lower wage tax is to reduce post-tax inequality, and as a result the gap between gross and net inequality increases.

Table 8.4 examines a number of alternative non-optimal policies. The first two lines are the benchmark case of no intervention, for a low-risk economy ($\sigma = 0.05$) and a medium/high risk economy ($\sigma = 0.20$), respectively. The welfare gains are measured as percentage increases over the welfare levels in these two base economies. We consider in turn the effects of financing a fixed 30 per cent investment subsidy through a capital income, wage tax or a consumption tax, respectively. Financing by a capital income tax generates the least positive impact on the mean growth rate and on welfare. It raises the pre-tax income inequality, but lowers the post-tax income inequality.

Employing a wage tax has a significantly larger effect on the mean growth rate and on welfare. It also has the opposite impacts on income distribution, reducing the before tax inequality, but raising it after tax. The consumption tax has the greatest benefit on the average agent and the most beneficial effect on the mean growth rate, and it increases the degree of income inequality (both pre- and post-tax) slightly. Interestingly, all three policies have virtually no adverse effect on aggregate volatility and in fact the wage tax, by increasing leisure, actually permits a substantial increase in the growth rate to be accompanied by a small reduction in its volatility.

The last two rows of the table consider financing the subsidy through a combination of wage and consumption taxes. In particular, we set $\tau_w = -\tau_c$; that is, these two taxes are optimally set, although the subsidy is below the

Table 8.4 Arbitrary taxation

		τ_k	τ_w	τ_c	l	ψ	σ_ψ	Gini (y)	Gini(ny)	ΔX
$s=0$	$\sigma = 0.05$	0	0	0	72.5	3.30	1.73	33.30	33.30	–
	$\sigma = 0.2$	0	0	0	72.4	3.79	6.96	33.82	33.82	–
$s=30$	$\sigma = 0.05$	10.07	0	0	71.6	4.73	1.76	34.80	32.83	7.64
	$\sigma = 0.2$	10.96	0	0	71.4	5.18	7.09	35.26	33.13	7.65
$s=30$	$\sigma = 0.05$	0	7.63	0	72.8	5.24	1.72	32.79	34.39	9.66
	$\sigma = 0.2$	0	8.28	0	72.6	5.71	6.89	33.09	34.79	10.18
$s=30$	$\sigma = 0.05$	0	0	5.42	72.3	5.32	1.74	33.75	33.75	10.05
	$\sigma = 0.2$	0	0	5.98	72.0	5.80	6.98	34.11	34.11	10.54
$s=30$	$\sigma = 0.05$	0	–18.76	18.76	71.2	5.47	1.78	35.55	32.44	10.71
	$\sigma = 0.2$	0	–21.63	21.63	70.8	6.00	7.16	36.08	32.60	11.13

first-best level. The effect of this policy on the growth rate is stronger than in the previous three cases, the reason being that this policy does not distort the allocation of time between labour and leisure. Employing only a wage or a consumption tax tends to reduce the supply of labour, partially offsetting the effect of the subsidy. When both are used, this effect is absent. Since setting $\tau_w = -\tau_c$ results in faster growth than using only one tax, and since the volatility of growth is only slightly higher, this policy generates larger welfare gains than any of the pure policies. The effect on distribution is quite significant. In contrast to financing the investment subsidy by either a wage or a consumption tax alone, financing through a combination of a consumption tax and wage subsidy reduces substantially post-tax inequality.

The investment subsidy in Table 8.4 is arbitrary. Table 8.5 summarizes a number of second-best policies, whereby the policy maker sets the optimal subsidy for each of the three modes of finance. In the case where it is financed with a tax on capital income, it is able to attain the first-best optimum. Focusing on $\sigma = 0.05$, setting s = 85.7 per cent and $\tau_k = 64.2$ per cent improves welfare by over 20 per cent and generates the distribution of income associated with the first-best optimum. Alternatively, setting s = 57.2 per cent, financed with 26 per cent tax on wages, or s = 60.2 per cent financed with a 25.5 per cent consumption tax yield second-best optima. The interesting aspect about these latter two alternatives is that they are fairly moderate policies, in contrast to the first-best, summarized in Table 8.3. In particular, the consumption tax yields the major portion of the welfare gains obtained in the first-best case (19 per cent out of a total increase in welfare of 20.6 per cent), while having only a minimally adverse impact on income distribution. Indeed, to a policymaker concerned with

Table 8.5 Second best

	s	τ_k	τ_w	τ_c	L	ψ	σ_ψ	Gini(y)	Gini(ny)	ΔX
$\sigma = 0.05$	0	0	0	0	72.5	3.30	1.73	33.30	33.30	–
$\sigma = 0.2$	0	0	0	0	72.4	3.79	6.96	33.82	33.82	–
$\sigma = 0.05$	85.7	64.17	0	0	67.1	11.53	1.92	41.02	27.06	20.56
$\sigma = 0.2$	85.5	68.82	0	0	66.7	12.20	7.76	41.57	26.42	20.92
$\sigma = 0.05$	57.2	0	25.96	0	74.1	9.07	1.67	30.20	36.56	16.79
$\sigma = 0.2$	57.4	0	27.39	0	74.1	9.54	6.66	30.20	36.97	17.72
$\sigma = 0.05$	60.2	0	0	25.52	72.0	10.40	1.75	34.19	34.19	18.87
$\sigma = 0.2$	60.1	0	0	27.02	71.9	10.87	7.01	34.39	34.39	19.72

maximizing the welfare of the average agent, with minimum distortionary effect on the distribution of income, this policy may be particularly attractive.

8.6. CONCLUSIONS

Stochastic shocks are a major source of income disparities, and an extensive literature has explored how 'luck' and the market's tendency towards convergence combine to create persistent inequality Yet this literature is concerned with idiosyncratic shocks that have no relation with aggregate shocks. The idea that aggregate uncertainty may also affect the distribution of income remains to be explored, and this chapter is a first step in that direction.

We have used an AK stochastic growth model to show that, when agents differ in their initial stocks of capital, greater growth volatility is associated with a more unequal distribution of income. Greater risk tends to increase the supply of labour, reducing wages and raising the interest rate. If capital endowments are unequally distributed, while labour endowments are not, the change in factor prices raises the return to the factor that is the source of inequality, and the distribution of income becomes more spread.

The endogeneity of the labour supply also implies that policies aimed at increasing the growth rate will have distributional implications, and we have examined how these differ depending on the particular form of the policies. In particular, we have compared financing an investment subsidy through a capital income tax, a wage tax, or a consumption tax.

Our analysis yields two main conclusions. First, it is possible simultaneously to increase the growth rate and reduce net income inequality. In many instances, we find that polices that generate faster growth are associated with a reduction in the Gini coefficient of post-tax income, allowing the policymaker to attain both efficiency and equity goals. Second, it is often the case that fiscal policy has opposite effects on the distribution of gross and net income. These results highlight the fact that rather than the usual trade-off between equity and efficiency, policymakers concerned with the distribution of income may face a trade-off between pre- and post-tax inequality. Moreover, the divergence between pre- and post-tax inequality is greater the more risky the economy is. Understanding which type of inequalities agents and the social planner care about becomes essential, in particular in high risk economies, and it raises the question of whether a slightly more unequal distribution of both gross and net incomes may, in certain cases, be a more viable policy than a huge, but offset, increase in pre-tax inequality.

We conclude by relating the implications of this model to a diverse range of relevant empirical evidence pertaining to the relationships between growth, volatility and income inequality. First, our model implies a *positive* equilibrium relationship between *volatility* and *income inequality*, consistent with the empirical evidence provided by Breen and García-Peñalosa (2005). Second, the increase in production risk also generates a *positive* relationship between *income inequality* and *growth*. Evidence on this relationship is mixed and indeed, given that both variables are endogenous, it would seem plausible to argue that their relationship should depend upon the source of the underlying change. From this viewpoint, our analysis suggests that macroeconomic volatility may provide one explanation of the positive relationship obtained in recent studies by Li and Zou (1998) and Forbes (2000).

There are also several studies relating *volatility* to both the *consumption growth rate* and the *output growth rate*, respectively. In this regard our model implies that an increase in macroeconomic volatility increases precautionary savings, that is reduces the level of consumption, but increases the consumption growth rate. This conclusion is consistent with Blanchard and Mankiw (1988) and empirical evidence discussed by Zeldes (1989), which argues that there have been long periods of time in the US when the consumption growth rate has been positive, even though the riskless rate of return was less than the rate of time preference and consumption would otherwise be falling in a riskless world.

On the other hand, apart from an early study by Kormendi and Meguire (1985), most empirical studies obtain a *negative* relationship between *volatility* and the *growth rate of output*; see Ramey and Ramey (1995), and more recently Barlevy (2004). Our equilibrium, which implies a *positive* relationship between volatility and the growth rate of capital and therefore output, contradicts this relationship. Moreover, it is impossible to reconcile a positive precautionary savings effect with a negative volatility–growth relationship in this full employment equilibrium framework. To the extent that production risk reduces consumption and increases saving, it increases investment, and the growth rate of output.

Finally, our conclusions are based on the assumption $\gamma < 0$, which we have interpreted as implying a coefficient of relative risk aversion, $R \equiv 1-\gamma$ say, in excess of unity, an assumption strongly supported by empirical evidence. But by employing the constant elasticity utility function we are imposing the restriction that the intertemporal elasticity of substitution, ε say, be inversely related to R, by $\varepsilon = 1/R$. In general ε and R are independent parameters, and if one adopts more general recursive preferences that allow them to be independently set, then given $R > 0$, ,the output growth–volatility relationship will remain positive if and only if

$\varepsilon < 1$ (see for example, Giuliano and Turnovsky, 2003). The empirical evidence on ε is less uniform. Early data, based on consumption behaviour, suggested values of ε around 0.3, well below unity, again implying a positive output–growth volatility relationship. But more recent studies, based on stock market data, suggest larger values of ε, closer to unity and even greater than unity in some instances (see Guvenen, 2003). In the latter case, we will obtain a *negative output growth–volatility* relationship, consistent with the Ramey–Ramey results. However, by the same token, there will now be negative precautionary savings, thus contradicting the Blanchard–Mankiw and Zeldes findings. To obtain both positive precautionary savings and a negative output growth–volatility relationship will require a more fundamental modification of the model, and is a direction in which the analysis may be usefully extended.

APPENDIX

This appendix provides some of the technical details underlying the derivations of the equilibrium conditions (8.9) and (8.16a) to (8.16g).

8.A.1 Consumer Optimization

Agent *i*'s stochastic maximization problem is to choose her individual consumption–capital ratio and the fraction of time devoted to leisure to maximize expected lifetime utility

$$\max E_0 \int_0^\infty \frac{1}{\gamma}\left[C_i(t) l_i^{\eta}\right]^{\gamma} e^{-\beta t} dt, \qquad -\infty < \gamma < 1, \eta > 0, \gamma\eta < 1 \quad \text{(8.A.1a)}$$

subject to her individual capital accumulation constraint

$$dK_i = \frac{(1-\tau_k) r K_i + (1-\tau_w) w (1-l_i) K - (1+\tau_c) C}{1-s} dt + K_i dk \quad \text{(8.A.1b)}$$

and the aggregate capital accumulation constraint

$$dK = \frac{\left[(1-\tau_k) r + (1-\tau_w) w (1-l)\right] K - (1+\tau_c) C}{1-s} dt + K dk \quad \text{(8.A.1b}')$$

together with the economy-wide shock

$$dk = \frac{1-\tau_k'}{1-s'}\Omega du \equiv \frac{1-\tau_k'}{1-s'} du_K \,. \tag{8.A.1c}$$

Since the agent perceives two state variables, K_i, K, we consider a value function of the form

$$V\left(K_i, K, t\right) = e^{-\beta t} X(K_i, K)$$

the differential generator of which is

$$\begin{aligned}\Psi\left[V(K_i,K,t)\right] \equiv \frac{\partial V}{\partial t} &+ \left[\left(\frac{1-\tau_k}{1-s} r - \frac{1+\tau_c}{1-s}\frac{C_i}{K_i}\right)K_i + w(1-l)\frac{1-\tau_w}{1-s}K\right]V_{K_i} \\ &+ \left[\frac{1-\tau_k}{1-s} r - \frac{1+\tau_c}{1-s}\frac{C}{K} + w(1-l)\frac{1-\tau_w}{1-s}\right]KV_K \\ &+ \frac{1}{2}\sigma_K^2 K_i^2 V_{K_iK_i} + \sigma_{K_iK} K_i K V_{K_iK} + \frac{1}{2}\sigma_K^2 K^2 V_{KK}\end{aligned} \tag{8.A.2}$$

where $\operatorname{cov}(dK_i, dK) = \sigma_{K_i,K} dt$, and so on.

The individual's problem is to choose consumption, leisure, and the rate of capital accumulation to maximize the Lagrangian

$$e^{-\beta t}\frac{1}{\gamma}\left(C_i l_i^{\eta}\right)^{\gamma} + \Psi\left[e^{-\beta t} X(K_i, K)\right]. \tag{8.A.3}$$

In doing this, she takes the evolution of the aggregate variables and the externality as given. Taking the partial derivatives with respect to C_i and l_i, and cancelling $e^{-\beta t}$ yields

$$\frac{1}{C_i}\left(C_i l_i^{\eta}\right)^{\gamma} = \frac{1+\tau_c}{1-s} X_{K_i} \tag{8.A.4a}$$

$$\frac{\eta}{l}\left(C_i l_i^{\eta}\right)^{\gamma} = \frac{1-\tau_w}{1-s} wKX_{K_i} \tag{8.A.4b}$$

In addition, the value function must satisfy the Bellman equation

$$\max\left\{e^{-\beta t}\frac{1}{\gamma}\left(C_i l_i^{\eta}\right)^{\gamma} + \Psi\left[e^{-\beta t} X(K_i, K)\right]\right\} = 0 \tag{8.A.5}$$

The Bellman equation is a function of two state variables, individual and aggregate capital, and hence it is a partial differential equation in these two variables. Using equations (8.A.1b) and (8.A.1b′), and given (8.A.2), the Bellman equation can be written as

$$\frac{1}{\gamma}\left(C_i l_i^{\eta}\right)^{\gamma} - \beta X(K_i, K) + \frac{E(dK_i)}{dt} X_{K_i} + \frac{E(dK)}{dt} X_K$$
$$+\frac{1}{2}\frac{E(dK_i)^2}{dt} X_{K_iK_i} + \frac{E(dK_i dK)}{dt} X_{K_iK} + \frac{1}{2}\frac{E(dK)^2}{dt} X_{KK} = 0 \tag{8.A.6}$$

Next we take the partial derivative of the Bellman equation with respect to K_i, noting that l_i is independent of K_i, while C_i is a function of K_i through the first-order condition (8.A.4a),

$$\frac{1}{C_i}\left(C_i l_i^{\eta}\right)^{\gamma} C_{i,K_i} - \beta X_{K_i} + \frac{E(dK_i)}{dt} X_{K_iK_i} + \left(r\frac{1-\tau_k}{1-s} - C_{i,K_i}\right) X_{K_i}$$
$$+\frac{E(dK)}{dt} X_{K_iK} + K_i X_{K_iK_i}\sigma_K^2 + \frac{1}{2}\frac{E(dK_i)^2}{dt} X_{K_iK_iK_i} \tag{8.A.7}$$
$$+\frac{E(dK_i dK)}{dt} X_{K_iKK_i} + K X_{K_iK}\sigma_{K_iK} + \frac{1}{2}\frac{E(dK)^2}{dt} X_{K_iKK} = 0$$

Consider now $X_{K_i} = X_{K_i}(K_i, K)$. Taking the stochastic differential of this quantity yields:

$$dX_{K_i} = X_{K_iK_i} dK_i + X_{K_iK} dK + \frac{1}{2} X_{K_iK_iK_i} (dK_i)^2$$
$$+ X_{K_iKK_i}(dK_i)(dK) + \frac{1}{2} X_{K_iKK}(dK)^2 \tag{8.A.8}$$

Taking expected values of this expression, dividing by *dt,* and substituting the resulting equation along with (8.A.4a) into (8.A.7) leads to:

$$\left(r\frac{1-\tau_k}{1-s} - \beta\right) X_{K_i} + \left(K_i X_{K_iK_i} + K X_{K_iK}\right)\sigma_K^2 + \frac{E(dX_{K_i})}{dt} = 0\,, \tag{8.A.9}$$

The solution to this equation is by trial and error. Given the form of the objective function, we propose a value function of the form:

$$X(K_i, K) = cK_i^{\gamma-\gamma_2} K^{\gamma_2} \tag{8.A.10}$$

where the parameters c, γ_2 are to be determined. From (8.A.10) we obtain:

$$\begin{aligned} &X_{K_i} = (\gamma-\gamma_2)X/K_i; X_K = \gamma_2 X/K; \\ &X_{K_iK_i} = (\gamma-\gamma_2)(\gamma-\gamma_2-1)X/K_i^2; \\ &X_{K_iK} = (\gamma-\gamma_2)\gamma_2 X/K_iK; X_{KK} = (\gamma_2-1)\gamma_2 X/K^2. \end{aligned} \tag{8.A.11}$$

We can now use equation (8.A.11) to re-express (8.A.9) as

$$\frac{E(dX_{K_i})}{X_{K_i}dt} = \beta - r\frac{1-\tau_k}{1-s} - (\gamma-1)\sigma_K^2 \tag{8.A.12}$$

Now, returning to the first-order condition (8.A.4a), computing the stochastic differential of this relationship and taking expected values yields

$$\frac{E\left(dX_{K_i}\right)}{X_{K_i}} = (\gamma-1)\frac{E(dC_i)}{C_i} + \frac{1}{2}(\gamma-1)(\gamma-2)E\left(\frac{dC_i}{C_i}\right)^2 \tag{8.A.13}$$

Along the balanced growth path, C_i/K_i is constant. Hence $dC_i/C_i = dK_i/K_i = \psi dt + dw,$ and thus

$$\frac{E\left(dX_{K_i}\right)}{X_{K_i}dt} = (\gamma-1)\psi + \frac{1}{2}(\gamma-1)(\gamma-2)\sigma_K^2 \tag{8.A.14}$$

As shown in equation (8.7a) of the text, the government's balanced budget implies that the stochastic component of the individual budget constraint is

$$dk = \Omega du. \tag{8.A.15}$$

Combining (8.A.13), (8.A.14) and (8.A.15) yields the mean growth rate of individual consumption

$$\psi = \frac{r(1-\tau_k)-\beta(1-s)}{(1-\gamma)(1-s)} - \frac{\gamma}{2}\Omega^2\sigma^2. \tag{8.A.16}$$

The labour supply is obtained from the first-order conditions (8.A.4a) and (8.A.4b), namely

$$C_i = \frac{w}{\eta}\frac{1-\tau_w}{1+\tau_c}Kl_i\,. \tag{8.A.17}$$

Dividing (8.A.17) by K_i we obtain the individual consumption to wealth ratio,

$$\frac{C_i}{K_i} = \frac{w}{\eta}\frac{1-\tau_w}{1+\tau_c}\frac{l_i}{k_i}\,, \tag{8.A.18}$$

and summing over all agents we have the aggregate consumption to wealth ratio,

$$\frac{C}{K} = \frac{w}{\eta}\frac{1-\tau_w}{1+\tau_c}l\,. \tag{8.A.19}$$

From the individual budget constraint, the rate of growth is

$$\psi = r\frac{1-\tau_k}{1-s} + w\frac{1-\tau_w}{1-s}\frac{1-l_i}{k_i} - \frac{1+\tau_c}{1-s}\frac{C_i}{K_i}\,, \tag{8.A.20}$$

which using (A.18) and rearranging gives

$$1-l_i = \frac{1}{1+\eta} - \frac{\eta}{1+\eta}\frac{r(1-\tau_k)-\psi(1-s)}{w}\frac{1}{1-\tau_w}k_i. \tag{8.A.21}$$

8.A.2 Macroeconomic Equilibrium

Note that the growth rate is the same for all agents, irrespective of their initial wealth holdings. Equation (8.A.16) is hence the mean growth rate of the economy. The dynamic evolution of the aggregate stock of capital is given by

$$\frac{dK}{K} = \left[\frac{r(1-\tau_k)-\beta(1-s)}{(1-\gamma)(1-s)} - \frac{\gamma}{2}\Omega^2\sigma^2\right]dt + \Omega du$$

and the standard deviation (volatility) of the growth rate is

$$\sigma_\psi = \Omega\sigma\,. \tag{8.A.22}$$

Summing (8.A.21) over all agents gives a relationship between the aggregate labour supply and the growth rate,

$$1-l=\frac{1}{1+\eta}-\frac{\eta}{1+\eta}\frac{r(1-\tau_k)-\psi(1-s)}{w}\frac{1}{1-\tau_w} \tag{8.A.23}$$

Goods market equilibrium requires $dK=\Omega K(dt+du)-Cdt,$ which, taking expectations and dividing by $K,$ yields,

$$\psi=\Omega-\frac{C}{K} \tag{8.A.24}$$

Equations (8.A.16), (8.A.22), (8.A.18), (8.A.19), (8.A.20), (8.A.24) and (8.7b) are the macroeconomic equilibrium conditions as specified in equations (8.16a)–(8.16g), respectively.

In the absence of taxation, the equilibrium reduces to

$$\frac{C_i}{K_i}=\frac{w}{\eta}\frac{l_i}{k_i}, \tag{8.A.25a}$$

$$\psi=\frac{r-\beta}{1-\gamma}-\frac{\gamma}{2}\Omega^2\sigma^2, \tag{8.A.25b}$$

$$\psi=r+w(1-l)-\frac{C_i}{K_i}, \tag{8.A.25c}$$

$$\frac{C}{K}=\frac{wl}{\eta}, \tag{8.A.25d}$$

$$\sigma_\psi=\Omega\sigma, \tag{8.A.25e}$$

which jointly determine the consumption–capital ratio, the average growth rate, the labour supply, and the volatility of growth.

8.A.3. Existence of a Balanced Growth Equilibrium

It suffices to focus on the economy without taxation; the introduction of taxes leads to minor modifications and can be analysed analogously. Differentiating the relations in (8.A.25) we obtain

$$\left.\frac{\partial\psi}{\partial l}\right|_{RR}=-\frac{\alpha\Omega(l)}{1-l}\left[\frac{1-\alpha}{1-\gamma}-\gamma\Omega(l)\sigma^2\right]<0, \tag{8.A.26a}$$

$$\left.\frac{\partial\psi}{\partial l}\right|_{PP}=-\frac{\alpha\Omega(l)}{\eta(1-l)}\left[1+\eta+(1-\alpha)\frac{l}{1-l}\right]<0, \tag{8.A.26b}$$

so that both schedules have a negative slope. Using the fact that $\Omega = A(1-l)^{\alpha}$, and under the assumption that $\alpha < 1/2$, both the (PP) and (RR) schedules can be shown to be strictly concave (see Turnovsky, 2000b, for more details). Also

$$\psi_{RR}(l=0) = \frac{A(1-\alpha)-\beta}{1-\gamma} - \frac{\gamma}{2}A^2\sigma^2, \qquad \psi_{RR}(l=1) = -\frac{\beta}{1-\gamma}$$

$$\psi_{PP}(l=0) = A, \qquad \psi_{PP}(l=1) = -\infty.$$

A necessary and sufficient condition for the existence of a unique equilibrium is $\psi_{PP}(l=0) > \psi_{RR}(l=0)$. In this case, the (PP) schedule is below (RR) for $l=1$, and the two schedules cross only once. This condition is satisfied when

$$\alpha - \gamma + \frac{\beta}{A} > -\gamma\frac{1-\gamma}{2}A\sigma^2, \tag{8.A.27}$$

that is when risk is not excessively high. When equation (8.A.27) is not satisfied either an equilibrium does not exists or there are two.

Note also that the PP schedule is steeper than RR if and only if

$$\frac{1+\eta}{\eta} + \frac{1-\alpha}{\eta}\frac{l}{1-l} > \frac{1-\alpha}{1-\gamma} - \gamma\Omega\sigma^2. \tag{8.A.28}$$

Since at $l=0$, $l/(1-l)$ has its lowest and $\Omega(l)$ its highest possible value, PP is everywhere steeper than RR if and only if

$$\frac{1+\eta}{\eta} - \frac{1-\alpha}{1-\gamma} > -\gamma A\sigma^2. \tag{8.A.29}$$

8.A.4. The Centrally Planned Economy

The social planner's problem (8.23), leads to the following Bellman equation

$$\frac{1}{\gamma}\left(C_i l_i^{\eta}\right)^{\gamma} - \beta X(K_i, K) + \frac{E(\Omega K_i - C_i)}{dt}X_{K_i} + \frac{1}{2}\frac{E(dK_i)^2}{dt}X_{K_iK_i} = 0. \tag{8.A.6'}$$

Taking the partial derivative of this equation with respect to K_i then yields

$$\frac{1}{C_i}\left(C_i l_i^{\eta}\right)^{\gamma} C_{i,K_i} - \beta X_{K_i} + \frac{E(dK_i)}{dt} X_{K_i K_i} + (\Omega - C_{i,K_i}) X_{K_i} + \tag{8.A.7'}$$
$$+ K_i X_{K_i K_i} \sigma_K^2 + \frac{1}{2}\frac{E(dK_i)^2}{dt} X_{K_i K_i K_i} = 0$$

and hence

$$\frac{E(dX_{K_i})}{X_{K_i} dt} = \beta - \Omega - (\gamma - 1)\sigma_K^2, \tag{8.A.12'}$$

which together with (8.A.14) and (8.A.15) above yield (8.17a′).

The first order conditions with respect to consumption and leisure together imply

$$\frac{C_i}{K_i} = \frac{wl_i}{\eta}, \tag{8.A.19'}$$

Goods market equilibrium is again given by equation (8.A.24). Using (8.A.19′), the equilibrium conditions can be expressed as

$$\psi = \frac{\Omega(l) - \beta}{1-\gamma} - \frac{\gamma}{2}\Omega(l)^2\sigma^2, \tag{8.A.30}$$

$$\psi = \Omega(l) - \frac{C}{K} = \Omega(l)\left[1 - \frac{\alpha}{\eta}\frac{l}{1-l}\right]. \tag{8.A.31}$$

(8.A.36) is strictly decreasing and concave in l. Differentiating (8.A.30), we obtain

$$\left.\frac{\partial \psi}{\partial l}\right|_{R'R'} = -\frac{\Omega(l)}{1-l}\left[\frac{1}{1-\gamma} - \gamma\Omega(l)\sigma^2\right] < 0,$$

$$\left.\frac{\partial^2 \psi}{\partial l^2}\right|_{RR''} = -\frac{\alpha\Omega(l)}{(1-l)^2}\left[\frac{1-\alpha}{1-\gamma} - \gamma\alpha\Omega(l)(1-2\alpha)\sigma^2\right].$$

(8.A.30) is thus decreasing in l and a sufficient condition for concavity is $\alpha < 1/2$. The necessary and sufficient condition for the existence of a unique equilibrium, $\psi_{PP}(l=0) > \psi_{RR''}(l=0)$, is now

$$-\gamma+\frac{\beta}{A}>-\gamma\frac{1-\gamma}{2}A\sigma^2. \qquad (8.A.27'')$$

Note also that (8.A.31) schedule is steeper than (8.A.30) if and only if

$$\frac{1+\eta}{\eta}-\frac{1}{1-\gamma}>-\gamma A\sigma^2. \qquad (8.A.29')$$

NOTES

* We are grateful to our discussant at the Conference, Thomas Renstrom, and two anonymous referees for constructive comments. We have also benefited from comments received from seminar presentations at Nuffield College, Oxford, the University of Munich, the 2003 meetings of the EEA in Stockholm, and the 2003 meetings of the IAES in Quebec City.

1. See Breen and García-Peñalosa (2005) for a description of the data sources used for these calculations.
2. See, among others, Aghion and Bolton (1997) and Galor and Tsiddon (1997), as well as the overview in Aghion, Caroli and García-Peñalosa (1999).
3. A more general study of heterogeneity and the dynamics of distribution in growth models can be found in Caselli and Ventura (2000).
4. In the United States, for example, the relative variability of stock returns over the period 1955–95 were around 32 per cent per annum, while the relative variability of wages over that same period was only 2 per cent.
5. Intuitively, in a growing economy, with the labour supply fixed, the higher income earned by labour is reflected in higher returns, whereas with capital growing at the same rate as output, returns to capital remain constant.
6. Some of the empirical estimates supporting this assumption are noted in Section 8.4.2.
7. Thus we may write $\Omega(l)=A(1-l)^{\alpha}$ and $r=(1-\alpha)\Omega(l)$, where $\Omega'(l)<0$.
8. This latter condition reduces to $C/K>0$ in the original Merton (1969) model, which abstracted from labour income.
9. Writing $\rho(l)=[1/(1+\eta)(1-l)]\{[\eta(1-l)-\alpha l]+(1-\alpha)(1-l)\}$. If the equilibrium is one of positive growth, (8.17b) implies that the first term in brackets is positive, thus ensuring that $\rho(l)>0$. The fact that $\rho(l)<1$ is immediate from its definition.
10. Her results also require that the economy be amenable to Gorman aggregation, which is the case in our setup.
11. The mean growth rate for OECD economies is around 2.2 per cent, with a standard deviation also around 2.2 per cent.
12. Most notably it is characteristic of Lucas's (1987) model of the cost of business cycles, and it is also discussed at more length for a model closer to the present by Turnovsky (2000b).
13. The transversality condition (8.12) for the central planner's problem again reduces to (8.13) but is now automatically satisfied without further restrictions being imposed.
14. The reason why the social planner chooses less leisure is that there are in fact two externalities in the model. On the one hand, a greater individual stock of capital increases the aggregate level of technology. On the other, a higher labour supply raises the marginal product of capital and induces greater accumulation of capital, thus increasing the level of technology.

15. The optimal tax rates set out in (8.24) are similar to those obtained by Turnovsky (2000a) in a pure deterministic representative agent endogenous growth model.
16. We can, however, see that when the subsidy rate matches the externality, $s=\alpha,\ \tau_w=1$ and $\rho^{NET}=1$ implying that the net income inequality is increased to that of the initial endowment of capital.
17. The implied percentage increase in labour supply is much larger, being of the order of 20 per cent.

REFERENCES

Aghion, P. and P. Bolton (1997), 'A trickle-down theory of growth and development with debt overhang', *Review of Economic Studies*, **64**:151–72.

Aghion, P., E. Caroli and C. García-Peñalosa (1999), 'Inequality and economic growth: the perspective of the new growth theories', *Journal of Economic Literature*, **37**:1615–60.

Aghion, P., A. Banerjee and T. Piketty (1999), 'Dualism and macroeconomic volatility', *Quarterly Journal of Economics,* **114**:1359–7.

Alesina A. and Rodrik (1994), 'Distributive politics and economic growth', *Quarterly Journal of Economics*, **109**:465–90.

Barlevy, G. (2004), 'The cost of business cycles under endogenous growth', *American Economic Review*, **94**:964–90.

Bertola, G. (1993), 'Factor shares and savings in endogenous growth', *American Economic Review*, **83**:1184–98.

Blanchard, O.J. and N.G. Mankiw (1988), 'Consumption: beyond certainty equivalence', *American Economic Review, Papers and Proceedings*, **78**:173–7.

Breen, R. and C. García-Peñalosa (2005), 'Income inequality and macroeconomic volatility: an empirical investigation', *Review of Development Economics*, **9**:380–98.

Caselli, F. and J. Ventura (2000), 'A representative consumer theory of distribution', *American Economic Review*, **90**:909–26.

Constantinides, G.M., J.B. Donaldson and R. Mehra (2002), 'Junior can't borrow: a new perspective on the equity premium puzzle', *Quarterly Journal of Economics*, **117**:269–96.

Correia, I.H. (1999), 'On the efficiency and equity trade-off', *Journal of Monetary Economics*, **44**:581–603.

Corsetti, G. (1997), 'A portfolio approach to endogenous growth: equilibrium and optimal policy', *Journal of Economic Dynamics and Control*, **21**:1627–44.

Deininger, K. and L. Squire (1996), 'Measuring income inequality: a new data base', *The World Bank Economic Review*, **10**:565–91.

Forbes, K. (2000), 'A reassessment of the relationship between inequality and growth', *American Economic Review*, **90**:869–87.

Gali, J. (1994), 'Government size and macroeconomic stability', *European Economic Review*, **38**:117–32.

Galor, O. and D. Tsiddon (1997), 'Technological progress, mobility, and economic growth', *American Economic Review*, **87**:363–82.

García-Peñalosa, C. and S.J. Turnovsky (2005), 'Production risk and the functional distribution of income in a developing economy: tradeoffs and policy responses', *Journal of Development Economics*, **76**:175–208.

Gavin, M. and J. Hausmann (1995), *Overcoming Volatility in Latin America*, Washington: Inter-American Development Bank.

Giuliano, P. and S.J. Turnovsky (2003), 'Intertemporal substitution, risk aversion, and economic performance in a stochastically growing open economy', *Journal of International Money and Finance*, **23**:529–56.

Grinols, E.L. and S.J. Turnovsky (1993), 'Risk, the financial market, and macroeconomic equilibrium', *Journal of Economic Dynamics and Control*, **17**:1–36.

Grinols, E.L. and S.J. Turnovsky (1998), 'Risk, optimal government finance, and monetary policies in a growing economy', *Economica*, **65**:401–27.

Guvenen, F. (2003), 'Reconciling conflicting evidence on the elasticity of intertemporal substitution: a macroeconomic perspective', mimeo, University of Rochester.

Kormendi, R. and P. Meguire (1985), 'Macroeconomic determinants of growth: cross-country evidence', *Journal of Monetary Economics,* **16**:141–63.

Li, H.Y. and H.F. Zou (1998), 'Income inequality is not harmful to growth: theory and evidence', *Review of Development Economics*, **2**:318–34.

Lucas, R. (1987), *Models of Business Cycles*, Oxford: Basil Blackwell.

Merton, R.C. (1969), 'Lifetime portfolio selection under uncertainty: the continuous-time case', *Review of Economics and Statistics*, **51**:247–57.

Obstfeld, M. (1994), 'Risk-taking, global diversification, and growth', *American Economic Review*, **84**:1310–29.

Persson, T. and G. Tabellini (1994), 'Is inequality harmful for growth?' *American Economic Review*, **84**:600–621.

Ramey, G. and V. Ramey (1995), 'Cross-country evidence on the link between volatility and growth', *American Economic Review*, **85**:1138–51.

Romer, P.M. (1986), 'Increasing returns and long-run growth', *Journal of Political Economy*, **94**:1002–37.

Smith, W.T. (1996), 'Taxes, uncertainty, and long-term growth', *European Economic Review*, **40**:1647–64.

Turnovsky, S.J. (2000a), 'Fiscal policy, elastic labor supply, and endogenous growth', *Journal of Monetary Economics*, **45**:185–210.

Turnovsky, S.J. (2000b), 'Government policy in a stochastic growth model with elastic labor supply', *Journal of Public Economic Theory*, **2**:389–433.

Wolff, E.N. (1998), 'Recent trends in the size distribution of household wealt', *Journal of Economic Perpectives*, **12**:131–50.

Zeldes, S. (1989), 'Optimal consumption with stochastic income: deviations from certainty equivalence', *Quarterly Journal of Economics*, **104**:275–98.

9. The saving–investment nexus in the debate on pension reforms*

Sergio Cesaratto

9.1. INTRODUCTION

The saving–investment nexus is central to the pension debate, an important aspect of which concerns the impact of the different pension schemes, unfunded or funded, on economic growth. We shall consider two saving-led growth models. The dominant, neoclassical approach considers transfer-based pay-as-you-go programmes (PAYG) as injurious to capital accumulation and favours the adoption of saving-based fully funded schemes (FF) that would instead encourage it. A second approach, based on a 'classical growth model', is also sympathetic to an FF reform by considering investment as determined by saving, presumably assuming the validity of Say's Law at least in the long run. This view has recently sparked off some debate among non-orthodox economists. A third alternative approach (let us label this 'classical-Keynesian'), based on the extension to the long period of the Keynesian postulate of the independence of investment from saving, regards PAYG as favourable (or at least neutral) with respect to accumulation, and any reform aimed at encouraging saving as detrimental to aggregate demand and growth.

Section 9.2 will summarise some results of the attempt by a number of Sraffian economists to reinforce the implications of Keynes's theory of effective demand for the explanation of accumulation (regarding the long period) in a direction that is, however, divergent from that taken by economists working in the tradition of the 'Cambridge equation'. We shall also single out the nature of autonomous expenditure of PAYG transfers in the theory of effective demand. Section 9.3 will criticise the proposal of a wider adoption of an FF scheme based on the neoclassical growth model. We shall also consider the neoclassical saving–investment causality in the open economy, rebuking the mainstream argument that investment in southern countries is such as to ensure the profitability of a larger amount of old-age saving. Section 9.4 will rebut the feasibility of an analogous reform proposal advanced by Michl and Foley, based on a 'classical' growth model.

9.2. THE CLASSICAL–KEYNESIAN THEORETICAL FRAMEWORK

9.2.1. Two Keynesian Positions

In *The General Theory*, Keynes endeavoured to show that, within the limits of full utilisation of output capacity, a larger amount of investment does not require a prior reduction in consumption, and that the higher level of output and income generated by the greater utilisation of capacity generates savings equal to decisions to invest. Michal Kalecki proposed a similar approach. The 'neoclassical synthesis' applied this mechanism to short-period situations of low business and financial confidence, arguing that in such circumstances fiscal and monetary policies were required in order to achieve full employment. This was the received view until the Monetarist revolution began to reaffirm the pre-Keynesian doctrines in the late 1960s. Although 'new classical economics' subsequently receded, current prevailing conventional wisdom still basically reflects pre-Keynesian views. In the meantime, however, a number of non-orthodox economists have tried the opposite course of extending what Kaldor (1955–56, p. 95) termed the Keynesian Hypothesis, that is the idea that investment is independent of savings in the long period, when productive capacity may vary significantly, as well as in the short period.

In this regard Garegnani (1992) distinguishes between two Keynesian positions. The former, well-known approach, synthesised by the Cambridge equation, suggests that income distribution – and in particular the real wage rate – is the variable that adjusts capacity saving (the amount of saving out of a fully employed productive capacity) to the investment decided by the entrepreneurs. We may distinguish two characteristics in the second, 'classical-Keynesian', position.

The first is the particular relevance attributed to the capital theory critique in the rescue of Keynes's theory of effective demand. Indeed, Hicks, Modigliani and others pointed out the limits of Keynes's criticism of the conventional theory soon after the publication of *The General Theory*. They circumscribed the role of effective demand to short period cases, in which the rigidity of the interest rate or the slow reaction of investment to a fall in the interest rate prevents the fall of nominal wages or the expansion of money supply from leading the economy back to full employment. Keynes himself paved the way to the neoclassical synthesis when he did not entirely reject the traditional approach, conceding in particular a downward sloping schedule of investment demand that was elastic to the rate of interest. For instance, in Chapter 16 of *The General Theory* Keynes repeatedly mentioned the possibility that, in principle, a higher saving rate may lead to more 'indirect' or capital-intensive production methods. Against this possibility, he argued that the new equilibrium 'might require a method of production so inconveniently "roundabout" as to have an efficiency well below the current rate of interest, [so that] the *immediate* effect of the saving would still be adverse to employment' (1936, p. 211). Keynes refers here to the rigidity of the *nominal* rate of interest in the presence of a fall in the *natural* rate (to use Wicksell's terminology) that follows the increased supply of saving. Garegnani (1983) has pointed out that Hicks and Modigliani actually took advantage of the *short-term* nature of that rigidity to re-establish the validity of the traditional theory, at least in the long run. In this regard, Garegnani has suggested that the results of the capital theory controversy inspired by Sraffa (1960) could validate the Keynesian principle of the independence of investment from saving.

This is so because the capital theory critique undermines the very possibility of deriving the demand curves for factors in a rigorous and general way. This derivation relied on two alternative mechanisms (for example, Solow, 1970, ch. 1): (a) *direct factor substitution* in production; and/or when there are fixed production coefficients, (b) *indirect factor substitution*, which operates through adjustments in the composition of consumers' optimal consumption baskets in response to relative price changes. In short, the first substitution mechanism predicts that when, for example, the interest rate falls, more 'capital intensive' methods of production will become more profitable and capital demand will increase. The capital critique has revealed that when there are a multiplicity of techniques and more than one type of capital good, the possibility of reswitching techniques undermines this neoclassical prediction (Sraffa, 1960, pp. 81–4; Garegnani, 1970). Indirect substitution by reshuffling consumption is also barred. According to marginalist theory, a fall in the relative price of any factor leads to a fall in the relative price of, and a rise in the demand for, the goods in whose production the factor is used relatively more intensely. However, as the real wage varies from zero to maximum, the price of any commodity *X* produced using a given technique may alternately rise and fall with respect to the price of another commodity *Y*, produced using a different given technique, so that no *a priori* expectations as to the direction of change, based on the 'factor intensity' in the production of the two commodities, can be justified (Sraffa, 1960, pp. 37–8).[1]

This criticism is injurious to both the neoclassical theory of distribution, based on the simultaneous determination of the rate of profit and the wage rate as reflecting the relative capital and labour scarcity; and to the associated theory of output and employment, based on the tendency towards full employment of both capital and labour – with the sole proviso that the financial and labour markets are competitive enough not to obstruct the full adjustment of factor demand and supply. The critique therefore complements the Keynesian hypothesis that savings are generated by investment and not vice-versa, as erroneously predicted by traditional theory.

The second characteristic is that the classical-Keynesian position departs from the first Keynesian position in the consideration of which mechanisms are behind the adjustment of savings to investments in the long run. In short, the idea is that even if there is a tendency in a market economy of capacity to adjust to long-period effective demand, 'the margins of unutilized capacity which are normal in a capitalistic system make it plausible to think that, in the long period, even more than in the Keynesian short period, autonomous changes in the incentive to invest will usually generate the corresponding amount of savings through changes in output rather than through changes in the real wage rate and normal rate of profits' (Garegnani, 1992, pp. 62–3).[2]

According to the theory of long-period effective demand, the engine of growth is represented by the *autonomous* or *final* components of aggregate demand: components that do not depend on the actual or expected level of real income generated by firms' decisions to produce. These comprise autonomous consumption, government spending and exports.[3] Autonomous expenditure also includes social transfers from the State that comprise PAYG pension payments. The theory of long-period effective demand therefore regards the payment of pensions as an autonomous decision to spend by the government and as a determinant of the social product, rather than as a mere transfer of income from a given level of output. It can thus be appreciated that in the classical-Keynesian approach social spending can represent an engine of growth, so that there is no necessary contradiction between greater social equity and growth.[4] Although favourable to stable economic growth, Welfare State expansion may stumble upon the obstacle of the political acceptance of changes in distribution, such as higher taxation on profits, by the most affluent classes. The Western economies accommodated this expansion in the historical conditions of post World War II, but since the late 1970s the distribution role of the eelfare state has been under constant attack (although at the price of feebler and unstable growth, particularly in Europe). Conversely, in the marginal theory the contradiction between 'equity' and 'efficiency' is more mechanical. In particular, according to this theory, pension transfers diminish disposable income and may reduce savings decisions and accumulation.

9.2.2. PAYG Pensions, Effective Demand and Economic Growth

According to conventional wisdom, the creation of a PAYG system is equivalent to an original and reiterated sin: potential savings out of, namely, wage growth that could be used for capital accumulation, are wasted in sustaining the older generation. This was not the dominant view in the US and other countries at the time of the inception of the public pensions programmes, which was well into the middle of the Keynesian revolution. At that point the prevailing view was that social transfer programmes were not only beneficial as such, but also favourable to effective demand and growth. This reflected Keynes's stance that 'measures for the redistribution of incomes in a way likely to raise the propensity to consume may prove positively favourable to the growth of capital' (1936, p. 373). Beveridge (1942) regarded full employment as a prerequisite for the Welfare State. This view subsequently took a more explicit Keynesian flavour in Beveridge (1944), where social transfers were explicitly regarded as a policy instrument for reaching full employment (Dimand, 1999, p. 232). Beveridge viewed 'State action in re-distributing income by measures of Social Security, and by progressive taxation' as favourable to the growth of 'private consumption outlay' (Beveridge, 1944, p. 30), since 'the income provided by the scheme to persons who are sick, unemployed, injured or past work, will almost invariably be spent to the full' (ibid., p. 160).[5] In 1943 Metzler put forward the distinction between the 'Investment Multiplier' and the 'Redistribution Multiplier'. The latter referred to the effects of a subsidy levied on the 'low propensity to consume group' in favour of the 'high propensity group'. The discussion later overlapped with that of the famous Balanced Budget Multiplier. In this debate PAYG transfers were clearly considered as an *autonomous expenditure*. One year after Metzler's contribution, Wallich published his important paper on the 'Income-Generating Effects of a Balanced Budget'.[6] He started by taking as 'generally assumed' the idea that 'the income-generating effect of a balanced budget depends upon its "progressiveness" i.e. upon the extent to which taxes and expenditures lead to a redistribution of monetary income from high-saving to low-saving groups and thus to a rise in the average propensity to consume' (1944, p. 78). The Balanced Budget Theorem actually set out to show that economic expansion was possible with a balanced government budget, even with a 'non-progressive budget'. The analysis was synthesised later by Musgrave (1945, 1959, chap. 18), who pointed out that a full employment policy based upon the Balanced Budget Theorem could take place either through an increase in public spending, or by an increase in social transfers that take advantage of the differing marginal propensity to consume of different income groups, those taxed and the beneficiaries of the transfers.[7]

In his celebrated 1958 paper in which the 'overlapping generations' method was launched, Paul Samuelson did not regard PAYG as harmful to capital accumulation. He limited himself to trying to rationalise PAYG within a neoclassical framework (the limits of this attempt were promptly denounced by Abba Lerner (1959); see Cesaratto 2002, 2005, ch. 1). It was only later, as part of the Monetarist revolution of the 1970s, that Martin Feldstein (1974) began to present PAYG as injurious to capital accumulation, since workers would interpret PAYG contributions as a substitute for private savings, while older people consume most of the transfers they receive. For instance, Feldstein remarked that: ‘The evidence presented in this paper seems ... consistent with the Keynesian view that the aggregate saving would increase as income rose if there were no offsetting government policies’ (1974, p. 922), that is payroll taxes (‘offsetting government policies’) would crowd out net saving out of rising wages. This suggestion has since been heeded, for instance, by the World Bank, which states: ‘suppose the government introduces a mandatory pay-as-you-go old age security plan that requires young people to contribute payroll taxes (equivalent to their previous saving) to the plan and pays them a pension (equivalent to their previous dissaving) later on. Introducing the system reduces national saving initially, because the first group to benefit from the program … receive a windfall gain … fewer resources are left to be saved and invested during the initial period, permanently reducing capital stock and national income’ (1994, p. 307). Noticeably, the World Bank does not take into consideration the uncertain destiny of the ‘first group’ of old people had they not received the ‘windfall gain’. The Bank is nonetheless forced to conclude that: ‘Despite the logic of this argument, numerous empirical investigations … have been unable to prove conclusively that saving did, indeed, drop once pay-as-you-go programs were established’ (ibid.).[8]

Drawing together the threads of this section, according to the theory of effective demand, pension transfers are an autonomous component of aggregate demand that may positively affect the level of the output, either by changing income distribution in favour of the social groups characterised by a propensity to consume higher than those who are called upon to finance them, or by being deficit-financed. Only when the propensity to consume of both the beneficiaries and the contributors is the same, and the social security budget is kept in balance, do PAYG pensions appear as mere transfers without any effect on the level of National Income. Conversely, according to the dominant theory, PAYG transfers may displace saving and investment, although the relevance of this displacement effect is questioned even by mainstream economists. The results of the capital controversy lend sufficient theoretical credibility to the rebuttal of the neoclassical stance and to the acceptance of the Keynesian interpretation, as we shall see better in the next section.

9.3. THE MARGINALIST FOUNDATIONS TO FF REFORMS

Mainstream economists regard an FF reform as an instrument to cope with demographic developments – the fall in the labour supply due to the falling fertility rates, and higher longevity.

9.3.1. FF Reforms and Capital Theory

According to marginal theory, capital consists of a fund of consumption goods through which consumption can be postponed (see Garegnani, 1990b, pp. 36–7). The existence of a decreasing demand schedule for capital that is elastic to the interest rate assures that a rise in 'foresight' (old-age motivated) saving is matched by additional capital accumulation that will lead, given the labour supply, to a higher per-capita capital endowment. In the case of a fall in the labour supply, the accumulated capital could either be absorbed by a higher capital–labour ratio, or reconverted into consumption goods. Increasing longevity can be dealt with by increasing the individual saving rate, working longer or spreading the consumption of the accumulated capital over a longer period.[9] Synthesising a longer argument (I refer the reader to Cesaratto 2005 and 2006), two problems with the mainstream suggestion of an FF reform can be envisaged.

(A) To begin with, the reform may fail to tangibly raise the marginal propensity to save even in economies in which there is no mandatory pension scheme in operation, since mandatory old-age saving may substitute similarly motivated discretionary saving. Mandatory savings may indeed raise the propensity to save of short-sighted workers, who are often also the poorest, for instance in the stagnating sectors of developing countries. However, it is socially difficult in their case to impose a cut in their consumption levels, which comprise the subsistence to the family, possibly including the elders. (Note in this regard that the old generation receives an instantaneous 'windfall gain' when a PAYG scheme is introduced, but not in the case of the institution of an FF scheme.) It is also politically difficult to impose an extra-mandatory saving upon workers that already contribute to a PAYG scheme. There are two policy options in this regard.

(i) The first is to divert to the FF scheme the part of real wage gains (due, say, to productivity growth) that would have normally gone to PAYG (which is consumed by the pensioners).[10] In this way, however, the current pension benefits are frozen in real terms and the social propensity to save is raised at the price of relative impoverishment of the current pensioners. This is not surprising since the extra-saving supply from workers must be matched by a cut in the consumption of some other social group.

(ii) Alternatively, given the real wage level, it has repeatedly been proposed that the government might divert part of the PAYG contributions to an FF programme and, at the same time, if it feels committed to honouring the current PAYG pension promises, it could rely on public debt to finance the transition. This reform is doomed to fail, since the higher workers' mandatory saving is compensated by lower government saving. *Prima facie*, what happens is that once workers' contributions to PAYG are diverted to the pension funds (PFs), they use the funds to finance the public debt that the government has in the meantime issued to finance the pension commitments. In practice the reform is a privatisation of PAYG. Although this is well known, these reforms are still proposed both by individual scholars and by international organizations as genuine shifts towards FF programmes.

The EU, for instance, in the recent revision of the Stability and Growth Pact 'will give due consideration' to 'pension reforms introducing a multi-pillar system that includes a mandatory, fully funded pillar', recognising that although 'the implementation of these reforms leads to a short-term deterioration of the budgetary position', nonetheless 'the long-term sustainability of public finances clearly improves' (Council of the European Union, 2005, p. 29). In the first approximation, however, the new workers' mandatory saving devoted to the 'funded pillar' is precisely equal to the additional government deficit, so that national saving is unaffected by the reform. Thus, even if we accept a saving-led model of accumulation, for the sake of argument, this kind of reform does not affect the saving supply (the PFs will *prima facie* hold Treasury Bonds). It might be countered that, as a result of a privatised PAYG, workers would obtain a rate of return ρ on those bonds that might be higher than the notional rate γ obtained by the traditional PAYG (γ is approximately equal to the growth rate of the total wages), so that the same level of pension can be obtained in the new system with lower contributions (and workers may save more). The question is whether this higher return (if any) will generate a further strain on public finances, and it is not clear who will be called upon to pay for it. If, once the system is fully under way, the additional cost is met by increasing taxation on the same workers, the net advantage for them is nil (Geanakoplos et al., 1998, pp. 14–17). If progressive taxation hits other social groups, there will be a favourable net effect for workers (but perhaps these other groups will save less, so the net effect on saving is nil). However, this increase in benefits vis-à-vis past contributions could just as well be obtained under the traditional PAYG without incurring the higher managerial costs of the privatised PAYG. The additional debts will not arise if the Treasury bonds only yield the rate γ, but in this case, alas, workers will probably receive less than γ, given the high administrative costs of the PFs (see, for instance, Murthi et al., 2001, and Mesa-Lago, 2002, on the experiences of the UK and Latin America, respectively), so that the reform appears damaging to them (but certainly not for the profits of the private financial sector).[11]

The World Bank, acknowledging these objections, has suggested that the real purpose of privatising PAYG does not lie in fostering capital accumulation, but in generating social alarm that an increasing social security deficit may engender about its sustainability, therefore creating a climate more favourable to its dismantling (World Bank, 1994, pp. 267–9).

(B) The second problem met by an FF reform is that, in the light of the classical-Keynesian criticism of the neoclassical saving–investment nexus, the attempt to increase the average community propensity to save by imposing *extra* mandatory savings on workers in favour of an FF scheme, if successful, will have deflationary effects. As we have seen, the argument that a given amount of 'free resources' would better satisfy the old-age motive if committed to capital accumulation – rather than being transferred to current pensioners through PAYG – crucially relies on the neoclassical proposition according to which a rise in the saving rate leads, *ceteris paribus*, to a higher investment rate due to the adoption of more capital intensive techniques. The capital critique shows that the theory underlying this prediction is flawed. According to many non-conventional economists, this result is the most fundamental analytical reason in support of the conclusions reached by Keynes in *The General Theory* (1936, for example, pp. 83–4, p. 211) not to confuse the desire by some individuals to hold more *financial* wealth with an increase in the capital stock. By negatively affecting effective demand and employment, this desire will negatively affect the income of other individuals and their saving supply. As a result, aggregate financial wealth and its real counterpart, capital stock, are unaffected.

The criticism of the neoclassical interpretation of the saving–investment relation is the ultimate challenge to the conventional view of an FF reform. In synthesis, not only is it difficult for policy makers to raise the propensity to save, but even if they are successful, the effects on investment will be nil and the reform will prove abortive (see also Palley, 1998, pp. 99–100).

9.3.2. The Saving–Investment Nexus in Open Economies and in Endogenous Growth Theory

We often hear the standard argument that *southern* countries will be the natural outlets for the abundant capital supply from 'grey' developed countries. According to this argument, current investment in southern countries will raise the rate of return on pension saving and help to meet the expected abundance of supply of financial assets when the 'baby-boom generation' retires. Although demographic changes are also occurring in southern countries, they will be much less dramatic and much slower there, so that a market for the capital assets supplied by the retiring *northern* baby-boom generation will not be lacking in the south. The international institutions warmly subscribe to this theoretical prescription.[12]

As is well known, in Solow's model of growth a decline in the growth rate of population leads to a new secular equilibrium characterised by a higher per-capita capital endowment and a lower interest rate. According to the standard view, southern countries are generally in an opposite situation, with a younger and abundant labour force accompanied by scarcer capital endowment. Hence, these countries offer a higher rate of return on capital investment than developed countries. As a result, if capital funds can move freely, the abundant saving supply from developed countries will tend to flow to more profitable southern countries, thus helping the latter to avoid the low saving trap, and the northern countries to cope with demographic imbalances without incurring excessively low interest rates on pension funds. Appropriately, Krugman and Obstfeld (1994) interpret foreign lending by northern to southern countries as inter-temporal trade, that is as a convenient channel through which the rich northern workers can postpone their consumption by exporting capital to those countries where this is relatively scarce.

It should be appreciated that in the case of foreign investment, no less than in the domestic case, the idea that domestic saving finds an automatic *debouche* in investment in southern countries depends on the neoclassical saving–investment nexus, and is therefore a victim of the capital critiques. In the light of this critique, there is no automatic mechanism that would translate the larger (potential) saving supply into domestic or foreign investment, since a fall in the rate of interest does not affect investment either in the domestic economy or in that of other countries. The result is therefore a fall in domestic (and foreign) income and employment.

This can easily be seen in terms of the well-known national account identity $S_n - I_n = (G_n - T_n) + (X_n - M_n)$ (where the subscripts indicate the region, north or south). Given government savings $(G_n - T_n)$, conventional economists would maintain that a rise in S_n, given domestic I_n, causes a rise of investment in the south I_s that will determine a rise in X_n. Conversely, according to Keynesian economists, a rise in the propensity to save in the north, given I_n, causes a fall in national income accompanied by a fall in M_n. The north's foreign saving has increased, but at the cost of a lower national income. The fall in M_n entails, in turn, a fall in exports and output in the south. This is shown by a simple example.

Supposing there are two countries, one in the north and one in the south. Using textbook notation, in the north we suppose: $c_n = 0.8$, $m_n = 0.2$, $I_n = 100$, $E_n = 200$. Output is therefore $Y_n = 750$, and foreign saving $S_n^f = X_n - M_n = 50$. In the south c_s and m_s have the same values, while $I_s = 250$ and $E_s = 150$. Output is therefore $Y_s = 1000$, and $S_s^f = X_s - M_s = -50$. If $c_n = 0.7$, and if we take into account the full repercussions of the fall in imports from the north on the exports from the south, we obtain the following new values: $Y_n = 500$, $S_n^f = 68.7$, $Y_s = 875$ and $S_s^f = -68.7$. This disappointing result is, so to speak, an extension of Keynes's saving paradox to an open economy: given the level of world investment, a rise in the marginal propensity to save in one country may lead to an increased amount of foreign saving in this country, but at the price of a lower domestic (and foreign) output. In an enlightening paper that was brought to my attention after this chapter was written, Dalziel and Harcourt (1997), taking inspiration from James Meade, also extend Keynes's multiplier analysis to an open economy.[13]

Alternatively, conventional economists have identified the outcomes of endogenous growth theory (EGT) as a response to the excessive old-age motivated supply of capital in northern countries and consequent decline in its marginal productivity, made worse by the decline of labour supply. The results of EGT are not, however, very robust (not to mention the traditional saving–investment nexus on which EGT is built). Only by employing an impressive battery of *ad hoc* devices can this approach assume away the traditional causes of a falling marginal product of capital.[14]

9.4. A 'CLASSICAL' MODEL OF FF REFORMS

The classical economists were quite open to different approaches to the theory of accumulation (Garegnani, 1983, pp. 24–8). For instance, on the one hand, mechanical forces leading to full employment are not to be found either in Ricardo or in Marx. On the other hand, Ricardo believed in Say's Law, which was instead discarded by Marx. The classical-Keynesian interpretation of the Keynesian Hypothesis suggested in Section 9.2 combines the rejection of automatic tendencies to full employment with the rebuff of Say's Law (basing this refusal on Keynes, Kalecki and the results of the capital controversy, rather than on Marx). An alternative stance, combining the rejection of automatic full employment but accepting Say's Law, is pursued by Michl and Foley (2004; Michl, 2001, 2002), who apply it to back an FF reform. This view has provoked some discussion that we shall extend in this section.

We shall examine Michl and Foley's (M&F) argument by first presenting their 'classical' model and social security reform proposal, then comparing it to the neoclassical model, and finally putting forward some criticism.

9.4.1. M&F's 'Classical' Model and Social Security Reform

The 'classical' model on which M&F build their proposal of social security reform is based on the assumption that, in the long run, capital accumulation and employment growth are determined by saving. In support of this assumption, however, the authors do not go much further than pronouncements like: 'The classical model we employ in this analysis assumes full utilisation of the capital stock, and thus excludes consideration of aggregate demand effects of social security' (M&F, 2004, p. 3); or: 'the mere possibility of a problem with aggregate demand does not absolve us from making decisions that are fundamentally long-run in nature' (Michl, 2001, p. 87).[15]

On these frail foundations M&F suggest a social security reform based, in synthesis, either on a social security 'off-budget' surplus obtained through an additional payroll tax on wages or on taxes on capitalists' profits or wealth, or on the utilisation of a possible 'on-budget' government surplus, which should preferably be the outcome of progressive taxation, to constitute a *reserve fund* held by the social security administration that should be used to foster (private) capital accumulation (M&F, 2004, p. 10; Michl, 2001). Supposing that the reserve fund is financed out of increasing payroll taxes – and assuming that workers do not cut their voluntary old-age saving – then constitution of the fund would demand 'a period of primitive accumulation, during which taxes exceed benefits' (M&F, 2004, p. 10). 'If this surplus' – they continue – 'is built up from the payroll taxes of workers, it implies a kind of forced saving in which one or more generations of workers contribute to a fund which benefits future generations' (ibid.). Since the rate of profit r is considered to be higher than the rate of income growth g (which, for a constant share of wages in income and PAYG contribution rate, is also the rate of growth of the wage bill, often interpreted as the rate of return γ on PAYG contributions), then this accumulation would enable workers to benefit from some of the results of capital growth and would even permit a reduction of the contribution rates necessary to obtain the same target replacement rate (the initial pension benefit over the last wage) obtained with PAYG. The idea is that the sacrifice of some generations, who are called upon to increase their mandatory saving – with the important proviso that they do not cut their discretionary saving – is invested by the social security fund in private capital accumulation, thus creating a fund of real capital assets that can later finance pensions out of profits: ‘With returns from the reserve fund subsidizing the system, a given level of benefits can be supported with lower taxes, increasing the money’s worth of the payroll tax’ (Michl, 2001, p. 85). Moreover, thanks to the sacrifice of some initial generations, through social security the working class (in particular retired workers) will end up possessing a share of the larger capital stock and earning profits.

9.4.2. ‘Classical’ and Neoclassical Models and Reform Proposals

A comparison with the neoclassical argument in support of an FF reform is timely here. The idea that the higher ‘profitability’ of an FF scheme and the continuous reinvestment of profits would consent a reduction of the contribution rates necessary to reach a final target replacement rate is in common with the reform proposals of, respectively, Feldstein (1996) and Modigliani et al. (1999) (see Cesaratto, 2002, 2005, Chapter 4, 2006). If the creation of the reserve fund is financed out of additional payroll taxes (additional to those necessary to finance the current payment of existing PAYG pensions and stipulating that workers do not reduce their discretionary saving), then although workers see their net wage fall initially, the same or subsequent generations of workers will later see their net wage increasing. This is because the higher returns from the invested reserve funds will allow a reduction of the contribution rate. There are also some dissimilarities between M&F and the neoclassical Modigliani–Feldstein reform proposals that may help to clarify what M&F have in mind.[16] According to Modigliani–Feldstein, a net increase in old-age mandatory saving does lead, given a fully employed labour supply, to a higher capital–labour ratio and consequently to a higher per-capita product. The advantage of the reform therefore lies in the workers’ higher endowment of real capital and productivity (see Section 9.3 above). In contrast, M&F assume a constant capital–labour coefficient and the existence of a perfectly elastic labour supply. In this case, the mandatory net saving is used to provide extra workers with the average capital endowment, but there is no increase in labour productivity. Profits out of the new capital, however, now pertain to the working class through the public fund. Thus, although the product per worker and the real wage rate have remained the same, this is not true for the total income accruing to the working class. The extra income is used to finance pension expenditure and reduces the pressure on active workers.

Alternatively, higher labour productivity can be obtained in M&F's model by introducing some sort of externality from capital accumulation, as in the AK models of endogenous growth theory, or a Kaldorian technical progress function.

Noticeably, M&F believe that their 'classical' saving-led model is not only analytically superior, but also shows the advantages of an FF reform better than the neoclassical growth theory. M&F regard the 'theory of population' as the main difference between the 'classical' and the neoclassical growth theory. The former would regard 'population growth as an endogenous response to the accumulation of capital' (M&F, 2004, p. 2; Michl, 2001, p. 88). In neoclassical theory, on the contrary, the rate of accumulation is adapted to the rate of growth of the population. In this approach, labour as an exogenous constraint to growth would discourage any endeavour to speed up growth by increasing the saving rate – in the neoclassical context the saving supply out of full employment income – due to the operation of the law of decreasing returns to capital accumulation. This would not happen in the 'classical' model. More precisely, in neoclassical growth theory a rise in the saving rate, given the labour supply growth rate, would bring about three frustrating outcomes: (a) it has only level effects on the output per worker but not on the rate of growth, which is exogenously determined by labour growth and technical progress; (b) the increase of output per worker is progressively less and less than proportional to the rise in the per-capita capital endowment – as shown by the standard decreasing slope of the well-behaved per-capita production function; and (c) the marginal product of capital tends to fall with the progressive abundance of capital with respect to labour. In M&F's view these implications would discourage any reform aimed at fostering old-age saving through an FF scheme reform (Michl, 2001, p. 88).[17] However, they argue that once the idea of labour supply as an exogenous constraint to growth is dismissed, the question of the decreasing returns would also be discarded.

9.4.3. Shortcomings of M&F's 'Classical' Model and Reform Proposal

We shall advance three main critiques of M&F's model. The first regards the treatment of population as the main distinctive element between classical and neoclassical theories. The second concerns the saving–investment nexus as postulated by these authors. Lastly, we shall point out the self-contradictory role that M&F attribute to economic policies in the context of their reform proposal.

(i) To begin with, M&F do not criticise the *analytical* consistency of the conventional theory. In the first place, they do not note that the mainstream conclusions that the rate of accumulation is conditioned by the growth rate of the labour supply (which therefore acts as a constraint) depend on the conviction that in a market economy competition brings about a tendency to full employment. This tendency is in turn based on the belief in decreasing demand functions for productive factors. In contrast, the absence of these demand functions has allowed classical economists to regard the determination of the accumulation rate and of labour demand as independent of labour supply: labour supply is not mechanically binding in classical theory because competition does not lead to full employment. The observation of reality, moreover, suggested to them that capitalism did not suffer from labour scarcity, and that the labour supply tended to be responsive to periods of faster accumulation, although the relation between labour demand and population developments was far from the mechanical views that have often been attributed to them (see Stirati 1994, Chapter 4 and pp. 175–6). In the second place, M&F do not recognise that the notion of decreasing returns to capital accumulation is also a theoretical result drawn from the same marginal apparatus. The two authors appear to argue as if decreasing returns were not a theoretical construct, but rather a factual result, similar to Ricardo's classification of lands according to their different fertility. M&F thus seem to limit themselves to taking issue with the *empirical* side of the hypothesis of labour constraint, that is whether *de facto* a situation of labour scarcity is currently biting or will be so in the near future.[18]

Given the present fertility trends in developed countries, the two authors regard immigration as the most responsive source of labour supply to the US accumulation patterns in the past as well as in the future, and not without reason. However, this prompts three questions. First, this may not apply in the long run for a number of developed non-settlement countries such as in Europe and Japan. This introduces new problems that cannot be addressed here.[19] More importantly, if labour supply is not a constraint to accumulation in the classical framework (for a given capital–labour ratio), this would also be true in a neoclassical setting. This shows that it is not sufficient to distance oneself from marginal theory in the way M&F do. Finally, even if the supply of material labour in developed countries was binding, investment in developing labour-rich regions and EGT have represented a safety measure for neoclassical economists against the three discouraging results listed above. However, M&F do not discuss foreign investment and EGT (at least in the pension context), as we have attempted to do in Section 9.3.

(ii) Secondly, Foley and Michl argue in their book that since in 'the long run, the economy would be attracted to the full utilization of capacity' ($u = 1$), then an independent investment function 'becomes superfluous at $u = 1$', and they endorse the idea that 'the economy is Keynesian in the short run but classical in the long run' (Foley and Michl, 1999, p. 192). They state that both the classical and neoclassical traditions 'see saving as the engine of capital accumulation and assume that saving decisions always lead to a corresponding decision to invest. In these models *Say's Law* holds, and there can be no discrepancy between aggregate demand and supply' (ibid., p. 193, italics in the text).[20] There are two problems in this regard. First, neoclassical economists advance a sophisticated argument to show the adequacy of investment decisions to full employment saving decisions. A corresponding argument is missing in the 'classical' approach proposed by M&F. Secondly, the extension to the long run of the Keynesian principle of the autonomy of investment from saving (see Section 9.2) does not depend on denying the tendency of capacity to a *normal* degree of utilisation, that is to adjust to long period aggregate demand in the long run – as in the neo-Kaleckian models that Foley and Michl (1999, chapter 10) take as the only example of long-period Keynesian models.[21] Whereas the adjustment of saving to investment takes place through *variations of output* in the short run, it can be argued that in the long run the tendency of saving to adjust to the long-run level of investment takes place precisely through the *variations of capacity* (see Garegnani, 1992; Garegnani and Palumbo, 1998, and for a criticism of the neo-Kaleckian models see Trezzini, forthcoming). In conclusion, a long-run tendency towards normal utilisation does not imply that, in the long run, increases in saving will convert into increased accumulation, as M&F seem to believe.

As we have seen, M&F back Say's Law in a rather naïve way, *assuming* that saving decisions determine investment. M&F are well acquainted with the capital theory critique (see Foley, 2001), so they avoid the marginalist decreasing demand curve for capital, but do not suggest any robust alternative mechanism of adjustment between supply and demand for capital. Ricardo's original view of accumulation was based on the *coincidence* of saving and investment decisions that were taken by an entrepreneurial capitalist class that saved in order to invest: a view that was perhaps more easily defensible at the time. The Ricardian view is echoed by M&F when they argue that 'growth depends critically on the existence of a capitalist class which accumulates wealth for bequest purposes' (2004, p. 19). Bequest may well explain saving decisions, as in some neoclassical models that M&F quote but, as noted above, the mainstream authors put forward a complex mechanism – the capital demand and supply apparatus which is the object of the capital theory critique – whereby saving decisions are translated into investment decisions. M&F do not have an analogous mechanism in place, and the decision to buy a piece of machinery cannot be easily defended by the wish to leave a bequest.

(iii) Thirdly, M&F circumscribe the role of aggregate demand in the determination of the growth rate to the short run (Michl, 2001, p. 87) and rely on fiscal and monetary policy to compensate the deflationary effects of a rising propensity to save (M&F, 2004, p. 3). The employment of fiscal policy is self-contradictory since it is equivalent to destroying with the right hand (the creation of a budget deficit) what the left hand has just done (the constitution of a budget surplus to finance the social security reserve fund). More consistently, Michl concedes that 'the fiscal surpluses needed to fund social security represent a deflationary policy, in so far as the short run is concerned', and relies on monetary policy as an effective instrument to avoid deflation in the short run (Michl, 2002, p. 113).[22] Thus, the peculiar policy mix that Michl proposes is a fiscal *surplus* to stimulate investment in the long run (ibid., p. 115), plus an expansionary monetary policy to avoid the deflationary effects of the same fiscal surplus in the short run. There is a conventional, Wicksellian echo in Michl's policy proposal: a higher saving rate implies a lower natural interest rate; in order to avoid deflation, it is the duty of the monetary authorities to guide the market interest rate towards the natural rate.[23] This stance is not acceptable in the light of the classical-Keynesian approach presented in this chapter.

Incidentally, M&F's saving-led growth view also leads them to regard PAYG as detrimental to capital accumulation 'by discouraging life cycle saving' (2004, p. 11). In the first place this displacement effect is not altogether obvious, as workers would not necessarily have saved the mandatory contributions to PAYG (see Section 9.3 above). Secondly, there is no reason why these savings would necessarily be invested.

To sum up, M&F recommend creating a Social Security fund to be used to foster capital accumulation. Our criticism has pointed in three directions. To begin with, M&F's criticism of neoclassical theory appears rather weak and not analytical, being based on a factual observation that the US will not suffer from labour scarcity in the near future (or at least that they will have fewer troubles than other developed countries). However, if this is true in their 'classical' context, it will also be true in the neoclassical framework, that will not therefore meet the mentioned 'decreasing marginal returns' to capital accumulation. Secondly, M&F's effort to support a transition reform towards an FF scheme by basing it on a classical model does not appear analytically robust since it lacks any plausible saving-based theory of accumulation. Finally, M&F acknowledge the necessity to compensate the deflationary effects of the proposed reform by using fiscal and monetary policy. However, an expansionary budget contradicts the very essence of their reform, which is based on increasing government saving, while monetary policy is ineffectual with respect to investment.

9.5. FINAL REMARKS

In the light of the theory of effective demand, PAYG is an institution that has a double linkage with full employment policies: its sustainability depends on them and it is an instrument of those policies. Whereas a neoclassical perspective sees PAYG as a mere sharing of full employment income and possibly detrimental to capital accumulation, in a classical-Keynesian perspective income transfers to older generations may positively affect aggregate demand and income by increasing the average propensity to consume. On the contrary, the attempt to raise the average community propensity to save by cutting PAYG and extending the FF schemes will have deflationary effects. An incomplete appreciation of the implications of the Keynesian–Sraffian critique of the neoclassical saving–investment nexus with regard to the secular determinants of capital accumulation, associated with a mistaken inference that a tendency of capacity to adjust to aggregate demand would restore the traditional causal relation between saving and investment, is, perhaps, behind the support given by non-conventional economists such as M&F to an FF reform.

The neoclassical capital theory critique and the theory of effective demand also dispute the conventional idea that a higher old-age saving supply from the northern countries will naturally flow to the capital-poor southern countries, stimulating investment there. On the contrary, from a Keynesian point of view, it is an investment decision (or autonomous consumption or public spending decisions) in the south, independent of any saving decisions in the north, that generates a higher output and savings in the north by raising exports to the south. *Ex post*, this saving will appear as a surplus in the north's trade balance that is matched, in the balance of payments, by a capital outflow to the south. This may give the illusion that foreign savings in the north have generated investment in the south. As James Meade put it in a letter to Dalziel and Harcourt (1997, p. 628), it is enough to think of the globe as a closed economy and the same Keynesian causality, if valid in a single country, will necessarily apply to the world as well.

To sum up, demographic developments are a serious fact (UN, 2002) and pose new challenges (Cesaratto, 2005, Chapter 8), but saving-led models of economic growth provide an analytically defective policy recipe for dealing with them.

NOTES

* This chapter is partly based on selected parts of Cesaratto (2005). I wish to thank, without implication, two referees and Duncan Foley for their valuable comments on previous versions, and Prue Kerr for help in improving this exposition.

1. General equilibrium models try to circumvent the problems of capital measurement by taking the vector of capital goods given in physical terms as exogenously known data. This approach, however, leads to further problems, both methodological and substantial (see Garegnani, 1990b).
2. Following Joseph Steindl, classical-Keynesian economists stress the distinction between normal and full-capacity utilisation. They argue that firms normally tend to install enough capacity to be able to respond, for instance, to sudden peaks of demand, so as not to lose customers to competitors. This implies that a speed-up in the accumulation rate can be accommodated by increasing the degree of capacity utilisation over the normal level, without the need for a fall of real wages, as implied by the models based upon the 'Cambridge equation' (see Ciccone, 1990).
3. Cesaratto et al. (2003) have argued that gross investment should be excluded from the autonomous components and rather be considered as induced by expected demand. Palumbo and Trezzini (2003) are of the opposite opinion
4. The role of the welfare state and PAYG pensions in the classical distribution theory are discussed in Cesaratto (2005, Chapter 7).
5. Beveridge regarded his proposed policy of full employment as a 'policy of socialising demand rather than production' (Beveridge, 1944, p. 190), and one that had to 'be undertaken if free society is to survive' (ibid., p. 192). Clearly, the failure of the market economies before the war and the hopes raised in vast sectors of the population by the Soviet experiment, explain why the best minds in the bourgeoisie urged for the design of an alternative both to extreme *laissez-faire* and to the socialization of the means of production.
6. As is well known, the Balanced Budget Multiplier was independently discovered by a number of economists. Wallich mentions A. Hansen, P. Samuelson and W. Salant. Haavelmo (1945, p. 311) mentions P. Samuelson, A. Hansen, N. Kaldor and H. Wallich.
7. Suppose that pension transfers are partially financed out of deficit spending. By generating a Keynesian income-multiplier process, a social security deficit finds its financing through the increase in private savings. Suppose then that savings come from the working class, motivated by the 'foresight' motive, and are collected by the pension funds that, *prima facie*, employ them to finance the public deficit. The social security deficit has then fostered the existence of a 'privatised PAYG': a pension scheme in which pension transfers are financed (at least in part) by issuing government bonds bought by workers through the pension funds (see Section 9.3.1 below). This particular scheme is not a strictly genuine FF programme since the reserves are constituted by public bonds and not by financial assets representative of real capital goods. As de Finetti (1956, p. 279) pointed out many years ago: '[A]lmost all the authors seem to agree that it would simply be window dressing if the capitalisation consisted in investments in public debt' (my translation; for full discussion see Cesaratto, 2005, Chapter 4).
8. To give an idea of this indeterminate prediction on saving, suppose that pensioners receive a transfer Tr. The consequences hinge upon the combinations of workers' possible saving behaviours and the possible reactions of the pensioners (see, for example, Engen and Gale, 1997). On the one hand, in the case of workers that are too poor and, probably connected with this, too improvident to save, Tr levied on their wage does not displace any saving. Workers that are wealthy enough to save might instead cut their savings by Tr, or by sTr (where s is the marginal propensity to save), or not cut them at all (if, for instance, savings

are motivated by bequest). On the other hand, affluent pensioners endowed with other income sources, who see their revenue rise by $Tr,$ may altruistically react by leaving their consumption level untouched, saving and bequeathing $Tr.$ This is of course Barro's much discussed Ricardian Equivalence case. Poorer pensioners, if previously supported by their descendants, may simply return Tr to their offspring. If previously lacking support, they may instead greedily consume the 'windfall gain' and survive a little longer.

9. The limits of the capacity to cope with demographic developments that neoclassical economists attribute to an FF scheme are discussed in Cesaratto (2005, Chapter 3).
10. It is generally recognised that the elderly have a higher propensity to consume than workers (for example, Engen and Gale, 1997, p. 111).
11. The recent reform proposal of the US President Bush is a mixture of options (i) and (ii): divert part of the Social Security contribution to an FF scheme and cover the Social Security deficit (a) by freezing pension benefits in real terms (currently they are indexed to real wage growth) in order to reduce future pension spending, and (b) by issuing public debt.
12. It is worth quoting a conventional economist here: 'In the absence of foreign pension investment into younger economies, what should we expect to happen to capital returns on funded pensions once the OECD baby boomers have started to retire? As the labour force declines, the existing capital stock becomes oversized relative to the labour force. The change in relative factor proportion reduces the rental return on capital relative to wages; this effect is reinforced if fully-funded pensions indeed stimulate savings. Simultaneously, the prior phase of asset accumulation would give way to a long period of asset decumulation, as the baby boomers start to draw on their pension assets to finance their retirement. Clearly, therefore, a fully funded pension scheme is bound to get under stress by population ageing, very much like an unfunded scheme. But the funded pensions, unlike the unfunded schemes, can partly beat demography in an open economy. The asset decumulation during the retirement period will not be confined to home assets, but to emerging-market assets that still will be benefiting from net pension contributions of the underlying younger population. And capital returns, unlike in a closed economy, will not be lowered by a declining labour force, but by the world capital market and the demand for capital by the younger non-OECD area' (Reisen, 2000, pp. 3–4).
13. Two interpretations of capital flows to developing countries may be tentatively distinguished here. The conventional approach to international capital flows echoes the time-honoured Robertsonian *Loanable Fund Theory*. Accordingly, behind capital flows there are surplus countries – those which present an excess of domestic saving over domestic investment – that lend excess saving to borrowing countries – those who invest more than the domestic saving supply. An alternative approach would maintain that a process of development does not meet a 'saving constraint', but a 'foreign liquidity constraint', that is, the scarcity of international liquidity necessary to finance the balance of payments deficits that result from the process of development (I owe this point to Franklin Serrano and Carlos Medeiros of the Federal University of Rio de Janeiro). The international financial institutions are able to create credit facilities for southern countries that are thus able to increase their imports – of investment goods in the most fortunate cases, and often of luxury goods. It is this increasing demand for exports from the northern countries that generates the 'twin deficits' (both negative foreign saving and trade balance) in the southern countries, and the corresponding 'twin surpluses' in the north (both positive foreign saving and trade balance), in which the increasing exports determine a rise in income and saving. In this alternative account, 'credit precedes investment and both precede savings', as well expressed by Kriesler and Halevi (1996, p. 309), and the saving–investment gap is 'nothing but the *ex post* accounting result of the operation'.

14. This author has shown elsewhere (Cesaratto, 1999a, 1999b; Serrano and Cesaratto, 2002) that the questions that EGT tries to solve are not new, since they were already discussed in the 1960s. The solutions were also well known, but discarded by the best neoclassical minds of the time as too *ad hoc*. As Stiglitz (1990, p. 55) has recalled: 'We knew how to construct models that "worked", but felt uneasy making these special assumptions'.
15. According to M&F, in the 'classical' model (as well as in the neoclassical model) 'aggregate accumulation will be determined by the saving of the most "patient" class of households', the class that the classical economist would identify with 'capitalists' (M&F, 2004, p. 2). This suggests that, in spite of the authors' contention that their model is in the tradition of the 'Cambridge equation' (M&F, 2004, p. 1; Michl, 2001, p. 90) – in which the 'animal spirits' determine the rate of accumulation g and, as a consequence, also the profit rate according to the well-known equation $g = s_c r$ (in which s_c is the saving propensity of the capitalists) – their model seems closer to the spirit of the EGT's AK model (for example, Rebelo, 1991). Indeed, the mentioned assumption of an exogenous wage rate (M&F, 2004, pp. 2 and 6) is at odds with the endogenous determination of distribution in the Cambridge equation. Solow (1992) has suggested that the AK model is a resumption of the Harrod–Domar model in which the saving rate determines the rate of capital accumulation according to the famous equation of the 'warranted' rate of growth: $g_w = s / v_d$ (in which v_d is the desired capital–output ratio) and distribution is exogenous. A consistent neoclassical foundation of the AK model is presented in an unfortunately forgotten paper by Frankel (1962), summarised in Cesaratto (1999a, 1999b).
16. I thank Franklin Serrano for valuable suggestions in this regard.
17. It must be acknowledged that these dismal implications of a larger saving supply when labour supply is binding are present in the mainstream literature on pensions. They are, for instance, reported in the following passages by a conventional economist in which the 'impact of ageing on macroeconomic performance' is assessed: 'As regard to growth, it is widely considered that it will decelerate as ageing proceeds, principally because of lower labour-force growth. There will also be lower growth in living standards … than has been the case in recent decades, reflecting the accompanying increase in the dependency ratio. Effects on growth of a fall in labour force growth are unlikely to be offset by higher investment. Indeed, investment is itself likely to decline given a lesser need for capital widening, while capital deepening is likely to be limited by diminishing returns. Moreover, slower growth will tend to reduce returns on capital directly, thus again putting downward pressure on investment' (Davis, 2004, p. 2).
18. M&F also argue that once the real wage rate is exogenously given, 'the capital intensity of the technique of production and the profit rate are invariant in the short as well as the long run' (M&F, 2004, p. 2). But in this way they base the absence of decreasing returns to capital accumulation within the classical framework on the hypothesis of a given wage rate – a hypothesis which is certainly closer to the classical economists – and not on the (weaker) hypothesis that in the classical context labour is not binding. M&F, however, do not integrate the more classical hypothesis of a given real wage rate with a general refutation of the marginalist apparatus. There is an analogy between M&F's stance and mainstream non-substitution theorem, according to which when the wage rate is exogenously given and there are constant returns to scale, then distribution is not affected by output levels. Garegnani (1990a, p. 128) shows that this is not enough to criticise neoclassical theory.
19. In most non-settlement countries immigration can at best help to stabilise the population in the working age (see UN, 2000), unless the population is allowed to reach explosive levels. The big 'Kaleckian' question to be explored in this context is how capitalism will be able to operate with a limited industrial reserve army in an ageing society (Cesaratto, 2005, Chapter 8).

20. Keynes himself in *The General Theory* repeatedly suggested that although his theory was valid when resources were less than fully utilised, saving-led growth resumed its validity once full employment was restored. For instance, he wrote: 'If the rate of interest were so governed as to maintain continuous full employment, Virtue would resume her sway; – the rate of capital accumulation would depend on the weakness of the propensity to consume' (Keynes, 1936, p. 112). It can be appreciated that Keynes endorses the neoclassical saving–investment nexus if the interest rate were suitably 'governed'.
21. In this regard Michl (2002, p. 114) writes: '[our] analytical choice (shared by economists across the theoretical spectrum) makes good sense in describing an economy whose behavior is Keynesian in the short run but classical in the long run. Macroeconomic models, such as that of Marglin and Bhaduri (1990), that attempt to give meaning to demand-constrained growth require firms to operate with unplanned excess capacity even in a long-run equilibrium when they are continuing to add new capital. Many find this to be intellectually unappealing … these models are more convincingly interpreted as representations of the short-run behavior of a system that has classical characteristics in the long run. Social security, fundamentally a long-run macroeconomic policy, demands long-run analysis'.
22. Here, Michl is replying to the Keynesian criticism by Palley (2002) who argued: 'Michl's vision is predicated on a "corn" model approach to economic process, whereby increased saving (corn not consumed) is immediately translated into investment (corn planted in the ground) to yield more income (corn) in the future. This is the implicit logic behind the claim that buying equities automatically results in investment. This vision of the investment process is at odds with the reality of a monetary economy. Buying equities already in issue simply transfers money from the bank account of the buyer to the bank account of the seller, and does not result in more investment … Dissecting the transaction in this fashion makes clear that it is investment that drives saving, rather than the other way around as assumed in corn model economics' (Palley, 2002, p. 106). Michl (2001, p. 114) recognises that he is using a corn model. The capital theory critique allows us to extend this criticism to the more sophisticated neoclassical explanation of the relation between saving and investment.
23. It might be thought that if the budget surplus is obtained by taxing profits or capitalists' capital assets, the deflationary effects are negligible since taxes will mainly hit savings or be financed out of decumulation of wealth. However, in this case the only effect would be a reshuffling in the property of capital but not higher growth (even if saving caused investment). In the case of a tax on profits, *ceteris paribus*, the larger government saving would compensate the lower capitalists' saving. In the case of a tax on capital wealth, the reserve fund would just buy the capital assets sold by the capitalist to pay the new taxes.

REFERENCES

Beveridge, W.H. (1942), *Social Insurance and Allied Services*, New York: Macmillan.

Beveridge, W.H. (1944), *Full Employment in a Free Society*, London: George Allen & Unwin.

Cesaratto, S. (1999a), 'New and old neoclassical growth theory: a critical assessment', in G. Mongiovi and F. Petri (eds), *Value, Distribution and Capital: Essays in Honour of Pierangelo Garegnani*, London: Routledge.

Cesaratto, S. (1999b), ‘Savings and economic growth in neoclassical theory: a critical survey’, *Cambridge Journal of Economics*, **23**:771–93.

Cesaratto, S. (2002), ‘The economics of pensions: a non-conventional approach’, *Review of Political Economy*, **14**:149–77.

Cesaratto, S. (2005), *Pension Reform and Economic Theory: A Non-Orthodox Analysis*, Cheltenham UK and Northampton, MA, USA: Edward Elgar.

Cesaratto, S. (2006), ‘The transition to fully funded pension schemes: a non-orthodox criticism’, *Cambridge Journal of Economics.*

Cesaratto, S., F. Serrano and A. Stirati (2003), ‘Technical change, employment and effective demand’, *Review of Political Economy*, **15**:33–52.

Ciccone, R. (1990), ‘Accumulation and capacity utilisation: some critical Council of the European Union, 2005, Presidency Conclusions, Brussels 22 and 23 March’, doc. 7619/05.

Council of the European Union (2005), *Presidency Conclusions*, Brussels 22 and 23 March, doc. 7619/05.

Dalziel, P.C. and G.C. Harcourt (1997), ‘A note on “Mr Meade’s Relation” and international capital movements’, *Cambridge Journal of Economics*, **21**:621–31.

Davis, E.P. (2004), ‘Demographic and pension-system challenges to financial and monetary stability’, Austrian Central Bank Conference, Wien.

de Finetti, B. (1956), ‘Sicurezza sociale e obiettivi sociali’, in B. de Finetti (1965), *Un matematico e l’economia*, Milano: Franco Angeli Editore.

Dimand, R.W. (1999), ‘The Beveridge report: Beveridge’s response to the Keynesian challenge’, in L.L. Pasinetti and B. Schefold (eds), *The Impact of Keynes on Economics in the 20th Century*, Cheltenham UK and Northampton, MA, USA: Edward Elgar.

Engen, E.M. and W.G. Gale (1997), ‘Effects of social security reform on private and national saving’, in S.A. Sass and R.K. Triest (eds), *Social Security Reform*, Boston: Federal Reserve Bank of Boston, Conference Series no. 41.

Feldstein, M. (1974), ‘Social security, induced retirement, and aggregate capital accumulation’, *Journal of Political Economy*, **82**:905–26.

Feldstein, M. (1996), ‘The missing piece in policy analysis: social security reform’, *American Economic Review*, *Papers and Proceedings*, **86**(May):1–14.

Foley, D.K. (2001), ‘Value, distribution and capital: a review essay’, *Review of Political Economy*, **13**:365–81.

Foley, D.K. and T.R. Michl (1999), *Growth and Distribution*, Cambridge, MA: Harvard University Press.

Frankel, M. (1962), ‘The production function in allocation and growth: a synthesis’, *American Economic Review*, **52**:995–1002.

Garegnani, P. (1970), ‘Heterogeneous capital, the production function and the theory of distribution’, *Review of Economic Studies*, **37**:407–436.

Garegnani, P. (1983), ‘Notes on consumption, investment and effective demand’, in J. Eatwell and M. Milgate (eds), *Keynes’s Economics and the Theory of Value and Distribution*, Duckworth: London.

Garegnani, P. (1990a), ‘Sraffa: classical versus marginalist analysis’, in K. Bharadwaj and B. Schefold (eds), *Essays on Piero Sraffa – Critical Perspectives on the Revival of Classical Theor*y, London: Unwin Hyman.

Garegnani, P. (1990b), 'Quantity of capital', in J. Eatwell, M. Milgate and P. Newman (eds), *Capital Theory*, *The New Palgrave*, London: Macmillan.

Garegnani, P. (1992), 'Some notes for an analysis of accumulation', in J. Halevi, D. Laibman and E. Nell (eds), *Beyond the Steady State*, London: Macmillan.

Garegnani, P. and A. Palumbo (1998), 'Accumulation of capital', in H.D. Kurz and N. Salvatori (eds), *Elgar Companion to Classical Economics*, Cheltenham UK and Northampton, MA, USA: Edward Elgar.

Geanakoplos, J., O.S. Mitchell and S.P. Zeldes (1998), 'Would a privatized social security system really pay a higher rate of return?', National Bureau of Economic Research Working Paper Series, no. 6713.

Haavelmo, T. (1945), 'Multiplier effects of a balanced budget', *Econometrica*, **13**:311–18.

Kaldor, N. (1955–56), 'Alternative theories of distribution', *Review of Economic Studies*, **23**:83–100.

Keynes, J.M. (1936), *The General Theory of Employment, Interest, and Money*, London: Macmillan.

Keynes, J.M. (1942), 'The Beveridge Report', in D. Moggridge (ed.), *The Collected Writings of John Maynard Keynes*, Vol. XXVII, London: Macmillan.

Kriesler, P. and J. Halevi (1996), 'Asia, Japan and the internalization of effective demand', *Economies and Sociétés – Monnaie et production*, **10**:301–20.

Krugman, P.R. and M. Obstfeld (1994), *International Macroeconomics*, 3rd edn., New York: HarperCollins College Publishers.

Lerner, A.P. (1959), 'Consumption-loan interest and money', *Journal of Political Economy*, **57**:512–18, with a 'Rejoinder', 523–25.

Marglin S.A. and A. Bhaduri (1990), 'Profit squeeze and Keynesian theory', in S.A. Marglin and J.B. Schor (eds), *The Golden Age of Capitalism*, Oxford: Clarendon Press.

Mesa-Lago, C. (2002), 'Myth and reality of pension reform: the Latino-America evidence', *World Development*, **30**:1309–21.

Metzler, L.A. (1943), 'Effects of income redistribution', *Review of Economic Statistics*, **24**:49–57.

Michl, T.R. (2001), 'Why we should fund social security permanently', *Challenge*, **44**:78–92.

Michl, T.R. (2002), 'Prefunding is still the answer', *Challenge*, **45**:112–16.

Michl, T.R. and D.K. Foley (2004), 'Social security in a classical growth model', *Cambridge Journal of Economics*, **28**:1–20.

Modigliani F., M.L. Ceprini and A.S. Muralidhar (1999), 'A solution to the social security crisis', Sloan Working Paper, no. 4951.

Murthi, M., J.M. Orszag and P.R. Orszag (2001), 'Administrative costs under a decentralized approach to individual accounts: lessons from the United Kingdom', in R. Holzmann and J. Stiglitz (eds), *New Ideas About Old Age Security*, Washington DC: World Bank, pp. 308–35.

Musgrave, R.A. (1945), 'Alternative budget policies for full employment', *American Economic Review*, **35**:387–99.

Musgrave, R.A. (1959), *The Theory of Public Finance*, New York: McGraw-Hill.

Palley, T.I. (1998), ‘The economics of social security: an old Keynesian perspective’, *Journal of Post Keynesian Economics*, **21**:93–110.
Palley, T.I. (2002), ‘Social security: prefunding is not the answer!’, *Challenge*, **45**:97–118.
Palumbo, A. and Trezzini A. (2003), ‘Growth without normal capacity utilisation’, *European Journal of Economic Thought*, **10**:109–35.
Rebelo, S. (1991), ‘Long-run policy analysis and long-run growth’, *Journal of Political Economy*, **99**:500–521.
Reisen, H. (2000), *Pensions, Savings and Capital Flows*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar in association with the OECD.
Samuelson, P.A. (1958), ‘An exact consumption–loan model of interest with or without the social contrivance of money’, *Journal of Political Economy*, **56**:467–82.
Serrano, F.L. and S. Cesaratto (2002), ‘As leis de rendimento nas teorias neoclasica de crescimiento: Una critica Sraffiana’, *Revista Ensaios FEE*, **23**:699–730 (English translation: ‘The laws of return in the neoclassical theories of growth: a Sraffian critique’, www.networkideas.org).
Solow, R.M. (1970), *Growth Theory: An Exposition*, Oxford: Clarendon Press.
Solow, R.M (1992), *Siena Lectures on Endogenous Growth Theory*, Collana del Dipartimento di economia politica, vol. 6, Università di Siena.
Sraffa, P. (1960), *Production of Commodities by Means of Commodities. Prelude to a Critique of Economic Theory*, Cambridge: Cambridge University Press.
Stiglitz, J.E. (1990), ‘Some retrospective views on growth theory’, in P. Diamond (ed.), *Growth/Productivity/Unemployment*, Cambridge, MA: MIT Press.
Stirati, A. (1994), *The Theory of Wages in Classical Economics*, Aldershot, UK and Brookfield, USA: Edward Elgar.
Trezzini, A. (forthcoming), ‘Steady state and the analysis of long-run tendencies: the case of Kaleckian models’, in R. Ciccone, C. Gehrke and G. Mongiovi (eds), *Proceedings of the Conference Sraffa and Modern Economics*, Routledge: London (forthcoming).
UN Population Division (2000), *Replacement Migration: Is it a Solution to Declining and Ageing Population?*, New York.
UN Population Division (2002), *World Population Ageing: 1950–2050*, New York.
Wallich, H.C. (1944), ‘Income-generating effects of a balanced budget’, *Quarterly Journal of Economics*, **59**:78–91.
World Bank (1994), *Averting the Old Age Crisis, Policies to Protect the Old and Promote Growth*, Oxford: Oxford University Press.

10. Income distribution and output change: a macro multiplier approach*

Maurizio Ciaschini and Claudio Socci

10.1. INTRODUCTION

The degree of income inequality affects the productive structure, the market, the households and the public sector. Personal income distribution is the result of microeconomic and social variables – such as the ownership of factors and the individual features which connect each industry to the various institutional sectors – and macroeconomic variables such as the functional income distribution and the sector's share in national income. In Italy the degree of income inequality increased in the 1990s as a result of many correlated factors: a redistribution of income from labour to capital, a high fiscal burden on labour, a system of monetary transfers based on retirement provision, a change in the output structure which has privileged some sectors over others. This process, however, has been different in the various regions. This is the reason why this analysis requires both a reliable regional data base – such as that provided by the social accounting matrix (SAM) – and a consistent model which follows the Leontief recommendation 'Empirical inquiry and theoretic model building have to be carried out hand-in-hand' (Leontief, 1988). From the SAM approach a model of circular income flow which is more articulated than the usual one emerges: each macroeconomic flow variable, conveniently disaggregated, generates a second flow variable through the use of a structural matrix and progressively so on until the loop is closed. Final demand determines total output and value added by industry; the latter generates domestic income by factors which compose disposable income by institutional sector; this gives rise to final demand closing the loop. The proposed model is an extension of the Miyazawa approach (Miyazawa, 1970) through the integration of secondary income distribution (Pyatt, 2001).

To address this progress in the design of a data base which provides meaningful sectorization of the major macroeconomic variables, flexible tools of analysis are needed. These tools will also allow for a deeper insight into the propagation phenomena characterizing sectoral and industrial interactions. In these phenomena it is not so much the scale as the structure of macroeconomic variables which plays a major role. The traditional tools for studying propagation are those provided by impact multipliers and linkage analysis. These tools, however, provide procedures that do not give a complete account of the changes in the structures of macro variables. Although the analysis we propose is based on the original ideas of Rasmussen from which linkage analysis originates, that is the notion of propagation analysis with the determination of summary measures and their use in statistical terms, it applies a new tool that overcomes the difficulty of considering only one predetermined structure of macro variables. Our propagation analysis is centred on a particular decomposition, singular value decomposition. The most popular spectral decomposition is given by the eigenvalues and eigenvector decomposition. However this procedure, while extremely interesting for the study of the powers of matrices in any structural analysis, is not convenient for our purposes, since it can be applied only to square matrices and produces roots which can be positive, negative or complex conjugated pairs. Singular value decomposition has no such drawbacks and produces roots (singular values) that can be easily interpreted as aggregated macro economic multipliers. We then define and quantitatively determine a set of aggregated macro multipliers, which leads the economic interactions, and the structures of macroeconomic variables, that either hide or activate these forces.

On the basis of these new tools we determine new indices of propagation, that we define as backward and forward dispersion, that give an evaluation of the propagation phenomena with reference to some notable structures of the variables implied (singular vectors) and to associated scales (singular values). The approach allows for a further development: if the decomposition is applied on the standardized structural matrix, we can have the picture of the degree of interaction between each row and column of the matrix in terms of correlations.

The analysis will be applied to an extended income–output loop, derived from a SAM, that can be quantitatively tested forwarding a shock on a given macro variable and observing the effects on another macro variable within the loop. It will identify the most efficient structure, without impinging on the equi-distributed unitary shock. 'Summary' measures will be found, consistent with the multi-sectoral and multi-industry framework, that will allow us to measure the degree of interaction between sector (row) and industry (column) components.

In Section 10.2 the concept of backward and forward dispersion as developed by Rasmussen is explained. It points out that these multipliers reflect a particular and unlikely structure of final demand. In order to circumnavigate this misconception, we propose to base the multipliers on a social accounting framework rather than on an input–output table. Section 10.3 explains the concept of a SAM and extended model: it provides a schematic representation of the complex interlinkages between final demand formation, production, value added generation and income allocation. In Section 10.4 the multipliers of disposable income changes (by seven institutional sectors) on outputs are developed by performing singular value decomposition of the corresponding matrix multiplier. Dispersion indices are then determined and finally the correlation analysis performed providing a graphical representation of the degree of interaction between institutional sectors and industries. Some conclusions are given in Section 10.5.

10.2. SOME CONSIDERATIONS ON MULTIPLIERS AND LINKAGE ANALYSIS

The original input–output (I–O) problem consists in the search for an equilibrium output vector for the *n* I–O sectors of the economy. Since, in the following section, income will be disaggregated by institutional sector, in order to avoid misinterpretation, we will use the term *industries* for producing sectors, and *sector*s for institutional sectors. Such a vector conveniently faces the predetermined final demand vector **f** by industries, and the induced industrial demand. The equilibrium output vector is given by

$$\mathbf{x} = \mathbf{Rf} \tag{10.1}$$

where $\mathbf{R} = [\mathbf{I} - \mathbf{A}]^{-1}$ and **A** is the technical coefficients matrix, and generally exists, as in general the technology can be expected to be productive, that is the technology is such that a part of total output is still available for final uses after the intermediate requirements have been satisfied. In this case, **A** satisfies the Hawkins–Simon conditions (Ciaschini, 1988). The matrix **R** is usually referred to as the Leontief multipliers matrix and its elements (r_{ij}) show the direct and indirect requirements of industry output *i* per unit of final demand of product at industry *j*. Extensive use is made of matrix **R** within traditional multipliers analysis and a substantial part of linkage and key sectors analysis is based on it.

The **R** matrix provides, in fact, a set of disaggregated multipliers that are recognized to be the most precise and sensitive for studies of detailed economic impacts. These multipliers recognize the evidence that total impact on output will vary depending on which industries are affected by changes in final demand. The *i*-th total output multiplier measures the sum of direct and indirect input requirements needed to satisfy a unit final demand for goods produced by industry *i* (Bulmer-Thomas, 1982).

Research on linkage analysis dates back to the definitions elaborated by Rasmussen (1956) of 'summary measures for the inverse matrix'. He noted that the sum, $r_{.j}$, of column elements $(r_{.j} = \sum_{i=1}^{m} r_{ij})$ corresponds to the total increase in output from the whole system of industries needed to match an increase in the final demand for the product of industry j by one unit. Similarly the sum, $r_{i.}$, of row elements $(r_{i.} = \sum_{i=1}^{m} r_{ij})$ gives the increase in output of industry i required to meet a unit increase in final demand for the product of each industry.

We can take the average, $(1/m)r_{.j}$, which represents an estimate of the (direct and indirect) increase in output to be supplied by an industry chosen at random if final demand for the products of industry j expands by one unit (Rasmussen, 1956, p. 130). Similarly, $(1/m)r_{.j}$ can be regarded as the average increase in output to be supplied by industry i if the final demand for the products of an industry chosen at random is increased by one unit.

To carry out consistent interindustry comparisons, we need to normalize these averages by the overall average defined as $(1/m^2)\sum_{i=1}^{m} r_{i.}$ and thus consider the indices

$$\pi_j = \frac{mr_{.j}}{\sum_{j=1}^{m} r_{.j}} \quad \text{and} \quad \tau_i = \frac{mr_{i.}}{\sum_{i=1}^{m} r_{.i}} \tag{10.2}$$

The aim of the direct and indirect backward dispersion index π_j, the *power of dispersion* in the Rasmussen definition (Rasmussen, 1956, p. 135), is to measure the potential stimulus to other activities from a demand shock in any industry j. The forward dispersion τ_i, the *sensitivity of dispersion* in the Rasmussen definition, measures the degree to which one industry output is used by other industries as an input.

It has to be stressed, however, that the Rasmussen definitions were of a statistical nature, since both measures were mean values of either outputs or final demands of industries chosen at random. For each of these measures, he elaborated a coefficient of variation, that is a standard deviation. Nevertheless, the original statistical approach of the Rasmussen analysis progressively disappeared and the interpretation of his measures have definitely become deterministic.

It has to be stressed, however, that all these measures, built starting from matrix $\mathbf{R}$, are not independent of the structure of either total output vector, on which we observe the effects, nor of the structure of the final demand vector on which we impose the unit demand shock. The column sum of matrix $\mathbf{R}$ in equation (10.1) implies the consideration of a set of final demand vectors of the type:

$$\mathbf{f}^1 = \begin{bmatrix} 1 \\ 0 \\ . \\ . \\ 0 \end{bmatrix}, \quad \mathbf{f}^2 = \begin{bmatrix} 0 \\ 1 \\ . \\ . \\ 0 \end{bmatrix}, \ldots., \quad \mathbf{f}^m = \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ 1 \end{bmatrix} \tag{10.3}$$

while the sum of row elements in equation (10.1) implies consideration of a final demand structure of the type:

$$\mathbf{f} = \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ 1 \end{bmatrix} \tag{10.4}$$

We can expect these measures to hold for demand vectors of varying scale but with the same structures of equations (10.3) or (10.4). However, neither the demand vector nor its changes will ever assume a structure of this type. This is why some authors come to the drastic conclusion that 'linkage should be never used' (Skolka, 1986).

On the other hand, it is commonly held that the structure of final demand produces the most different effects on the level of total output (Ciaschini, 1993). Given a set of nonzero final demand vectors, whose elements sum up to a predetermined level, but with varying structures, we will have to expect that the corresponding level of total output will also vary considerably.

For these reasons we cannot confine our knowledge of the system to the picture emerging from measures which can only show what would happen if final demand assumed a predetermined and unlikely structure.

10.3. THE EXTENDED OUTPUT–INCOME CIRCULAR FLOW

The results attained in social accounting encourage the attempt to build an extended version of the income circular flow where the interactions between industries and institutions may be specified and evaluated.

Figure 10.1 shows a diagram where the fundamental mechanism of production and distribution is shown in terms of interaction between industries, sectors and factors (value added components).

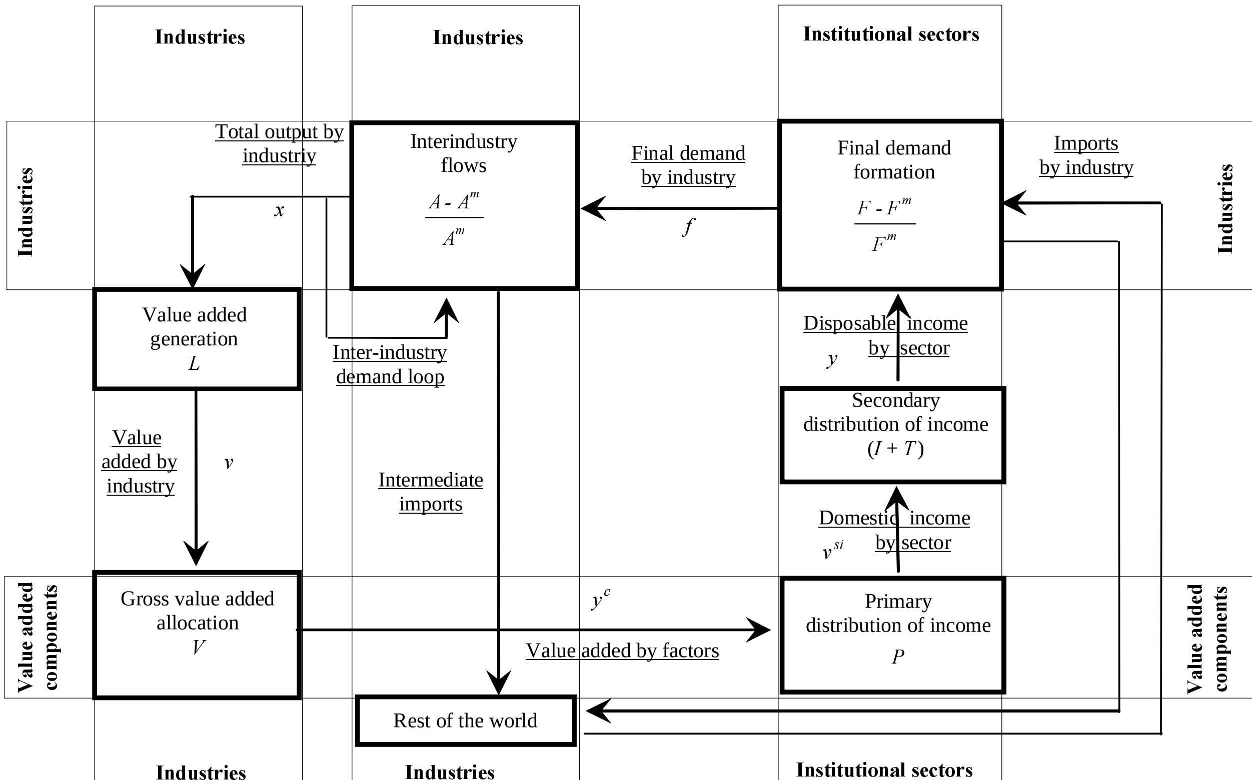


Figure 10.1 Extended output–income circular flow

In Figure 10.1 each arrow identifies an expenditure flow while each box a matrix transformation of one flow variable into another. This loop is built through various logical phases. The production process, that takes place at industry level, generates total output, **x**, and gross value added by the 11 I–O industries (see Table 10.1), $\mathbf{v}(x)$, (Gross value added generation). Value added by I–O industry is then allocated to the three value added components (factors),[1] $\mathbf{v}^c(x)$ (gross value added allocation). Value added by components is then allocated to the seven institutional sub-sectors,[2] $\mathbf{v}^c(x)$ (primary distribution of income). Value added by institutional sectors is then redistributed among them through taxation to generate disposable incomes by the seven institutional sub-sectors, $\mathbf{y}(x)$ (secondary distribution of income). Finally, disposable income will generate final demand by institutional sub-sectors which will be transformed into final demand by I–O industries, $\mathbf{f}(x)$ (final demand formation).

On this logical sketch of the extended circular flow, we can define the structural parameters representing the distribution matrices. If we introduce institutional sectors and income distribution in the interindustry model, final demand will no longer be exogenous but explained by income distribution. The model proposed is built under the assumption of fixed prices and constant distributive shares and coefficients (Pyatt and Round, 1979). The distributive shares and the distributive coefficients describe the behaviour of the agents that we consider in the model, that is consumption shares, primary and secondary income distribution shares, shares of valued added generated, import and technical coefficients of production. This procedure is a common practice in multisectoral applied models for policy and in particular in the case of SAM models (Round, 2003). Usually the constancy of the shares provides the model with flexibility in the phase of determining the results. These are obtained through the simulation of parametric changes on specific distributive shares which constitute the scenario of the specific simulation, keeping the rest constant.

Income originates from value added, which is determined by subtracting from industry output x_i the sum of intermediate purchases by industry i. The gross value added generation (by industry) will be given by

$$\mathbf{v}(x) = \mathbf{L}\,\mathbf{x} \qquad (10.5)$$

where $\mathbf{L}[11,11]$ is a diagonal matrix whose elements give the constant residual shares of value added by industry, $l_j = 1 - \sum_i a_{ij}$ and a_{ij} is the technical coefficient, which is usually defined import inclusive, and x_i is the i-th industry output.

Gross value added is allocated to factors through the following equation

$$\mathbf{v}^c(x) = \mathbf{V}\mathbf{v}(x) \tag{10.6}$$

where the elements of matrix $\mathbf{V}[3,11]$, v_{ij}, represent the constant shares of wage and salaries, other incomes (gross operating surplus) and indirect taxes generated in the j-th industry.

Primary distribution of incomes (by institutional sectors) requires the attribution of each factor income to the owners of those factors, namely households by income category, firms and government. Primary income by institutional sectors is then given by

$$\mathbf{v}^{si}(x) = \mathbf{P}\mathbf{v}^c(x) \tag{10.7}$$

where the elements of matrix $\mathbf{P}[7,3]$, p_{ij}, represent the constant share of value added by factor attributed to the i-th institutional sector.

In order to determine the disposable income of institutional sectors (secondary distribution of income) we need to correct sectoral incomes by income and tax transfers. This will be determined as follows

$$\mathbf{y}(x) = (\mathbf{I} + \mathbf{T})\mathbf{v}^{si}(x) \tag{10.8}$$

where matrix $\mathbf{T}[7,7]$ represents the net income and tax transfers constant shares that each institutional sector receives/forwards from/to the remaining institutional sectors and matrix $\mathbf{I}$ is the identity matrix.

Final demand by industry – in the two components, endogenous and exogenous – will then be given by

$$\mathbf{f}(x) = \mathbf{F}^0\mathbf{y}(x) + \mathbf{K}\mathbf{y}(x) + \mathbf{f}^0 \tag{10.9}$$

where $\mathbf{F}^0 = \mathbf{F}^1\mathbf{C}$ is given by the product of two matrices. Each element of matrix $\mathbf{F}^1$ $[11,7]$, f^1_{ij}, represents the consumption demand of the commodity produced by the i-th industry as a constant share of the consumption expenditure of the j-th institutional sector, while the diagonal matrix $\mathbf{C}[7,7]$ represents the consumption propensities by institutional sector.

Each element of matrix $\mathbf{F}^0$, f^0_{ij}, then provides the share consumption expenditure for the commodity produced by the i-th industry relative to disposable income of institutional sector j-th. The consumption structures are fixed within each institutional sector. However, if the distribution of incomes among sectors changes, say from sector 1 to sector 2, we will observe a corresponding shift from consumption structure f^0_{i1} $i = 1, \ldots, n$ to consumption structure f^0_{i2} $i = 1, \ldots, n$.

The matrix $\mathbf{K}$ represents the investment demand and is given by $\mathbf{K} = \mathbf{K1}\, s\, (\mathbf{I} - \mathbf{C})$ where each element $\mathbf{K1}[11,7]$, k^1_{ij}, represents the investment demands to I–O industry as a share of investment expenditures by institutional sectors. Scalar s represents the share of private savings which is transformed into investment, that is 'active savings'. Finally, $\mathbf{f}^0$ is a vector of 11 elements which represents exogenous demand.

If we put $\mathbf{F} = [\mathbf{F}^0 + \mathbf{K}]$ equation (10.9) becomes

$$\mathbf{f}(x) = \mathbf{F}\mathbf{y}(x) + \mathbf{f}^0 \tag{10.10}$$

Substituting through equations (10.5)–(10.9) in (10.10) we get

$$\mathbf{f}(x) = \mathbf{F}[\mathbf{I} + \mathbf{T}]\mathbf{PVLx} + \mathbf{f}^0 \tag{10.11}$$

Let us consider now the interindustry output generation process. Given a matrix of constant technical coefficients $\mathbf{A}$, that implies no choice of technique, we get

$$\mathbf{x} + \mathbf{m} = \mathbf{A}\,\mathbf{x} + \mathbf{f}(x) \tag{10.12}$$

where $\mathbf{m}$ represents imports, $\mathbf{A}$ the technical coefficients matrix, and $\mathbf{f}(x)$ represents the demand vector in equation (10.11).

Imports can be modelled according their main components, intermediate consumptions, endogenous demand and exogenous demand:

Import

$$\mathbf{m} = \mathbf{A}^m\mathbf{x} + \mathbf{F}^m(\mathbf{I} + \mathbf{T})\mathbf{PVLx} + \mathbf{f}^m \tag{10.13}$$

where $\mathbf{A}^m$ [11,11] represents the intermediate coefficients imports matrix, $\mathbf{F}^m$ [11,7] represents import shares of endogenous demands and $\mathbf{f}^m$ represents imports generated by an exogenous shock.

Substituting equations (10.11) and (10.13) in (10.12) we finally get:

$$\mathbf{x} = [\mathbf{I} - (\mathbf{A} - \mathbf{A}^m) - (\mathbf{F} - \mathbf{F}^m)(\mathbf{I} + \mathbf{T})\mathbf{PVL}]^{-1}(\mathbf{f}^0 - \mathbf{f}^m) \tag{10.14}$$

Assuming

$$\mathbf{R} = [\mathbf{I} - (\mathbf{A} - \mathbf{A}^m) - (\mathbf{F} - \mathbf{F}^m)(\mathbf{I} + \mathbf{T})\mathbf{PVL}]^{-1} \tag{10.15}$$

we can rewrite equation (10.14) as

$$\mathbf{x} = \mathbf{R}[\mathbf{f}^0 - \mathbf{f}^m] \tag{10.16}$$

which gives the reduced form of the model, that is the solution of the model with respect to the exogenous variables. Here we have only considered the exogenous component of final demand. However the model is such that, at the level of each box in the diagram in Figure 10.1, we can introduce an exogenous component which plays the role of policy instrument.

Irrespective of its matrix formalization the model is very simple. Its simplicity mainly resides in the assumptions of fixity in coefficients and shares as well as in the absence of a price side (fix-price) and in its uni-periodicity. However, it gives rigorous account of the various phases of income generation and distribution and of the industry versus sectoral coherence of the economic aggregates. Some additional information may emerge from the comparison of the results obtained imposing different configurations of the structural parameters. On the other hand we think that a simple model may allow for easier data interpretation and we are encouraged by Leontief's recommendation of keeping the model very near to the data base.

10.4. SEARCHING FOR A 'SUMMARY' APPROACH: THE DECOMPOSITION OF THE STRUCTURAL RELATIONSHIP

In this section we will explicitly consider the interaction between industries and institutional sectors operating on the structural matrices composing the loop in equations (10.5)–(10.14). We will also utilize singular value decomposition in our attempt to find a 'summary' measure of propagation.

The interactions between industries and institutional sectors can be appreciated if one considers the direct and indirect effects of disposable incomes on industry outputs. For this type of evaluation we introduce an exogenous shock on disposable income $(\mathbf{y}^0)$ in equation (10.8) that becomes

$$\mathbf{y}(x) = (\mathbf{I} + \mathbf{T})\mathbf{v}^{si}(x) + \mathbf{y}^0 \tag{10.17}$$

From the extended income output circular flow we determine the link between a unit change in disposable income by institutional sectors to total output by industries that is given by

$$\mathbf{x} = \mathbf{R}[\mathbf{f}^0 + \mathbf{F}\mathbf{y}^0] \tag{10.18}$$

where $\mathbf{F} = [\mathbf{F}^0 + \mathbf{K}]$ gives the link between disposable income and final demands shown in equation (10.10) and $\mathbf{R}$ is given in equation (10.15). If $\overline{\mathbf{R}} = \mathbf{RF}$, equation (10.18) becomes

$$\mathbf{x} = \bar{\mathbf{R}}\mathbf{y}^0 \tag{10.19}$$

Equation (10.19) provides the solution of the model with respect to an exogenous control (shock) on disposable income.

We will now perform a spectral decomposition of matrix $\bar{\mathbf{R}}$. The most popular spectral decomposition is given by the eigenvalues and eigenvector decomposition. However this procedure, while extremely interesting for the study of the powers of $\bar{\mathbf{R}}$, does not suit our aims, since it can be applied only to square matrices and produces roots which can be positive, negative or complex conjugated pairs.

A further type of decomposition may be derived which has no such drawbacks and produces roots that can be easily interpreted as aggregated macro economic multipliers (Ciaschini, 1989). The decomposition proposed, singular value decomposition, can be applied both to square and to non-square matrices. Here the general case of non-square matrix $\bar{\mathbf{R}}$ will be shown. The square matrix case is easily developed along the same lines.

In order to determine the singular values of matrix $\bar{\mathbf{R}}$ we need to perform eigenvalue decomposition of the square of matrix $\bar{\mathbf{R}}$. Let us consider matrix $\mathbf{W}$, the square of our [11,7] structural matrix $\bar{\mathbf{R}}$:

$$\mathbf{W} = \bar{\mathbf{R}}^T\bar{\mathbf{R}}$$

Matrix $\mathbf{W}$ has a positive definite or semi-definite square root. Given that $\mathbf{W} \geq 0$ by construction, its eigenvalues (λ_i) $i = 1, ..., 7$ will be all real non-negative (Lancaster and Tiesmenetsky, 1985). Importantly, matrix $\bar{\mathbf{R}}^T\bar{\mathbf{R}}$ does not coincide with matrix $\bar{\mathbf{R}}\bar{\mathbf{R}}^T$. However the nonzero eigenvalues of matrices $\bar{\mathbf{R}}^T\bar{\mathbf{R}}$ and $\bar{\mathbf{R}}\bar{\mathbf{R}}^T$ coincide and their square roots are defined as singular values of matrix $\bar{\mathbf{R}}$ that is $s_i = \sqrt{\lambda_i}$ with $i = 1, ..., 7$.

The system of eigenvectors [$\mathbf{u}_i$ $i = 1, ..., 11$] for $\bar{\mathbf{R}}^T\bar{\mathbf{R}}$ and [$\mathbf{v}_i$ $i = 1, ..., 7$] for $\bar{\mathbf{R}}\bar{\mathbf{R}}^T$ are orthonormal bases. Hence

$$\bar{\mathbf{R}}^T\mathbf{u}_i = s_i\mathbf{v}_i \qquad i = 1,...,7$$

and

$$\bar{\mathbf{R}}^T\mathbf{u}_i = \mathbf{0} \qquad 8 \leq i \leq 11$$

We can construct the two matrices

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{11}] \quad \text{and} \quad \mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_7]$$

and we obtain

$$\bar{\mathbf{R}}^T\mathbf{U} = \left[s_1\mathbf{v}_1, s_2\mathbf{v}_2, \ldots, s_7\mathbf{v}_7, 0, \ldots, 0\right] = \mathbf{VS}$$

The main result of these mathematical manipulations is that we can re-write the reduced form of our model, shown in equation (10.19), as the product of three matrices where the dimensional (scale) effects are completely quantified in diagonal matrix $\boldsymbol{S}$ while the other two matrices, **U** and **V**, are unitary and represent structures:

$$\mathbf{x} = \mathbf{USV}^T\mathbf{y}^0 \tag{10.20}$$

V is a [7,7] unitary matrix whose columns, as we will see, define the seven reference structures for disposable income:

$$\mathbf{v}_1 = \begin{bmatrix} v_{1,1} & v_{1,2} & v_{1,3} & . & . & . & v_{1,7} \end{bmatrix}$$

$$\mathbf{v}_2 = \begin{bmatrix} v_{2,1} & v_{2,2} & v_{2,3} & . & . & . & v_{2,7} \end{bmatrix}$$

....

$$\mathbf{v}_7 = \begin{bmatrix} v_{7,1} & v_{7,2} & v_{7,3} & . & . & . & v_{7,7} \end{bmatrix}$$

U is a [11,11] unitary matrix whose columns, as we will see, define 11 reference structures for output:

$$\mathbf{u}_1 = \begin{bmatrix} u_{1,1} \\ u_{2,1} \\ u_{3,1} \\ . \\ u_{11,1} \end{bmatrix}, \; \mathbf{u}_2 = \begin{bmatrix} u_{1,2} \\ u_{2,2} \\ u_{3,2} \\ . \\ u_{11,2} \end{bmatrix}, \ldots, \; \mathbf{u}_{11} = \begin{bmatrix} u_{1,11} \\ u_{2,11} \\ u_{3,11} \\ . \\ u_{11,11} \end{bmatrix}$$

and $\boldsymbol{S}$ is a [7,7] diagonal matrix of the type:

$$\mathbf{S} = \begin{bmatrix} s_1 & 0 & 0 & . & 0 \\ 0 & s_2 & 0 & . & 0 \\ 0 & 0 & s_3 & . & 0 \\ . & . & . & . & . \\ 0 & 0 & 0 & . & s_7 \end{bmatrix}$$

Scalars s_i, the set of singular values, are all real and positive and can be ordered by order of magnitude as $s_1 > s_2 >, \ldots, > s_7$.

Now the economic interpretation of these results is based on the graph in Figure 10.2. We have all the elements to show how this decomposition correctly represents the macro multipliers that quantify the aggregate scale effects and the associated structures of the impact of a shock in disposable income on total output.

The vector of disposable income by institutional sector $\mathbf{y}^0$ can be expressed in terms of the structures identified by matrix $\mathbf{V}$, where each row suggests a new income distribution pattern among sectors. We obtain a new disposable income vector, $\boldsymbol{y}$, expressed in terms of the new income distribution structures suggested by $\overline{\mathbf{R}}$:

$$\overline{\mathbf{y}} = \mathbf{V}^T \mathbf{y}^0 \tag{10.21}$$

On the other hand, we can also express total output according to the output structures implied by matrix $\overline{\mathbf{R}}$, identified by each column of matrix $\mathbf{U}$:

$$\overline{\mathbf{x}} = \mathbf{U}^T \mathbf{x} \tag{10.22}$$

Equation (10.20) then becomes through equations (10.21) and (10.22):

$$\overline{\mathbf{x}} = \mathbf{S}\overline{\mathbf{y}} \tag{10.23}$$

which implies:

$$\overline{x}_i = s_i \overline{y}_i \tag{10.24}$$

where $i = 1, ..., 7$. We note that matrix $\overline{\mathbf{R}}$ hides seven fundamental combinations of the outputs. Each of them is obtained multiplying the corresponding combination of incomes by a predetermined scalar which has in fact the role of aggregated macro multiplier. Whatever the amount of the

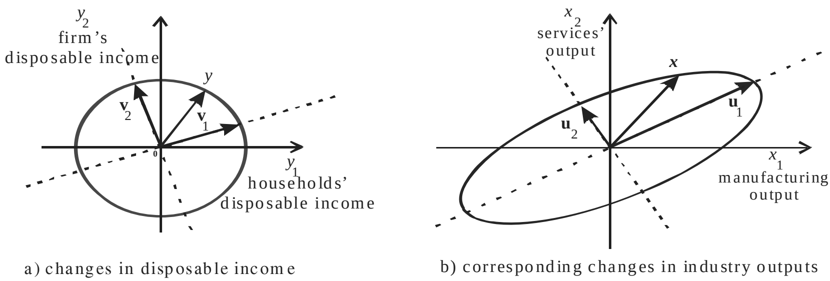


Figure 10.2 Unit income circle and corresponding ellipsoid for disposable income

income change, its multiplicative effect: 1) is given by a combination of macro multipliers; 2) can be equal to one of them, under appropriate conditions and in this case, 3) it can not be greater than the greatest macro multiplier (s_1). No other multiplicative effects are possible.

If we introduce changes in equation (10.19), and interpret vector $\mathbf{y}^0$ as a change from the initial distribution of disposable incomes among sectors, the complex effect on the industry outputs can be reduced to a multiplication by constants s_i.

The structures we have identified play a fundamental role in determining the change of industry outputs corresponding to all the possible changes in income distribution among sectors. We can in fact evaluate the change effect on output of all income structures.

This is easily done by imposing in equation (10.20) a change in income distribution vector whose modulus is kept constant but whose structure can assume all possible configurations. If vector $\mathbf{y}^0$ in equation (10.20) is such that

$$\sqrt{\sum_i (y^0)^2} = 1 \tag{10.25}$$

then geometrically we mean that the income vector describes a sphere of unit radius: the unit ball. It rotates around the origin – as in Figure 10.2(a) for a two-dimensional case – assuming all the possible structures, including those implied by the columns of matrix **V**.

Correspondingly the vector of total output will describe an ellipsoid with semi-axes of length s_1, oriented according to the directions designated by the columns of matrix **U** – as in Figure 10.2(b) for a two-dimensional case.

When the income vector crosses a structure in **V**, the vector of total output crosses the corresponding structure in **U** and the ratio between the moduli of the two vectors is given by the corresponding scalar s_i, such that s_i is the macro multiplier which transforms a change in the *i*-th income distribution configuration into a change in the *i*-th industry output structure.

Table 10.1 shows matrix $\bar{\mathbf{R}}$ elaborated from SAM data (Socci, 2004). The element in the first row and first column equals 0.13 and represents the direct and indirect effect of a unit income increase in the first institutional sector (first household income class) on agricultural output. Two additional rows and columns show totals and quadratic moduli of the row (column).

We can perform the singular value decomposition of data in the table and determine the macro multipliers, which are shown in the first column of Table 10.2. For example, the scalar 6.73 represents the greatest multiplicative effect, detectable on output along the output structure $\mathbf{u}_1$, that is obtained when the income change is given according to the first structure $\mathbf{v}_1$.

Table 10.1 Direct and indirect effects of disposable incomes on industry outputs

	I	II	III	IV	V	VI	VII	Totals	Moduli
x_1 Agriculture	0.13	0.11	0.1	0.1	0.1	0.09	0.06	0.71	0.27
x_2 Oil	0.15	0.15	0.14	0.14	0.15	0.14	0.11	0.97	0.37
x_3 Energy	0.08	0.07	0.06	0.06	0.06	0.05	0.04	0.41	0.16
x_4 Metal & Chem.	0.29	0.36	0.4	0.45	0.51	0.61	0.23	2.85	1.12
x_5 Machinery and Cars	0.26	0.46	0.61	0.77	0.98	1.33	0.27	4.67	2.01
x_6 Food	0.34	0.31	0.28	0.28	0.28	0.25	0.18	1.93	0.74
x_7 Tobacco & Alc. Bever.	0.06	0.06	0.05	0.05	0.05	0.04	0.03	0.35	0.13
x_8 Manufacturing	0.83	1.07	1.23	1.45	1.73	2.16	0.55	9.01	3.66
x_9 Transport & Trade	2.04	1.95	1.86	1.92	1.98	1.95	1.29	13.0	4.95
x_{10} Marketable Services	0.3	0.27	0.25	0.25	0.25	0.22	0.16	1.68	0.65
Non-marketable Serv.	0.4	0.42	0.43	0.47	0.52	0.58	1.49	4.29	1.88
Totals	4.86	5.23	5.43	5.95	6.61	7.43	4.41		
Moduli	2.33	2.39	2.43	2.64	2.93	3.33	2.09		

A generic income change will activate all the macro multipliers. From Table 10.2, column 2, we obtain further information that introduces the possibility of simplifying our analysis. If we determine the cumulated percentage shares of macro multipliers, we see that the first two singular values cover 89 per cent of total output change. This means that we can confine our analysis of intersectoral and interindustry interactions to the first two macro multipliers with a degree of significance (*ds*)

$$ds = \frac{\sum_{i=1}^{2} s_i}{\sum_{j=1}^{7} s_j}$$

Table 10.2 Macro multipliers and cumulative percent sum

	Singular values	Cumulative percent sum
s_1	6.73	74%
s_2	1.34	89%
s_3	0.98	100%
s_4	0.02	100%
s_5	0.00	100%
s_6	0.00	100%
s_7	0.00	100%

and obtain results valid in 89 per cent of the total output change. We know that matrix $\bar{\mathbf{R}}$ can be decomposed into the sum of seven 'impact' components, each determined by a macro multiplier

$$\bar{\mathbf{R}} = s_1\mathbf{u}_1\mathbf{v}_1 + s_2\mathbf{u}_2\mathbf{v}_2 + \ldots + s_7\mathbf{u}_7\mathbf{v}_7 \tag{10.26}$$

Rather than considering matrix $\bar{\mathbf{R}}$, we can refer to the following matrix

$$\bar{\mathbf{R}}^0 = s_1\mathbf{u}_1\mathbf{v}_1 + s_2\mathbf{u}_2\mathbf{v}_2 \tag{10.27}$$

where components greater than two have been neglected with the aim of obtaining 'summary' measures (Basilevsky, 1983). Now the economic interactions are completely determined by the first two aggregated macro multipliers s_1 and s_2.

We note that in matrix $\bar{\mathbf{R}}^0$, vectors

$$s_1\mathbf{u}_1 = \begin{bmatrix} s_1u_{1,1} \\ s_1u_{2,1} \\ s_1u_{3,1} \\ \cdot \\ s_1u_{11,1} \end{bmatrix}, \qquad s_2\mathbf{u}_2 = \begin{bmatrix} s_2u_{1,2} \\ s_2u_{2,2} \\ s_2u_{3,2} \\ \cdot \\ s_2u_{11,2} \end{bmatrix} \tag{10.28}$$

are the result of splitting the two macro multipliers into the eleven output industries. These two vectors quantify how each of the macro multipliers affects outputs.

As we stressed in Section 10.2, the aim of the *sensitivity of dispersion* is to measure the extent to which industries draw upon industry *i* and the degree of importance of each industry as a supplier. As may be seen in Table 10.3, the change in the *i*-th industry output is quantified by vectors [$s_1\ u_{1i}$, $s_2\ u_{2i}$]. The first column shows how the first macro multiplier affects the eleven industries and quantifies the importance of each industry as a supplier with respect to the others. For example, 0.27 represents the change in the first industry output stimulated by the first macro multiplier (6.73). We note that the modulus of the first column is equal to the first macro multiplier (6.73). The same reasoning applies to the second column in relation to the second macro multiplier. It is to be noted that the industry change effect is measured with reference to the two macro multipliers independently from the fact that such multipliers have been activated by a change in disposable incomes. This feature allows for a generalization of the sensitivity of dispersion concept. In order to avoid misinterpretation, we will define forward dispersion, fd_i, as the change in the value of the sales by industry *i* (to face a demand vector generated by an increase in disposable income in all sectors). The generic index is obtained as

$$fd_i = \sqrt{(s_1 u_{1i})^2 + (s_2 u_{2i})^2}$$

It should be noted that the forward dispersion index, being defined as the modulus of two impacts, quantifies the dispersion away from zero. The last column of Table 10.3 shows the percentage forward dispersion that can be easily obtained on dividing forward dispersion by its total value.

The result produces a ranking of key industries according to forward dispersion: Industry x_9 Transport and Trade (31.3 per cent), x_8 Manufacturing (23.2 per cent), x_5 Machinery and Cars (12.6 per cent), x_{11} Non-marketable Services (11.2 per cent), x_4 Metal & Chem. Products (7.1 per cent), x_6 Food (4.6 per cent), x_{10} Marketable Services (4.1 per cent), x_2 Oil (2.4 per cent), x_1 Agriculture (1.7 per cent), x_3 Energy (1.0 per cent), x_7 Tobacco and Alcoholic Beverages (0.8 per cent).

On the other hand, in matrix $\bar{\mathbf{R}}^0$, vectors

$$s_1\mathbf{v}_1 = \left[s_1 v_{1,1}, \ldots, s_1 v_{1,7} \right] \quad s_2\mathbf{v}_2 = \left[s_2 v_{2,1}, \ldots, s_2 v_{2,7} \right] \tag{10.29}$$

split the same two macro multipliers into the seven institutional sectors and represent how the change in sectoral disposable income influences the two macro multipliers.

The aim of the *power of dispersion* is to measure the extent to which a change in disposable income for institutional sector *j* is dispersed throughout the system of industries. The first column of Table 10.4 shows how the first macro multiplier is affected (activated) by the income of the seven institutional sectors and quantifies the relevance of each sector as a purchaser with respect to the others. For example, 2.22 represents the share of the first macro multiplier activated by the first institutional sector. We note that the modulus of the first column, also in this case, is equal to the first macro multiplier (6.73). The second column has the same meaning in relation to the second macro multiplier. This feature allows for a generalization of the power of dispersion concept. In order to avoid misinterpretation, we will define backward dispersion, bd_j, as the change in the value of the purchases by those industries that produce goods according to the disposable income of institutional sector *j*. The generic index is obtained as

$$bd_j = \sqrt{(s_1 v_{1j})^2 + (s_2 u_{2j})^2}$$

Backward dispersion can also be determined in percentage terms. The last column of Table 10.4 shows that the percentage backward dispersion can be easily obtained by dividing backward dispersion by its total value. The result produces a ranking of key institutional sectors according to backward

Table 10.3 Forward dispersion, that is impacts on industry outputs of intersectoral interactions, in terms of macro multipliers

	First impact component $u_1 s_1$	Second impact component $u_2 s_2$	Forward dispersion (Modules) $fd_i = \sqrt{(s_1 u_{1i})^2 + (s_2 u_{2i})^2}$	Percent forward dispersion
x_1	0.27	0.03	0.27	1.7
x_2	0.37	0.04	0.37	2.4
x_3	0.15	0.02	0.15	1.0
x_4	1.12	–0.11	1.12	7.1
x_5	1.9	–0.55	1.98	12.6
x_6	0.72	0.08	0.73	4.6
x_7	0.13	0.02	0.13	0.8
x_8	3.59	–0.67	3.65	23.2
x_9	4.91	0.42	4.93	31.3
x_{10}	0.63	0.08	0.64	4.1
x_{11}	1.5	0.92	1.76	11.2
Modules	6.73	1.34		

Table 10.4 Backward dispersion, that is impacts of a unit disposable income shock on economic interactions, in terms of macro multipliers

	First impact component $\mathbf{v}_1 s_1$	Second impact component $\mathbf{v}_2 s_2$	Backward dispersion (Modules) $bd_j = \sqrt{(s_1 v_{1j})^2 + (s_2 u_{2j})^2}$	Percent backward dispersion
I – Income class households	2.22	0.41	2.26	12.6
II – Income class households	2.35	0.19	2.36	13.2
III – Income class households	2.42	0.02	2.42	13.5
IV – Income class households	2.64	–0.11	2.64	14.7
V – Income class households	2.91	–0.3	2.93	16.3
VI – Firms	3.24	–0.63	3.3	18.4
VII – Administration	1.73	1.05	2.02	11.3
Modules	6.73	1.34		

dispersion: institutional sector firms VI (18.4 per cent), Income class V households (16.3 per cent), Income class IV households (14.7 per cent), Income class III households (13.5 per cent), Income class II households (13.2

per cent), Income class I households (12.6 per cent) and Administration VII (16.3 per cent). Forward and backward dispersion may be graphically represented starting from impact components. We will define the axis of the first macro multiplier, on which we measure the elements of vectors $s_1\mathbf{u}_1$, $s_1\mathbf{v}_1$ and the axes of the second macro multiplier, where we measure the elements of vectors $s_2\mathbf{u}_2$, $s_2\mathbf{v}_2$. Then we will represent the pairs $(s_1 v_{1,i},\ s_2 v_{1,i})$ $i = 1, ..., 7$, with seven arrows, showing how the change in disposable income impacts on intersectoral interactions in terms of the two macro multipliers. The backward dispersion is geometrically given by the length of each arrow. We will then represent the pairs $(s_1 u_{1,i},\ s_2 u_{1,i})$ $i = 1, ..., 11$, with eleven dots, showing how intersectoral interactions impact on industry outputs. The forward dispersion is given by the distance of each dot from the origin.

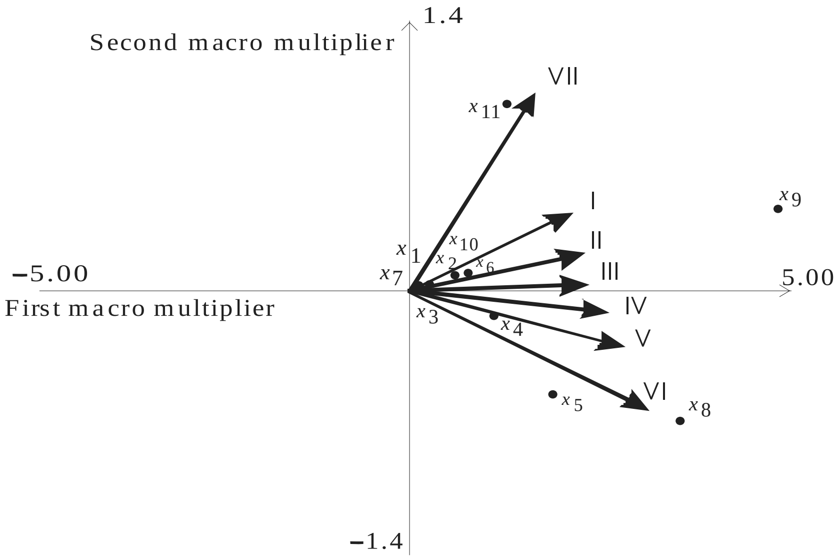


Figure 10.3 Sector and industry interactions – backward and forward dispersions (absolute levels)

This representation allows for one further extension of the analysis in the statistical sense. In order to perform consistent comparisons let us standardize data in Table 10.1 taking the deviations from the mean values and dividing by the standard deviations. This will produce the matrix $\tilde{\mathbf{R}}$ of standardized data. We note that the singular value decomposition of matrix $\tilde{\mathbf{R}}$ will result in the eigenvalue decomposition of matrices $\tilde{\mathbf{R}}^T\tilde{\mathbf{R}}$ and $\tilde{\mathbf{R}}\tilde{\mathbf{R}}^T$

which represent the correlation matrices of sectoral incomes and industry outputs respectively. We will then get the diagram in Figure 10.4.

Figure 10.4 identifies clusters of industries that move together, that is respond linearly, to intersectoral interactions as quantified by the two macro multipliers. The angular distance of two dots will represent the correlation coefficient since:

$$Corr(\mathbf{x}_i, \mathbf{x}_j) = \cos\beta = \frac{\mathbf{x}_i \cdot \mathbf{x}_j}{|\mathbf{x}_i||\mathbf{x}_j|}$$

Indeed, if two industries are located on the same line in Figure 10.4 they will be stimulated in the same proportion by the two macro multipliers.

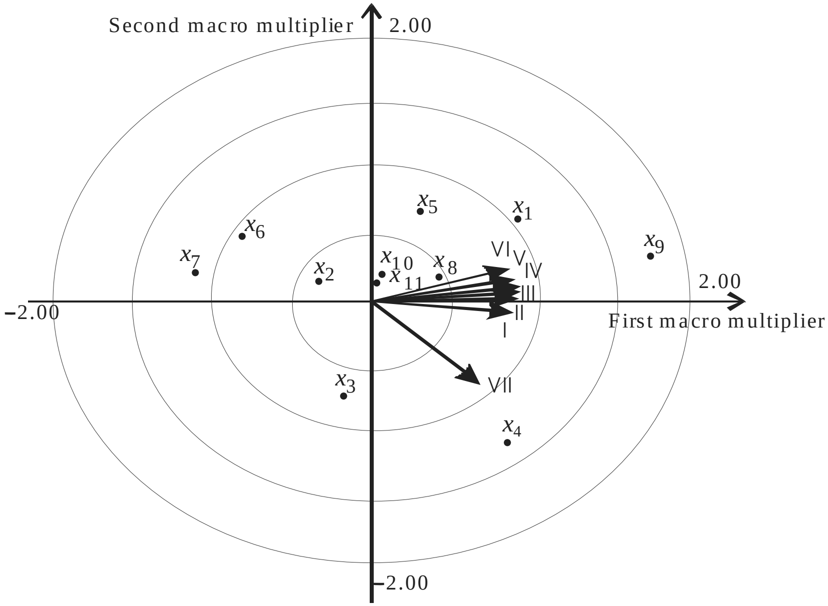


Figure 10.4 Sector and industry interactions – Backward and forward dispersions standardized

From data in Figure 10.4, a correlation table is derived as shown in Table 10.5. For correlation coefficients greater than 90 per cent we can identify a set of six industry clusters: 1st cluster: positive correlation characterizes industry x_1 (Agriculture) with respect to x_8 (Manufacturing) and x_9 (Transport and Trade); 2nd cluster: positive correlation between industry x_2

Table 10.5 Correlation coefficients between industries

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
x_1	1										
x_2	–0.57	1									
x_3	–0.72	–0.16	1								
x_4	0.1	–0.88	0.62	1							
x_5	0.85	–0.06	–0.98	–0.43	1						
x_6	–0.45	0.99	–0.3	–0.93	0.09	1					
x_7	–0.73	0.98	0.05	–0.75	–0.27	0.94	1				
x_8	0.99	–0.7	–0.59	0.27	0.75	–0.59	–0.83	1			
x_9	0.92	–0.84	–0.39	0.48	0.58	–0.76	–0.94	0.97	1		
x_{10}	0.78	0.06	–1	–0.54	0.99	0.2	–0.15	0.67	0.48	1	
x_{11}	0.66	0.24	–1	–0.68	0.95	0.38	0.03	0.52	0.31	0.98	1

(Oil) and x_6 (Food), x_7 (Tobacco & Alcoholic Beverages): 3rd cluster: negative correlation is observed between industry x_3 (Energy), and x_5 (Machinery and Cars), x_{10} (Service market), x_{11} (Service non-market); 4th cluster: negative correlation between industry x_4 (Metal and chemical Products) and x_6 (Food); 5th cluster: negative correlation between industry x_7 (Tobacco and Alcoholic Beverages) and x_9 (Transport and Trade); 6th cluster: negative correlation between industry x_7 (Tobacco and Alcoholic Beverages), x_9 (Transport and Trade). Four other clusters are implied by the previous as x_5, x_{10} and x_{11}; x_6 and x_7; x_8 with x_9, x_{10} and x_{11}.

As far as backward dispersion is concerned, the modulus of each vector labelled I, II, III, IV, V, VI, VII, represents the stimulus forwarded to the interindustry interactions by a unit change in disposable income by institutional sector. From Figure 10.4 we note that in our example the effects of disposable incomes of institutional sectors from I to VI are highly correlated, more than 90 per cent in terms of the correlation coefficient. Only sector VII, Administration, seems to exhibit a different pattern. Its correlation with the other sectors decreases from 80 per cent with sector I to 53 per cent with sector VI.

Figure 10.4, in addition, allows for a sector/industry cross comparison which can identify the 'strength' of the link between sectors and industries in terms of cross correlation coefficients.

Table 10.6 shows high positive correlations between sector I and industry x_9; sector II and industries x_8 and x_9; sector III and industries x_8 and x_9; sector IV and industries x_8 and x_9; sector V and industries x_1, x_8 and x_9; sector VI and industries x_1, x_8 and x_9; sector VII and industries x_4.

Table 10.6 Cross-correlation coefficients between industries and sectors

	I	II	III	IV	V	VI	VII
x_1	0.78	0.83	0.87	0.89	0.92	0.95	0.24
x_2	–0.96	–0.93	–0.9	–0.88	–0.85	–0.8	–0.93
x_3	–0.12	–0.22	–0.28	–0.33	–0.38	–0.46	0.5
x_4	0.7	0.63	0.58	0.54	0.49	0.41	0.99
x_5	0.33	0.42	0.48	0.53	0.57	0.65	–0.31
x_6	–0.91	–0.87	–0.83	–0.8	–0.77	–0.71	–0.97
x_7	–1	–0.99	–0.97	–0.96	–0.94	–0.91	–0.84
x_8	0.87	0.92	0.94	0.96	0.97	0.99	0.4
x_9	0.96	0.98	0.99	1.0	1.0	1.0	0.6
x_{10}	0.22	0.31	0.38	0.42	0.47	0.55	–0.42
x_{11}	0.04	0.13	0.2	0.25	0.3	0.39	–0.58

While high negative correlations are observed between sector I and industries x_6 and x_7; sector II and industries x_2 and x_7; sector III and industries x_2 and x_7; sector IV and industry x_7; sector V and industry x_7; sector VI and industry x_7 Tobacco and Alcoholic Beverages; sector VII and industries x_2 and x_6.

Among sectoral and industrial interactions these emerge as the strongest. Here it seems that sectoral disposable income has a direct influence on industrial output. In these cases the structure of backward dispersion is the same as that of forward dispersion since the sectoral disposable-income change activates the macro multipliers in the same combination in which industrial outputs are stimulated.

11.5. CONCLUSIONS

Income inequality comes from production, where value added is distributed to factors, develops through the interactions among institutional sectors in the primary and secondary distribution of incomes and, finally, feeds back into the production process through disposable income which determines the consequent final demand. On the other hand, the long-established measures of dispersion can be applied both to a traditional Leontief framework and to an enlarged model, where income distribution may also be taken into consideration. Propagation phenomena through industries were studied to find 'summary' measures of dispersion and apply them statistically to interindustry data. However, in later developments, the original statistical approach has been progressively abandoned and the interpretation of these measures has definitely become deterministic. Moreover all these measures are built under the assumption of few unlikely structures of the macro variable changes while we show that all possible compositions have to be investigated. We have sought to take inspiration from some of these developments to design new measures of dispersion, both 'summary' and 'statistical', based on all the possible structures of a change in policy variables. The results have been discussed on the basis of a specific regional model whose data base we have tried to render consistent, having in mind a social accounting scheme. The emerging enlarged income flow has been analysed, identifying the macro multipliers that 'govern' the flow, through the singular value decomposition of the structural matrix. Having identified these multipliers that represent the potential scale of all the possible types of dispersions through industries and sectors, we evaluated both backward and forward dispersions with reference to them. This procedure generates a set of indices – in absolute and percent values – for industry forward dispersion and sector backward dispersion which quantify, respectively, the change in the value of the sales by an industry to meet a demand vector generated by an increase in disposable income in all sectors, and the change in the value of the purchases by those industries that produce goods according to the consumption patterns of an income sector. An extension of the method has also been provided in terms of 'summary' graphical representation. The standardization of data, in fact, produces a representation, explainable in terms of correlation analysis, which allows for an immediate interpretation of the strength of the mutual links among and between the disaggregated components of total output and disposable income. A synthetic picture of the working of sector–industry interactions is then attained in graphical and quantitative terms.

NOTES

* The authors are indebted to Neri Salvatori and two anonymous referees for helpful suggestions.

1. Wage and salaries, other income and indirect tax.
2. Institutional sectors are: household income class I, household income class II, household income class III, household income class IV, household income class V, firms and public administration.

REFERENCES

Basilevsky, A. (1983), *Applied Matrix Algebra in the Statistical Sciences*, New York, Oxford and Amsterdam: North Holland.

Bulmer-Thomas, V. (1982), *Input–Output Analysis in Developing Countries*, USA: John Wiley and Sons Ltd.

Ciaschini, M. (1988), *Input–Output Analysis Current Developments*, London: Chapman and Hall.

Ciaschini, M. (1989), 'Scale and structure in economic modelling', *Economic Modelling*, **6**(4):355–73.

Ciaschini, M. (1993), *Modelling the Structure of the Economy*, London: Chapman and Hall.

Lancaster, P. and M. Tiesmenetsky (1985), *The Theory of Matrices*, New York: Academic Press, 2nd edn.

Leontief, W. (1988), 'Foreword', in M. Ciaschini (ed.), *Input Output Analysis Current Developments*, London: Chapman and Hall.

Miyazawa, K. (1970), 'Input–output analysis and structure of income distribution', *Notes in Economics and Mathematical Systems*, 116.

Pyatt, G. (2001), 'Some early multiplier models of the relationship between income distribution and production structure', *Economic Systems Research*, **13**(2):139–64.

Pyatt, G. and J. Round (1979), 'Accounting and fixed price multipliers in a social accounting matrix framework', *Economic Journal*, **89**:850–73.

Rasmussen, P. (1956), *Studies in Intersectoral Relations*, Amsterdam: North Holland.

Round, J. (2003), 'Social accounting matrices and sam-based multiplier analysis', in L.A.P. da Silva and F. Bourguinon (eds), *Techniques for Evaluating the Poverty Impact of Economic Policies*, ch. 14, World Bank and Oxford University Press.

Skolka, J. (1986), *Input–Output Multipliers and Linkages*, paper presented at the 8th International Conference on Input-Output Techniques, Sapporo.

Socci, C. (2004), 'Una sam biregionale per le Marche', in A. Fossati and R. Targetti (eds), *Matrici regionali di contabilità sociale e analisi di politiche economiche: il caso della Liguria, Toscana e Marche*, Milan: Franco Angeli.

11. A dynamic AGE model from a classical–Keynesian–Schumpeterian approach

Oscar De-Juan

11.1. INTRODUCTION

This chapter lays the foundations of a dynamic and computable 'applied general equilibrium model' (AGE), useful for policy evaluation and growth analysis. Its suitability for analysing economic growth makes a difference with the usual AGE model which focuses on comparative statics from a neoclassical standpoint.[1]

Our dynamic AGE model is rooted in classical, Keynesian and Schumpeterian traditions (Clakesch, for short). Classical political economy was concerned with the analysis of the processes of production, distribution, consumption and accumulation (which brings about economic growth). The social accounting matrix of Section 11.2 is a fair reflection of this scheme, where distribution (and redistribution) of income plays a crucial role. Our AGE model purports to explain one of the possible systems of prices and quantities embedded in social accounting matrices (SAM) and input–output tables (IOT), describing an economy during a given period. Reading the first block vertically we can get the Sraffian prices of production, or an equivalent, when physical units are unknown (Section 11.3). Reading horizontally we can obtain the level and composition of output as a multiple of autonomous demand (Section 11.4). This is nothing but Keynes's principle of effective demand rewritten as a fully disaggregated multiplier–accelerator mechanism. The driving force of the system (what ultimately explains growth) lies in the introduction of new products and new methods of production by innovative entrepreneurs. This is the Schumpeterian contribution expounded in Section 11.5. To link all these traditions in a coherent and useful way might be the first achievement of this chapter. A second is the provision of an alternative to the standard AGE model. In the last section we summarize the differences between our Clakesch–AGE model and the neoclassical one.

11.2. A SOCIAL ACCOUNTING MATRIX FOR A GROWING ECONOMY

A SAM reflects the transactions between activities, factors and institutions. The advantage of SAM over national accounts and IOT, is that the former considers a variety of institutions, and presents in full detail not only production activities but also distribution, redistribution and final expenditures. The basic design of a SAM proposed by the United Nations Statistics Division (1993) fits with our main purpose perfectly, that is the analysis of a growing economy. Table 11.1 illustrates the SAM from which our theoretical and empirical analysis will be derived.

Our starting point is a 'symmetrical' input–output table, where n homogeneous *industries* represented in each column are producing n goods represented in the rows.[2] Joint production is possible but has already been removed from the table, using any of the accepted methods.

The value added in the production process is divided into wages and profits (operating surplus) and assigned to the *factors of production*: labour and fixed capital.[3] In the *satellite accounts*, that may accompany the SAM, it would be convenient to gather information about the different types of labour and different capital goods allocated in each industry. The fact that such information is not generally available is no excuse for ignoring it, but an opportunity to urge its production.

Among the *institutions* we include households (H), government (G), and corporations or enterprises (E, which encompasses both financial and non-financial corporations). The 'rest of the world' (RW) deserves special treatment. Each institution can be disaggregated according to the particular interests of the researcher. For our purposes it is useful to separate *households* according to the main source of income or the level of income. H1 would stand for non-qualified labour; H2 for qualified labour; H3 for managers; H4 for pensioners, and so on. *Enterprises* could be separated by industries (1 to n) to facilitate the analysis of accumulation.[4]

The flows among industries, factors and institutions are classified into four accounts (plus the *balance of payments*).

Production account. The traditional input–output table is included in the first block of Table 11.1. Rows register the proceeds from the sale of outputs; columns, expenditure derived from purchase of inputs. 'Value added' (*VA*), that is the payment of primary incomes to the factors of production, sets the balance.

Income or current account. It is divided into two subsets (blocks 2 and 3 of Table 11.1). Rows in the second block show the *primary distribution of*

Table 11.1 SAM

		1 Production Purchase of inputs	2 Income Secondary distribution	3 Consumption Use of income	4 Accumulation investment	6 Rest World (+ , -)	Total
		1, 2, ..., n	L(W), K(B)	H; G; E	H; G; E	RW	
1. Production Sale of output	1 2 … *n*	Intermediate Consumption Λ		Final Consumption C	Final Investment I	+ Exports – Imports X'	Total output T1
2. Income Primary Distribution	L(W) K(B)	Value Added VA					T2
3. Consumption Disposable income	H G E		Income to institutions Y Disposable income, Y_d	Current transfers (includes d. taxes) T		$Y_{(rw)} + T_{(rw)}$	T3
4. Accumulation Savings	H G E			Savings S	Capital Transfers Tk	$Tk_{(rw)}$	T4
5. Finance (+ , -)	F				Net Lending lending (+);borrowing (-) FF	$FF_{(rw)}$	T5 (NL)
Total		Total inputs, T1	T2	T3	T4	T6 (BP)	

income into wages and profits (operating surplus). A vertical reading shows the allocation of VA among institutions. A portion of the operating surplus is retained in corporations as 'reserves', the rest is distributed as 'property incomes'. Wages, apart from social contributions to government, reach households. A horizontal reading of the third block (after redistribution among institutions has taken place) shows the disposable income of any institution (Y_d). Reading the third block vertically we observe the use of disposable income between final consumption (C) and savings (S).

Accumulation or capital account. Savings plus capital transfers (Tk) allow institutions to finance their investments in a variety of capital goods as expressed in Table 11.1 of the SAM. One of the advantages of classifying corporations according to the industry they belong to is that the resulting SAM visualizes the investment pattern of each industry (that is the structure of sectoral investment by capital goods).

Finance account. It gives information about the flows of funds (FF) from creditors to debtors (usually intermediated by banks). To simplify the design of Table 11.1, we have grouped these flows into a single row with positive figures for lending and negative figures for borrowing. Each row of the finance account is identified with the traditional financial assets: cash and bank deposits, bills, bonds, equities and bank loans. The financial account could be completed with a balance of outstanding financial assets (as part of the satellite accounts). This would constitute a flow-fund financial subset in the post-Keynesian tradition (Godley, 2004).

Balance of payments. Transactions with the rest of the world (current, capital and finance accounts) are gathered together in the last vertical block. The inflow of foreign currency (exports of goods and services, for instance) bears a positive sign. The outflow of foreign currency (associated with imports, purchase of foreign financial assets and so on) is treated as negative. Note that in the last two accounts we have moved away from the usual double accounting method: inflows in the row and outflows in the corresponding column. Apart from being a device to simplify Table 11.1, this has several advantages: it allows the reader to visualize the balance of payments (BP) and it shows how the current account surplus matches the net lending (NL) of the national economy.

A SAM is an accounting technique that presents economic flows in a meaningful way. To make sense of such flows we have to postulate functional relationships explaining intermediate consumption, final consumption, investment, exports and so on. AGE models rely strongly on *calibration* to explain the coefficients of such functions. Under our technological assumptions, if industry 1 needs 4000 units of capital and 100 units of labour to produce 1000 units of output, it is supposed that the 'optimal' capital coefficient is 4, and the optimal labour coefficient is 0.1. Obviously this is not always true. The data may have corresponded to a recession period in which firms were operating with excess capacity and excess hired labour, that could not be fired at once. Such a risk urges us to examine the data carefully and complete calibration with alternative techniques, including econometrics.

11.3. THE PRICE SYSTEM AND DISTRIBUTION IN A CAPITALIST ECONOMY

As a long-run tendency, competition forces firms to introduce the best available techniques, to use capacity at the optimal level and to adjust prices to production costs (which includes a 'normal' rate of profit on the capital advanced). Sraffa (1960) built the system of equations leading to such prices. He proved that, for a given technology and distribution, there is a unique vector of relative production-prices. Before expounding the price equations let us comment on the 'givens', that is on technology and distribution.

11.3.1. Technology

Classical or Leontievian production functions are consistent with the first vertical block of the SAM that corresponds to the columns of a symmetrical input-output table (IOT). Technical coefficients are fixed, which imply constant returns to scale and no input substitution. Entrepreneurs are free to choose among different techniques, but, once the choice has been made, they cannot combine inputs and factors of production at will. In the short run, however, entrepreneurs may change the degree of capacity utilization (capital–output ratio) in order to adjust to demand fluctuations. But using capital more hours a day implies hiring extra labour-time, so the degree of mechanization (capital–labour ratio) remains constant.

Technology is materialized in the following sets of data.

A matrix of technical coefficients: $A = \Lambda\,\hat{q}^{-1}$, Λ being a square matrix for intermediate consumptions or inter-industry transactions, and q, the column vector of the total value produced by each industry (here presented as a diagonal matrix). The result is a square matrix $n \times n$, n being the number of homogeneous industries. We should separate domestic from imported intermediate consumptions (Λ_d, Λ_m) and compute two different matrices of technical coefficients (A_d, A_m). Let $\hat{\theta}$ be a diagonal matrix indicating the percentage of each good that is imported. These percentages reflect price elasticity of imports and are bound to change with the ratio 'domestic price to international price', both expressed in euros. Tariffs and the nominal exchange rate play a role here. Our previous matrix A should be segmented into two: $A_m = A\hat{\theta}$ and $A_d = A\langle II - \hat{\theta}\rangle$.[5]

A rectangular matrix of labour coefficients: $l = L\hat{q}^{-1}$. Matrix L is part of the satellite accounts; it has as many columns as industries and as many rows as types of labour: non-qualified labour, qualified labour, managers … l will have the same dimensions. It is an inverse measure of sectoral labour productivity.

A square matrix of capital coefficients: $k = K\hat{q}^{-1}$. K is the fixed capital matrix with a column for each industry and a row for each good, although only the rows containing capital goods will have positive figures. k has the same dimensions and content as K but refers to a unit of production. The figures are supposed to reflect the *normal* or *desired* capital–output coefficients.

11.3.2. Distribution

The real *wage* can be presented as a fraction of labour productivity. Workers consider the real wage achieved in the past as a social conquest and try to improve it, absorbing productivity increases. Historically, both variables have increased *pari passu.* In the yearly agreements, trade unions will also press for wage increases to catch up with inflation and to take advantage of the ongoing tensions in the labour market. Usually, such claims (not justified by productivity improvements) result in wage inflation. Equation (11.1) summarizes the forces causing nominal wage increases:

$$\Delta w = f(\hat{\pi}) + \tau + f'(\hat{\varepsilon}) \qquad (11.1)$$

w refers to the nominal wage for the basic labour category (let's say, 'non-qualified labour'). Other types of labour will earn a multiple of w. To obtain the 'real wage' we divide by a price index. $\hat{\pi}$ stands for productivity increases; τ, for the expected rate of inflation, which nowadays can be proxied by the one targeted by central banks; $\hat{\varepsilon}$, for the deviations of the employment rate over its historical-conventional value.[6]

In classical political economy, *profits* appear as an 'operating surplus' belonging to the owners of capital. The role of prices of production is to distribute this surplus among industries in such a way that the 'representative' or 'regulating' firm of each industry gets the same rate of profit (r) on the capital advanced.[7] Abstracting from short-term deviations from normal capacity utilization, the 'regulating' rate of profit has been quite stable through decades. This fact allows us to take the rate of profit as a datum in most of the applications of our AGE model.[8]

11.3.3. Prices of Production and Input–Output Prices

A vertical reading of the coefficients of the first block of a SAM allows us to obtain the Sraffian system of *prices of production* enforced by competition. 'Competition' is used in the classical sense that simply implies the free flow of savings towards the industries yielding a rate of profit above average.

To begin with, let us assume that we know the quantities of inputs and factors employed in each industry and are able to compute A_d, A_m, l, k in physical units. The price of production of any commodity would be the result of adding up the following 'unit costs'. (1) Value of intermediate domestic commodities: pA_d; (2) Value of intermediate imported commodities: $p_m A_m$ (p_m being the international price in euros plus tariffs); (3) Unit labour cost, $wl = w/\pi$. (4) Unit profit as r times the value of fixed capital invested (rpk).

$$p = pA_d + p_m A_m + wl + r(pk) \tag{11.2a}$$

Alternatively we can compute relative prices as a multiple of the unit costs of 'non-produced inputs'. The 'multiplier' would be a modified Leontief's inverse matrix.

$$p = (wl + p_m A_m)[II - A_d - rk]^{-1} \tag{11.2b}$$

In the preceding equations we have n unknowns $(p_1, p_2, \ldots, p_n)$ and n equations (one for each industry). The two distributive variables (r, w) can be taken as given for most of our purposes. The rate of profit (r) is fixed at its historical level (although we are free to move it at any moment). The nominal wage (w) (and wage dispersion), is taken as given, but it is supposed to change yearly. Dividing the (row) vector of prices by w we get 'labour commanded prices', that is the hours of basic labour that can be hired selling one unit of q_1, $q_2 \ldots$. It is a way to separate absolute (nominal) price increases from relative (real) price movements. Notice, however, that changes in nominal prices may have 'real' effects if they alter import propensities. (p_m, A_m, is included among the data and revised yearly.)

Unfortunately national accounts do not show physical units so we are obliged to find prices in a different, less transparent way. Instead of dividing each column j by the physical output (q_j) we divide by the money value of output $(p_j q_j)$. The new coefficients will add up to one in each column, implying that all *input–output prices* $(p'_j q'_j)$ are always unity.

$$p' = \alpha + \mu + \omega + \beta = p'A'_d + p'_m A'_m + w'l' + rp'k' = [1,1\ldots1] \tag{11.3a}$$

$$p' = (w'l' + p'_m A'_m)[II - A'_d - rk']^{-1} = [1,1\ldots1] \tag{11.3b}$$

$\alpha = p'A'_d$ stands for the share of intermediate inputs, A'_d being an undefined quantity or intermediate inputs whose prices (p') equal one. $\mu = p'_m A'_m$ stands for the share of intermediate imports in the value of output. $\omega = w'l'$ stands for the share of wages, l' being an undefined quantity of labour whose initial price (w') can also be fixed equal to one. $\beta = rp'k'$ is the share of profits, k' being an undefined quantity of capital goods whose prices (p') equal one, as any other commodity price. r is the rate of profit that should be uniform across industries and similar to the one encountered in (11.2a) and (11.2b).

Despite such an odd result (all the input–output prices being unity), the model is useful to compute the impact (on the relative prices of the same physical units) of a change in wages, tariffs, productivity and so on. We should be careful, however, with the way we represent the shocks and the transmission mechanism. The traditional presentations of input–output prices and neoclassical AGE prices do not compute 'profits' as r times the value of the commodities used as 'capital'. As a consequence, the new prices resulting after the 'shock' do not warrant a uniform rate of profit on capital invested.

11.3.4. Market Prices and Demand Fluctuations

Prices of production consider only supply forces. Market prices are supposed to reflect both supply and demand. In principle, excesses in demand will push prices up. Nevertheless, this is a transient phenomenon since higher prices and profits in industry j will attract investment and production will rise cancelling out the excess of demand in the output of j. After the adjustment of quantities, relative prices will return to the long-run equilibrium determined by production costs.

This is the theoretical scheme. In practice only a handful of primary products (oil and raw materials, in particular) are sensitive to demand, as post-Keynesians have repeatedly shown after Kalecki (1971) and Sylos-Labini (1957). Such prices are determined abroad and are taken as data in our model. In an advanced industrial economy, the bulk of industries is prepared to accommodate demand shocks by piling inventories and adjusting capacity utilization. In services there is no such possibility, but the risk of losing customers by continuous changes in prices has convinced entrepreneurs to maintain prices in their long-run equilibrium, determined by costs of production.[9] We can take it for granted – a key conclusion for our purpose– that in an advanced industrial economy relative prices are rarely influenced by the ordinary ups and downs of demand.

11.4. THE QUANTITY SYSTEM IN A DEMAND-CONSTRAINED SYSTEM

The quantity system may be explained either from the supply side or from the demand side. A *supply-led* quantity system was implicit in the classical equations and in von Neumann's 'general equilibrium model' (Kurz and Salvadori, 1998; Nell, 1998, 2004; von Neumann 1945–46). Both approaches provide useful hints for understanding certain equilibrium conditions and certain technological limits. But they do not adequately describe the working of a capitalist economy. The same can be said about neoclassical economics in general and neoclassical AGE models in particular that continue to rely on Say's law: supply creates its own demand, saving causes investment. What we find more compelling is the description of capitalism as a *demand-constrained system* (Keynes, 1936; Kalecki, 1971; Kornai, 1979). According to the *principle of effective demand,* the equilibrium level of output at any moment does not depend on the productive capabilities of the economy but on expected demand at normal prices. More precisely, it can be expressed as a multiple of the autonomous demand expected for the period under consideration. In the simplest Keynesian model, the 'multiplicand' (autonomous demand) is identified with investment, and the multiplier with the inverse of the propensity to save. Following an increase in investment expenditure, output will grow until the savings stemming from the new incomes match the new investments: $\Delta S = \Delta I$.

In this section we are going to extrapolate the principle of effective demand into a multisectoral growing economy. Our first task is to separate autonomous demand from induced demand. In the second step we will endogenize the bulk of consumption and investment to obtain a 'super-multiplier'.[10] Autonomous demand is independent of income. The main components (to be analysed in Section 11.5) are: exports, real government expenditure and 'modernization investment'. To provide for the expected increases in autonomous demand, firms are supposed to purchase intermediate goods, to hire labourers (who will consume a significant portion of their accruing incomes) and to buy new capital goods in order to match the expected increase in demand efficiently. Production will adjust to aggregate demand, autonomous plus induced. Output will rise until the new 'uncommitted incomes' (Σ = incomes not devoted to induced consumption or expansionary investment) match the value of autonomous demand (Z). Table 11.2 illustrates the process of separation of autonomous from induced demand to emphasize that it is Z which determines Σ.

Table 11.2 A compact SAM

		Induced demand 1, 2, ..., n	Autonomous demand
Induced incomes	1 2 . . . n	Λ^* Intermediate consumption Induced final consumption Induced final investment	Z Exports Government real expenditure Modernization investment
Uncommitted incomes		Λ $(Z \rightarrow \Lambda)$	

11.4.1. Induced Consumption

In *The General Theory*, Keynes assumed that the bulk of private consumption was a percentage (rather high and stable) of household disposable income. This hypothesis was verified in the 1930s and has been confirmed ever since. Kalecki contributed to the debate suggesting that the aggregate propensity to consume was a weighted average of the propensities of different income groups. Our SAM allows us to represent a variety of social groups, each with a particular propensity to consume and a particular consumption basket. De-Juan, Cadarso and Córcoles (1994) explain the process of endogenizing final consumption.

The first step consists in showing how the income generated in the n industries is eventually distributed and redistributed among the h institutions.

$$[Y_d]_{hn} = [t]_{hh} [y]_{hf} [VA]_{fn} \tag{11.4}$$

[VA] gathers the primary incomes, that is payments to factors of production. It is an $f \times n$ matrix (f = factors; n = industries). [y] is an $h \times f$ matrix (h for institutions, in particular, household groups). The first column shows the shares of each household group in total wages. The second column shows the portion of profit retained in firms, and the portion distributed to households. The product [y] [VA] informs about the allocation of value added to institutions. Premultiplying this result by [t] we obtain the disposable income of institutions. [t] is an $h \times h$ matrix, whose columns add up to one. It gathers income tax rates and the share of transfers in government receipts.[11]

Pre-multiplying $[Y_d]$ by $\langle PC\rangle$ we obtain the incomes that are systematically consumed. In the diagonal of $\langle PC\rangle$ we find the consumption propensities of the different households. (Other institutions are represented but their final induced consumption is nil). $[DC]$ indicates the distribution of consumption expenditure among goods.[12] By construction, any column of $[DC]$ adds up to 1. To obtain *domestic* induced consumption $[C_{i,d}]$ we have to subtract the portion of consumption goods imported from abroad.

$$
\begin{aligned}
\left[C_i\right]_{nn} &= \left[CD\right]_{nh} \left\langle CP\right\rangle_{hh} \left[Y_d\right]_{hn} \\
\left[C_{i,d}\right]_{nn} &= \left\langle II - \hat{\theta}\right\rangle_{nn} \left[C_i\right]_{nn}
\end{aligned}
\qquad (11.5)
$$

Our model is ready to introduce the influence of prices in the allocation of consumption among different goods or the influence of interest rates on the consumption (and saving) propensities. We are not going to do so because empirically these new variables add very little to the explanation of consumption. The Cambridge multisectoral model has shown that linear expenditure functions, similar to those we have used here, explain consumption better than any other (Barker and Peterson, 1987, following Stone's suggestions, 1981). Changes in prices might affect the substitution in consumption between, say, two different types of meat, but not between food and clothing, which is the aggregation level we are considering in a SAM.

11.4.2. Induced or Expansionary Investment

Firms undertaking gross investment have three purposes in mind: (1) replacement of used capacity by means of *fixed capital consumption*; (2) expansion of productive capacity to match efficiently the expected increases in the demand of traditional goods; (3) transformation of productive capacity in order to introduce new processes of production or to produce new commodities. Here we are going to focus on the second category, the so-called *expansionary net investment* (I_i). In the next section we shall deal with the third one, *modernization investment* (Z).[13]

First of all we should clarify that the kind of inducement of 'expansionary investment' is quite different from induced consumption. The 'propensity to consume' explains the increase in final consumption associated to an increase in disposable income. Rising income also means additional savings. But it would be incorrect to link increasing savings to increasing investment via a *propensity to invest*. For Keynesians the causality runs the other way round. Firms expecting higher demand for their products decide to expand capacity, that is to invest in equipment. The production of this equipment will raise incomes and savings. Part of the new savings (whether retained by firms or borrowed in the financial markets) can be 'associated' to expansionary investment.

Expansionary investment is explained by the *acceleration principle.* Though a well-known macroeconomic tool, it has rarely been applied to multisectoral models.[14] An acceleration investment function could be captured by the following $n \times n$ matrices

$$\begin{aligned} I_i &= [k]\langle g\rangle\langle q\rangle \\ I_{i,d} &= \left\langle II - \hat{\theta}\right\rangle I_i \end{aligned} \tag{11.6}$$

Investment decisions (I_i) are taken at the end of the period of production. In the diagonal of $\langle q\rangle$ we find the current levels of production of the different industries. $\langle g\rangle$ is a diagonal matrix gathering the expected rate of growth in each industry. To simplify the model it would be better to suppose that g_i refers to the rate of sectoral growth in the recent past, and allow for an adjustment afterwards (in Z). As we already know, $[k]$ is a square matrix of *normal* capital–output ratios by industries and goods. To obtain induced *domestic* investment $[I_{i,d}]$ we should subtract imports of capital goods, according to the information provided by $[\theta]$.

11.4.3. The Structural Multiplier

We are now ready to compute the multiplier that links all types of induced demand. First we obtain the enlarged inter-industry transaction table (Λ^*), adding up the tables of intermediate consumptions (Λ_d), final induced consumption (C_{id}) and final induced investment (I_{id}). Second, we divide the cells of each column by the total output of the industry to obtain the enlarged matrix of coefficients (A_d^*). Then we compute a Leontief's inverse matrix to obtain the structural multiplier of total output (MQ).

$$\Lambda_d^* = \Lambda_d + C_{i,d} + I_{i,d} \tag{11.7}$$

$$A_d^* = \Lambda_d^* \hat{q}^{-1} \tag{11.8}$$

$$MQ = \left[I - A_d^*\right]^{-1} \tag{11.9}$$

Economists are generally less interested in total output (that involves the problem of double counting of intermediate goods) than in final output or value added. They are most of all interested in employment. Table 11.3 explains how to obtain the corresponding super-multipliers of income (value added) (MV) and labour (ML).

Table 11.3 Super-multipliers

Multiplier of total output	$MQ = \left[I - A_d^* \right]^{-1}$	$\Lambda_d \rightarrow \Lambda_d^* \rightarrow A_d^*$ (all of them are $n \times n$ matrices, n being the number of industries)
Multiplier of income (VA)	$MV = v \left[I - A_d^* \right]^{-1}$	v is a rectangular $f{\cdot}n$ matrix with as many rows as primary factors.
Multiplier of employment	$ML = l \left[I - A_d^* \right]^{-1}$	l is a rectangular matrix with as many rows as types of labour and a column for each industry.

Each column of any super-multiplier matrix informs us about the direct and indirect effects of a unitary expansion of industry j over the output, income or employment of all the industries providing resources to j. Provisions may be direct or indirect, and the 'resources' are defined in the broadest sense so as to include intermediate goods, final consumption goods resulting from new incomes, and fixed capital goods to expand capacity at the required rate.

The *structural multiplier* presented herein combines several strands in an original fashion. Like Miyazawa and Masegi (1963) and Kurz (1985), our multiplier links classical and Keynesian traditions altogether. Like Hicks' *super-multiplier* (1950) it adds up induced consumption and induced investment. Like Pyatt and Round (1979, 1985) and Pyatt (1991) it is derived from a SAM and presented in disaggregated fashion. There is a formal difference, however. Instead of adding additional columns and rows, we increase the value of the cells of the original $n \times n$ matrix. Similar to the neoclassical AGE multipliers, prices are embedded in the multiplier matrix. But we do not over-emphasize the flexibility of prices. Changes in quantity will not cause a general movement in relative prices, as neoclassical economists claim.

11.5. AUTONOMOUS DEMAND (OR THE 'DRIVING FORCES' OF THE ECONOMY)

The AGE model sketched so far allows us to compute the level of output and employment at a given moment and their increases after a supply or demand 'stimulus'. Output at year t can be presented as a multiple of the expected autonomous demand in that year (Z_t). An increase in any of the components of the (column) vector of autonomous demand will bring about an increase in output, compounded by the structural multiplier.

$$q_t = [MQ] Z_t.$$
$$\Delta q_t = [MQ] \Delta Z_t \tag{11.10}$$

Similar expressions can be found for income and labour, applying the corresponding multipliers (*MV* and *ML*). In the multiplicand we can introduce any increase in demand whose effects we wish to analyse. The model can also account for the impacts associated to a rise in wages (or in wage dispersion), the abolition of tariffs, the introduction or increase in value added tax, a rise in productivity, a change in the pension funding system, and so on. In the last cases we should modify both the vector of autonomous demand and the multiplier matrix.

In the analysis of a dynamic economy, the rates of growth of autonomous demand (vector g_z) are the key element, since induced demand adapts passively to the former. What does this vector contain? As a general rule, autonomous demand should include expenditure not funded by national income and expenditure that is not systematically related to national income. The first category refers to *exports*. The second refers to *real public expenditures*, that is public consumption and public investment.[15] Autonomous demand also stems from entrepreneurs, and it would exist even in a closed, private economy. According to Schumpeter (1912) the driving force of capitalist economies lies in the decisions of innovative entrepreneurs to launch into *new markets, new products and new processes*. In our model, Keynesian 'animal spirits' are replaced (or reinforced) by Schumpeterian innovative firms. Let us summarize these phenomena under the label of *modernization investment* and add it to the vector of autonomous demand. As a practical rule we could identify it with (1) R&D expenditures; (2) production of new goods; (3) 'excess growth' (that is growth above average) in the production of traditional goods devoted to the domestic market.

The vector of autonomous demand could also host the adjustments of investment to the oncoming disequilibria reflected in excess capacity and abnormal profit rates. When explaining induced investment we said that firms tend to expand capacity at the rate that has prevailed in the recent past. Whenever the economy follows the warranted and balanced path of growth, these investment decisions will be proved right. But usually the items of the autonomous demand grow at different rates and new goods crowd out old ones. In such unbalanced dynamics some firms will have excess capacity and ought to slow down the investments decided by the accelerator principle; others will lack capacity and should speed investment up. Let us also remember that the introduction of new methods of production increases productivity especially in a handful of industries. They will earn, temporarily, a higher profit rate than normal and will attract new investments at a very fast rate. The interest rate deserves a final comment. Contrary to most investment theories (but in accordance with empirical evidence), the interest rate is not a key determinant of investment. The main factor – to restate the point – is the expected growth of demand for any particular commodity. When taking investment decisions entrepreneurs consider the 'conventional' rate of interest, that is the one that has prevailed in the past and is expected to endure in the foreseeable future. A sharp change in this rate may speed up or slow down implementation of investment plans. Macroeconomic conditions and monetary policy become relevant at this point.

11.6. COMPARISON BETWEEN THE CLAKESCH AND NEOCLASSICAL AGE MODELS

Our Clakesch–AGE model is purported to be an alternative to the dominant neoclassical–AGE model, whose standard references were given in footnote 1. The diversity of the information taken as data by the two models (see Table 11.4) reflects deep differences in the vision of a capitalist economy, in the concept of equilibrium and in the theories proposed to explain it.

Neoclassical AGE models, no less than Walras' seminal book (1889), are static and supply constrained. They take as given certain endowments of capital and labour and solve the system of equations for prices that warrant full capacity and full employment[16] in the production of the set of goods which maximizes consumer utility. Prices are supposed to be 'market clearing' and 'efficient'. The model is mainly used in 'comparative static' analysis. It shows the new equilibrium prices and quantities corresponding to the abolition of tariffs, the introduction of a new tax and other natural or political shocks. Relative prices play the key role in the process of adjustment. They are quite sensitive to changes in the quantities demanded. At the same time, the structure of demand is also fairly sensitive to changes in relative prices. Contrary to the original fixed-price multipliers derived directly from a SAM (Pyatt and Round, 1979 and 1985), AGE models rely on flexi-price multipliers.

The Clakesch–AGE model highlights the fact that most adjustments do occur via quantities. Production adjusts to the expected demand at 'normal' prices. *Normal prices* or *prices of production* depend just on technology and distribution. Market prices may be influenced by demand; but the induced adjustments in the quantity produced will push market prices towards

Table 11.4 Data in neoclassical and Clakesch–AGE models

Neoclassical–AGE model	Clakesch–AGE model
Endowments of capital and labour Factors of production are supposed to be fully employed; prices adjust for that purpose.	The stock capital and labour supply are inherited from the past, but there is no presupposition of full employment.
Technology Malleable production functions of Cobb–Douglas type.	**Technology** Leontief's linear production function (fixed coefficients).
Individual preferences Malleable consumption functions of Cobb–Douglas type. There is no investment function.	**Expenditure patterns of social groups** Linear expenditure model to determine final consumption. Acceleration-type function of investment.
	Distribution Several alternatives are available. We can fix the rate of profit at its historical level and allow for yearly changes in the basic nominal wage (w_t). (After dividing prices by w_t we obtain 'labour commanded prices').
	Autonomous demand and its rate of growth.

production prices. The emphasis on 'quantity adjustments' should not, however, blur the importance of prices and the prices–quantities transmission mechanisms. Import propensities constitute the outstanding example.

The Clakesch–AGE model, no less than its neoclassical counterpart, is useful for policy evaluation. In addition it is especially suited to the analysis of the dynamics of a demand-constrained system. The vector of autonomous demand emerges as the driving force of output and employment. It refers to innovative entrepreneurs launching new goods either for consumption or investment. New capital goods mean new methods of production that are generally linked to increases in productivity and, after a time, in the real wage. A cluster of innovations with important diffusion and dragging effects (the latter captured by the super-multiplier) will bring about a long wave of prosperity. When the market for new products becomes saturated, and no other innovations take over, a long-lasting recession is on the way.

Any model is particularly well suited to certain purposes. If we were asked to explain welfare gains we would refer the reader to the neoclassical studies derived from Walras and Pareto, traditional AGE models included. But when we are asked to analyse the dynamics of a demand-constrained system, we believe that our Clakesch–AGE model has comparative advantages. To use a sporting metaphor, it is like playing a match of European football at home and on a pitch built for that purpose. Neoclassical AGE practitioners seem to play American football on a European football pitch. There are so many talented players, that one can expect some interesting results. It is our contention, however, that with the same intellectual investment the yield will be greater in a Clakesch–AGE model.

NOTES

1. AGE models have been a successful branch of neoclassical economics since the 1980s. Standard references are Scarf and Shoven (1984), Kehoe and Kehoe (1994), Ginsburgh and Keyzer (2002), Kehoe, Srinivasan and Whalley (2004). They were presented as an improvement on fixed-price multipliers derived from a SAM (Pyatt and Round, 1979, 1985; Pyatt, 1991). Among the few alternatives to the neoclassical model we will highlight out the 'structuralist' AGE model by Gibson and Saventer (2000).
2. In the UN National Accounts Manual (1993) it corresponds to a square 'commodity by commodity table'. Each good is produced exclusively by a unique homogeneous industry. The industry may produce a basket of related goods (for instance, textiles and shoes), but the combination of these commodities should be kept constant throughout the analysis, as we do with technology.
3. Value added tax (VAT) can be appended to obtain purchaser's prices. In order to avoid the formal complexity introduced by VAT, we are going to dispense with it.
4. Institutions engaged in a specific activity (say, provision of social services by the Government) should appear among industries.
5. In this chapter *II* stands for the identity matrix. A diagonal matrix is represented either by a circumflex (^) or an angular bracket (⟨⟩). Relative prices (p) appear as a row vector, quantities (q) as a column vector.
6. Alternatively, we could refer (after changing the sign) to deviations from the conventional unemployment rate. Two warnings are in order: (1) there is no 'natural' employment (or unemployment) rate, determining a long-period equilibrium. It is just a historical position that is bound to change with aggregate demand fluctuations; (2) there is no absolute limit of employment. Labour supply adjusts to permanent increases in labour demand.
7. We assume that only fixed capital is properly advanced; intermediate consumption and wages are paid regularly out of sales proceeds. By 'representative' or 'regulating' firm in each industry we mean the one using the best available technology. Probably a handful of innovative firms are using more productive technologies protected by patents and the like. Other firms may be using old-fashioned technology until they replace capital or quit the industry. Whenever we compute technological coefficients by calibration from an IOT, we obtain the average technology in the industry.

8. The ‘operating surplus’ is subdivided into interest payments, rents and so on. At this point we depart from the major Sraffian stream, which makes the rate of profit dependent on the rate of interest (Pivetti, 1991). In the classical, Marxian and Kaleckian tradition, changes in the rate of interest do not affect the profit rate but the distribution of profits between ‘finance capital’ and ‘real capital’.
9. Neo-Keynesian literature explains this phenomenon under the heading of ‘menu costs’.
10. The theoretical basis of the super-multiplier model are explained in Hicks (1950), Serrano (1995), Trezzini (1995) and De-Juan (2004).
11. In the first cell in column 1 of [t] we write ‘$1 - t_1$’ being t_1 the effective tax rate on the income of the first household group (H1). In the last cell (corresponding to government receipts) we write t_1. In the first cell of the last column, we write tr_1 which stands for the ratio ‘transfers to H1 / total transfers from the government’.
12. Information about propensities to consume and expenditure patterns can be obtained from family budget statistics. Unfortunately the consumption groups of such statistics do not coincide with the consumption goods contemplated in input–output tables. A bridge is necessary to join both sets of statistics. Econometrics will help to fill up certain gaps.
13. Capital consumption deserves, at least, a footnote. A convenient capital matrix should provide information about the structure of capital goods in each industry and the production capabilities of such goods. When production speeds up, the stock of capital is going to be used up above its normal level resulting in higher capital consumption. Were these data available, capital consumption could be easily mixed with intermediate consumption. In the absence of such information, capital consumption should be added to I_i now defined as *gross* expansionary investment.
14. Leontief (1970) and Lager (1997) are outstanding exceptions, although their objectives and methodology are different from ours.
15. Public transfers can be treated in different ways. Let us reflect on pensions, which nowadays absorb the bulk of public transfers. Retired people spend systematically on consumption almost all their pensions, but the amount of money they receive does not depend strictly on current national income. This justifies the traditional treatment of private consumption financed by transfers (*cTr*) as autonomous demand. In practice, governments are committed to raise pensions in parallel to other incomes. Under this perspective, consumption out of pensions would be ‘induced’ and should be part of the ‘multiplier’. In the vector of autonomous demand (the ‘multiplicand’) we should include just atypical transfers and atypical variations of ordinary transfers.
16. Actually they do not warrant full employment but an employment rate similar to the initial one, assuming that existing unemployment was ‘voluntary’.

REFERENCES

Barker, T. and W. Peterson (1987), *Cambridge Multisectoral Dynamic Model of the British Economy*, Cambridge: Cambridge University Press.

De-Juan, O. (2004), ‘Paths of accumulation and growth. Towards a Keynesian long-period theory of output’, *Review of Political Economy*, **17**(2):231–52.

De-Juan, O., M.A. Cadarso and C. Córcoles (1994), ‘Multiplicadores input–output kaleckianos: una estimación a partir de la tabla input–output española de 1990’, *Economía Industrial*, **298**:129–44.

Dietzenbacher, E. and M. Lahr (eds) (2004), *Wassily Leontief and Input–Output Economics*, Cambridge: Cambridge University Press.

Gibson, B. and D. Seventer van (2000), 'A Tale of two models: comparing structuralist and neoclassical computable general equilibrium models for South Africa', *International Review of Applied Economics*, **14**(2):100–121.

Ginsburgh, V. and M. Keyzer (2002), *The Structure of Applied General Equilibrium Models*, Cambridge, MA: The MIT Press.

Godley, W. (2004), 'Towards a reconstruction of macroeconomics using a stock flow consistent (SFC) model', Unpublished paper presented at CERF, Cambridge.

Hicks, J.R. (1950), *A Contribution to the Trade Cycle*, Oxford: Clarendon Press.

Kalecki, M. (1971), *Selected Essays on the Dynamics of the Capitalist Economy*, Cambridge: Cambridge University Press (papers written between 1933 and 1970).

Kehoe, P. and T. Kehoe (1994), 'A primer on AGE models', *Federal Reserve Bank of Minneapolis Quarterly Review,* **18**(2):2–16. Also available at http://www.minneapolisfed.org.

Kehoe, T., T. Srinivasan and J. Whalley (eds) (2004), *Frontiers in Applied General Equilibrium Modelling*, Cambridge: Cambridge University Press.

Keynes, J.M. (1936), *The General Theory of Employment, Interest and Money*, London: Macmillan.

Kornai, J. (1979), 'Resource-constrained versus demand-constrained systems', *Econometrica*, **47**(4):801–19.

Kurz, H.D. (1985), 'Effective demand in a "classical" model of value and distribution: the multiplier in a Sraffian framework', *Manchester School of Economic and Social Studies*, **LIII**(2):121–37.

Kurz, H. and N. Salvadori (1998), *Understanding 'Classical' Economics: Studies in Long-period Theory*, London: Routledge.

Lager, C. (1997), 'Treatment of fixed capital in the Sraffian framework and in the theory of dynamic input–output models', *Economic Systems Research*, **9**(4):357–73.

Leontief, W. (1970), 'The dynamic inverse', in A.P. Carter and A. Bródy (eds), *Contributions to Input–Output Analysis*, Amsterdam: North Holland, ch. 1, pp. 17–46.

Miyazawa, K. and S. Masegi (1963), 'Interindustry analysis and the structure of income distribution', *Metroeconomica*, **15**(2–3):161–95.

Nell, E.J. (1998), *The General Theory of Transformational Growth*, Cambridge: Cambridge University Press.

Nell, E.J. (2004), 'Monetising the classical equations: a theory of circulation', *Cambridge Journal of Economics*, **28**:173–203.

Pivetti, M. (1991), *An Essay on Money and Distribution*, London: Macmillan.

Pyatt, G. (1991), 'Fundamentals of social accounting', *Economic Systems Research*, **3**:315–41.

Pyatt, G. and J. Round (1979), 'Accounting and fixed price multipliers in a social accounting matrix framework', *The Economic Journal*, **89**:850–73.

Pyatt, G. and J. Round (eds) (1985), *Social Accounting Matrices. A Basis for Planning*, Washington, DC: The World Bank.

Scarf, H.E. and J.B. Shoven (1984), *Applied General Equilibrium Analysis*, Cambridge: Cambridge University Press.

Schumpeter, J. (1912), *Theorie der Wirtschaftlichen Entwicklung*, Munich: Verlag Dunker and Humbolt.

Serrano, F. (1995), 'Long period effective demand and the Sraffian supermultiplier', *Contributions to Political Economy*, **14**:67–90.

Sraffa, P. (1960), *Production of Commodities by Means of Commodities. Prelude to a Critique of Economic Theory*, Cambridge: Cambridge University Press.

Stone, R. (1981), *Aspects of Economics and Social Modelling*, Geneve: Libraire Droz.

Sylos-Labini, P. (1957), *Oligopolio e progresso tecnico*, Milan: Giuffrè.

Trezzini, A. (1995), 'Capacity utilisation in the long run and the autonomous components of aggregate demand', *Contributions to Political Economy*, **14**:33–66.

United Nations Statistics Division (1993), *National Accounts*, New York. Available at: http://unstats.un.org/unsd/sna1993/tocLev8.asp?L1=20&L2=7.

von Neumann, J. (1945–46), 'A model of general equilibrium', *The Review of Economic Studies*, **XIII**:1–9. (Original German paper of 1938.)

Walras, L. (1889), *Eléments d'économie politique pure*, Lausanne: F. Rouge.

12. Heterodox business cycles*

Lance Taylor, Nelson H. Barbosa-Filho and Codrina Rada

12.1. INTRODUCTION

Contemporary heterodox macroeconomics combines classical, Marxist, structuralist, Kaleckian and Keynesian approaches to output determination, growth and distribution. In so doing, it directly addresses the principal concerns of this volume. In contrast to the mainstream, heterodox macroeconomics attempts to build historical and institutional features of the economy directly into its models, as we illustrate in this chapter.

Three key principles – determination of total income by effective demand, determination of the functional income distribution by processes of bargaining and social conflict, and recognition of the sensitivity of the macroeconomic system to potentially destabilizing interactions of its real and financial components – undergird heterodox theory and data analysis which have made great strides over the past two decades.[1]

From Keynes and Kalecki and on through the years, heterodox analysts studied economic fluctuations. But somewhat surprisingly, not much recent work from within the camp has been devoted to a long-established and central macroeconomic concern: the analysis of business cycles in industrialized economies. Drawing on scattered cycle models that do exist, in this chapter we outline an approach (some parts of it already underway) that we think can make cycles a major component of the contemporary heterodox tradition. As will be seen, work in progress is being described, and not a set of consolidated results.

Our strategy is to sketch low order (one- or two-dimensional) models of recurrent oscillations observed in rich economies, with emphasis on the USA. The cycles we consider involve variations in net borrowing (investment minus saving) by government, the rest of the world, business and households as compared to peaks and troughs in output; the level of economic activity (proxied by capacity utilization u) and the wage share (ψ) as a measure of income distribution; a financial cycle patterned after Minsky (1975) involving the profit rate (r) and interest rate (i, with real value j);

longer-term linked fluctuations in the business sector debt–capital ratio (λ) and the equity–capital ratio (Tobin's q); and an apparent post-Bretton Woods cycle in the dollar exchange rate (nominal level e in units of dollars per euro) with a period of a decade or so. Keyed to several of these cycles are movements in the investment–capital ratio (g, and/or the capital stock growth rate $g-\delta$ with δ as the 'radioactive' rate of depreciation).

By construction, variables that are either ratios of real or financial quantities such as u, ψ, r, λ, q, e and g or rates per unit time such as such as i and j may be (at least quasi-) stationary in the data and certainly have movements contained within fairly narrow ranges. For purposes of illustration and understanding, it is helpful to assume that their dynamics can be examined in simple phase diagrams. That is the approach adopted here, treating time as continuous so the analysis can be set up in terms of ordinary differential equations. We hint at systems involving more than two variables which may generate higher order oscillations (or chaos?). Building up an inclusive model of moderate dimensionality combining aspects of models like the ones presented in this chapter is the long-term goal of our research.

12.2. CYCLES IN NET BORROWING

The heterodox tradition has always emphasized that different groups of economic actors behave differently. For example, how do levels of saving and investment for households, the business sector, government and the rest of the world vary over the cycle? If I_i is investment by 'institutional sector' i and S_i is its saving, its 'net borrowing' B_i in national accounts terms is $B_i = I_i - S_i$. Accounting consistency enforces the condition $\Sigma B_i = 0$, but the B_i are otherwise free to vary. Insofar as they are affected by movements in output, the functional distribution, interest and exchange rates, and asset prices, borrowing levels are linked to cycles analysed in following sections of this chapter.

Broadly following Godley (1999) and ultimately the 'New Cambridge' economists of the 1970s with their emphasis on flow of funds accounting, Figure 12.1 presents evidence for the American economy, with net borrowing levels for households, business, government and the rest of the world expressed as shares of capacity or potential output as in the model of the next section. The solid lines running upward from the horizontal axis represent cyclical peaks according to the National Bureau of Economic Research (NBER) methodology, and the lines running downward mark troughs. As is well-known, peak-to-trough periods are longer than troughs-to-peaks.

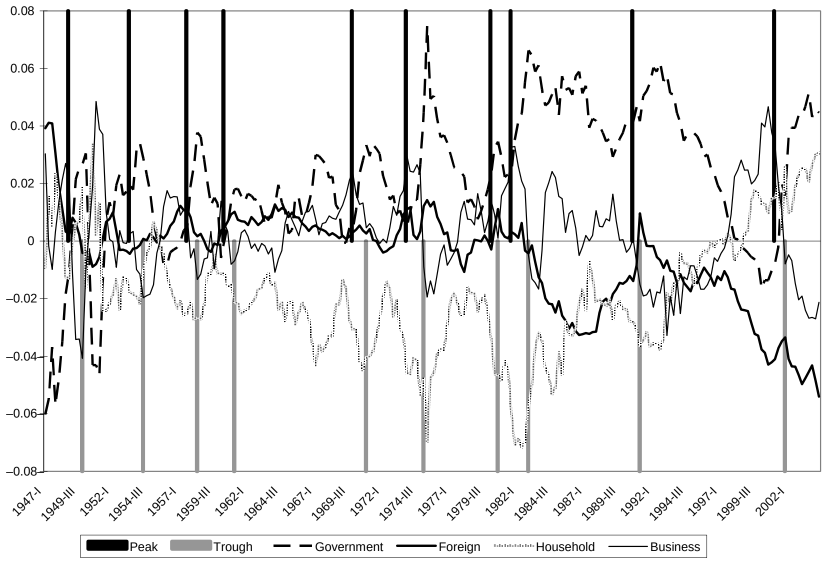


Figure 12.1 NIPA-based financial needs of the government, foreign sector, household and rest of the private sectors as scaled to potential output. Upper and lower vertical lines mark NBER peaks and troughs respectively

Net borrowing levels seem to be reasonably clear lead-lag relationships among themselves and in relation to the NBER benchmarks. There are also possible trends and structural breaks. It is simplest to look at the curves after the early 1950s when war echoes had died down.[2]

Households. Until very recently, cyclical output troughs and lows in household net borrowing virtually coincide – historically households have built up assets and paid off debt as the economy contracted. Peaks in borrowing tend to lead output peaks as households begin to run debt up and assets down early in the upswing. After around 1980, household net borrowing on a flow basis shows a strong upward trend. Between 1980 and 2003 the ratio of household liabilities to GDP rose from around 0.5 to over 0.85, amply collateralized until 2001 by an increase of financial assets/GDP from around 2.1 to 3.2. But even after the stock market crash the usual decrease in household borrowing at a trough was very weak in 2002. Such apparent changes in household behaviour have implications as discussed below.

Government. Note the peaks in borrowing after the NBER cyclical peaks. The government deficit ramps up between peak and trough and then typically continues to grow early in the upswing. Much of the extra borrowing finances transfers to households near the bottom of the cycle (reduced taxes, unemployment insurance payments, and so on). Such automatic stabilizers as well as the shift in the functional distribution toward labour during the downswing (as analysed in the next section) probably underlie the leading role in household borrowing just discussed.

Policy also influences government borrowing. There was a spike in the late 1960s under Lyndon Johnson's presidency; a continued high in the mid-1980s under Ronald Reagan and then a decline thereafter with the Plaza accords; the Bill Clinton fiscal soundness in the 1990s; and probably a reprise of Reagan under George W. Bush (Clinton redux is still in the wings).

Business. Troughs in business net borrowing tend to coincide with or lag the output trough; evidence not presented here also suggests that they are associated with business capital gains. Peaks lead or coincide with the output peak, that is there is a sharp rise in net borrowing in the latter phase of the output upswing, in line with likely effects on investment demand of movements in the profit share and interest and profit rates as discussed below. In contrast to households this pattern persists into the 2000s. The troughs of net borrowing (or peaks in net lending) are deeper post-Reagan, with a possible link to the household pattern shift noted above.

Foreign. Troughs tend to occur midway between output troughs and peaks. The real exchange rate cycle discussed below is no doubt playing a role in affecting the cyclical pattern, for example the depreciation after the 1985 Plaza accords. Troughs in foreign net borrowing (peaks in the current account deficit) appear to lag the government peak, casting doubt on the convenient 'twin deficits' explanation of the external deficit. In cyclical terms, foreign troughs track more closely to business sector peaks. Over time, the trend increases in foreign net lending and the increase in household net borrowing are just about the same magnitude. Rather than fiscal or business deficits, the deterioration in the US external position mirrors the shift in household financial behaviour discussed above.

A final conjecture – intriguing but probably impossible to verify at the macro level – is that greater household debt contributed to the lengthening of the cycle after 1980 that is evident in Figure 12.1. By permitting more effective consumption-smoothing over time, borrowing may have permitted households to sustain their spending levels for more extended periods of time.

12.3. OUTPUT AND DISTRIBUTION CYCLES

The first two principles mentioned at the outset – determination of the functional income distribution by bargaining and social conflict and determination of total income by effective demand – lead to the analysis in this section. Both capital stock and labour productivity growth rates are implicit in the model. In this and subsequent sections, we concentrate on the economics and rudiments of stability analysis for the models concerned, leaving fuller development of the mathematics to other cited presentations.

Richard Goodwin (1967) introduced a growth cycle involving the wage share ψ and the employment ratio (ε or employed labour/labour supply). It was based on Lotka–Volterra dynamics with ψ as 'predator' and ε as 'prey'. His own specification had investment determined by available saving along (at least on some interpretations) Marxist lines, and real wage determination by a Phillips' curve reflecting changes in the size of the reserve army of unemployed. It generated counterclockwise closed orbits in the (ε, ψ) phase plane and thus was a model of Marxist cyclical growth.[3]

More than a decade later, Bob Rowthorn and Amitava Dutt independently introduced a Kaleckian two-dimensional non-cyclical temporary equilibrium macro model in which the output–capacity ratio u was a function of ψ (or the real wage ω) in a 'demand' relationship and ψ could be interpreted as a function of u along a 'distributive' schedule.[4] Adopting terminology introduced subsequently, a positive dependence of u on ψ (or $\partial u / \partial \psi > 0$ so that the effective demand curve slopes upward when ψ rises) means that demand is 'wage-led'; in the opposite case it is 'profit-led'. Along the distributive curve, a positive response of ψ to u represents a 'profit-squeeze'. A negative response involves 'forced saving' as in Nicholas Kaldor's vintage-1960s growth models, typically on the part of workers as price inflation outruns wage inflation in the short to medium run.

In the American business sector, these two variables (both normalized around unity) in fact enter into Goodwinesque counterclockwise cycles as illustrated in Figure 12.2. There have been nine such oscillations since World War II, with the last two having periods of around ten years and the earlier ones running about three to five years (as with the NBER cycles in Figure 12.1).

Barbosa-Filho and Taylor (2005) analyse these data using a dynamic version of the Dutt–Rowthorn model. It basically replaces Goodwin's saving-determined investment with output determination by effective demand, and his Phillips' curve with a distributive relationship as just discussed.

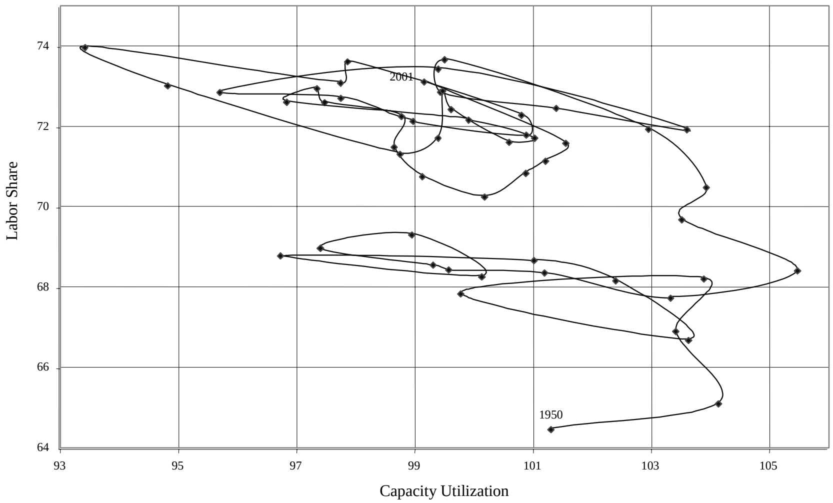


Figure 12.2 Capacity utilization and the broadly defined wage share in the US economy, 1950–2001

We set $u = X/Q$ with capacity Q treated as a function of the capital stock K. The wage share is also a ratio variable, $\psi = \omega/\xi$ with $\xi = X/L$ being labour productivity (L is the employed labour force). Let $\dot{X} = dX/dt$ and $\hat{X} = \dot{X}/X$. Then one immediately has that $\hat{u} = \hat{X} - \hat{K}$ (if for simplicity Q and K are assumed to be proportional) and $\hat{\psi} = \hat{\omega} - \hat{\xi}$.

To set up a model in the (u, ψ) phase plane, let the growth rates $\hat{X}$, $\hat{K}$, $\hat{\omega}$ and $\hat{\xi}$ all be linear functions of u and ψ. Then the analysis is reduced to two dimensions involving a pair of linked differential equations

$$\dot{u} = u(\phi_0 + \phi_u u + \phi_\psi \psi) \tag{12.1}$$

and

$$\dot{\psi} = \psi(\theta_0 + \theta_u u + \theta_\psi \psi). \tag{12.2}$$

This system has four potential equilibrium points, one at $u = \psi = 0$, two with one of the variables zero and the other non-zero, and the fourth obtained by solving linear equations for u and ψ emerging from the terms in parentheses in (12.1) and (12.2) when $\dot{u} = \dot{\psi} = 0$. An economy with no output and/or a zero labour share does not make a lot of sense so we concentrate on an equilibrium with positive values of u and ψ.[5] The slopes of the nullclines through such a point are

$$\left.\frac{d\psi}{du}\right|_{\dot{u}=0} = \frac{-\phi_u}{\phi_\psi}$$

and

$$\left.\frac{d\psi}{du}\right|_{\dot{\psi}=0} = \frac{-\theta_u}{\theta_\psi}.$$

VAR estimates of a difference equation analogue of (12.1) and (12.2) are presented by Barbosa-Filho and Taylor (2005). They suggest that $\phi_u < 0$ (u is self-stabilizing in (12.1)) and $\phi_\psi < 0$. The upshot is that in Figure 12.3 the 'effective demand' schedule (or the $\dot{u} = 0$ nullcline) has a negative slope – a temporary equilibrium relationship in the Rowthorn–Dutt model becomes a description of steady state behaviour in the present specification.

Turning to the 'distributive' schedule (the $\dot{\psi} = 0$ nullcline), the econometrics shows that $\theta_\psi < 0$ and $\theta_u > 0$ so that in steady state a higher level of activity increases the wage share in a profit squeeze. The trajectory sketched in the diagram demonstrates that this configuration of parameters can generate a counterclockwise cycle involving u and ψ. The oscillations could be either damped or divergent, with weak damping appearing to be the empirically relevant case (as in most estimated business cycle models).[6]

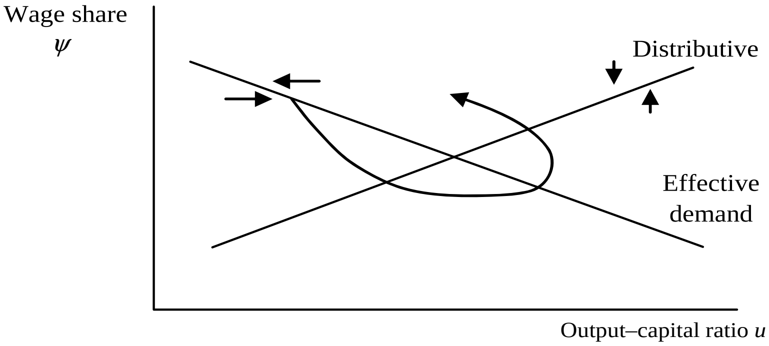


Figure 12.3 A structuralist Goodwin model with stable wage share dynamics

In a bit more detail, the diagram suggests that in an economy starting from a point toward the NW on the effective demand schedule the wage share will tend to fall, basically due to acceleration in productivity growth (the real wage is only weakly procyclical in the US so most variation in ψ is due to changes in ξ). The profit share $\pi = 1 - \psi$ and profit rate $r = \pi u$ will rise (consistent with the pattern in Figure 12.2), stimulating demand in a profit-led system so u swings up. As the trajectory crosses the distributive schedule productivity growth slows and the wage share starts to increase until it forces growth in demand to turn negative at the effective demand curve. These processes reverse during the subsequent downswing in u and upswing (until the Distributive curve is crossed again) in ψ. In other words, a Goodwin-style demand–distributive cycle exists in the US economy. It can be rationalized by profit-led demand in the long run combined with a profit squeeze as economic activity goes up.

12.4. INTEREST AND PROFIT RATE CYCLES

Effective demand and recognition that the macroeconomic system is sensitive to potentially destabilizing interactions of its real and financial components – two of the principles stated above – lead naturally into analysis of cycles.

For example, one way of thinking about Hyman Minsky's (1975) interpretation of *The General Theory* is that the liquidity trap is transformed from a *floor under* to the *ceiling above* the interest rate. Combined with fluctuating animal spirits in the investment function, this feature of interest rate determination can generate cycles. Following Taylor (2004), Figure 12.4 illustrates short-term relationships in a Minskyan IS–LM system.

A simple specification of the asset market can be based on the balance sheets in Table 12.1. Firms borrow L and issue equity $P_e E$ with E as an index of shares outstanding and P_e as the equity price. They also carry non-zero net worth Ω_f (the Modigliani–Miller theorem does not apply). Their debt is held by households (L_h) and banks (L_b). The banks form a pure credit system in which L_b equals the money supply M – 'loans create money' *à la* Wicksell. Households hold money, business debt and equity; their net worth is Ω_h. The value of the capital stock at current prices P is PK, and $\Omega_h + \Omega_f = PK$. Let $L / PK = \lambda$ (the debt to capital stock ratio) and in standard fashion let Tobin's q be defined as $q = P_e E / PK$.

Shares of household wealth respectively held as money, firm debt and equity are μ, η and σ with $\mu + \eta + \sigma = 1$. Because the equity price can jump, Ω_h is endogenous in the short run. The interest rate can be viewed as principally affecting households' portfolio decisions between business debt on one hand, and money and equity on the other. The portion of debt L they do not hold is held by the banking system and thereby monetized. The implication is that $(\mu + \eta)\Omega_h / PK = (1 - \sigma)\Omega_h / PK = \lambda$ meaning that the

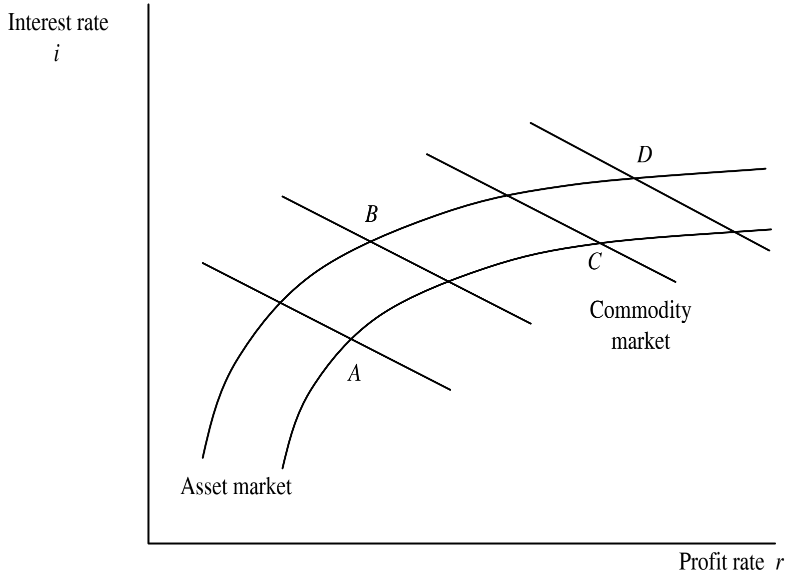


Figure 12.4 Effects of an increase in investment demand in macroeconomic equilibrium at relatively low (A to B) and high (C to D) levels of economic activity in the Minsky model

debt to capital ratio can be expressed as the ratio of non-equity holdings of households to capital stock. The household balance sheet itself states that $\lambda + q = \Omega_h / PK.$ Combining expressions gives

$$\lambda \frac{\sigma}{1-\sigma} = \lambda \frac{1-(\mu+\eta)}{\mu+\eta} = q \tag{12.3}$$

for asset market equilibrium, with λ determined in the short run by the history of business borrowing.

A standard gross substitutes assumption would make $\mu+\eta$ an increasing function of *i,* with a higher interest rate leading households to substitute toward debt from both loans and equity. The effect of a higher *r* (profit rate) on σ is presumably positive, or negative on $\mu+\eta.$ A very simple specification (replaced by an arbitrage equation below) would set $q = r/i$ or the profit rate capitalized by borrowing costs.

Minsky appears to argue that short-run portfolio choices are non-linear. On his view, the 'asset market' curve (12.3) is concave because of shifts in liquidity preference over the cycle. When *u* and *r* swing up, Keynesian 'transactions demands' for money and liquid assets rise less rapidly than 'speculative demands' decline as investors seek to shift toward equity, hedge fund positions, and so on. The outcome is that the interest rate responds inelastically to an increase in r and u while asset prices (or q) sharply increase. With a modest increase in i, higher values of r and q feed back into the level of activity, bidding up r even more. Hence the flattening of the asset market schedule shown in the diagram. The effect of a change in λ on the position of the asset market schedule is ambiguous but one would normally assume that $\partial i / \partial \lambda > 0$.[7]

The 'commodity market' schedule in Figure 12.4 is almost a run-of-the-mill IS curve, with an exception involving the dynamics of investment demand that is described below. The economic intuition is that a reduction in the interest rate leads to an increase in effective demand which bids up both capacity utilization and the profit rate.

The IS curve moves upward in response to the investment–capital ratio $g = I / K$, thereby increasing r and i if the asset market schedule does not shift. But how does g get determined? Two strands of thought intertwine in the heterodox literature. Rowthorn, Dutt, and many subsequent authors make the *level* of g depend on u, r and/or q (positively), i (negatively), and other variables. On the other hand, authors as diverse as Roy Harrod, Kaldor, Steindl and Joan Robinson can be interpreted as arguing on expectational grounds that the *change* in investment (or $\dot{g}$) is the more appropriate endogenous variable. Minsky seems to be in this camp. A convenient explanatory variable for $\dot{g}$ is r/i or profits per unit of capital, capitalized by the interest rate:

$$\dot{g} = f\left(\frac{r}{i}\right). \tag{12.4}$$

Because it is determined by a differential equation, g is constant in the short run. Hence, any downward slope in the commodity market schedule of Figure 12.4 has to be due to effects of r and i on savings – this is the schedule's 'exceptional' feature noted above.

From Figure 12.4, we have $\partial \dot{g} / \partial g > 0$ in (12.4) because r responds more strongly than i to higher effective demand. A change in λ affects $\dot{g}$ through several channels. For a given level of g it increases firms' payment obligations, reducing their saving and thereby stimulating effective demand. As discussed above, at a given level of effective demand a higher λ may make the interest rate move either way. If more debt does drive up interest rates (the standard assumption, of course) we could easily have growth in investment being 'debt-burdened', or $\partial \dot{g} / \partial \lambda < 0$.[8]

The growth rate of firms' capital stock is $\hat{K} = g - \delta$. Using this fact, a differential equation for the loan–capital ratio λ can be written out on the basis of their flow-of-funds balance $\dot{L} = I + iL + \theta K - rK - \chi I$ in which θ is the share of the value of capital paid out as dividends and χ is the proportion of investment financed by new equity issues (the aggregate level of χ is typically negative in the US). Using this expression, it is easy to show that

$$\dot{\lambda} = [i-(g-\delta)]\lambda + (1-\chi)g + \theta - r \ . \qquad (12.5)$$

This differential equation permits various 'regimes' for steady state values of λ, based on inequalities among *g*, *i* and *r* (Foley and Taylor, 2004; Rada, forthcoming). For present purposes we concentrate on a simple case in which $\partial\dot{\lambda}/\partial\lambda < 0$ (the equation is locally stable). The investment rate *g* affects $\dot{\lambda}$ both directly and indirectly via *i* and *r* in (12.5). For illustrative purposes, we assume that $\partial\dot{\lambda}/\partial g > 0$. More investment does not drive up profits enough to lead to an overall reduction in firms' borrowing requirements.

Figure 12.5 illustrates linear approximations to equations (12.4) and (12.5) around an initial steady state at point A.[9] Suppose that there is a sudden shock to business confidence or animal spirits, making investment jump down from *A* to *B*. The downward jump leads into further steady declines in *g* and λ until enough of the debt burden is worked off to permit *g* to start to rise at point *D*. If it is damped, the resulting cycle will eventually spiral back to *A*. From Figure 12.4, fluctuations in *r* will be wider than those in *i* – a direct test of Minsky's formulation.

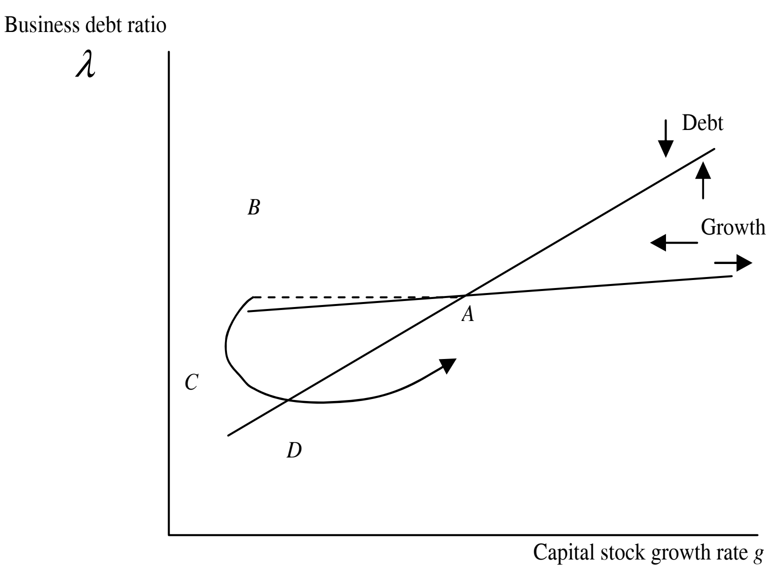


Figure 12.5 Growth rate and debt ratio dynamics in a Minskyan financial cycle

12.5. DEBT–EQUITY CYCLES

Figure 12.6 shows the long-term evolution of the US business sector's debt–capital ratio λ and equity valuation ratio q (data from the US Flows of Funds accounts). Most but not all viewers of the diagram perceive two 'long cycles' in q and λ, the first from the immediate post-WWII period to the early 1980s and the second (at higher levels of the two variables) from the late 1980s through 2002. A similar pattern shows up after the early 1960s in the UK. Such debt–equity cycles do not feature in the academic economics literature but are familiar in bits of Wall Street lore such as the 'Dow Theory' or 'Elliott Wave Principle'. Almost needless to say, they are consistent with the movements in business and household net borrowing shown in Figure 12.1.

Assuming that the cycles exist, how can they be rationalized? Taylor and Rada (2003) set out a real-financial Kaleckian macro model incorporating equation (12.5) for the evolution of the business debt burden. Because they are dealing 'with the long run', the authors replace the simple approximation $q = r/i$ of the last section with a differential equation for q based on a standard formula (routinely used in pension fund calculations) for equity valuation,

$$\rho = \hat{P}_e + \frac{\theta K}{P_e E} = \hat{P}_e + \frac{\theta}{q}$$

in which ρ is the 'required' or 'long-term' return to equity (a famous 7 per cent in the US). After the first equality, the return is shown to stem from capital gains $\hat{P}_e$ and a term incorporating the dividend yield. From the assumption above about issues of new equity we further have that

$$\hat{E} = \chi \frac{g}{q}.$$

Because $\hat{q} = \hat{P}_e + \hat{E} - \hat{K}$ these equations can be combined to give

$$\dot{q} = [\rho - (g - \delta)]q + \chi g - \theta. \tag{12.6}$$

Working with an investment function in which the level of g depends positively on u and q and negatively on the business debt load $i\lambda$, Taylor and Rada show that a clockwise (λ, q) cycle like the one in Figure 12.6 comes out in equations (12.5) and (12.6) if the investment–capital ratio g is 'debt-led' in the long run (higher debt stimulates growth via its effects on the saving–investment balance), and the debt ratio is 'equity-accelerated' in the sense that $\partial\dot{\lambda}/\partial q > 0$ via investment and savings effects of changes in q.

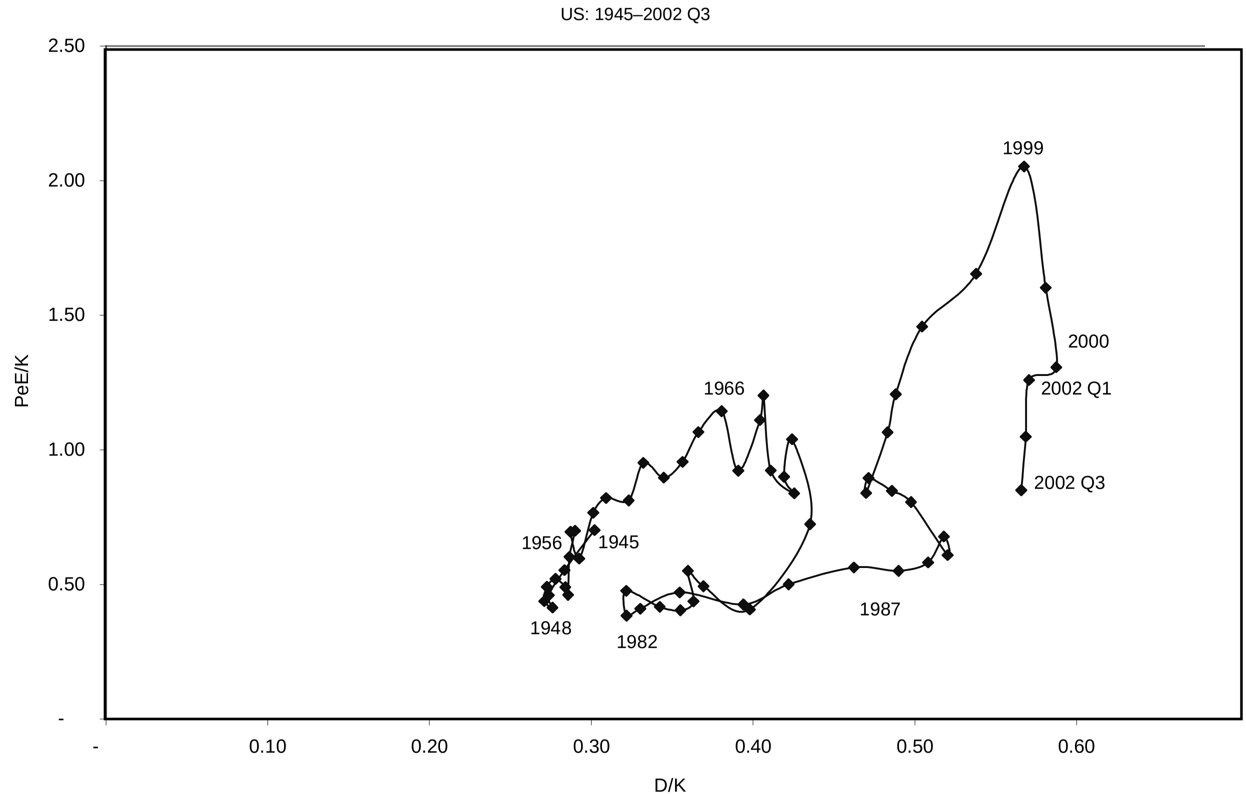


Figure 12.6   Postwar debt–equity cycles in the US

Along Kindleberger's (2000) 'manias, panics, and crashes' lines, the sharp upswings in q at the peaks of Figure 12.6's cycles can be brought into the picture if a dynamic specification is added for ρ. For example, the price–earnings ratio R can be expressed in present notation as $R = q/(r - i\lambda)$, an increasing function of λ and q. Now suppose that the 'required' equity return follows the rule

$$\dot{\rho} = \mu(\bar{R} - R)(\rho - \bar{\rho})(\bar{\rho} + \sigma - \rho) \tag{12.7}$$

in which the function $\mu(\bar{R} - R)$ takes the same sign as its argument. This formulation permits a simple 'transcritical bifurcation' in ρ.

To begin to tease out the implications, note that when $\mu > 0$ the right-hand side of (12.7) looks like Figure 12.7a. A positive value of μ pushes ρ toward an upper bound $\bar{\rho} + \sigma$ where $\bar{\rho}$ is a base level return to equity and σ reflects market exhilaration. The implication is that in a mania a high value of ρ makes q rise rapidly until R surpasses its crisis level $\bar{R}$. The equity–capital ratio passes a cyclical turning point, and then drops off quickly as Figure 12.7b comes into effect. This is the beginning of Kindleberger's panic phase with economic agents rushing to exit the stock market and selling their share holdings en masse. Bankers behave in a similar manner and stop lending, making the crash inevitable. Economic agents, now bears, switch asset demands from shares toward liquidity. Sooner or later investment drops substantially, slowing the capital stock growth rate enough to set up a turnaround in q as recovery gets underway.[10]

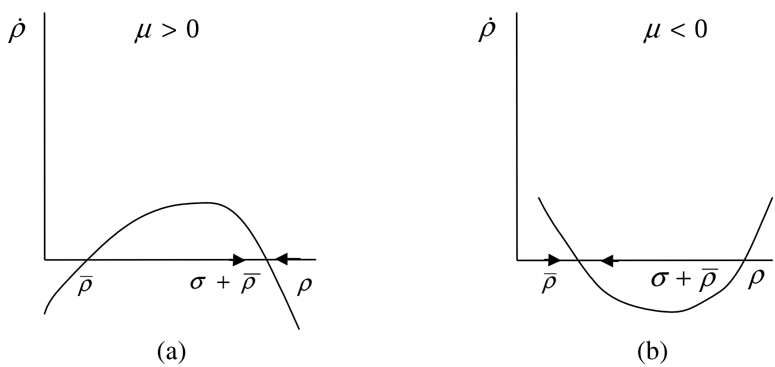


Figure 12.7 Fast dynamics and bifurcations in debt–equity cycles

12.6. EXCHANGE RATE CYCLES

A final question regards possible cyclical behaviour of a floating exchange rate – how does it respond to changes in domestic activity and monetary policy? There is little consensus in the literature in this regard. We address the question using a bare-bones model emphasizing the exchange rate's role as an asset price (thereby ignoring its many other effects on effective demand, inflation, trade performance, and so on).

Since the end of the Bretton Woods era, the dollar exchange rate has gone through two long depreciation/appreciation oscillations, lagging behind periods of loose and then more restrictive monetary policy. In more formal terms, exchange rate dynamics appears to take the form

$$\dot{e} = \alpha e(i^* - i) \qquad (12.8)$$

with i^* as the foreign interest rate. A relatively 'low' value of the domestic rate i 'creates pressure' for the (\$/€) exchange rate to depreciate or rise. This response was foreseen by Minsky (1983) in a prescient paper; it has also recently been common in developing countries.

All this is suspicious from the mainstream point of view. Orthodox economic theory postulates that arbitrage across futures markets will enforce uncovered interest rate parity (UIP). It states that the *expected* change in the spot rate $\dot{e}^{\exp}$ is related to the spot rate and the interest rates by the formula $\dot{e}^{\exp} = e(i - i^*)$. If one further postulates myopic perfect foresight (MPF), $\dot{e} = \dot{e}^{\exp}$, then the differential equation

$$\dot{e} = e(i - i^*) \qquad (12.9)$$

emerges.

We thus have alternative dynamic theories – the heterodox Minsky story (12.8) and UIP/MPF (12.9) – which seem to predict responses of $\dot{e}$ to the interest rate spread with *opposite* signs. An immediate question is whether they can be brought under one theoretical umbrella. To sort out scenarios we have to state how interest rates get determined. We adopt a standard open economy macroeconomics formulation based on asset market equilibrium.

For a 'small' home country (even the United States) it is simplest to treat the foreign rate i^* as predetermined. The domestic rate follows from a bond market equilibrium condition that could be written as

$$i = i_0 + a\dot{e} - be - cM \;. \qquad (12.10)$$

More rapid expected (= actual) depreciation makes wealth-holders desire to shift out of dollar securities, driving their prices down and thus increasing the interest rate i. An upward jump depreciation in e 'makes the dollar cheaper' and reduces i. If the Federal Reserve increases the money supply M by open market purchases of local bonds (or more generally pursues expansionary policy), bond prices rise or the interest rate goes down.

In (12.8) and (12.9) the bond market equilibrium condition (12.10) effectively makes the change in the exchange rate $\dot{e}$ depend on the spot rate e, monetary policy M, and itself. A simple dynamical system to investigate exchange rate cycles can be constructed if we add a differential equation for M.

Minsky provides a hint about how to proceed: 'Any move by owners of dollar-denominated deposits to deposits denominated in other currencies will put down pressure on dollar exchange rates … A movement out of the dollar can quite quickly escalate into a run on the dollar … In order to break the run the Federal Reserve ha[s] to move to increase the income available from holding dollars …'. In other words, the domestic interest rate has to rise. A simplified description focusing only on the exchange rate is

$$\dot{M} = m(\bar{e} - e), \tag{12.11}$$

with $m > 0$ as a response coefficient and $\bar{e}$ as a long-term reference exchange rate (possibly even set by purchasing power parity). A high (depreciated) value of e pushes the Federal Reserve in the direction of contractionary monetary policy.

To trace through the implications, it is instructive to begin with the UIP/MPF differential equation (12.9). In other words, exchange futures arbitrage and perfect foresight are assumed to apply. Together with (12.11), (12.9) defines a two-dimensional dynamical system with a steady state at $i = i^*$ and $e = \bar{e}$. After substitution from (12.10), the reduced form for the UIP/MPF equation is

$$\dot{e} = \frac{e}{1 - ea}(i_0 - be - cM - i^*). \tag{12.12}$$

Linearized around the steady state, the Jacobian for (12.12) and (12.11) can be written as

$$J = \begin{bmatrix} -\dfrac{be}{1-ea} & -\dfrac{ce}{1-ea} \\ -m & 0 \end{bmatrix}. \tag{12.13}$$

The key term is $1-ea,$ which measures the strength of the interest rate response to an increase in the expected rate of depreciation in (12.10). There are two cases:

(a) For a 'high' value of a, $1-ea<0,$ expectational effects are strong and the signs of the elements of the Jacobian take the pattern

$$\begin{bmatrix} + & + \\ - & 0 \end{bmatrix}. \tag{12.14a}$$

(b) A 'low' value of a means that $1-ea>0$ and the signs take the pattern

$$\begin{bmatrix} - & - \\ - & 0 \end{bmatrix}. \tag{12.14b}$$

Case (a) is close to Minsky's story in (12.8) because $\dot{e}$ in (12.12) responds *positively* to the determinants of i – in effect the coefficient $\alpha=-1/(1-ea)$. By itself, the exchange rate has unstable dynamics because depreciation (an increase in e) reduces i in (12.10) and thereby increases $\dot{e}$ in (12.8). Despite the fact that $\partial\dot{e}/\partial e>0$, the system can be partially (but not completely) stabilized by the oppositely signed off-diagonal terms in (12.14a).

Cyclical dynamics are illustrated in Figure 12.8. With an initial equilibrium at A, suppose that foreign financial institutions switch portfolio preferences toward dollars (after the break-up of the Bretton Woods system, for example). The 'exchange rate' locus along which $\dot{e}=0$ shifts to the left as i_0 in (12.10) declines. The exchange rate starts to depreciate and in response the Federal Reserve starts to reduce the money supply. At point B, i rises above i^* in (12.8) and the exchange rate starts to appreciate although monetary contraction continues. At C, e falls below $\bar{e}$ in (12.11) and M starts to rise, setting off a new depreciation phase at D.

Consistent with Minsky's views about the intrinsic instability of the exchange rate, the equilibrium point E in Figure 12.8 is an *un*stable focus. Formally, this result follows from our assumption that $\dot{M}$ depends only on e in (12.11). Adding a damping term to make $\partial\dot{M}/\partial M<0$ (which may or may not be realistic) could transform the diverging spiral in the diagram to one that converges to E.

Finally, with weaker expectational effects in case (b), somewhat similar counterclockwise dynamics emerges. The equilibrium is a saddle point, so along mainstream lines the dynamics would involve e jumping upward to a

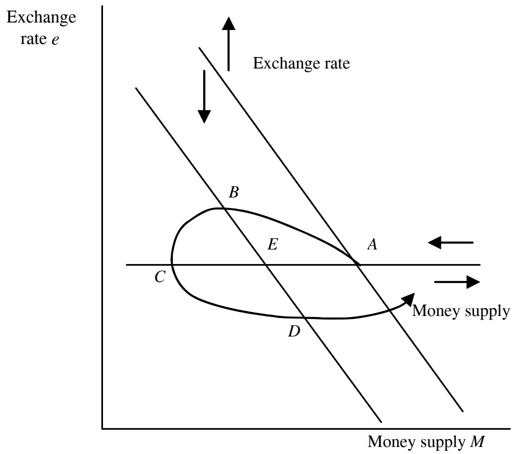


Figure 12.8 Cyclical exchange rate dynamics with strong expectational effects in the bond market

saddlepath along the lines of the Dornbusch overshooting model and then dropping slowly down to the steady state as M declines. As in case (a), exchange rate dynamics are strongly linked to domestic monetary considerations.

12.7. NEXT STEPS

As emphasized at the outset, this chapter describes a research project in its early stages. It is certainly not a finished product. Our next step will be a thorough statistical analysis of the net borrowing oscillations illustrated in Figure 12.1, to quantify trend-cycle and lead-lag relationships and calculate impulse response functions.

We also plan to investigate other low-dimensionality cycles, at least those involving employment and distribution, the interest and profit rates, and the exchange rate as discussed in previous sections. The follow-up would be a more complicated model, perhaps involving the four net borrowing series, capacity utilization, labour productivity, the employment ratio, price and wage levels, exchange rate, short- and long-term interest rates, profit rate, and debt–capital and equity–capital ratios.

As discussed above, we have some guidance from heterodox theory about all these variables, and hope to illuminate how they have interacted in the US over the past five decades.

NOTES

* Support from CEPA and The Ford Foundation is gratefully acknowledged, as well as comments from the editor and referees and assistance by Luca Zamparelli.

1. Foley and Taylor (2004) give a synthetic presentation, drawing on book-length versions in Foley and Michl (1999) and Taylor (2004). Names are dropped freely in this chapter but for reasons of space only a few key references are formally cited – the others can be checked in the sources just mentioned.
2. Cross correlograms between net borrowing and GDP based on series smoothed with bandpass filters confirm the informal observations that follow.
3. Or in other words, a counterclockwise trajectory involving *x* (horizontal axis) and *y* (vertical axis) variables signifies that *x* 'leads' *y*. The roles reverse if the oscillations run clockwise.
4. Because they think a concept like 'full employment' is largely irrelevant, heterodox modellers tend often to use capacity utilization instead of the (un)employment rate as an indicator of economic activity. Nevertheless, fluctuations in employment are important economically and we are currently extending the model presented here to build them in.
5. The (0,0) equilibrium is of some interest because it occurs in Goodwin's original model. In the present case, stability properties depend on the signs of the intercept or intrinsic growth rate terms ϕ_0 and θ_0. They are the entries on the main diagonal of the relevant Jacobian matrix with the off-diagonal entries equal to zero. If both are positive (negative) the origin is an unstable (stable) node, and opposite signs give a saddlepoint (as in Goodwin's own model). In formal terms, we assume that the case of the stable node with two negative intrinsic growth rates does not exist.
6. Strictly speaking the equilibrium in Figure 12.2 could be either a node or a focus. Parameter estimates rule out the former possibility. If the estimated model is perturbed from its steady state, it cycles back to a near-equilibrium position in 10–20 quarters.
7. In a bit more detail, let $f(i,r)=\sigma/(1-\sigma)$ in (12.3). Then with $q=r/i$ one has $[f_i+(r/i^2)]di=-fd\lambda+[(1/i)-f_r]dr$ with signs of the partial derivatives being $f_r>0$ and $f_i<0$. Even if the term multiplying *di* is negative so that a higher volume of loans drives up the interest rate, the effect of *r* on *i* is ambiguous. Minsky's reasoning suggests that the term multiplying *dr* is negative and that the ratio $[(1/i)-f_r]/[f_i+(r/i^2)]=\partial i/\partial r$ decreases as *r* goes up due to a decrease in the absolute value of f_i and (especially) an increase in f_r.
8. A form of 'debt-led' growth comes into the discussion below.
9. Both differential equations are non-linear in *g* and λ, directly and through Figure 12.4. But since we are only undertaking local stability analysis the nullclines in Figure 12.5 are sketched as straight lines.
10. Strong cyclical behaviour can also emerge from (12.5) and (12.6) when the investment function is set up in terms of the growth rate derivative $\dot{g}$ as in (12.4) instead of *g*. See Rada (forthcoming).

REFERENCES

Barbosa-Filho, N.H. and L. Taylor (2005), 'Distributive and demand cycles in the US economy – A structuralist Goodwin model', *Metroeconomica,* forthcoming.

Foley, D.K. and T.R. Michl (1999), *Growth and Distribution*, Cambridge, MA: Harvard University Press.

Foley, D.K. and L. Taylor (2004), 'A heterodox growth and distribution model', New York, NY: Department of Economics, New School University.
Godley, W. (1999), 'Seven unsustainable processes: medium-term prospects and policies for the US and the world', Annandale-on-Hudson, NY: Jerome Levy Economics Institute, Bard College.
Goodwin, R.M. (1967), 'A growth cycle', in C.H. Feinstein (ed.), *Socialism, Capitalism, and Growth*, Cambridge: Cambridge University Press.
Keynes, J.M. (1936), *The General Theory of Employment, Interest, and Money*, London: Macmillan.
Kindleberger, C.P. (2000), *Manias, Panics, and Crashes: A History of Financial Crises*, New York: John Wiley & Sons.
Minsky, H.P. (1975), *John Maynard Keynes*, New York: Columbia University Press,
Minsky, H.P. (1983), 'Monetary policies and the international financial environment', St. Louis, MO: Department of Economics, Washington University.
Rada, C. (forthcoming), 'A Keynesian real-financial model with a potentially unstable investment function', New York, NY: Center for Economic Policy Analysis, New School University.
Steindl, J. (1952), *Maturity and Stagnation in American Capitalism*, Oxford: Basil Blackwell.
Taylor, L. (2004), *Reconstructing Macroeconomics: Structuralist Proposals and Critiques of the Mainstream*, Cambridge, MA: Harvard University Press.
Taylor, L. and C. Rada (2003), 'Debt–equity cycles in the 20th Century: empirical evidence and a dynamic Keynesian model', New York, NY: Center for Economic Policy Analysis, New School University.

13. Technological innovation, financial fragility and complex dynamics*

Alberto Russo, Domenico Delli Gatti and Mauro Gallegati

13.1. INTRODUCTION

In this chapter we develop a heterogeneous interacting agents (HIAs) model suitable to replicate well-known empirical regularities of industrial dynamics, such as the power law distribution of firm size (Okuyama et al., 1999; Ramsden and Kiss-Haypal 2000; Axtell, 2001; Gaffeo et al., 2003) and the Laplace distribution of firm growth rates (Stanley et al., 1996; Bottazzi and Secchi, 2003).

In our model the analysis of the business cycle is related with the properties and evolution of the power law distribution of firm size. The power law distribution was originally discovered by Vilfredo Pareto who maintained that the distribution of personal incomes above a certain threshold y_0 follows a heavy-tailed distribution (Pareto, 1897). In particular, he found that the probability of observing an income Y greater than or equal to y is proportional to a power of y, that is $Pr(Y \geq y) \propto y^{-\alpha}$, with α close to 1.5. Our aim is to propose a suitable agent-based model to analyse the evolution of power law distributions and discover how interaction among heterogeneous agents works and influences the relation between the distribution of agents' variables and business cycle fluctuations. For instance, a major consequence of a heavy-tailed firm size distribution is that small idiosyncratic shocks can generate large aggregate fluctuations even in the absence of aggregate shocks; therefore, understanding how firm size distribution changes and interacts with the business cycle is an important goal for economic research with considerable consequences for policy intervention.[1]

In general, in order to account for the scaling-type stylized facts emerging from studies on industrial dynamics (Stanley et al., 1996; Axtell, 2001), business cycles (Delli Gatti et al., 2004a), and financial markets (Mantegna and Stanley, 2000), we believe that economists have to adopt a methodological approach based on heterogeneous interacting agents (HIA), rejecting the 'reductionist' approach centred on the 'representative agent' hypothesis. In particular, two works have stressed the limits of this approach: Kirman (1992), from a theoretical point of view, and Stoker (1993), from an empirical perspective. In sum, the practice of combining heterogeneity and interactions is at odds with mainstream economics which reduces the analysis of the aggregate to that of a single representative agent and which is unable, by construction, to explain non-normal distributions, scaling behaviour, self-similarity, self-organizing criticality (Bak, 1996) or the occurrence of large aggregate fluctuations as a consequence of small idiosyncratic shocks.

Starting from Gallegati et al. (2003) and Delli Gatti et al. (2005), we have developed an agent-based model by extending the initial framework, in which a large number of firms interact with a banking sector giving rise to complex dynamics, through the introduction of a labour-saving technological progress and a wage–firm size relationship.

In this model, in discussing a scaling approach to business fluctuations, we are particularly interested in the analysis of the evolution and shifts of the distribution of firm size. Although some work on this topic has been pursued in the last decade in physics, econophysics literature has only sporadically dealt with the issue. Scarce attention has been paid so far to establishing a link between power law shifts and the business cycle theory, mainly because mainstream economics lacks adequate conceptual and analytical tools to accomplish such an endeavour.

The analysis of how firm-size distribution changes and interacts with the business cycle is performed by introducing new elements, mainly related to the role of labour-saving technological progress, into the financial fragility framework with heterogeneous firms proposed by Gallegati et al. (2003) and Delli Gatti et al. (2005). We provide a tentative interpretation of the shifting behaviour of the power law distribution along business cycles based on the interplay among R&D investments, technological progress, wage dynamics, firms' productivity and financial factors.[2] In particular, we focus on the shifts of the production function towards the origin as an indicator of ongoing technological development, that is, a sequence of periodic arrival of innovations that leads to a permanent improvement in the production function (Schumpeter, 1939). Therefore, firms' productivity is proxied by the capital–labour ratio increasing over time due to labour-saving technological innovation. In addition, our agent-based model reproduces some of the growth-type stylized facts provided by Kaldor (1961) and a Goodwin-like growth cycle (1967).[3]

In general, our analysis suggests that there are significant changes in firms' distribution during different phases of the business cycle and that the power law scaling behaviour, emerging as an invariant feature of size distribution of firms, is at the basis of the understanding of business fluctuations.

The model is presented in Section 13.2. The model simulation and the discussion of the results are in Section 13.3. Section 13.4 concludes.

13.2. THE MODEL

The structure of the model can be divided into two parts: the real side and the financial side of the economy. The real (supply) side of the model – goods market – is characterized by the behaviour of firms that sell all output at a stochastic price and invest resources in R&D so as to obtain innovations. The financial side – credit market – is constituted by a banking sector that, in the presence of asymmetric information, allocates the supply of credit among firms on the basis of the collaterals they can provide.

13.2.1. The Goods Market

Firms produce a homogeneous good by means of the following production function

$$Y_{it} = \phi K_{it} \qquad (i = 1, \ldots, F \text{ and } t = 1, \ldots, T) \tag{13.1}$$

where K_{it} is the stock of capital of the i-th firm, ϕ is the productivity of capital, constant and uniform across firms, F is a 'large' number of firms and T is the length of the period of time considered.

In order to produce the output, firms need a given amount of labour N_{it} depending on its capital–labour ratio

$$\lambda_{it} = \frac{K_{it}}{N_{it}} \tag{13.2}$$

Consequently, each firm has a labour requirement function $N_{it} = K_{it} / \lambda_{it}$.[4] There are no constraints on the labour market, that is, firms can hire (and fire) all the workers they need at the wage:

$$w_{it} = \rho\left(\delta K_{it}^{\varepsilon}\right) + (1-\rho) w_{it-1} \tag{13.3}$$

where $0 < \rho < 1$, $0 < \delta < 1$, and $0 < \varepsilon < 1$. We simply assume that there is a wage–size relation[5] combined with an adaptive term in the setting of the wage paid to workers.[6]

Due to limited knowledge of market conditions, firms sell their output at an uncertain (relative) price. The individual selling price, p_{it}, is a random variable with $E(p_{it}) = 1$ and finite variance; therefore p_{it} is an idiosyncratic shock.[7]

The balance sheet of the firm is:

$$K_{it} = A_{it} + L_{it} \tag{13.4}$$

where A_{it} is the equity base and L_{it} is the demand for credit.

The firm's profit is equal to:

$$\pi_{it} = p_{ti} Y_{ti} - r_{ti} K_{it} - w_{it} N_{it} = p_{it} Y_{it} - \left(r_{it} + \frac{w_{it}}{\lambda_{it}} \right) K_{it} \tag{13.5}$$

where r_{it} is the interest rate faced by the i-th firm.

Firms invest a portion of retained profits in R&D activity so as to obtain innovations in the upcoming periods:

$$RD_{it} = \begin{cases} \sigma \pi_{it-1} & if \quad \pi_{it-1} > 0 \\ 0 & otherwise \end{cases} \tag{13.6}$$

where $0 < \sigma < 1$.[8]

Consequently, profits after R&D expenditure are

$$\pi'_{it-1} = \pi_{it-1} - RD_{it} = (1 - \sigma) \pi_{it-1} \tag{13.7}$$

Firms' technological level z_{it} enhances due to a Poisson distributed process depending on the R&D investments and due to the possibility of imitating other firms.

Accordingly, the evolution of technology due to the internal innovations made by firms is given by

$$z'_{it} = \begin{cases} z_{it}^{P} & if \quad \pi_{it-1} > 0 \\ 0 & if \quad \pi_{it-1} \le 0 \end{cases} \tag{13.8}$$

where z_{it}^P is the number of innovations at time t for firm i, that is, the realization of a Poisson process with mean $\mu_{it} = RD_{it-1} / K_{it-1}$.[9]

The imitation process is simply characterized by a mean-interaction term, that is, by a term proportional to the average technological level of firms. Then, the level of technology due to imitation is equal to

$$z_{it}'' = v\bar{z}_{t-1} \tag{13.9}$$

where $\bar{z}_{t-1}$ is the average technological level in the past period and $0 < v < 1$.

The technological level of the i-th firm at period t is equal to

$$z_{it} = z_{it-1} + z_{it}' + z_{it}'' \tag{13.10}$$

Finally, the capital–labour ratio is a function of the technological level :

$$\lambda_{it} = \gamma z_{it-1} \tag{13.11}$$

where $0 \leq \gamma \leq 1$.

Each firm maximizes an objective function:

$$\Gamma_{it} = E(\pi_{it}) - E(BC_{it}) \tag{13.12}$$

where $E(\pi_{it})$ is the expected profit and $E(BC_{it})$ is the expected bankruptcy cost. We assume a quadratic functional form for the bankruptcy cost:[10]

$$BC_{it} = cY_{it}^2 \tag{13.13}$$

A firm goes bankrupt if the net worth becomes negative. Then, the bankruptcy condition is[11]

$$A_{it+1} = A_{it} + \pi_{it} < 0 \tag{13.14}$$

Substituting (13.5) in (13.14), we obtain

$$p_{it} < \frac{\left(r_{it} + w_{it} / \lambda_{it}\right) K_{it} - A_{it}}{Y_{it}} \tag{13.15}$$

where

$$\bar{p}_{it} \equiv \frac{\left(r_{it} + w_{it} / \lambda_{it}\right) K_{it} - A_{it}}{Y_{it}} \tag{13.16}$$

is the critical value for the relative price of a firm below which bankruptcy occurs. For the sake of simplicity, we assume that p_{it} is a uniformly distributed variable with support (0,2); consequently, the probability of bankruptcy is

$$BP_{it} = \frac{\left(r_{it}\lambda_{it} + w_{it}\right)K_{it} - A_{it}\lambda_{it}}{2Y_{it}\lambda_{it}} \tag{13.17}$$

and the expected bankruptcy cost is

$$E(BC_{it}) = \frac{c}{2}\phi K_{it}\left[\left(r_{it} + \frac{w_{it}}{\lambda_{it}}\right)K_{it} - A_{it}\right]$$

After that, the firm's objective function, $\Gamma_{it} = E(\pi_{it}) - E(BC_{it})$, becomes:

$$\Gamma_{it} = \phi K_{it} - r_{it}K_{it} - \frac{w_{it}}{\lambda_{it}}K_{it} - \frac{c}{2}\phi K_{it}\left[\left(r_{it} + \frac{w_{it}}{\lambda_{it}}\right)K_{it} - A_{it}\right] \tag{13.18}$$

From the maximization of the objective function (13.18), we obtain the optimal capital stock; the first order condition is:

$$\frac{\partial \Gamma_{it}}{\partial K_{it}} = \phi - \left(r_{it} + \frac{w_{it}}{\lambda_{it}}\right) - c\phi\left(r_{it} + \frac{w_{it}}{\lambda_{it}}\right)K_{it} + \frac{c}{2}\phi A_{it} = 0$$

hence

$$c\phi\left(r_{it} + \frac{w_{it}}{\lambda_{it}}\right)K_{it} = \phi - \left(r_{it} + \frac{w_{it}}{\lambda_{it}}\right) + \frac{c}{2}\phi A_{it}$$

and finally

$$K_{it} = \frac{(\phi - r_{it})\lambda_{it} - w_{it}}{c\phi\left(r_{it}\lambda_{it} + w_{it}\right)} + \frac{A_{it}\lambda_{it}}{2\left(r_{it}\lambda_{it} + w_{it}\right)} \tag{13.19}$$

Investment is equal to

$$I_{it} = K_{it} - K_{it-1}. \tag{13.20}$$

The demand for credit is:

$$L_{it}^{d} = L_{it-1} - \pi_{it-1} + I_{it} = K_{it} - A_{it} \tag{13.21}$$

Finally, substituting (13.19) in (13.21) we have the following relation for firms' loans:

$$L_{it}^{d} = \frac{(\phi - r_{it})\lambda_{it} - w_{it}}{c\phi(r_{it}\lambda_{it} + w_{it})} + \frac{\left[\lambda_{it} - 2(r_{it}\lambda_{it} + w_{it})\right]A_{it}}{2(r_{it}\lambda_{it} + w_{it})} \tag{13.22}$$

13.2.2. The Credit Market

The banking sector is modelled as in Gallegati et al. (2003). Thus there is a bank[12] that allocates the total supply of credit among firms according to relative firm size:

$$L_{it}^{s} = L_t\left(\frac{K_{it-1}}{K_{t-1}}\right) \tag{13.23}$$

where L_t is the total supply of credit at time *t*, K_{it-1} is the capital of the *i*-th firm, and K_{t-1} is the aggregate stock of capital in the precedent period of time. This rule of credit allocation is one way of tackling asymmetric information in the credit market: the bank does not know the 'true' financial conditions of the heterogeneous borrowers and uses collaterals, proxied by the capital stock of the firm relative to the aggregate stock of capital, to determine the individual supply of credit.

The supply of credit is vertical (it is independent of the interest rate) at the level

$$L_t = \frac{E_{t-1}}{\alpha} \tag{13.24}$$

where α is a coefficient of risk (for example, a prudential rule set up by a regulatory institution) that the bank has to respect and E_{t-1} is the equity base of the bank in the previous period of time.

The balance sheet of the bank is

$$L_t = E_t - D_t \tag{13.25}$$

where D_t are deposits.

The bank's equity base is equal to

$$E_t = E_{t-1} + \Pi_t - \sum_i B_{it-1} \tag{13.26}$$

where Π_t is the bank's profit and B_{it} is the 'bad debt' of a bankrupted firm, that is

$$B_{it} = \begin{cases} -A_{it} & if \quad A_{it} < 0 \\ 0 & if \quad A_{it} \geq 0 \end{cases} \tag{13.27}$$

The bank's profit is

$$\Pi_t = \sum_i r_{it} L_{it} - (1-\omega)\overline{r}_t D_{t-1} - \overline{r}_t E_{t-1} \tag{13.28}$$

where $\overline{r}_t$ is the average interest rate and ω is the mark-up for the bank.[13]

Finally, the individual rate of interest is endogenously determined when $L^d_{it} = L^s_{it}$, that is, when (13.22) is equal to (13.23):

$$r^*_{it} = \frac{2 + cA_{it}}{2c\left[L_t\left(\frac{K_{it-1}}{K_{t-1}}\right) + \frac{1}{c\phi} + A_{it}\right]} - \frac{w_{it}}{\lambda_{it}} \tag{13.29}$$

13.3. SIMULATING THE MODEL

We simulate an artificial economy with $F = 1000$ firms and a banking sector under the assumption that if a firm goes bankrupt it is replaced by a new firm (with initial conditions), so that F is fixed.[14]

As we can see in Figure 13.1, the aggregate output fluctuates showing phases of smooth growth and periods of large variability;[15] in addition, sudden drifts and different slopes appear from time to time.[16]

The growth process is due to the growth of firm size and to productivity enhancements. Note that we model a supply-driven economic system in which all output produced by firms is demanded at a stochastic price. Hence, the growth of firms due to investment choices and financial factors has no (quantity) constraints from the demand side, even if the volatility of prices has important consequences on firms' dynamics.

Importantly, a domino effect through a balance sheet contagion may develop because of firms' bankruptcies. In fact, when a firm goes bankrupt it leaves the market and it does not pay back the debt to the bank.

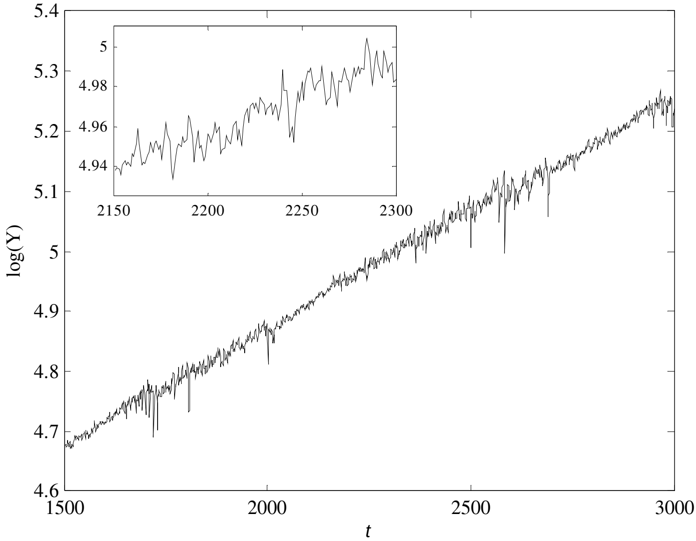


Figure 13.1 Aggregate output

Consequently, the bank has a 'bad debt' and the total supply of credit diminishes, producing an increase in the interest faced by surviving firms. Since debt commitments rise, firms' insolvencies increase even further, thus self-reinforcing this vicious circle.

In addition, firms' growth is due to technological progress, that is, labour-saving innovations due to R&D investments: when a firm obtains an innovation it can produce the same output with a smaller amount of labour input. It can then accumulate more capital and grow faster than firms that have not innovated. In particular, we focus on the shifts of the production function towards the origin as an indicator of ongoing technological development (that is, labour-saving innovations that allow the same output to be produced using less input), instead of analysing the shifts along the production function due to factor substitution (Schumpeter, 1939).[17]

The capital–labour ratio grows along time due to diffused technological progress that saves labour inputs in the production process. We can see the time course of this ratio in Figure 13.2. The model simulation reproduces some important growth-type stylized facts (Kaldor, 1961):[18]

- the capital–labour ratio (Figure 13.2) and the output–labour ratio increase over time (due to labour-saving technological progress);
- the capital–output ratio is constant;[19]
- the investment–output ratio is roughly constant;

- the rate of return on capital is roughly constant;[20]
- the real wage increases over time;
- the relative share of capital and the relative share of labour are roughly constant.[21]

Firms are characterized by an asymmetric distribution of capital–labour ratios (Figure 13.3). Since productivity improvements are due to an incremental innovation process and to an imitation term, older firms that have had positive profits for many periods of time are more likely to have higher capital–labour ratios with respect to young firms with no R&D experience and a limited time to imitate others. Firm size distribution (FSD) is right-skewed and it is distributed according to a power law (Figures 13.4 and 13.5). The model simulation thus ably replicates a stylized fact that the empirical literature on industrial dynamics has recently highlighted (Okuyama et al., 1999; Ramsden and Kiss-Haypal, 2000; Axtell, 2001; Gaffeo et al., 2003).

Moreover, Gaffeo et al. (2003) find that there are significant shifts of the FSD during different phases of the business cycle. In other words, power law is a persistent but not time-invariant feature of the FSD. In the following, we will show that our agent-based model is able to replicate also the shifting behaviour of the FSD.

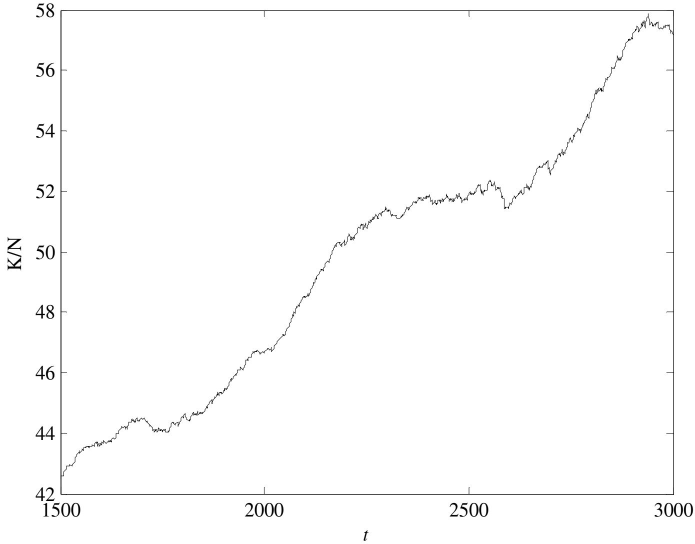


Figure 13.2 The time evolution of the capital–labour ratio of the economy

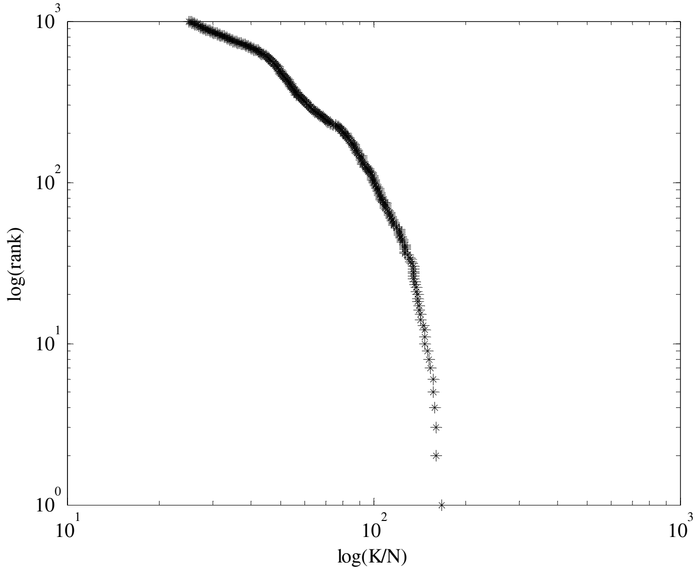


Figure 13.3 The distribution of the capital–labour ratio across firms

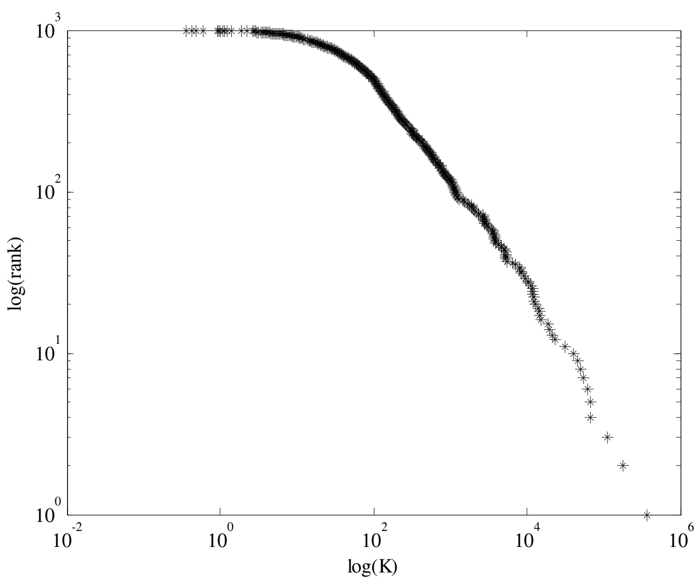


Figure 13.4 Power law distribution of firm size (proxied by capital)

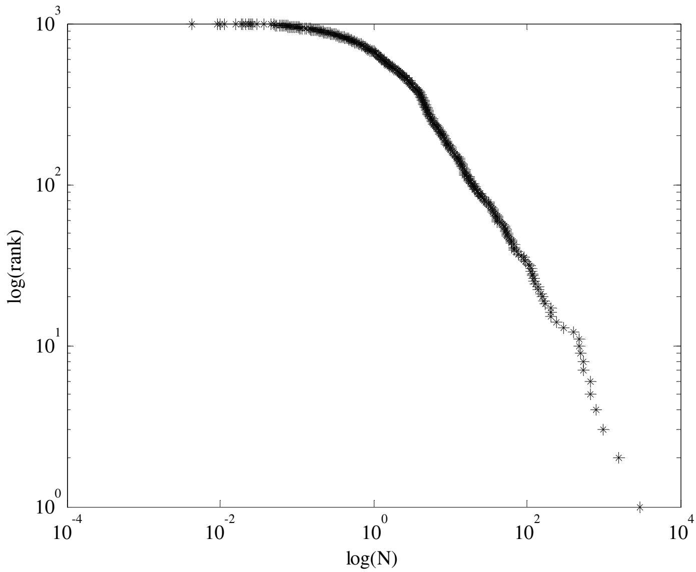


Figure 13.5 Power law distribution of firm size (proxied by number of employees)

Another important stylized fact (Stanley et al., 1996; Bottazzi and Secchi, 2003) is about the firm growth rates that follow a Laplace distribution (Figure 13.6a). Aggregate output growth rates are also tent-shaped (Figure 13.6b). Interestingly enough, simulations show that behaviour of the largest units (the industrial sector) reproduces the behaviour of smaller units (firms) (Lee et al., 1998).[22]

In order to obtain an explanation of the shifting behaviour of the FSD, we will propose a simple economic mechanism based on the interplay among R&D investment, technological innovation, firms' productivity and wage dynamics.

In this model we assume a firm size–wage relationship as a simply way to determine the wage that firms pay to workers (jointly with an adaptive term).[23] Since there is capital growth in the economy, the average wage level increases in time.

Let us analyse the joint behaviour of wage levels and productivity dynamics, given that there are different implications for firms, depending on the size and the capital–labour ratio. In particular, we examine the behaviour of the ratio between the average wage paid to workers and the labour productivity. Figure 13.7 shows that the wage–productivity ratio fluctuates and presents many cycles of different length. Clearly, an increment

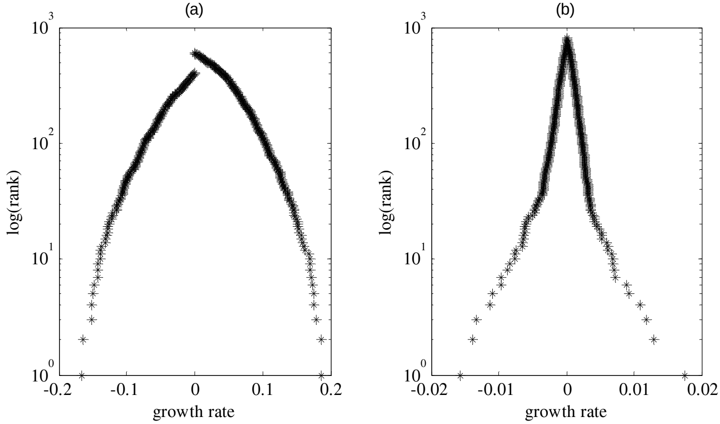


Figure 13.6 Laplace distribution of (a) firm growth rates, and (b) aggregate growth rates

(decrement) in the ratio can be due to an increase (decrease) in the average level of wage or to a decline (increase) in firms' productivity.

The typical shape of a business cycle that we analyse has the following structure. Firms accumulate capital due to technological progress that allows the production of the same output using less labour as input (for example, labour productivity increases). The growth of firm size implied by labour-saving innovations and financial factors generates an increase in wages, due to the wage–firm size relationship and a shift towards north-east of the firm size power law distribution in the double logarithmic space (from distribution A to distribution B in Figure 13.8). Since the wage–productivity ratio increases in this phase of the cycle, wage levels grow faster than firms' productivity (firms' productivity increases throughout the cycle, as we can see in Figure 13.9). This process continues until the wage level reaches the peak of the cycle, after which the capital size of firms starts to diminish.[24] Consequently, wages decrease and the power law moves towards south-west (from B to C in Figure 13.8).[25]

The fluctuating behaviour of the wage–productivity ratio suggests that in the model there is also a Goodwin-like growth cycle at work, that is a cyclical relationship between workers' wage and firms' profit. In fact, we can see the wage–productivity ratio as equivalent to the relative share of labour (see Figure 13.7).[26]

What happens in the model is that firms accumulate capital (due to technological and financial factors) and, because of a wage–firm size relationship, growing firms pay higher wages. In other words, firms' capital

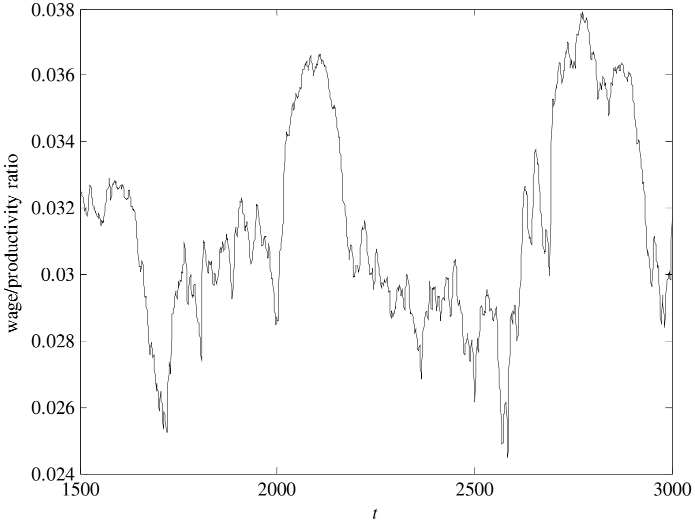


Figure 13.7 Joint evolution of average wage and labour productivity

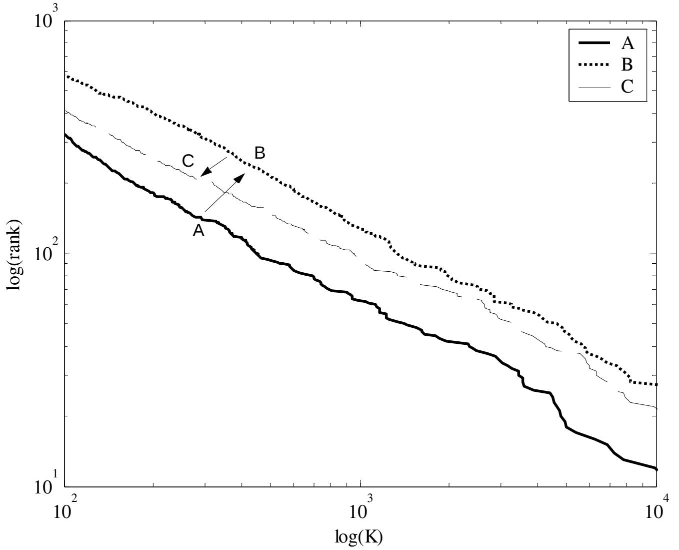


Figure 13.8 Power law shifts along the business cycle. The distributions are relative to periods: 2570 (A); 2775 (B); 2973 (C). In Figure 13.7 these three periods correspond to the beginning (A), the peak (B) and the end (C) of a business cycle.

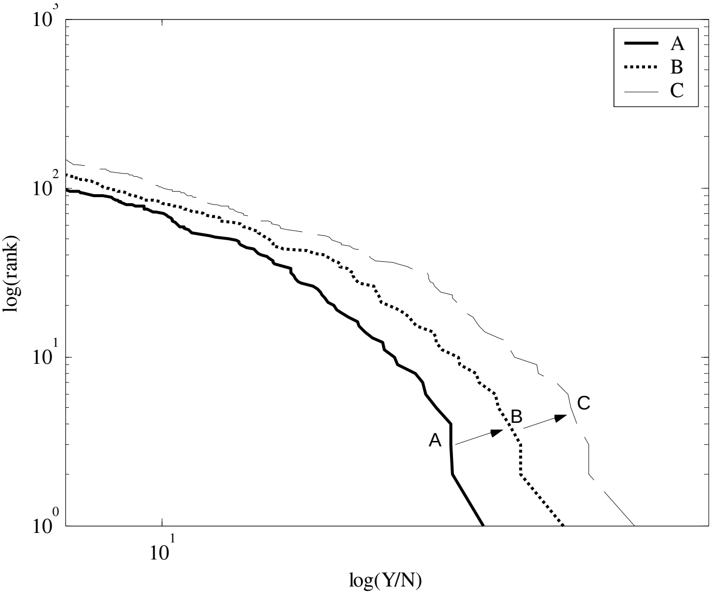


Figure 13.9 Firms' productivity shifts

accumulation increases the labour demand and, at the same time, the wage levels.[27] Consequently, the relative share of labour increases, while firms' profits and investments diminish. This implies a lower capital accumulation that generates a decrease in wages and thus a decline in the relative share of labour, producing the condition for the capital accumulation to re-start.

13.4. CONCLUSION

In this chapter we propose a heterogeneous interacting agents (HIAs) model in which a large number of financially fragile firms interact with a banking sector giving rise to complex dynamics. In particular, we extend the framework proposed by Gallegati et al. (2003) and Delli Gatti et al. (2005) focusing on the role of technological progress and its effects on firms' dynamics. We propose a simple economic mechanism based on the interplay among R&D investments, technological progress, wage dynamics, firms' productivity and financial factors, providing a tentative explanation of the shifting behaviour of firm size distribution (FSD) along business cycles. Assuming a wage–firm size relationship and considering that firms obtain productivity enhancements by means of labour-saving innovations, we find that FSD shifts are linked to the co-movement of wages and labour productivity. We find that the model simulation also replicates important growth-type stylized facts (Kaldor, 1961) and a dynamic relationship between workers' wages and firms' profits (Goodwin, 1967).

APPENDIX A: PARAMETER SETTING AND INITIAL CONDITIONS.

The parameter values and initial conditions used in the numerical simulations of the model are shown below.[28]

(i) Firms' specific parameters:
$\phi = 0.1$ (productivity of capital); $c = 1$ (bankruptcy function parameter); $\sigma = 0.5$ (weight in the wage equation); $\delta = 0.01$ (wage equation parameter); $\varepsilon = 0.5$ (wage equation parameter); $\nu = 0.001$ (imitation coefficient); $\gamma = 1$ (technological level vs. capital–labour ratio); $\sigma = 0.075$ (percentage of retained profit invested in R&D).

(ii) Firms' initial conditions:
$A_{i0} = 20$ (equity base); $L_{i0} = 80$ (loan); $B_{i0} = 0$ (bad debt); $\lambda_{i0} = 25$ (capital–labour ratio); $w_{i0} = 0.1$ (individual wage).

(iii) Bank's specific parameters:
$\alpha = 0.45$ (risk coefficient); $\omega = 0.01$ (mark-up).

(iv) Bank's initial conditions:
$L_0 = \Sigma L_{i0} = 8000$ (total supply of credit); $E_0 = \alpha L_0 = 2000$ (equity base); $D_0 = L_0 - E_0 = 6000$ (deposits); $\Pi_0 = 0$ (profit).

NOTES

* We are grateful for comments and suggestions to the participants at the Conference. We have also benefited from comments and suggestions received at the '9th WEHIA, Workshop on Economics and Heterogeneous Interacting Agents', May 27–29, 2004, Kyoto University, Kyoto, Japan, the '10th International Schumpeter Society Conference – Schumpeter 2004', June 9–12, 2004, Bocconi University, Milan, Italy; the '1st Workshop on Financial Fragility, Business Fluctuations and Monetary Policy', September 13–14, 2004, Graduate College Santa Chiara, University of Siena, Siena, Italy. Special thanks to Rainer Andergassen. We also thank two anonymous referees. The usual disclaimers apply.

1. For a more detailed discussion of power law distribution of firm size and, in general, scaling properties of industrial dynamics see Delli Gatti et al. (2004b, 2005). A good introduction to

power law behaviour is provided by Reed (2001). Some basic concepts about scaling behaviour in economics and its importance for empirical and theoretical researchers is provided by Brock (1999). Furthermore, an interesting approach to understand the 'granular origin' of business cycle fluctuations is proposed by Gabaix (2004). This author argues that, in a world with heavy-tailed firm size distributions, small idiosyncratic firm-level shocks can aggregate up to large aggregate fluctuations.

2. To be more precise, the distinctive features of this chapter with respect to the basic framework proposed by Gallegati et al. (2003) and Delli Gatti et al. (2005) are the following: the introduction of the labour input in the production function (the papers mentioned have one productive input, for example, capital) and of a simple wage-setting mechanism, a labour-saving innovation process based on firms' R&D investments, a different mechanism of interaction among variables and the business cycle with a focus on the role of labour productivity and wage dynamics.
3. Analysing the statistical regularities regarding financial markets and personal wealth distribution, Solomon and Richmond (2001) show that the analogy of the prey-predator model, that is a generalized Lotka–Volterra (GLV) model, can be used to explain economic and financial scaling phenomena as the emergence of a power law tail in the distribution of individual variables.
4. Alternatively, we can view the production function in the following way: $Y_{it} = \phi\lambda_{it}N_{it}$
5. Many empirical studies have found the existence of a strong positive relationship between employer's size and wages, emphasizing different aspects of wage formation – labour quality, efficiency wages, and so on – and institutional factors – working conditions, the role of unions, and so on. Brown and Medoff (1989), in their seminal paper, find a positive and significant employer's size–wages effect using US data. A possible theoretical explanation is related to the role of unions: larger firms are subject to greater union influence in wage determination with respect to smaller firms. We simply assume this wage–firm size relationship instead of reproducing it by means of our agent-based model. This is only a preliminary step toward a more complete model (see also the following footnote).
6. A further improvement in the model will be a matching mechanism between firms and workers to jointly determine wage and employment levels, in a way that allow us to explain the wage–size effect as one of the emerging properties of the model.
7. As in the *leveraged aggregate supply* class of models first developed by Greenwald and Stiglitz (1990, 1993), our model describes a supply-side economy in which consumers buy all the produced output at a stochastic price, the difference across heterogeneous prices for buying a homogeneous good being due to asymmetric information across individuals. For example, Salop (1977) shows that firms producing a homogenous good can set different prices if consumers have different costs to obtain or process information.
8. This implies that the initial equity base of the firm is equal to $A_{it} = A_{it-1} + (\pi_{it-1} - RD_{it})$.
9. According to this relation, the effect of R&D investments on innovations is scaled by a factor K_{it-1}.
10. As in Greenwald and Stiglitz (1993), the probability of bankruptcy is incorporated directly into the firm's profit function because going bankrupt costs, and such a cost is increasing in the firm's output. This assumption is made largely for analytic reasons, but one of the possible economic justifications for it is the following: 'as firms become larger, they presumably involve managers whose loss of position, income, and power in the event of insolvency is likely to increase. Bankruptcy should, therefore, be a more serious matter for General Motors than for a local grocery store' (Greenwald and Stiglitz, 1993, p. 89). Thus, following Greenwald and Stiglitz (1993), we stress the role of risk considerations (firms are averse to the bankruptcy risk that increases as they produce more) on firms' production decisions'.

11. Remember that the initial equity base is $A_{it} = A_{it-1} + \pi'_{it-1}$ because a fraction of retained profits obtained in period $t - 1$ is invested in R&D.
12. We can interpret the only bank in the model as a vertically integrated banking sector.
13. We assume that returns on a bank's equity are given by the average lending interest rate and that a spread, $1 - (1/\omega)$, between lending and borrowing rates exists.
14. The parameter values and the initial conditions are illustrated in Appendix A.
15. All figures presented below concern simulations from period 1500 to period 3000. The first part of the simulation is considered as a 'transition phase'. Then, we analyse simulation data after the relevant random variables (for example, firm size) converge to a power law distribution, that is a stable distribution that once emerged continues to characterize firm size distribution well beyond the period 3000 (even though it may shift along business cycles, as we will see below); thus, the choice of analysing the simulation period 1500–3000 is made for expositional convenience, given that the simulation (qualitative) results are the same even after period 3000.
16. For a discussion of other important features of the model (without considering technological progress) see Gallegati et al. (2003) and Delli Gatti et al. (2004b, 2005). In particular, Delli Gatti et al. (2004b, 2005) show that the total output standard deviation and its autocorrelation are within the 5 per cent confidence interval with respect to quarterly real data.
17. Here we are analysing only the effect of a diffused innovation process on the labour quantity used in the production process. A further improvement in the model concerns the possibility of modelling the monopolistic power of a firm that can diminish the price of the output as a consequence of an innovation, that is the competitive advantage of an innovation for a firm that can sell its output at a price lower than that of its competitors.
18. It is worth noting that 'Kaldorian' stylized facts emerging from simulations are long run with respect to the time scale of the model.
19. By construction, see equation (13.1).
20. The rate of return on capital is equal to the ratio between the profit and the capital of firms.
21. This result holds for the setting of the model's parameters that guarantees the 'viability condition' of the prey–predator cycle at work between firms and workers.
22. For a general comparison between simulation and empirical data relative to scaling, industrial, financial and business cycle stylized facts, see Delli Gatti et al. (2004b).
23. See equation (13.3).
24. In the following we provide a Goodwinian interpretation of this fact.
25. In addition, shifts of the firm size power law can also affect the slope of the distribution. See Delli Gatti et al. (2004a) on different slopes of power law in expansions and recessions.
26. The relative share of labour is equal to wN/Y, where wN is the wage bill (w is the wage and N the number of employees) and Y is the output. We can consider it in the following way: $w/(Y/N)$, where Y/N is the labour productivity.
27. In other words, if we consider that larger firms are subject to higher union power with respect to smaller ones, the wage–firm size relationship that we assume in the model implies that the bargaining power of workers improves when firm size increase.
28. In order to assess the robustness of our model we think that a further step could be taken in the direction of empirical calibrating simulations as suggested by Werker and Brenner (2004): 'In economics simulation models are used quite a lot to carry out mathematical experiments. However, the specification of the parameter set with which to run these simulations is, in general, quite an adventure into the unknown. Criticism is easily found with the procedure, as it is difficult to justify why to choose one specification of parameters and not another – especially if the results found in the simulation models are striking. Then, the audience cannot help but think that there has been quite some arbitrary trial and error

going on to achieve this. To avoid this impression, we suggest to empirically calibrating (sic) simulation models in a way that makes their results more acceptable. However, compared with models of mainstream economics that usually can be solved analytically, simulation models have only recently been opened to empirical data. It is fair to say that also analytically solvable models have quite some problems concerning the integration of empirical data (Kydland and Prescott, 1996). These problems also emerge when working with simulation models.'

REFERENCES

Axtell, R. (2001), 'Zipf distribution of U.S. firm sizes', *Science*, **293**:1818–20.

Bak, P. (1996), *How Nature Works*, Oxford: Oxford University Press.

Bottazzi, G. and A. Secchi (2003), 'Why are distributions of firm growth rates tent-shaped?', *Economic Letters*, **80**:415–20.

Brock, W. (1999), 'Scaling in economics: a reader's guide', *Industrial and Corporate Change*, **8**:409–46.

Brown, C. and J.C. Medoff (1989), 'The employer size wage effect', *Journal of Political Economy*, **97**:1027–59.

Delli Gatti, D., C. Di Guilmi, E. Gaffeo, G. Giulioni, M. Gallegati and A. Palestrini (2004a), 'Business cycle fluctuations and firms size distribution dynamics', *Advances in Complex Systems*, **7**:223–40.

Delli Gatti, D., E. Gaffeo, G. Giulioni, M. Gallegati, A. Kirman, A. Palestrini and A. Russo (2004b), 'Complex dynamics and empirical evidence', http://ssrn.com/abstract=545223.

Delli Gatti, D., C. Di Guilmi, E. Gaffeo, G. Giulioni, M. Gallegati and A. Palestrini (2005), 'A new approach to business fluctuations: heterogeneous interacting agents, scaling laws, and financial fragility', *Journal of Economic Behavior and Organization*, **56**:489–512.

Gabaix, X. (2004), 'The granular origins of aggregate fluctuations', MIT, mimeo.

Gaffeo, E., M. Gallegati and A. Palestrini (2003), 'On the size distribution of firms: additional evidence from the G7 countries', *Physica A*, **324**:117–23.

Gallegati, M., G. Giulioni and N. Kichiji (2003), 'Complex dynamics and financial fragility in an agent-based model', *Advances in Complex Systems*, **6**(3):267–82.

Goodwin, R.M. (1967), 'A growth cycle', in C.H. Feinstein (ed.), *Socialism, Capitalism and Economic Growth*, Cambridge: Cambridge University Press, pp. 54–8.

Greenwald, B. and J. Stiglitz (1990), 'Macroeconomic models with equity and credit rationing', in R. Hubbard (ed.), *Information, Capital Markets and Investment*, Chicago: Chicago University Press, pp. 15–42.

Greenwald, B. and J. Stiglitz (1993), 'Financial markets imperfections and business cycles', *Quarterly Journal of Economics*, **108**:77–113.

Kaldor, N. (1961), 'Capital accumulation and economic growth', in F.A. Lutz and D.C. Hague (eds), *The Theory of Capital*, New York: St. Martin's Press, pp. 177–222.

Kirman, A. (1992), ‘Whom or what does the representative agent represent?’, *Journal of Economic Perspectives*, **6**:117–36.
Kydland, F. and E. Prescott (1996), ‘The computational experiment: an econometric tool’, *Journal of Economic Perspectives*, **10**:69–85.
Lee, Y., L. Amaral, D. Canning, M. Meyer and E. Stanley (1998), ‘Universal features in the growth dynamics of complex organizations’, *Physical Review Letters*, **81**:3275–78.
Mantegna, R. and E. Stanley (2000), *An Introduction to Econophysics*, Cambridge: Cambridge University Press.
Okuyama, K., M. Takayasu and H. Takayasu (1999), ‘Zipf’s law in income distribution of companies’, *Physica A*, **269**:125–31.
Pareto, V. (1897), *Course d’economie politique*, Vol. 2, Paris: Pichou.
Ramsden, J. and G. Kiss-Haypal (2000), ‘Company size distribution in different countries’, *Physica A*, **277**:220–27.
Reed, W. (2001), ‘The Pareto, Zipf and other power laws’, *Economic Letters*, **74**:15–19.
Salop, S. (1977), ‘The noisy monopolist: imperfect information, price dispersion, and price discrimination’, *The Review of Economic Studies*, **44**:393–406.
Schumpeter, J. (1939), *Business Cycles*, New York: McGraw Hill.
Solomon, S. and P. Richmond (2001), ‘Stability of Pareto-Zipf law in non-stationary economies’, in A. Kirman and J.-B. Zimmermann (eds), *Economics with Heterogeneous Interacting Agents*, Berlin: Springer.
Stanley, M., L. Amaral, S. Buldyrev, S. Havlin, H. Leschorn, P. Maas, M. Salinger and E. Stanley (1996), ‘Scaling behavior in the growth of companies’, *Nature*, **379**:804–6.
Stoker, T. (1993), ‘Empirical approaches to the problem of aggregation over individuals’, *Journal of Economic Literature*, **21**:1827–74.
Werker, C. and T. Brenner (2004), ‘Empirical calibration of simulation models’, *Working Paper* 04.13, June, Eindhoven Centre for Innovation Studies.

14. Growth, history and institutions*

Graziella Bertocchi

14.1. INTRODUCTION

This contribution aims to illustrate a selection of applications of the ongoing research which has developed in recent years around the combination of three main ingredients: growth theory, the theory of institutions and their interrelationship with history. We shall refer to this body of the economic literature as the Growth, History and Institutions (GHI) research line. The revival of growth theory during the 1980s, building on Solow's (1956) seminal contribution, is where this research line has its deepest roots. Romer (1986) and Lucas (1988) transform the Solowian model into a more useful tool to comprehend the post-war persistence of differences in the growth performances across countries of the world. Endogenous technical progress and human capital are the key new concepts of this earlier stage. The two additional features that are added to this basis within the GHI research line are, first, a historical dimension and, second but no less important, an institutional one, the latter owing an intellectual debt to the theory of institutions based on North's (1981, 1990) earlier contributions.

The rest of the chapter is organized as follows. Section 14.2 illustrates how each of the two new key ingredients – history and institutions – adds to the field of economic growth, both at the theoretical and empirical level. Section 14.3 summarises the motivations and implications of the GHI research line, and presents a few applications to specific issues. Building on these applications, Section 14.4 suggests the emergence of a new standard, for future research on long-term economic growth, which extends the post-WWII current benchmark as far back as the second half of the 19th century, as the new relevant time period for economic investigations both at the empirical and theoretical level, and indicates directions along which future research on GHI may fruitfully evolve. Section 14.5 concludes with policy implications.

14.2. THE INTERACTIONS BETWEEN GROWTH AND HISTORY, AND GROWTH AND INSTITUTIONS

14.2.1. The Historical Dimension

Before the GHI research line started to develop, the time perspective for the vast majority of the theoretical and empirical investigations on growth and development was that of the post-WWII period. Data availability constituted a crucial constraint that conditioned this perspective. Extended data collection projects starting from the 1950s, such as the Penn World Table (Heston et al., 2002) – albeit a huge advance over the previous situation – not only limited the time horizon of the empirical investigation but also the choice of the issues that could be posed, thus representing a boundary even for the questions that theoretical economists would seek to address in their models. Economic historians formed a network of their own, which was often marginalized in economics departments and in graduate programmes. More recently, joint effort among economists, economic historians and historians has broadened the time horizon for data availability, thus opening the way for empirical investigations over a longer time span. As a result, not only can the old questions be addressed within longer time series, but also new questions can be raised and answered. Among the recent contributions to the construction of data sets that are playing a crucial role in current growth theory, we find Mitchell (2003), Maddison (2001) and Williamson (1995). On the theoretical front, an influential stream of this literature has focused on the determinants of growth over the long run, with the goal of finding a unified explanation of very different phases of the history of human development, going back to the Malthusian era and beyond. This work is summarised in Galor (2005). However, the institutional dimension is not simultaneously addressed in this stream of contributions.

14.2.2. The Institutional Dimension

As data availability allows the time horizon to be extended retrospectively, the role of institutional factors increasingly appears at least as important as that of purely economic ones. Within the field of economics, North (1981, 1990) has become the standard reference for the idea that institutions shape economic outcomes and are in turn affected by them. Going further back, the classical economists – Smith, Ricardo and Malthus – were already profoundly aware of the relationship between institutions and economic activity. For example, Smith ([1776] 1994) provides a clear analysis of the importance of legal order for the development of a country, although his attempt at developing a general treatise on law and government was never completed. More recently, the issue has also been addressed by other influential thinkers such as Schumpeter, Abramovitz and Kuznets, whose contributions are surveyed in Rostow (1990). Outside of economics, similar concepts had already been expressed, for example, in the writings of Marx and Weber. The attention paid by economists to contributions from other social sciences such as political science and sociology has since increased considerably. As economic theories were shaped by contact with other disciplines, new data sets on political and social variables became increasingly common in economic investigations, as exemplified by Banks's (2001) Cross National Time Series, and the Polity (2002) project initiated by Gurr. While part of the GHI literature has focused specifically on economic institutions, such as property rights protection and financial contracts, other forms of institutions have also been extensively investigated, running from political institutions such as suffrage and constitutional design, to social contracts such as welfare and educational systems, down to other norms that belong to the private – rather than the public – sphere, such as family structure, inheritance laws and women's rights.

Within broader GHI research, part of it has specifically focused on the causation between institutions and growth. Acemoglu et al. (2005) advance a general theoretical argument, supported by empirical tests, according to which institutions cause growth. They carefully distinguish between economic and political institutions, where the former are viewed as the determinants of the incentives and constraints on economic actors, and thus the determinants of economic outcomes. As such, they are social decisions, taken in often conflictual contexts and depending on the distribution of political power, which is in turn determined by political institutions. The resulting dynamic interactions determine the joint evolution of political and economic institutions. It is only when political institutions allocate power to groups which benefit from property rights enforcement that we observe the emergence of economic institutions that foster growth. Engerman and Sokoloff (2003) take a more cautious view and stress how very different institutional structures have often been found to be reasonable substitutes for each other; the historical record does not support the notion that any particular institution, narrowly defined, is indispensable for growth. By contrast, Sachs (2003) argues that geography is a more important determinant of income than institutions. More generally, the difficulty in 'unbundling' institutions is the focus of Acemoglu and Johnson (2003), who evaluate the relative importance of property rights institutions, that protect the property rights of producers and investors against expropriation, and contracting institutions, that are determined by the alternative legal traditions studied in La Porta et al. (1998). Cervellati et al. (2004) also analyse the interdependence of economic and political institutions, focusing on their interaction with inequality. It is also worth noting the connection between a country's level of social capital, its political institutions, and its growth potential, as illustrated for example by Putnam (1993). While the debate on how to construct a general theory of growth and institutions is still open, some recent work has indeed focused on the way specific institutions are formed in response to specific economic and non-economic factors. Engerman and Sokoloff (2002) show how factor endowments shaped economic and political institutions in the Americas. Barro (1999) studies the determinants of democracy, Barro and McCleary (2004) the emergence of state religion, and Bertocchi and Strozzi (2004) the evolution of citizenship laws.

14.3. SUMMARY AND APPLICATIONS

To sum up, the GHI research line, rooted on growth theory, stems both from availability of new data and from the emergence of new questions. The historical and institutional dimensions complement each other, since the economic impact of institutions tends to manifest itself more clearly in the long run. The following applications illustrate the motivations and the implications of research on GHI. Far from representing an exhaustive survey, the topics and contributions that we shall present aim at exemplifying a few of the major issues raised within the recent GHI literature. The first application deals with the relationship between growth and colonization. Given the relevant time horizon, which starts from the first colonization wave in the 16th century, goes up to 19th century colonization and, finally, to the 20th century decolonization phase, this issue is intrinsically addressed within a long time horizon which is required for an understanding of the historical dimension. At the same time, due to the influence of colonization not only on economic outcomes, but also on institutions, it is clear that the institutional dimension is also crucial for capturing it fully. The second application focuses on franchise extension and its relationship with the welfare state. Here again, to capture the full length of the evolution of the issue, it is necessary to go back to that time period in which institutional and political reforms were introduced. The next application concerns the evolution of educational systems, which shares the same basic time profile as the previous issue. The fourth application will analysze the relationship between economic structural reallocation and the evolution of political systems, also over a necessarily extended time horizon. Finally, we will discuss the question of international migration and its economic and institutional determinants, starting from the time period which witnessed the mass migration waves of the 19th century.

14.3.1. Colonization and Growth

Although Smith ([1776] 1994, Book 4, Ch. 7) and Malthus ([1798] 1826, Book 3, Ch. 4), both addressed the economic and social implications of colonization, at least initially new growth theory – following classical growth theory in this respect – paid the issue very little attention, despite its obvious potential relevance in light of the recent dismal performances of areas like Africa, where the more recent phase of colonial expansion had occurred. But this shortcoming should come as no surprise, as data availability, for the time period during which 19th century colonial domination was introduced and established, represents a formidable obstacle to any empirical investigation. In addition, it is only in a subsequent, more recent phase of the research on growth that the role of institutions – with their crucial impact on the countries that were colonized – has received proper consideration. Early exceptions are represented by Lucas (1990) and Grossman and Iyigun (1995), who develop static models where foreign intervention is optimally determined from the vantage point of the metropolitan country. Bertocchi (1994) represents the first attempt to analyse the historical determinants of underdevelopment by introducing colonization within an otherwise standard growth model of the colonial economy, which identifies two main features of colonial economic domination: a restricted inflow of foreign direct investment, which is controlled by the metropolitan country, and direct exploitation. This is therefore a model of extractive colonization which deliberately excludes *a priori* the experience of the so-called white colonies. The model shows that colonization can promote output growth but at the same time depress living standards in the colonies. It also predicts that decolonization may deliver disappointing, or even disastrous, economic performances, when investment withdrawal is accompanied by persistent economic and institutional damage. An extension with a threshold externality *à la* Azariadis and Drazen (1990) shows the potential long-term effect of colonization on human capital. Bertocchi and Canova (2002) test the implications of Bertocchi (1994) using standard growth regressions. To overcome the obstacle of data availability for the colonial period proper, they employ the Penn World Table to identify the consequences of colonial domination for the post-war period. While their main focus is on Africa, with a view to uncovering the mystery of Barro's (1991) sub-Saharan dummy, they also extend their investigation to the rest of the world. Their main findings are that colonial heritage, as measured by the identity of the metropolitan ruler and by the degree of economic penetration, matters for the heterogeneity of growth performances. In particular, they find evidence that colonization did exert a direct effect on the growth pattern of African countries and that it also affected physical and human capital accumulation and the sociopolitical factors typically thought to explain growth.

A related but separate research line has developed around the issue of colonialism following the original contribution by Acemoglu et al. (2001). With the more general goal of establishing the importance of institutions for growth, they find in colonial history, and in a specific data set on the colonists' survival rate, a tool to instrument for institutions and solve their endogeneity problem. Engerman and Sokoloff (2002) also focus on colonial migration to understand the differential patterns of development among New World economies. A further stream of the literature on colonization, pioneered by La Porta et al. (1998), has stressed that it represented the vehicle through which western legal systems spread around the world.

14.3.2. Franchise Extension and the Welfare State

Machiavelli ([1515] 1981) in *The Prince* was perhaps the first to address the question of how to implement major reforms despite the presence of opponents. Between the end of the 19th century and the beginning of the 20th the voting franchise was progressively expanded in Europe. A natural question to ask is why a political elite would voluntarily agree to give up its power and extend the franchise. The answer has generated an important stream of research, starting with Acemoglu and Robinson (2000), who model franchise extension as a commitment device that a ruling elite can choose on a once-and-for-all, universal basis, when faced by the threat of an upheaval. Whereas simple redistribution to the disenfranchised would not be a credible deterrent against upheaval, franchise extension puts the choice of future redistribution in the hands of a median voter, thus assuring preemption. However, if the fear of a revolution was probably the correct motivation of Bismarck's policies, in England the franchise was extended without massive social unrest. Lizzeri and Persico (2004) therefore develop an alternative rationale according to which an elite may wish to expand the franchise even in the absence of serious threats to the established order. They provide a theoretical model in which, when a majority within the elite is dissatisfied with the functioning of current political institutions because of the inadequate provision of public goods, it is in the elite's interest to extend the franchise. Another aspect not captured by Acemoglu and Robinson (2000) is the fact that extensions have often been partial and gradual. Jack and Lagunoff (2003) formulate a model where in each period the enfranchised group can choose, through its median voter, to expand the set of citizens with voting rights. The resulting equilibria generate paths that display a gradual, sometimes uneven history of enfranchisement that is roughly consistent with observed patterns of extensions and can accommodate both explanations for franchise extension, that is, the threat of insurrection and the ideological conflict within the elite. Lee (2003) examines the incentive effect of political democracy on the tax rate by defining a political regime over two dimensions, the extent of the franchise and the extent that the redistribution of tax revenues is biased towards the rich, and finds that when a bias is present democratization tends to reduce exploitation by the rich of the disenfranchised poor. Bertocchi and Spagat (2001) show that the extension of a franchise can be modelled as a co-optation policy to avoid upheaval, since franchise extension implies redistributive transfers. Their focus is on the connection between franchise extension and size of the welfare state. In this light the creation of the welfare state by Bismarck can be viewed as a response to the mobilization of the working class, in an effort to undercut more radical demands by co-opting it into the prevailing political order. They characterize welfare transfers and show how societies with stronger upheaval technologies co-opt more people with larger offers than societies with weaker upheaval technologies, thus encompassing both Germany and England as special cases. They also show conditions for a co-optation strategy to imply the creation of a new, privileged group that separates itself from its group of origin, inducing an asymmetric treatment of a co-opted group and a non-coopted one. The emergence of a middle class is therefore explained by a 'divide and rule' strategy which can usefully be applied to other contexts such as ethnically divided societies.

14.3.3. The Formation of School Systems

Another topic which has received increasing attention, under the influence of contributions from the fields of history, sociology and economic history, is the evolution of educational systems in a political economy perspective. Mueller et al. (1977) provide a socio-historical and comparative account of that decisive period, in the history of European education, that goes from 1870 to WWI. Hitherto, the dominant form of secondary education had been centred upon the classical languages and literature. The subsequent change revolved around the introduction of less prestigious institutions which were supposed to provide so-called modern, or technical, training. On the one hand, this transformation was meant to bring schools into closer interaction with the occupational system of the high industrial era. On the other, specific social roles and ranks were associated with different institutions, with technical curricula being ranked very low. Segmentation produced parallel, non-communicating tracks, with marked differences both in their curricula and the social origin of the students enrolled. The system that emerged, meant to perpetuate the hierarchical structure of these societies, was adopted by all European countries and, with some crucial differences, also by the

United States. However, after WWII, the American secondary school system was the first to move in the direction of a mass preparatory, rather than a terminal, system based on the comprehensive school (see Trow, 1967, and Goldin, 1998). The sociology literature also has a long tradition of studying curriculum differentiation. The 'functionalist' perspective emphasizes the role of technology in driving it, while the 'conflict' perspective – inspired by Weber (1921) – sees it as an attempt by the elite to protect its exclusiveness. Within GHI research, Bertocchi and Spagat (2004) formulate a theoretical model that replicates the evolution of an educational system founded on a hierarchical differentiation between vocational and general education. The dynamics are best summarized by the ratio of the fraction of the population in vocational to that in general education, which can be interpreted as a measure of the degree of stratification of the society. In the model, this ratio first rises and then declines with the level of development, displaying an inverted U-shape which reflects the complex interaction between economic and political forces. By incorporating technological progress and socio-political change, this approach integrates aspects of both the functionalist and the conflict approach while being closer in spirit to the latter since curricula differentiation serves the purpose of perpetuating the pre-existing social order through a process of exclusion of the emerging middle classes from the more prestigious, academically oriented institutions. Accordingly, in an early stage of economic development, vocational curricula tend to expand, while the subsequent rise of the economic and political power of the middle class causes an expansion of general education at a later stage. While the model is designed to capture the specific dynamics of the education mix, it also offers consistent and useful predictions for the evolution of aggregate income, wealth distribution and political participation. Aggregate income grows throughout the process, inequality initially rises and eventually declines, and political participation gradually expands. The theoretical results are supported by cross-section evidence for the post-war period. Bertocchi and Spagat (1997) demonstrate that this theory fits well the historical evolution of the school system in Italy starting from the country's unification in 1861.

Within the political economy field, a growing recent literature on education has addressed other questions, without adding, however, a historical dimension. Glomm and Ravikumar (1992), Saint-Paul and Verdier (1993), Fernandez and Rogerson (1995), Benabou (1996), and Gradstein and Justman (1997) focus on several specific aspects such as public vs. private financing and local vs. state provision. Moreover, in an explicit long-term perspective, a number of other papers have focused – like Bertocchi and Spagat (2004) – on the role of a self-interested elite group in shaping educational systems. Bourguignon and Verdier (2000) analyse the dynamics of inequality, democratization and economic development in a political economy model of growth where education is both the engine of growth and a determinant of political participation. They investigate the incentives for an educated oligarchy to subsidize education of the poor and initiate a democratic transition, and show that the elite may promote the endogenous emergence of a middle class for purely political economy reasons. Galor and Moav (2003) hypothesize that the demise of the 19th century's European class structure reflects a deliberate transformation of society orchestrated by capitalists as an optimal reaction to the increasing importance of human capital in sustaining their profit rates. Due to the complementarity between physical and human capital in production, capitalists were among the prime beneficiaries of the accumulation of human capital by the masses and therefore had the incentive to support public education, although this would ultimately undermine their position in the social ladder. Galor et al. (2003) stress the negative attitude of great landowners towards public expenditure on education.

14.3.4. Industrialization and Democratization

Another issue at the heart of GHI research is the relationship between economic development and political change or, more precisely, the evolution of countries from land-based aristocracies to industrialized democracies. A few attempts have been recently made to explain the long-term determinants of class structure and political participation and their connection with the process of structural reallocation from agriculture to manufacturing. This work has a debt to the literature on interest group politics, first initiated by Kuznets (1968), and expanded by Olson (1982) and Mokyr (1990). More specifically, Bertocchi (2006) looks at the historical evolution of the relationship between an economy's structure and the corresponding political system, with the focus on Europe and the agrarian basis of the feudal system. The paper develops a dynamic specific-factor model with an agricultural sector and a manufacturing sector, and two types of individuals, landlords and landless workers. Since political power is determined by wealth, land ownership immediately entitles the aristocracy to political power, while there exists a minimum capital bequest requirement for active participation of landless workers in the political process. Individuals vote on the removal of feudal privileges, modelled as the appropriation of a portion of the agriculture product by the landlords. The model generates the following results. In the initial, primarily agrarian, phase landlords hold all political power and impose feudal rights on peasants. As capital accumulates, workers start migrating towards the manufacturing sector, and the agricultural sector shrinks. As the workers' capital increases, the process of democratization begins. The expansion of political participation is therefore endogenously determined by the process of economic development. Once the distortive feudal rights are abolished by a bourgeois majority, capital accumulation and income growth accelerate, so that wealth equalization and democratization further progress. Since feudal rents constitute a growth-retarding factor, their removal makes agriculture more efficient, retarding its decline. To sum up, the model establishes a connection between the evolution of society from an aristocratic political system into a democracy, and the process of capital accumulation and sectoral reallocation from agriculture into manufacturing, and generates a theory of endogenous determination of the voting franchise, which is linked to the evolution of wealth composition and distribution. Importantly, this model best captures the transition from feudalism to the triumph of bourgeois power at the end of the 18th century, before formal suffrage legislation is introduced.

Recent work on the role of agriculture in development – outside of a historical perspective – includes Caselli and Coleman (2001), who focus on regional growth patterns in the US during the past century, Gollin et al. (2002), who stress the role of agricultural productivity in explaining cross-country income disparities, and Horowitz (1993), who focuses on the related issue of land reform. In an explicitly long-term perspective, Acemoglu and Robinson (2002) link economic backwardness to political considerations in the spirit of Gerschenkron (1962). In their set up, political elites may block technological and institutional development because innovations tend to erode their incumbency advantage, by increasing the likelihood that they will be replaced. They show that it is only when political competition is limited that elites will block development, and in this light they offer an interpretation for why Britain, Germany and the US industrialized during the 19th century, while the landed aristocracy in Russia and Austria-Hungary blocked development. Llavador and Oxoby (2005) show that growth and democratization occur when the following conditions are simultaneously met: there is an economic conflict among the elite, the landed classes are not politically strong, and there is a critical mass of industrial workers. In their analysis, the absence of the first two conditions resolves in stagnant autocracies, while the absence of the third drives growth-deterring democratic expansions. Justman and Gradstein (1999) also analyse the interdependent processes of economic and political development in a dynamic perspective. Their model provides a characterization of the endogenous transition from oligarchic to democratic institutions under different potential scenarios. Democratization can be initiated by a strong elite aiming for economic benefits of democracy, or can be forced by a politically disenfranchised majority under the shadow of conflict, while revolutions can also arise under different conditions. Boix (2003) develops a unified model to explain the distribution of different political regimes, as a function of equality and capital mobility, which is systematically tested on two data bases for the periods 1950 to 1990 and 1850 to 1980. Cervellati et al. (2004) analyse the joint evolution of economic and political institutions and establish conditions under which a state of law can be implemented under oligarchy. Inequality in endowments and income is shown to crucially affect the development process.

A theme that this literature developed on the side, as a by-product of the more general implications for growth and democracy, is that of family structure and inheritance laws. Bertocchi (2006) observes that in a feudal society, in the presence of a politico-economic constraint on the minimum size of estates, land cannot be alienated so that primogeniture emerges as the norm regulating the intergenerational transmission of property rights on land. Therefore, as agrarian aristocratic societies evolve towards industrialization and democratization, we also witness an evolution in the legal system regulating the intergenerational transfers of property rights, with primogeniture being replaced by equal partition. As divisible capital replaces indivisible land, partition promotes equalization, whereas primogeniture amplified wealth inequality. The connection between other aspects of family structure and the economic and political environment has also been explored recently by Botticini and Siow (2003) and Guner (1999), who focus on dowries and marriage systems, respectively. More generally, gendered work patterns and family planning can affect the relationship between growth and institutions by playing a role in triggering the demographic transition, another crucial component of the growth process which is explored for instance by Galor and Weil (1996).

14.3.5. The Political Economy of Migration

The history of international migration is also a history of political decisions and institutional development. A specific literature has focused on 19th century mass migration and its political economy implications. During the first half of the century, Britain was the main source of emigration. Germany, Scandinavia and then Southern and Eastern Europe joined in during the second half. The main destination was North America, followed by South America and Australia. Most of the migrants were young, many took advantage of friends and relatives networks, economic incentives were crucial determinants of their decision to migrate, and their skill level tended to decline over time (see Taylor and Williamson, 1997, and Hatton and Williamson, 1998). While the period that precedes WWI is often referred to as the era of free migration, in fact policy started to respond as early as the 1890s, when the US frontier was officially declared closed, soon followed by other New World countries. Timmer and Williamson (1998) construct an index of immigration barriers in the main destination countries from 1850 to 1930 and find that the most important determinant of increasing policy restrictiveness was a measure of inequality given by the ratio of the unskilled wage to per capita income, while they find no evidence in support of alternative motivations such as racism or xenophobia. During the interwar period restrictive immigration policies became the norm, soon involving, besides the traditional lands of immigration, also the countries of Europe that were newly exposed to inflows. After WWII, restrictions persisted and took a variety of forms (quotas, points systems, and the like), often in the attempt to attract high-skill workers and discourage the low-skill. From a 24-country survey conducted in 1995, O'Rourke and Sinnott (2003) find that voters' attitudes towards immigration negatively respond to factors such as patriotism and chauvinism, so that economics alone cannot explain the hostility towards immigration which is expressed in many countries. On the other hand, economic factors also play a role, with high-skill individuals tending to be less anti-immigration in rich countries than in poor countries and in countries with more equal income distributions. Bertocchi and Strozzi (2004) focus on a specific form of institution, citizenship acquisition, and its relationship with migration. Citizenship laws come from two broad traditions, common and civil law, the former applying the *jus soli* principle, according to which a child is a citizen as long as he/she was born in a given country, the latter the *jus sanguinis* principle, according to which a child simply inherits his/her parents' citizenship. They show that the impact of the original, exogenously-given laws on international migration is insignificant for the early, mass migration waves, which they confirm as being driven primarily by economic incentives. For post-war data, they investigate the determinants of citizenship laws evolution and find that their convergence can be linked to legal tradition and international migration, but is also affected by border stability, the degree of democracy, the welfare burden, colonial history, and cultural factors such as religion and ethno-linguistic fractionalization.

14.4. A NEW STANDARD

On appraising the five applications reviewed above, one common factor clearly emerges: their timing. Each traces the evolution of a specific economic and social issue starting, in all cases, from the second half of the 19th century. Indeed it is around 1850 that the nature of colonial expansion starts changing, with the old colonialists – Portugal, Spain, the Netherlands – being replaced by much stronger new ones such as England, France, Belgium and, later, also Germany and Italy. With 1870 the age of imperialism begins its course. In Britain, the franchise was extended first in 1832, and then again in 1867 and 1884, while the school reform process leading to the modern system started with the Education Act in 1870. The rest of the European countries followed a similar pattern, opening the way for a new and increasingly important role of governments. While the French revolution is commonly viewed as the turning point with respect to the elimination of feudal privileges, agrarian relations survived in Europe well beyond the Restoration, and only towards the end of the 19th century was the process of industrialization completed in most of Europe. Finally, the second half of the same century is also the age of mass migration from the Old to the New World. To sum up, the period starting around the year 1870, both because of the relative availability of data and because of its sufficient homogeneity, has the potential to become the new standard for theoretical and empirical work in growth theory. Indeed, historians identify 1870 as a decisive turning point, both politically and economically: but why 1870?

Taking a step back, we can identify three crucial events during the century that preceded 1870. These events provoked a worldwide economic, social and political upheaval (see Caracciolo, 1989), and continued to be felt well after 1870. We are referring, first of all, to the Industrial Revolution, starting its slow course in England between 1760 and 1790, and to the two political revolutions occurring in America in 1776 and in France in 1789. These events marked the end of the *Ancien Régime* and the rise of the bourgeoisie as the ruling class. However, the apex of bourgeois power would only be reached in the second half of the century when, after the turning point of the Restoration and then the instabilities of 1848, a consolidated institutional background was established and a new economic impulse given to the emerging middle class. The period between 1850 and 1870 also saw huge changes. Economic growth, despite short cyclical recessions in 1857 and 1866, reached its highest level ever. Technological progress, affecting industrial production but also agriculture, transportation and trade, intensified. Manufacturing accelerated – to the detriment of agriculture – and the tertiary sector also started its expansion. Rural areas are extended and urbanization spread. New demographic trends and unprecedented social and geographical mobility also characterize this period. The new social class emerging from this process occupied the ground between the high bourgeoisie, which was always close to the aristocracy, and the menial working class proper. This middle class, which included merchants, professional workers and clerks, was to expand exponentially in the following decades. With 1871, the end of the Franco-Prussian war marked the beginning of a prolonged period of relative peace that was to last until WWI, and Germany's unification settled the geography of Europe. In the same year the *Commune* in Paris changed the connotations of class conflict.

In the next 40 years, the Western world was to witness a profound political transformation with the development of the trade unions and the creation of socialist parties, as the establishment of the industrial bourgeoisie was accompanied by the emergence of the urban proletariat. Political reforms would follow, together with reforms of the educational system, the bureaucracy and the welfare state. Trade and the international financial system also developed, together with the process of colonization and with international migration. Therefore, the early 1870s indeed represented a decisive turning point, politically, socially and economically, that marks the beginning of the contemporary era.

All this suggests a broad agenda for future research on GHI, which is ambitious but, judging from the progress previously reviewed, also feasible. This agenda involves re-visiting all the issues for which a post-WW2, Penn-Table consensus has been reached – starting from the causes of growth and prosperity – taking the year 1870 as the beginning of the new reference period. This will allow us not only to capture long-term time variations and patterns of historical evolution previously uncovered, but also to recognize in full the role of institutional factors and their interaction with purely economic factors. We have the data, we have the models, we can do it.

14.5. CONCLUSION

His 20 February, 2003 obituary in *The Economist* described the role of Walt Rostow in actual policymaking. As a development economist, in 1960 he had published 'The Stages of Economic Growth: A Non-Communist Manifesto', where he had expounded his thesis that prosperity brings democracy, and that rapid economic growth should induce poor countries to embrace capitalism rather than communism. When Kennedy became president in 1961, Rostow became one of his advisers, and soon became particularly involved with Vietnam. The South of the country was under the economic and political influence of the United States and was developing fast, but was under the threat of communist guerrillas from the North. Rostow's advice was to fight the guerrillas to protect the development of the South and help contain the spread of communism in the area. Under Johnson, Rostow rose to the rank of National Security Advisor and masterminded the escalation of the intervention. Unlike McNamara (1995), who had served as Defence Secretary under Kennedy and Johnson but later admitted the failure of the Vietnam war and attributed it to the United States' 'profound ignorance of the history, culture and politics of the people in the area', Rostow always remained convinced that the intervention had achieved its purpose in delaying the advance of communism in South-East Asia long enough for the area to take off economically and join the path towards democracy. Thus, it was the Rostow thesis that justified the Vietnam war. To these days, the question of the long-term relationship between democracy and prosperity is still an open one. The point of this story is not to evaluate the Rostow thesis, or its consequences, but simply to convince the reader of the potential practical relevance, for domestic policymaking and international affairs, of current and future research aimed at uncovering the interrelationship among growth, history and institutions.

NOTE

* This chapter was presented within the panel session on 'Growth and Distribution: Different Perspectives' at the Conference. I would like to thank two anonymous referees and conference participants for helpful comments and suggestions.

REFERENCES

Acemoglu, D. and S. Johnson (2003), 'Unbundling institutions', mimeo.

Acemoglu, D. and J.A. Robinson (2000), 'Why did the West extend the franchise? Democracy, inequality, and growth in historical perspective', *Quarterly Journal of Economics*, **115**:1167–99.

Acemoglu, D. and J.A. Robinson (2002), 'Economic backwardness in political perspective', *NBER Working Paper 8831*.

Acemoglu, D., S. Johnson and J.A. Robinson (2001), 'The colonial origins of comparative development: an empirical investigation', *American Economic Review*, **91**:1369–401.

Acemoglu, D., S. Johnson and J.A. Robinson (2005), 'Institutions as the fundamental cause of long-run growth', in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Amsterdam: North-Holland, pp. 385–472.

Azariadis, C. and A. Drazen (1990), 'Threshold externalities in economic development', *Quarterly Journal of Economics*, **55**:501–26.

Banks, A. (2001), *Cross-National Time Series, 1815-1999*, Binghamton: Center for Comparative Political Research State University of New York.

Barro, R.J. (1999), 'Determinants of democracy', *Journal of Political Economy*, **107**(Part 2):S158–S183.

Barro, R.J. (1991), 'Economic growth in a cross section of countries', *Quarterly Journal of Economics*, **106**:407–44.

Barro, R.J. and R.M. McCleary (2004), 'Which countries have state religion?', *NBER Working Paper 10438*.

Benabou, R. (1996), 'Heterogeneity, stratification, and growth: macroeconomic implications of community structure and school finance', *American Economic Review*, **86**:584–609.

Bertocchi, G. (1994), 'Colonialism in the theory of growth', *Brown University Working Paper 94-14*.

Bertocchi, G. (2006), 'The law of primogeniture and the transition from landed aristocracy to industrial democracy', *Journal of Economic Growth*, forthcoming.

Bertocchi, G. and F. Canova (2002), 'Did colonization matter for growth? An empirical exploration into the historical causes of Africa's underdevelopment', *European Economic Review*, **46**:1851–71.

Bertocchi, G. and M. Spagat (1997), 'Il ruolo dei licei e delle scuole tecnico-professionali tra progresso tecnologico, conflitto sociale e sviluppo economico', in N. Rossi (ed.), *L'istruzione in Italia: solo un pezzo di carta?*, Bologna: Il Mulino, pp. 421–59.

Bertocchi, G. and M. Spagat (2001), 'The politics of co-optation', *Journal of Comparative Economics*, **29**:591–607.

Bertocchi, G. and M. Spagat (2004), 'The evolution of modern educational systems. Technical vs. general education, distributional conflict, and growth', *Journal of Development Economics*, **73**:559–82.

Bertocchi, G. and C. Strozzi (2004), 'Citizenship laws and international migration in historical perspective', CEPR *Discussion Paper 4737*.

Boix, C. (2003), *Democracy and Redistribution*, Cambridge: Cambridge University Press.

Botticini, M. and A. Siow (2003), 'Why dowries?', *American Economic Review*, **93**:1385–98.

Bourguignon F. and T. Verdier (2000), 'Oligarchy, democracy, inequality and growth', *Journal of Development Economics*, **62**:285–313.

Caracciolo, A. (1989), *Alle origini della storia contemporanea*, Bologna: Il Mulino.

Caselli, F. and J.W.II Coleman (2001), 'The U.S. structural transformation and regional convergence: a reinterpretation', *Journal of Political Economy*, **109**:584–616.

Cervellati, M., P. Fortunato and U. Sunde (2004), 'From Hobbes to Rousseau: inequality, development and democratization', mimeo.

Engerman, S. and K.L. Sokoloff (2002), 'Factor endowment, inequality, and paths of development among new world economies', *NBER Working Paper 9259*.

Engerman, S. and K.L. Sokoloff (2003), 'Institutional and non-institutional explanations of economic differences', *NBER Working Paper 9989*.

Fernandez, R. and R. Rogerson (1995), 'On the political economy of education subsidies', *Review of Economic Studies*, **62**:249–62.

Galor, O. (2005), 'From stagnation to growth: unified growth theory', in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Amsterdam: North-Holland, pp. 171–293.

Galor, O. and O. Moav (2003), 'Das human kapital: a theory of the demise of the class structure', mimeo.

Galor, O. and D.N. Weil (1996), 'The gender gap, fertility, and growth', *American Economic Review*, **86**:374–87.

Galor, O., O. Moav and D. Vollrath (2003), 'Land inequality and the origin of divergence and overtaking in the growth process: theory and evidence', mimeo.

Gerschenkron, A. (1962), *Economic Backwardness in Historical Perspective*, Cambridge, MA: Harvard University Press.

Glomm, G. and B. Ravikumar (1992), 'Public versus private investment in human capital: endogenous growth and income inequality', *Journal of Political Economy*, **100**:818–34.

Goldin, C. (1998), 'America's graduation from high school: the evolution and spread of secondary schooling in the Twentieth Century', *Journal of Economic History*, **58**:345–74.

Gollin, D., S. Parente and R. Rogerson (2002), 'The role of agriculture in development', *American Economic Review*, **92**:160–64.

Gradstein, M. and M. Justman (1997), 'Democratic choice of an education system: implications for growth and income distribution', *Journal of Economic Growth*, **2**:169–84.

Grossman, H.I. and M. Iyigun (1995), 'The profitability of colonial investment', *Economics and Politics*, **7**:229–42.

Guner, N. (1999), 'An economic analysis of family structure: inheritance rules and marriage systems', mimeo.

Hatton, T.J. and J.G. Williamson (1998), 'The age of mass migration: causes and economic impact', Oxford: Oxford University Press.

Heston, A., R. Summers and B. Aten (2002), 'Penn World Table Version 6.1', Center for International Comparisons at the University of Pennsylvania (CICUP), October.

Horowitz, A.W. (1993), 'Time paths of land reform: a theoretical model of reform dynamics', *American Economic Review*, **83**:1003–10.

Jack, W. and R. Lagunoff (2003), 'Dynamic enfranchisement', mimeo. Available at: http://www.georgetown.edu/faculty/lagunofr/franch10.pdf

Justman, M. and M. Gradstein (1999), 'The industrial revolution, political transition, and the subsequent decline in inequality in 19th Century Britain', *Explorations in Economic History*, **36**:109–27.

Kuznets, S. (1968), *Towards a Theory of Economic Growth*, New Haven: Yale University Press.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer and R. Vishny (1998), 'Law and finance', *Journal of Political Economy*, **106**:1113–55.

Lee, W. (2003), 'Is democracy more expropriative than dictatorship? Tocquevillian wisdom revisited', *Journal of Development Economics*, **71**:155–98.

Lizzeri, A. and N. Persico (2004), 'Why did the elites extend the suffrage? Democracy and the scope of government, with an application to Britain's "age of reform"', *Quarterly Journal of Economics*, **119**:707–65.

Llavador, H. and R.J. Oxoby (2005), 'Partisan competition, growth and the franchise', *Quarterly Journal of Economics*, **120**:1155–89.

Lucas, R.E. Jr. (1988), 'On the mechanics of economic development', *Journal of Monetary Economics*, **22**:3–32.

Lucas, R.E. Jr. (1990), 'Why doesn't capital flow from rich to poor countries?', *American Economic Review*, **80**:92–6.

Machiavelli, N. [1515] (1981), *The Prince*, Harmondsworth: Penguin.

Maddison, A. (2001), *The World Economy. A Millennial Perspective*, Paris: OECD.

Malthus, T.R. [1798] (1826), *An Essay on the Principle of Population*, London: John Murray.
McNamara, R. (1995), *In Retrospect: The Tragedy and Lessons of Vietnam*, New York: Times Books.
Mitchell, B.R. (2003), *International Historical Statistics, 1750–2002*, Vol. I: *Europe*, Vol. II: *The Americas*, Vol. III: *Africa, Asia & Oceania*, New York: Palgrave Macmillan.
Mokyr, J. (1990), *The Lever of Riches: Technological Creativity and Economic Progress*, New York: Oxford University Press.
Mueller, D.K. (1977), 'The process of systematization: the case of German secondary education', in D.K, Mueller, F. Ringer and B. Simon (eds), *The Rise of the Modern Educational System: Structural Change and Social Reproduction*, Cambridge: Cambridge University Press, pp. 15–52.
North, D.C. (1981), *Structure and Change in Economic History*, New York: W.W. Norton & Co.
North, D.C. (1990), *Institutions, Institutional Change, and Economic Performance*, New York: Cambridge University Press.
Olson, M. (1982), *The Rise and Decline of Nations: Economic Growth, Stagflation, and Economic Rigidities*, New Haven and London: Yale University Press.
O'Rourke, K. and R. Sinnott (2003), 'Migration flows: political economy of migration and the empirical challenge', mimeo, Trinity College Dublin.
Polity IV (2002), 'Political regime characteristics and transitions, 1800–2002', Center for International Development and Conflict Management, University of Maryland.
Putnam, R.D. (1993), *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton: Princeton University Press.
Romer, P.M. (1986), 'Increasing returns and long-run growth', *Journal of Political Economy*, **94**:1002–37.
Rostow, W.W. (1960), *The Stages of Economic Growth: A Non-Communist Manifesto*, Cambridge: Cambridge University Press.
Rostow, W.W. (1990), *Theorists of Economic Growth from David Hume to the Present: With a Perspective on the Next Century*, Oxford: Oxford University Press.
Sachs, J. (2003), 'Institutions don't rule: direct effects of geography on per capita income', *NBER Working Paper 9490*.
Saint-Paul, G. and T. Verdier (1993), 'Education, democracy, and growth', *Journal of Development Economics*, **42**:399–407.
Smith, A. [1776] (1994), *An Inquiry into the Nature and Causes of the Wealth of Nations*, New York: The Modern Library.
Solow, R.M. (1956), 'A contribution to the theory of economic growth', *Quarterly Journal of Economics*, **70**:65–94.
Taylor, A.M. and J.G. Williamson (1997), 'Convergence in the age of mass migration', *European Review of Economic History*, **1**:27–63.
The Economist (2003), Walt Rostow, February 20th.
Timmer, A. and J.G. Williamson (1998), 'Immigration policy prior to the thirties: labor markets, policy interactions and globalization backlash', *Population and Development Review*, **24**:39–71.

Trow, M. (1967), 'The second transformation of American secondary education', in R. Bendix and S.M. Lipset (eds), *Class, Status, and Power*, London: Routledge & Kegan Paul, pp. 437–49.

Weber, M. (1921), *Economy and Society*, Berkeley: University of California Press.

Williamson, J.G. (1995), 'The evolution of global labor markets since 1830: background evidence and hypotheses', *Explorations in Economic History*, **32**:141–96.

15. Religious culture, institutions and growth

Michele Bagella, Leonardo Becchetti and Stefano Caiazza

15.1. INTRODUCTION

The empirical literature on the determinants of economic development has progressively tested the significance of the impact on growth of several factors other than traditional production inputs. Durlauf and Quah (1998) in a survey of the literature identify 87 different proxies of variables tested in empirical analyses. Among the most important are human capital (Mankiw, Romer and Weil, 1992), the government sector (Hall and Jones, 1997), social and political stability (Alesina and Perotti, 1994), corruption (Mauro, 1995), social capital (Knack and Keefer, 1997), income inequality (Persson and Tabellini, 1994; Perotti, 1996) and financial institutions (Pagano, 1993; Levine, 1997).[1]

While this literature often claims the importance of looking at institutions and at 'deep fundamentals' in order to understand the roots of economic growth, the relationship between cultural heritage from religious beliefs, institution building and economic development has not been thoroughly explored.

Only a few empirical papers tackle the issue of the relationship between religion and economic growth.[2] The traditional approach adopted is a test on the impact of religion as an additional regressor in exogenous or endogenous growth models (Sala-i-Martin, 1987).[3] Its main limit is that of testing the marginal contribution of religious affiliation on levels or growth of per capita income without investigating the more complex links between religious beliefs and the same factors affecting growth. This broader perspective has been recently taken by Stultz and Williamson (2001) investigating whether religious background significantly affects financial institutions net of the effect of language, trade openness and the origin of the country's legal system. The authors find that creditors' rights are significantly less protected in Catholic than in Protestant countries.[4]

We believe that this is a fruitful direction in which the growth literature (and, in our particular case, the finance-growth literature) may be implemented. It is reasonable to assume that 'deep fundamentals', such as cultural *ethos* or religious beliefs, have an indirect impact on growth by shaping and ordering the set of values which crucially influence institution building and individual behaviours and, through them, the 'wealth of nations'.[5]

This chapter aims to extend this recent investigation in new directions. It analyses the effects of religious backgrounds on institutions in an enlarged set of countries and evaluates the comparative impact of two religious environments (Catholicism and Protestantism) on institution building. Finally, it verifies whether the different development of market rules and financial institutions affects the relationship between them and growth. In this respect, a final advantage of our approach is that the previously identified relationship between religious background and institutional development allows us to tackle the endogeneity problem in the relationship between growth and institutions by instrumenting the latter with religious background.

The chapter is divided into five sections including the introduction and conclusions. In the second section we outline our theoretical hypotheses and we discuss how religious beliefs are systems of value rankings which crucially affect rewards to talent (Murphy et al., 1991) and economic decisions of individuals. In the third section we analyse the effects of religious backgrounds on a set of institutional indicators, net of the impact of several control variables and verify whether some convergence in institutional shaping across religious worlds has occurred in the last decades. In the fourth section, we test whether the previously measured differences in institutions affect the relationship between institutions and growth.

15.2. RELIGIOUS BELIEFS AND INSTITUTIONS: IS THERE A DIFFERENCE BETWEEN PROTESTANT AND CATHOLIC COUNTRIES?

The hypothesis that religious and cultural backgrounds significantly affect institutions and growth has often been supported by sociologists and historians but much less explored by economists.

Among the few of them, Landes (2000) finds that cultural factors contribute to explain differences in human capital accumulation and in rates of economic development. This conclusion does not contradict the Weberian hypothesis that the Protestant Reformation divided Europe into two different areas with different standards for productive effort, contractual dealings and accumulation of capital. This hypothesis (Weber, 1930) postulates that Protestantism gives relatively higher social value to entrepreneurship and to the accumulation of 'productive' human capital. For these reasons, gifted individuals in Protestant societies invested more in human capital and directed it more toward productive activities.[6]

Delacroix (1992) takes Weber's argument further by saying that 'the worldview propagated by Protestantism broke with traditional psychological orientations through its emphasis on personal diligence, frugality and thrift, on individual responsibility, and through the moral approval it granted to risk-taking and to financial self-improvement'. In the same direction Blum and Dudley (2001) state that 'a representative Protestant would be more inclined to participate more actively in economic life and would be more likely to refrain from consuming fruits of its labour than would a representative Catholic'.

The rationale for the Weberian hypothesis is rooted in the theological differences between Catholic and Protestant creed. In the Protestant (Calvinist) ethos, wealth is often regarded as a sign of 'predestination' by God.[7] By contrast, the Catholic ethos remained much more related in the past to the idea that God's providence is what matters and that wealth accumulation is a sign of greed, and of misguided ranking of values.

One of the most insightful syntheses of the Weberian hypothesis is from Marshall (1982, p. 74) who advocates that

> protestant asceticism (according to Max Weber) restricted consumption of luxuries and prohibited spontaneous enjoyment of the world. On the other hand, it insisted upon relentless efforts in one's lawful vocation as the duty of all Christians. This unique combination, Weber felt, was almost certain, ceteris paribus, to lead to the accumulation and reinvestment of capital by those involved in business activities … Thus, through the entirely unintended consequences of the double injunction to diligence in lawful callings and asceticism in the world, ascetic Protestantism created the modern capitalist mentality.[8]

According to Blum and Dudley (2001) Protestant ethics also raises the costs of defection from contractual relationships. For Catholics this cost is low because it is always possible to obtain pardon with the intermediation of a priest. For a Calvinist, or for believers of other ascetic Protestant denominations influenced by Calvinism, there is no such intermediation and defection weakens the belief in one's own predestination. This factor leads to a stronger defence of property rights which also fosters returns and reduces uncertainty of entrepreneurial activity.

This hypothesis has been strongly criticised by Novak (1994) who argues that quality of relationships and not individualism is the root of capitalist entrepreneurial activity: all capitalist organisational forms (companies, investment funds, and so on) arise from the convergence of several individuals into creative activities. From this point of view, also Catholic culture with its strong sense of community is not in contrast. A reconciliation of these two conflicting perspectives comes from the insightful historical analysis of Trevor-Roper (1969) who recognises that Catholic southern European cities were flourishing before the Reformation but not after it. The author finds that the emphasis against the Reformation increased the political and economic weight of rent seeking activities (priests, civil servants) and consequently the tax burden on Catholic entrepreneurs forcing most of them to flee to northern Europe.

While Trevor-Roper's perspective is observationally equivalent to Weber's, it has the advantage of being consistent with the fact that Catholic countries (with many Catholic merchants) flourished before the Reformation.

In the light of the debate summarised above, we argue that religious beliefs – and the cultural environment generated by them – are not only likely to affect individual behaviour, but also to shape the institutional framework in which decisions are taken. More specifically, we postulate that: i) the higher social value of entrepreneurial risk-taking in Protestant societies, emphasised by Weber, underlies the higher development of financial markets in which people can share risk intertemporally or cross-sectionally; ii) the higher value of individual responsibility has led to greater emphasis on economic freedom and to a limitation of the role of the state in financial and non-financial markets.

On the first point, the four main differences between the two established archetypes of financial markets (the market-oriented and bank-oriented system) may be summarised as follows: i) the limited role of the state in bank ownership; ii) the quality of information; iii) the protection of minority shareholders and iv) the repression of insider trading (Allen and Gale, 1997). From a descriptive point of view, cultural backgrounds induced by religious beliefs and geographical location (Europe or the US) are two partial but not mutually exclusive explanations for these differences.[9]

On the second point, it is plausible to conclude that the increased emphasis on individual responsibility against tradition and the role of Church authorities led to a general reduction in the pressure of both religious and civil authorities on individuals' lives and hence fostered the creation of a society in which the role of the state was more limited than in Catholic countries (Trevor-Roper, 1969).

In conclusion, even though the Protestant cultural background, with its higher social support for entrepreneurial risk-taking and for private (rather than government-controlled) economic activity, may have influenced the development of freedom on exchange and financial markets, the hypothesis needs to be tested accurately since counterexamples may be easily found.

15.3. EMPIRICAL ANALYSIS

15.3.1. Descriptive Evidence for the Link Between Religious Heritage and Institutions

In the previous section we presented arguments on the potential effects of religious beliefs on the shaping of institutions. To verify the significance of them we analyse the relationship between religious beliefs and different indicators of civil, legal and economic freedom.[10] Religious affiliation is taken from two sources: the *CIA Economic Factbook* and the *Italian De Agostini Atlas* which collects historical information from domestic Census data.

Scores for the two indexes of civic, legal and economic freedom obtained for each country are weighted according to the following formula $I_{jk} = \sum_i \omega_{ki} I_{ji} / \omega_{ik}$ where I_{jk} is the weighted average of the j-th institutional indicator for the religious creed k, ω_{ki} is the share of the population affiliated to the religious creed k in country i and I_{ji} is the i-th country score on the j-th institutional indicator.

We define this as the *fuzzy approach* since each country may be fractionally attributed to different religious worlds.[11] We may argue, though, that this weighting is too generous toward minority beliefs. We therefore choose to attribute alternatively to a given creed only those countries in which the chosen religion has at least the relative majority. Under this second approach (*dichotomous approach*) ω_{ki} is therefore equal to one if the relative majority of the population of country i is affiliated to the creed k and zero otherwise.[12]

With both approaches we find descriptive support for the Weberian hypothesis when we consider average 1970–97 dichotomous indicators (Table 15.1). Protestant countries are characterised by reduced state interference in the economy and in market rules (EFW2) and by higher freedom to exchange in capital and financial markets (EFW7) with respect to Catholic countries. Furthermore, stock market capitalisation accounts on average for around 31 per cent of GDP in Protestant countries against around 19 per cent in Catholic countries.

T-stats on mean values averaged across the sample period (with the dichotomous approach) show that the difference in means between Catholics and Protestants is not significant in either of the above-mentioned indicators. Alternative determinants of institutional heterogeneity such as civil law/common law and the English language variables generate non-significant subgroup differences as well (Table 15.1). When we pass from the dichotomous to the fuzzy approach (that is Germany is considered half

Table 15.1 Institutions, economic freedom and religious affiliation (dichotomous classification) (mean values 1970–97)

	EFW (2)	EFW (7)
Catholic	4.647	6.016
Protestant	5.048	6.717
Christian	4.150	5.526
Civil Law	3.788	5.252
Common Law	4.233	5.266
English Language	4.062	5.107
Non-English Language	3.710	5.204
T-test on the significance of the difference in means (1.76 and 2.38 are respectively the 90% and the 99% significance threshold)		
Catholic vs. Protestant	–0.928	–1.112
Civil Law vs Common Law	–1.164	–0.028
English vs Non-English Lang.	0.950	–0.185

Note:
Any of the considered indicators has a 0–10 value range. A higher value means a higher level in the item considered by the indicator. Countries are assigned to the religion followed by the majority of inhabitants according to national Census data. In the dichotomous approach scores for the seven indexes of civic, legal and economic freedom obtained for each country are weighted according to the following formula $I_{jk} = \Sigma_i (\omega_{ki} I_{ji}) / \omega_{ik}$ where I_{jk} is the weighted average of the j-th institutional indicator for the religious creed k, I_{ji} is the i-th country score on the j-th institutional indicator and ω_{ki} is therefore equal to one if the relative majority of the population of country I is affiliated to the creed k and zero otherwise.

Catholic and half Protestant) all the differences between Catholics and Protestants become significant (Table 15.2).[13]

If we repeat the test at the beginning and end of the observation period we observe that (if we use the dichotomous approach) in both cases Catholics, starting from significantly lower levels, converge to Protestants at the end of the observation period (Table 15.2).

These institutional differences are consistent with the divergences arising from the two different religious backgrounds in terms of the social value of entrepreneurial risk-taking, individual responsibility and respect of contractual obligations described in the previous section.

By summarising the results on the dynamics of institutions under the two religious and cultural heritages we may conclude that Catholics converged to Protestant institutional quality in the last two decades. We may interpret the substantial convergence between the two Christian creeds in the light of the cultural revolution ignited by the Second Vatican Council,[14] which promoted convergence in terms of organisation (supply) and participation (demand) of cult activities (Iannacone, 1998) and, above all, in terms of earthly values with the rise of Catholic movements which emphasise the calling to professional activities and partially to entrepreneurship (*Opus Dei*, *Comunione e Liberazione*, *Focolarini*). These movements have increased the emphasis on horizontal coordination of believers in society as opposed to the pressure from the vertical hierarchy which has been typically considered as a determinant of the differences between Protestant and Catholic culture.

Table 15.2 Institutions, economic freedom and religious affiliation (fuzzy classification). Test t for difference in means

	EFW2	EFW7
Catholics vs protestants	4.775	4.577
Christians '75 vs christians '97	–4.244	–4.508
Catholics '75 vs catholics '97	–4.406	–4.943
Protestant '75 vs protestant '97	–2.615	–2.345
Catholics '75 vs protestant '75	3.040	2.433
Catholics '97 vs protestant '97	4.657	4.650
	EFW2	EFW7
Civil Law '75 vs Civil Law '97	–1.077	–3.531
Civil Law '75 vs Common Law '75	1.211	0.514
Common Law '75 vs Common Law '97	–3.686	–2.783
Civil Law '97 vs Common Law '97	–1.880	0.417
Christians '75 vs Christians '97	–3.041	–4.097
Catholics '75 vs Catholics '97	–2.957	–3.219
Catholics '75 vs Protestant '75	1.565	0.217
Protestant '75 vs Protestant '97	–5.088	–3.858
Catholics '97 vs Protestant '97	–0.735	–0.894
English '75 vs English '97	–3.589	–2.836
English '75 vs Non-English '75	–1.666	–1.085
Non English '75 vs Non-English '97	–0.923	–2.262
English '97 vs Non-English '97	1.309	–0.721

Note:
In the fuzzy index scores for the seven indexes of civic, legal and economic freedom obtained for each country are weighted according to the following formula $I_{jk} = \Sigma_i(\omega_{ki}I_{ji})/\omega_{ik}$ where I_{jk} is the weighted average of the j-th institutional indicator for the religious creed k, ω_{ki} is the share of the population affiliated to the religious creed k in country i and I_{ji} is the i-th country score on the j-th institutional indicator.
For the definition of institutional indicators see Table 15.1.
The table reports the value of t-Student for the significance of the difference across means.

15.3.2. Econometric Findings on the Relationship Between Religious Backgrounds and Institutions

Descriptive statistics on institutional quality in countries with different religious traditions may obviously experience composition effects. The most likely candidates are per capita GDP, language and origin of the legal system. A clear distinction among the effects of each of these variables on institutional quality is problematic. Religious backgrounds may have started affecting institutional quality and, through it, per capita income far before our estimation period. Hence, the insignificance of religious backgrounds once we introduce (beginning of period) per capita income as an explanatory variable does not necessarily imply that religion does not affect institutions.

In spite of these problems, to provide additional evidence on the existence of a significant impact of religious cultural background on institutions, we estimate the following equation

$$INST_j = \alpha_0 + \alpha_1 Christian + \sum_{i=2}^{n} \alpha_i X_i + \varepsilon \tag{15.1}$$

where $INST_j$ is the selected institutional variable among the following two indicators: EFW(2) Structure of the Economy and Use of Markets *(Production and allocation via governmental and political mandates rather than private enterprises and markets*); EFW(7) Freedom of Exchange in Capital and Financial Markets.

Regressors include Christian, a variable measuring the share of Christians, plus additional controls (the X-variables). Among these controls we include four variables (Muslim, Buddhist, Hindu and Confucianist) measuring respectively the share of believers in the four different religions. Additional controls are: gdpl60, GDP per working age population in 1960; civlaw (comlaw) a dummy which assumes a value of one for countries in which civil law (common law) prevails; engl a dummy which has a value of one for countries in which English is one of the main languages spoken; d_oecd a dummy for OECD countries and geography a variable measuring the log of the distance from the equator of the country capital.[15]

Our empirical findings show a significant and positive effect of the Christian cultural background on the two institutional indicators (EFW2 and EFW7) measuring, respectively, lack of government interference on market mechanisms[16] and freedom of access and exchange in financial markets.[17] The impact of Christian cultural background is significant even after controlling for the effect of the beginning of period per capita GDP, language, OECD/non-OECD country affiliation, origins of the legal systems, geography and different religious backgrounds (Table 15.3).[18] On the other hand, a separate estimate reveals that the difference between Catholics and Protestants in terms of impact on quality of institutions is not significant.[19] This result is consistent with the descriptive evidence on institutional convergence (between Catholics and Protestants) in the last two decades. Overall, the result on the impact of Christian religious creed on institution does not refute the hypothesis outlined in Section 15.2.

Two additional expected results from the estimates are the positive effects of OECD membership and the beginning of period level of per capita GDP on the quality of institutions. Among other religious creeds we find a significant negative impact of Buddhism with the interference of government in the market (EFW2).

Our findings on the impact of Christian cultural background on institutional quality suggest that the effects from institutional quality may be transmitted to economic growth to the extent that reduced government interference in market mechanisms and developed financial institutions[20] are supportive of economic development.

The results of the relationship between religious backgrounds and the EFW7 indicator of institutional quality evidenced a significant effect of the Christian culture on financial institutions, consistently with the arguments developed in Section 15.2. Once we recognise such a difference, we wonder whether traditional cross-country analyses on the relationship between finance and growth should take it into account. A recent generation of finance-growth theoretical models definitely suggests that we should.

Saint Paul (1992) identifies a trade-off between technological diversification (which implies despecialisation and no choice of the more specialized technology) and financial diversification. The development of financial markets allows entrepreneurs to achieve diversification on financial markets and hence reduce technological diversification by choosing the riskier and more profitable technology. The model has multiple equilibria. If financial markets are insufficiently developed and it is too costly for entrepreneurs to access them, a low growth–low financial development equilibrium is achieved. If, on the contrary, the development of financial markets is such that entrepreneurs' costs of access are lower than a given threshold,[21] a high growth–high financial market equilibrium is achieved.[22]

To test the institution–growth relationship we adopt the traditional Mankiw–Romer–Weil (1992) approach, including in the estimate measures of bottlenecks reducing ICT (only when the regression is run on the 1985–97 sample in which these data are available), economic freedom and religious affiliation.[23] The estimated level and conditional convergence equations are respectively:

Table 15.3 Institutions, economic freedom and religious affiliation. The institution–growth nexus

Variable	EFW2	EFW7
Christian	3.530	2.499
	[3.83]	[2.15]
Muslim	1.330	0.742
	[1.57]	[0.69]
Hindu	1.426	1.186
	[1.17]	[0.78]
Buddhist	3.820	2.591
	[2.61]	[1.41]
Confucianist	–9.543	8.762
	[–0.70]	[0.48]
Gdpl60	0.549	0.665
	[2.48]	[2.39]
Civlaw	–0.509	0.437
	[–0.68]	[0.46]
Comlaw	0.337	1.292
	[0.47]	[1.42]
Engl	0.891	0.207
	[2.02]	[0.37]
D_oecd	–0.365	1.464
	[–0.73]	[2.33]
Geography	0.055	–0.196
	[0.26]	[–0.75]
Constant	–3.145	–1.148
	[–1.54]	[–0.45]
	15.7	5.34
LR Test*	(.000)	(.002)
Observations	74	74
Log Likelihood	–119.15	–136.14

Notes:

Variable legend: EFW(2) *Structure of the Economy and Use of Markets* EFW(7) *Freedom of Exchange in Capital and Financial Markets* (for details see Table 15.1); Christian: share of Christians; Muslim: share of Muslims; Hindus: share of Hindus; Buddhist: share of Buddhists; Confucianist: share of Confucianists; Gdpl60: GDP per ILO worker in 1960; Civlaw: dummy which assumes unity for countries in which civil law prevails, Comlaw: dummy which takes value of one for countries in which common law prevails; Engl: dummy which assumes unity for countries in which English is among the main spoken languages; D_oecd: dummy for OECD countries; Geography: log of the distance from Equator.

T-stats are in square parentheses. * LR test null hypothesis: the inclusion of the Christian variable does not improve goodness of fit of the model.

Table 15.4 Level and growth equations. Institutional variable: unweighted average of EFW2 and EFW7 – sample period 1960–97

	OLS level estimate		IV level estimate		OLS growth estimate		IV growth estimate	
$\text{Ln}(s_k)$	0.549	0.568	0.374	0.077	0.834	0.818	0.769	0.740
	[2.23]	[2.57]	[2.73]	[0.12]	[6.06]	[6.25]	[2.54]	[2.75]
$\text{Ln}(s_h)$	0.811	0.671	1.001	0.781	0.149	0.146	0.196	0.183
	[7.23]	[6.36]	[4.39]	[2.45]	[1.80]	[1.86]	[2.37]	[2.97]
$\text{Ln}(n+g+d)$	–0.911	–0.962	–0.243	–0.047	–0.482	–0.434	–0.663	–0.997
	[–2.25]	[–2.15]	[–0.37]	[–2.08]	[–1.77]	[–1.68]	[–3.00]	[–1.61]
Efw2chr	0.191		0.215		0.072		0.069	
	[5.34]		[2.12]		[3.37]		[1.93]	
Efw7chr		0.197		0.260		0.081		0.137
		[7.57]		[2.11]		[4.60]		[1.69]
$\text{Ln}(gdp_0)$					–0.253	–0.300	–0.175	–0.052
					[–4.76]	[–5.67]	[–2.01]	[–0.19]
Constant	4.058	3.771	5.223	4.868	–1.441	–1.028	–1.931	–3.194
	[3.09]	[3.20]	[2.20]	[1.96]	[–1.72]	[–1.28]	[–1.32]	[–1.24]
Implied α	0.22	0.254	0.157	0.041				
Implied β	0.34	0.300	0.421	0.420				
Countries	105	105	67	67	88	88	65	65
Joint significance of regressors	–91.949	–81.303	–17.443	–14.241	–15.234	–13.044	–5.152	–7.315
Instruments			$\text{Ln}(s_k), \text{Ln}(s_h), \text{Ln}(n+g+d)$, chris60				$\text{Ln}(s_k), \text{Ln}(s_h), \text{Ln}(n+g+d)$, chris60	

Note:
The table reports results from the following level equation

$$\ln\left(\frac{Y_t}{L_t}\right)=\lambda_0 c+\lambda_1\ln\left(s_k\right)+\lambda_2\ln\left(s_h\right)+\lambda_3\ln\left(n+g+\delta\right)+\lambda_4 EFW\_+\varepsilon_i$$

and growth equation

$$\ln\left[\frac{Y}{L}(t)\right]-\ln\left[\frac{Y}{L}(0)\right]=c'+(1-e^{-\lambda t})\frac{\alpha}{1-\alpha-\beta}\ln\left(s_k\right)+(1-e^{-\lambda t})\frac{\beta}{1-\alpha-\beta}\ln\left(s_h\right)+$$

$$-(1-e^{-\lambda t})\frac{\alpha+\beta}{1-\alpha-\beta}\ln\left(n+g+\delta\right)+(1-e^{-\lambda t})\ln\left[\frac{Y}{L}(0)\right]+(1-e^{-\lambda t})\gamma EFW\_+u_i$$

where s_h, s_k and n (gross enrolment ratio, investment to GDP ratio and rate of population growth) are calculated as estimation period averages, while the dependent variable (gross domestic product per working age population in purchasing power parity) is measured at the end of the period. EFW_: values of, alternatively, EFW7 (Efw7chr) or EFW2 (Efw2chr) if the share of Christians is higher than zero, zero otherwise. In IV estimates chris60 is the share of Christians in 1960.

$$\ln\left(\frac{Y_t}{L_t}\right)=c+\lambda_1\ln\left(A_{BR-ICT(0)}\right)+\lambda_2 g_{BR-ICT}t+\frac{\alpha}{1-\alpha-\beta}\ln\left(s_k\right)+\frac{\beta}{1-\alpha-\beta}\ln\left(s_h\right)+$$

$$-\frac{\alpha+\beta}{1-\alpha-\beta}\ln\left(n+g+\delta\right)+\gamma EFW_j \tag{15.2}$$

and:

$$\ln\left[\frac{Y}{L}(t)\right]-\ln\left[\frac{Y}{L}(0)\right]=c'+g_{BR-ICT}t+(1-e^{-\lambda t})\frac{\alpha}{1-\alpha-\beta}\ln\left(s_k\right)+$$

$$+(1-e^{-\lambda t})\frac{\beta}{1-\alpha-\beta}\ln\left(s_h\right)+(1-e^{-\lambda t})\frac{\alpha+\beta}{1-\alpha-\beta}\ln\left(n+g+\delta\right)+$$

$$+(1-e^{-\lambda t})\ln\left[\frac{Y}{L}(0)\right]+(1-e^{-\lambda t})\ln(A_{BR-ICT(0)})+(1-e^{-\lambda t})\gamma EFW_j \tag{15.3}$$

where $c'=g_{KP}t+(1-e^{-\lambda t})\ln(A_{KP(0)})$.

Variables for our empirical analysis are taken from the World Bank database. The dependent variable *Y*/*L* is the gross domestic product per working-age person[24] converted into international dollars using purchasing power parity rates,[25] *L* is the number of people in the working age cohort (population aged between 15 and 64). s_k is gross domestic investment over GDP, s_h is the (secondary education) ratio of total enrolment, regardless of age, to the population of the age group that officially corresponds to the level of education shown (generally the 14–18 age cohort).[26] $(n + g + d)$ is the usual variable summing up the rate of growth of the population, the rate of growth of technological progress and the depreciation rate of physical capital. $A_{BR-ICT(0)}$ and g_{BR-ICT} are respectively the beginning of period stock and the rate of growth of factors reducing ICT bottlenecks.[27] We consider an unweighted average of the four different proxies: i) the number of main telephone lines per 1000 inhabitants;[28] ii) internet hosts (per 10 000 people) or the number of computers with active Internet Protocol (IP) addresses connected to the Internet, per 10 000 people; iii) mobile phones (per 1000 people); iv) personal computers (per 1000 people). Finally, EFW_j is the institutional variable represented, alternatively, by the EFW2 or EFW7 indicator when the country has a non-zero share of Christian believers, and zero otherwise.[29]

When we inspect the distribution and correlation among variables used to estimate equations (15.1) and (15.2) (Tables 15.5 and 15.6), what really

Table 15.5 Variance–covariance matrix of regressors included in the estimates

	Efw2	Efw7	Cat	Prot	Muslims	A_{BR-ICT}	Engl	Comlaw	Civlaw	Hindu	Budd	Confuc	Distance	s_k	s_h
Efw2	1														
Efw7	0.930	1													
Cath	0.648	0.591	1												
Prot	0.403	0.518	–0.216	1											
Muslims	–0.637	–0.653	–0.566	–0.303	1										
A_{BR-ICT}	0.571	0.622	0.293	0.422	–0.485	1									
Engl	–0.017	–0.087	–0.241	0.074	–0.046	–0.070	1								
Comlaw	0.128	0.170	0.390	–0.072	–0.067	0.037	–0.614	1							
Civlaw	0.017	–0.017	–0.297	0.209	–0.054	0.103	0.643	–0.832	1						
Hindu	–0.248	–0.258	–0.214	–0.101	–0.070	–0.171	0.201	–0.307	0.319	1					
Budd	–0.257	–0.264	–0.206	–0.113	–0.101	0.016	0.171	–0.197	0.206	0.043	1				
Confuc	–0.134	–0.138	–0.106	–0.059	–0.065	–0.022	0.174	–0.186	0.192	–0.030	0.624	1			
Distance	0.165	0.204	–0.088	0.317	–0.104	0.474	–0.057	–0.006	0.087	0.023	–0.075	–0.043	1		
s_k	–0.062	–0.090	–0.107	0.010	–0.029	0.454	0.081	–0.090	0.114	0.013	0.314	0.213	0.181	1	
s_h	0.469	0.509	0.196	0.379	–0.494	0.872	0.032	–0.029	0.139	–0.049	0.082	–0.037	0.506	0.549	1

Legend:

Cath: share of Catholics, Prot. Share of Protestants: Muslim: share of Muslims; Hindu: share of Hindus; Budd: share of Buddhists; Confuc: share of Confucianists; Civlaw: dummy which assumes unity for countries in which civil law prevails, Comlaw: dummy which assumes unity for countries in which common law prevails; Engl: dummy which is unity for countries in which English is one of the main spoken languages; Distance: log of distance from Equator of country capital; s_h: secondary school gross enrolment ratio; s_k : investment to GDP ratio. A_{BR-ICT} is the stock of BR–ICT factors. Efw(II) *Structure of the Economy and Use of Markets* (Production and allocation via governmental and political mandates rather than private enterprises and markets). Efw(VII) *Freedom of Exchange in Capital and Financial Markets*. Any of the economic freedom indicators has a 0–10 value range. A higher value means a higher level in the item considered by the indicator .

Table 15.6 Distribution of the main quantitative variables used in growth estimates (1985–97 averages)

Percentiles	Investment in physical capital (% of GDP)	Investment in human capital (Gross enrolment ratio as a % of GDP)	Income in PPA per working age population (per capita 1985 US $)	Income in PPA per worker	Trend income in PPA per working age population	Trend income in PPA per worker	Buddhists
1	6.620	4.675	195.834	216.245	200.082	216.267	0
10	13.058	8.873	446.769	503.295	448.290	503.070	0
20	15.671	17.278	671.443	764.644	675.713	764.719	0
30	17.602	23.725	1 047.948	1 260.976	1 040.689	1 261.166	0
40	19.640	37.000	1 688.625	2 325.163	1 685.002	2 321.192	0
50	21.089	46.150	2 704.084	3 844.928	2 698.030	3 844.621	0
60	22.126	55.864	4 193.635	6 432.814	4 194.923	6 433.710	0
70	23.714	73.167	6 661.203	9 600.666	6 706.979	9 596.848	0
80	25.912	85.991	22 572.370	33 133.670	22 577.790	33 133.680	0
90	29.419	101.073	38 623.740	47 839.680	38 701.350	47 827.540	0.00676
99	37.950	124.473	64 473.260	82 969.690	64 325.350	83 904.640	0.9
Percentiles	**Confucianists**	**Hindus**	**Muslims**	**EFW(2)**	**EFW(7)**	**Catholics**	**Protestants**
1	0	0	0.000	0.000	0.000	0	0
10	0	0	0.000	1.868	2.435	0	0
20	0	0	0.000	2.450	3.385	0	0
30	0	0	0.000	2.885	4.053	0	0
40	0	0	0.000	3.530	4.820	0.014	0
50	0	0	0.005	4.163	5.425	0.139	0.006
60	0	0	0.058	4.560	6.270	0.260	0.037
70	0	0	0.135	5.115	7.283	0.650	0.066
80	0	0	0.777	5.700	7.867	0.845	0.200
90	0	0.0176	0.934	6.548	8.858	0.912	0.468
99	0.013	0.9	0.990	9.675	9.950	0.980	0.946

stands out is the difference in the distribution of physical and human capital investment in our sample. The reduced distance between the 10th and the 90th percentile values (respectively 13 and 29 per cent) indicates that convergence in the investment to GDP ratio is almost achieved, while this is not the case when we look at human capital investment (where the 10th and 90th percentile values are respectively 8 and 101 per cent).[30]

Besides the already analysed relationship between religious backgrounds and institution building, the correlation matrix indicates that the share of Protestants has a more positive relationship than the share of Catholics with the stock of factors easing access to ICT (0.42, against 0.29) and with human capital (0.38, against 0.19).[31]

Results from econometric estimates on all countries when the institutional variable is represented by, alternatively, the EFW2 or the EFW7 indicator multiplied by the share of Christians are reported in Table 15.4. An initial significant empirical finding is that the effect of the institutional variable on growth is positive and significant (Table 15.4) and therefore our hypothesis that, by significantly affecting rules and institutions, religious creeds indirectly affect economic growth is not rejected. The result is confirmed in both level and growth estimates, with OLS and IV approaches.[32] Under closer scrutiny, we observe that the first column in Table 15.4 meets all predictions of the MRW model. Coefficients of physical capital, human capital and of the $n+g+d$ variable are significant and correct. Implied factor shares are in the range of those traditionally estimated.[33] The weak significance of the human capital coefficient in growth equations is a typical finding in growth estimates (columns 5 and 6). Interestingly, the adoption of the religion variable among instruments reinforces the significance of this variable (columns 7 and 8).

15.5. ROBUSTNESS CHECK

To reduce the risk that our results are sample-specific we performed a robustness check where: i) we work on an alternative and shorter sample period (1975–97 and 1985–97); ii) ICT factors are added in the 1985–97 sample estimate and iii) regressions are also run on the subsample of non-OECD countries.[34]

In Table 15.7 we show that the significance of our institutional variable for Christian countries is confirmed when we restrict the sample to non-OECD countries and augment the specification with ICT factors. We also observe that the significance of the coefficient in growth estimates tends to be stronger for non-OECD countries when we consider the extended (1960–97) time period.

*Table 15.7 Robustness test on the significance of institutional indicators in different sample periods, ICT augmented estimates and in the non-OECD sample**

Sample period (1960–97)

	NON-OECD countries only							
	OLS level estimate		IV level estimate		OLS growth estimate		IV growth estimate	
Efw2chr	0.166		0.186		0.088		0.058	
	[4.75]		[2.65]		[3.54]		[3.97]	
Efw7chr		0.143		0.260		0.076		0.084
		[4.55]		[2.22]		[3.44]		[3.86]

Sample period (1975–97)

	All countries							
	OLS level estimate		IV level estimate		OLS growth estimate		IV growth estimate	
Efw2chr	0.058		0.057		0.093		0.0913	
	[1.97]		[2.06]		[2.01]		[1.99]	
Efw7chr		0.092		0.103		0.084		0.101
		[2.13]		[2.01]		[2.16]		[1.96]
	NON-OECD countries only							
	OLS level estimate		IV level estimate		OLS growth estimate		IV growth estimate	
Efw2chr	0.079		0.080		0.056		0.073	
	[2.71]		[0.86]		[2.79]		[1.96]	
Efw7chr		0.102		0.133		0.065		0.149
		[4.33]		[0.86]		[3.76]		[1.69]

Sample period (1985–97)

	All countries							
	OLS level estimate		IV level estimate		OLS growth estimate		IV growth estimate	
Efw2ch	0.141		0.142		0.064		0.043	
	[3.25]		[2.10]		[2.14]		[1.88]	
Efw7chr		0.131		0.214		0.053		0.078
		[3.12]		[1.99]		[2.14]		[1.87]
	ICT augmented estimates – All countries							
	OLS level estimate		IV level estimate		OLS growth estimate		IV growth estimate	
Efw2chr	0.092		0.090		0.088		0.055	
	[2.15]		[1.99]		[2.10]		[1.86]	
Efw7chr		0.192		0.103		0.097		0.103
		[3.13]		[1.66]		[2.36]		[1.90]

Note:
The table reports coefficient magnitude and t-stats (in parentheses) for the coefficient of the institutional variable in the specification of level and growth estimates described in Table 15.3 legend. Full estimate results are available from the authors upon request.

15.6. CONCLUSIONS

It is reasonable to imagine the wealth of a nation as being determined by a complex pattern of relationships involving its legacy from religious traditions, institutional shaping, capacity to spread innovation and investment in physical and human capital. This chapter provides a theoretical interpretation and an empirical test of these patterns.

The chapter's results suggest that Christian religious and cultural backgrounds have a strong impact on the way institutions are designed. The reduced role of state interference on markets, the greater emphasis on economic freedom and on the development of financial institutions are shown to be significantly related to the specific cultural background originating from the Reformation and recently assimilated also by Catholic countries.

We believe that these results and, in particular, that regarding the significance of the indicators measuring the freedom and development of the financial and banking system, provide a contribution to the empirics of growth in several respects and reconcile empirical findings with the most recent theories on finance and growth. In countries with religious and cultural backgrounds unfavourable to the development of financial institutions fixed costs of accessing debt and equity markets are high and sources of external finance are costly. These countries do not possess enough instruments to diversify risk, thereby reducing the capacity of entrepreneurs to invest in risky activities. They are therefore trapped in a low growth equilibrium in which financial institutions are underdeveloped. In countries in which cultural and religious backgrounds fostered the development of financial institutions, intertemporal and cross-sectional risk-sharing induce easier access to external finance and allow entrepreneurs to invest in risky activities. In these countries a virtuous positive relationship between financial development and growth exists.

NOTES

1. According to these authors, financial institutions improve the screening and monitoring of investment projects, provide mobilisation and aggregation services to savings and enhance opportunities for risk management and liquidity.
2. Most of them focus on a more specific issue in this literature, the Weberian hypothesis, which postulates a positive relationship between protestant ethics and growth (Iannaccone, 1998). The Weberian hypothesis has been criticised on theoretical grounds by historians (Tawney, 1926; Viner, 1978; Trevor-Roper, 1969; Novak, 1994) and has not yet found adequate support in the empirical evidence.

3. Within such literature, Sala-i-Martin (1987) tests the marginal impact of religious affiliation together with a large set of potential regressors and finds a negative and significant impact of both Catholicism and Protestantism. Grier (1997) finds that former British colonies grew more than former Spanish colonies but he also shows that controlling for Protestantism does not help to close the development gap between them. Blum and Dudley (2001) compare Catholic and Protestant cities between the 16th and the 18th century and conclude that the difference in growth is partially attributable to a difference in religious beliefs.
4. The difference with our approach is that the authors: i) focus on four religious backgrounds (including Islam and Buddhism); ii) attribute a country to a given church or religion if the religion is practised by the largest fraction of the population (and therefore do not use our alternative fuzzy approach explained in Section 15.3.1); iii) consider culture and law interaction effects; iv) look at different dependent variables considering individual and not composite governance indicators.
5. In his famous work Weber argues that orthodoxy is not crucial. What is crucial is the moral imprinting which can be still alive even when religious practice has disappeared.
6. The existence of a link between capitalism and the Protestant Reformation is also implicitly advocated by the *Encyclopedia Britannica* which gives the following explanation to the entry for *Capitalism:* 'CAPITALISM: also called free market economy, or free enterprise economy. Economic system, dominant in the Western world since the breakup of feudalism, in which most of the means of production are privately owned and production is guided and income distributed largely through the operation of markets. Although the continuous development of capitalism as a system dates only from the 16th century, antecedents of capitalist institutions existed in the ancient world, and flourishing pockets of capitalism were present during the later European Middle Ages. The development of capitalism was spearheaded by the growth of the English cloth industry during the 16th, 17th, and 18th centuries. The feature of this development that distinguished capitalism from previous systems was the use of the excess of production over consumption to enlarge productive capacity rather than to invest in economically unproductive enterprises, such as pyramids and cathedrals. *This characteristic was encouraged by several historical events. In the ethic encouraged by the protestant Reformation of the 16th century, traditional disdain for acquisitive effort was diminished, while hard work and frugality were given a stronger religious sanction. Economic inequality was justified on the grounds that the wealthy were also the virtuous.*'
7. We argue that religious beliefs still breed local cultures even though actual catholic and protestant societies are largely 'secularised'. To quote an example, obituaries in protestant societies are focused on the professional qualities and life activities of the defunct, while catholic burial ceremonies emphasise much more the moral qualities of the dead.
8. The above-described cultural differences were likely to affect not only human capital investment but also the extent to which human capital was directed toward productive activities. A well known paper finds that 'social' rewards to talent have a crucial influence on growth (Murphy et al., 1991). Religious or cultural beliefs may indeed place high social value on rent-seeking activities such as army membership. In these cases country talents will be oriented toward these activities and not toward entrepreneurship.
9. The Netherlands and the UK are in Europe but their financial markets are more similar to the market-oriented archetype, so that their protestant cultural background may have influenced their institutional choice. On the other hand, financial markets in Scandinavian countries are not so similar to the bank-oriented archetype and Germany (a country which is divided between Protestants and Catholics) is the most relevant example of the bank-oriented archetype.

10. Among indexes of economic freedom published in the *Economic Freedom of the World: 2000 Annual Report* we select two indicators measuring respectively i) government interference in the economy and in market rules and ii) development of capital and financial markets. The two selected indicators (among the seven available in the *Report*) are EFW(II) *Structure of the Economy and Use of Markets (*Production and allocation via governmental and political mandates rather than private enterprises and markets) i) Government Enterprises and Investment as a Share of the Economy (32.7 per cent); ii) Price Controls: Extent to which Businesses Are Free to Set Their Own Prices (33.5 per cent); iii) Top Marginal Tax Rate (and income threshold at which it applies) (25.0 per cent); iv) The Use of Conscripts to Obtain Military Personnel (8.8 per cent). EFW(VII) *Freedom of Exchange in Capital and Financial Markets*, i) Ownership of Banks: Percent of Deposits Held in Privately Owned Banks (27.1 per cent); ii) Extension of Credit: Percent of Credit Extended to Private Sector (21.2 per cent); iii) Interest Rate Controls and Regulations that Lead to Negative Interest Rates (24.7 per cent); iv) Restrictions on the Freedom of Citizens to Engage in Capital Transactions with Foreigners (27.1 per cent).
11. Fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth–truth values between ‘completely true’ and ‘completely false’. It was introduced by Lotfi Zadeh in the 1960s. In short with fuzzy logic, the *tertium non datur* of conventional logic does not apply. This means that one object does not belong either to one set or another (in our case one country does not belong either to the group of Catholic or to the group of Protestant countries) but it may be partly attributed to both sets in given proportions. In our case this partial affiliation is easily given by the share of Catholic and Protestant believers in a given country. The fuzzy logic helps us to solve the problem of countries where the difference between Catholics and Protestants (in terms of share of the population) is small.
12. The difference between the two approaches is particularly relevant for some countries. Consider for instance the case of Switzerland, the Netherlands and the United States which all have two almost equal shares of Catholics and Protestants. The results which follow have been shown to be nonsensitive to observations from these three countries.
13. The move from the dichotomous to the fuzzy approach generates an increase in the degrees of freedom with consequences on the significance of tests since variance does not change much.
14. The Second Vatican Council generated a revolution which reduced the gap between Catholic and Protestant creeds on different perspectives: i) ecumenism; ii) access to the Bible of individuals iii) responsibility of laics in church life; iv) cult.
15. Durlauf and Quah (1996), in their survey on the empirics of economic growth, identify distance from the equator as one of the relevant factors of conditional convergence. Acemoglu (2000) proposes an interesting historical explanation for the effects of latitude on growth, arguing that colonisers settled down (and invested in infrastructures and institutional rules) chiefly in those countries in which the climate was temperate. This is why we introduce the variable among controls of the determinants of institutional quality.
16. Consider that the EFW2 index is such that a higher value indicates less government interference on price and other market mechanisms.
17. These findings are confirmed if we repeat our estimates with dichotomic indicators of religious cultural background (that is dummies which assume the value of one if the majority of believers follow that religion and zero otherwise). Estimates are omitted for reasons of space and are available from the authors upon request.
18. Results from the 1985–97 sample are omitted for reasons of space and are available upon request.
19. Estimates are omitted for reasons of space and are available from the authors upon request.

20. More specifically, EFW7 measures a crucial variable, such as government control of the banking system, which has been shown to affect growth. La Porta et al. (2001) find that government ownership of banks does not lead to rapid growth of financial intermediation and has negative effects on growth.
21. Note that the model assumes that costs of access to financial market are fixed while costs of technological diversification are proportional to the output. Therefore the relationship between growth and financial markets is biunivocal and, after a given threshold, growth positively affects financial market development which, in turn, has positive feedback effects on growth.
22. An almost observationally equivalent result of biunivocal finance–growth relationship after a given threshold is found by Harrison, Sussman and Zeira (1999) and Deidda (2001). The former find that economic growth increases banks' activity and promotes new entries. Entries reduce costs of financial intermediation and in turn boost investment and growth. The second argues that, in an economy with risk-averse savers and learning by lending, transition from financial repression to full financial liberalisation may initially lead to a recession while, with the increasing level of expertise and institutional quality, it guarantees a growth-inducing allocation of financial resources. Therefore, after a learning period, financial development leads to a high growth equilibrium.
23. The theoretical foundation of this specification augmented with BR–ICT variables is described in Adriani and Becchetti (2004).
24. We perform the estimate with four different specifications which alternatively consider the ILO labour force and population in working age as labour inputs and observed income or trend income as a dependent variable. The ILO labour force includes the armed forces, the unemployed, and first-time job-seekers, but excludes homemakers and other unpaid caregivers and workers in the informal sector. We use trend income as an alternative to observed income to avoid that our results be influenced by cyclical effects on output (Temple, 1999). Full estimates results are available upon request.
25. An international dollar has the same purchasing power over GDP as the US dollar in the United States.
26. It is also defined as gross enrolment ratio to compare it with the net enrolment ratio in which the denominator is the enrolment ratio only of the age cohort officially corresponding to the given level of education.
27. The empirical literature on growth usually neglects the impact of technological progress on the differences between rich and poor countries by implicitly assuming that knowledge and its incorporation into productive technology is a public good, freely available to individuals in all countries (Temple, 1999). This approach cannot be applied to one of the most important sources of innovation in the last few decades (Information and Communication Technology) since ICT is a bundle of quasi-public knowledge products and non-public goods, needed for the fruition of the *knowledge products* themselves. Knowledge products are in fact weightless, expansible and infinitely reproducible (software, databases). They may be considered almost as public goods since expansibility and infinite reproducibility make then non-rival in consumption, and copyright protection makes them much less excludable than other innovations such as new drugs which are protected by patents (Quah, 1999). If ICT were to consist only of knowledge products, it would be available everywhere almost immediately regardless of the country in which it was created. This does not occur, though, since the immediate diffusion and availability of knowledge products is prevented by some 'bottlenecks'. In our opinion these '*bottlenecks*' are: i) the capacity of the network to carry the largest amount of knowledge products in the shortest time; ii) the access of individuals to the network in which knowledge products are immaterially transported; and iii) the power and availability of terminals which process, implement and exchange

knowledge products which flow through the network. We therefore argue that bottleneck-reducing factors such as the diffusion and power of personal computers, the diffusion of Internet access and the capacity of the network have been crucial determinants of the wealth of nations in these last two decades and we wish to establish how deep fundamentals have affected domestic diffusion of ICT technology.

28. Telephone mainlines are telephone lines connecting a customer's equipment to the public telephone network. Data are presented per 1000 people for the entire country.
29. Results which follow are substantially unchanged if we raise the threshold to a 10 per cent share of believers.
30. Secondary school gross enrolment ratio may be more than 100 per cent as the denominator is represented by the age class while the nominator may include students who are more than 18 years old.
31. The table is omitted for reasons of space and is available from the authors upon request.
32. The rationale for using IV estimators is that, even though the distance between institutional variables and the dependent variable should reduce the risk of endogeneity typical of the finance–growth relationship, the risk does not disappear completely if we consider that the dependent variable is highly serially correlated. Therefore we use our previously tested relationship between institutions and religious background, instrumenting the former with the latter. Religious backgrounds vary just slightly across decades (even under the impact of economic growth) but can definitely be considered exogenous by construction here since country participation in a given religious world was invariant in the last half of the century and not affected by economic growth of the sample period (or by previous economic growth correlated with it).
33. Mankiw, Romer and Weil (1992) expect a human capital share between 1/2 and 1/3, while their estimate of the physical capital factor share ranges from 0.4 to 0.14 when considering OECD countries only.
34. Detailed results of these estimates are available upon request.

REFERENCES

Acemoglu, D., S. Johnson and J. Robinson (2001), 'The colonial origins of comparative development: an empirical investigation', *American Economic Review*, **91**:1369–401.

Adriani, F. and L. Becchetti (2004), 'ICT bottlenecks and the wealth of nations: a contribution to the empirics of economic growth', *Economics of Innovation and New Technology*, **14**(6):435–53.

Alesina, A. and R. Perotti (1994), 'The political economy of growth: a critical survey of the recent literature', *World Bank Economic Review*, **8**(3):351–71.

Allen, F. and D. Gale (1997), 'A welfare comparison of intermediaries and financial markets in Germany and the US', *European Economic Review*, **39**:179–209.

Blum, U. and L. Dudley (2001), 'Religion and economic growth: was Weber right?', *Journal of Evolutionary Economics*, **11**:207–30.

Brennan, M. and A. Kraus (1987), 'Efficient financing under asymmetric information', *Journal of Finance*, **42**:1225–43.

Costantinides, G. and B. Grundy (1989), 'Optimal investment with stock repurchase and financing signals', *The Review of Financial Studies*, **2**:445–66.

Deidda, L. (2001), ‘Financial institutions expertise and growth effects of financial liberalisation’, paper presented at the X Tor Vergata Financial Conference.

Delacroix, J. (1992), ‘A critical empirical test of the common interpretation of the protestant ethic and the spirit of capitalism’, Paper presented at meeting of Int, Ass. Business and Society, Leuven, Belgium.

Delacroix, J. (1993), ‘Religion and economic action: the protestant ethics, the rise of capitalism and the abuses of scholarship’, *Journal of the Scientific Study of Religion*, **34**(1):126–7.

De Soto, H. (2001), *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else,* Milan: Garzanti.

Durlauf, S.N. and D.T. Quah (1998), ‘The new empirics of economic growth’, *Center for Economic Performance Discussion Paper 384*.

Grier, R. (1997), ‘The effect of religion on economic development: a cross national study of 63 former colonies’, *Kyklos*, **50**(1):47–62.

Hall, R.E. and C.I. Jones (1997), ‘Levels of economic activities across countries’, *American Economic Review*, **87**(2):173–7.

Harrison, P., O. Sussman and J. Zeira (1999), ‘Finance and growth: theory and new evidence; Joseph Board of Governors of the Federal Reserve System’, *Finance and Economics Discussion Series 99/35*.

Iannaccone L.R. (1998), ‘Introduction to the economics of religion’, *Journal of Economic Literature*, **36**:1465–95.

Jensen, M. (1986), ‘Agency costs of free cash flow, corporate finance and take-overs’, *American Economic Review*, **76**:323-39.

Jensen, M.C. and W.H. Meckling (1976), ‘Theory of the firm: managerial behaviour, agency costs and ownership structure’, *Journal of Financial Economics*, **3**:305–60.

Knack, S. and P. Keefer (1997), ‘Does social capital have an economic payoff? A cross-country investigation’, *Quarterly Journal of Economics*, **112**:1251–88.

Landes, D. (2000), ‘Culture makes almost all the difference’, in L.E. Harrison and S.P. Hungtington (eds), *Culture Matters*, New York: Basic Books.

La Porta, R., F. Lopez-de-Silanes and A. Shleifer (2001), ‘Government ownership of banks’, mimeo.

Lehr, B. and F. Lichtenberg (1999), ‘Information technology and its impact on productivity: firm-level evidence from government and private data sources, 1977–1993’, *Canadian Journal of Economics*, **32**(2):335–62.

Leland, H. and D. Pyle (1987), ‘Information asymmetries, financial structures and financial intermediation’, *Journal of Finance*, **32**:321–38.

Levine, R. (1997), ‘Financial development and economic growth: views and agenda’, *Journal of Economic Literature*, **35**(2):688–726.

Lipford, J., R.E. McKormick and R.D. Tollison (1993), ‘Preaching matters’, *Journal of Economic Behaviour and Organisation*, **21**(3):235–50.

Mankiw, N.G., D. Romer and D. Weil (1992), ‘A contribution to the empirics of economic growth’, *Quarterly Journal of Economics*, **107**:407–37.

Mauro, P. (1995), ‘Corruption and growth’, *Quarterly Journal of Economics*, **110**(3):681–712.

Marshall G. (1982), *In Search of the Spirit of Capitalism: An Essay on Max Weber’s Protestant Ethic Thesis*, London: Hutchinson.

Murphy, K., A. Shliefer and R. Vishny (1991), 'The allocation of talent: implications for growth', *Quarterly Journal of Economics*, **106**:503–36.

Noe, T. (1988), 'Capital structure and signalling game equilibria', *Review of Financial Studies*, **1**:331–56.

Novak, M. (1994), *Catholic Ethics and the Spirit of Capitalism*, Milan: Comunità.

Oliner, S.D. and D.E. Sichel (1994), 'Computers and output growth revisited: how big is the puzzle?' *Brookings Papers on Economic Activity*, **2**:273–334.

Pagano, M. (1993), 'Financial markets and growth: an overview', *European Economic Review*, **37**:613-22.

Perotti, R. (1996), 'Growth, income distribution and democracy: what the data say', *Journal of Economic Growth*, **1**:149–87.

Persson, T. and G. Tabellini (1994), 'Is inequality harmful for growth?', *American Economic Rev*iew, **84**(3):600–621.

Quah, D. (1999), 'Technology and growth, the weightless economy in economic development', *LSE Discussion Paper 417*.

Rousseau, P.L. and P. Watchel (2001), 'Inflation, financial development and growth', paper presented at the X Tor Vergata Financial Conference.

Saint-Paul, G. (1992), 'Technological choice, financial markets and economic development', *European Economic Review*, **36**(4):763–81.

Sala-i-Martin, X. (1987), 'I just run two million regressions', *American Economic Review*, **87**(2):178–83.

Stein, J.C. (1992), 'Convertible bonds as backdoor equity financing', *Journal of Financial Economics*, **32**(1):3–21.

Stultz, R.M. and R. Williamson (2001), 'Culture, openness and finance', *NBER Working Paper 8222*.

Tawney, R.H. (1926), *Religion and the Rise of Capitalism*, New York: Harcourt, Brace & World, Inc.

Temple, J. (1999), 'The new growth evidence', *Journal of Economic Literature*, **37**:112–56.

Trevor-Roper, H. (1969), *Protestantesimo e trasformazione sociale*, Bari: Laterza.

Viner, J. (1978), *Religious Thought and Economic Society*, Durham: Duke University Press.

Watchel, P. (2001), 'Globalisation of banking: why does it matter?', in *Presente y Futuro del Sistema Financiero en Paises Emergentes*, Caracas, Venezuela.

Weber, M. (1930), *The Protestant Ethics and the Spirit of Capitalism*, NewYork: Scribner's.

16. Trade, poverty and growth: two perspectives, one message?*

Gabriella Berloffa and Maria Luigia Segnana

16.1. INTRODUCTION

Between 1990 and 1998, the headcount index of poverty fell from 29.3 per cent to 24.2 per cent, but the decline in the actual number of people living below the $1-a-day poverty line was more modest, falling from 1.3 billion to 1.2 billion. Clearly, these figures are sensitive to the data used and the time periods chosen; however, it seems clear that, although the proportion of the world's population living in poverty is falling, the actual number of the poor displays more limited change (Besley and Burgess, 2003). Furthermore, the poverty trajectories of different regions between 1990 and 1998 greatly diverged. In this period, for example, the poverty rate in East Asia dropped from 27.5 per cent to 15.3 per cent, whereas in sub-Saharan Africa poverty rates remained stagnant at between 47.6 per cent and 46.3 per cent. Understanding the main reasons for these different poverty-reduction performances is therefore of crucial importance.

The traditional view on how to reduce poverty has focused closely on boosting economic growth, and therefore on the policy variables which have proved significant in explaining cross-country differences in growth rates: trade openness, monetary and fiscal policy, financial development, and the rule of law. The empirical fact supporting this view was that the income share of the bottom quintile appeared to be insensitive to growth and to known growth determinants, so that growth affected the income of the poor in the same proportion as that of other groups, whatever the nature of the policy variables behind it and its sectoral structure.

This led to the result often summarized as 'Globalization is good for the Poor', with evidence from individual cases and cross-country analyses supporting the view that open trade regimes lead to faster growth and poverty reduction in poor countries (Dollar and Kraay, 2001, 2002). The core argument behind this result is that openness is associated with higher growth, but with no systematic tendency for inequality to increase, on average.

Therefore, the acceleration in growth rates that accompanies expanded trade usually translates into proportionate increases in the incomes of the poor.

Analytically, this view is very convenient because it severs the link between policies and poverty reduction into two steps:

- identify the main determinants (and therefore policies) which maximize the growth rate;
- identify the speed at which poverty would be reduced by any growth rate by estimating the poverty/growth elasticity.

This view has been recently criticized at both theoretical and empirical levels. This chapter starts from the separated agenda for growth-enhancement and poverty-reduction strategies and analyses their linkages rather than their separation from two different perspectives: that of sustaining economic growth[1] and that of poverty reduction. These linkages, supported by theoretical and empirical analyses, show that we can neither separate the problem of reducing poverty from the way in which growth is achieved[2] nor address the problem of supporting economic growth while ignoring the distributional consequences of the policies we wish to implement.

We derive our conclusions by examining the results obtained by two different currents of thought in the literature.[3] The first focuses on trade as a growth-supporting policy and its impact on poverty and inequalities. The second current focuses on the impact of initial inequality on growth (through different micro channels), and on the resulting effects on poverty and inequalities. Analysis of the results obtained by these two strands of the literature converges on similar findings, and allows some common conclusions to be drawn:

1. Initial as well as trade-induced inequalities matter for both the sustainability of growth processes, and the speed at which poverty is reduced.
2. As a consequence, the problem of supporting economic growth and the problem of reducing poverty are closely linked: we cannot address the former while ignoring its distributional consequences and we cannot address the latter while ignoring the way in which growth is achieved.
3. Specific conditions (sectoral composition, wealth and land distribution, distribution of schooling, convexity of earning profiles, specialisation of income sources, and so on), can explain why, at the macro level, similar growth rates may have such different impacts on poverty and why the same policies may have different effects on growth performance. Thus, the problem of poverty reduction cannot be separated from the context in which trade is liberalized.

The structure of the chapter is as follows. Section 16.2 focuses on the problem of supporting economic growth and ensuring its sustainability in the long run. It examines the trade–growth relationship, the impact of trade on within-country inequality, and the effects of the latter on growth. It shows that, whenever the effects of trade on inequality are considered, the latter may exert a negative impact on growth, and therefore that the response of growth to trade liberalization may be reduced or even reversed in the long run. Thus, the empirical evidence on the positive association between trade and growth is questioned by the effects of trade-induced inequalities on growth. Section 16.3 discusses trade as a poverty reduction strategy by examining its long-run effects on poverty via poverty/growth elasticity and its short-run effects via factor markets. It shows that, taken growth as given, its effects on poverty are affected by the level of initial or trade-induced inequality. Therefore, the time span for poverty reduction is definitely longer in the presence of growth-induced inequalities. Section 16.4 draws some conclusions from the two different perspectives, pointing out their linkages: the problem of reducing poverty cannot be separated from the way in which growth is achieved, and the problem of supporting growth cannot be separated from its distributional consequences.

16.2. THE PERSPECTIVE OF ECONOMIC GROWTH

Trade liberalization, openness and international linkages have often been viewed as the key growth strategies for countries, and they have acquired the status of ‘stylized facts’ in the growth literature.[4] On the other hand, trade openness has played a role in the growth (income)-inequality literature as a determinant of within-country inequality in that it acts through changes in commodity prices and thus affects factor income prices.[5] The effects of trade on growth and the effects of trade on inequality cannot be simply considered as two distinct issues reflecting two distinct concerns: increasing a country’s GDP and ensuring that the gains are equitably distributed. Indeed the distributional consequences of a given trade policy are a critical issue even for the first concern, that is for the prospects of economic growth in the long run. There is increasing evidence that inequality may be harmful for growth; therefore, if trade worsens within-country inequality, the process of growth that trade itself might enhance will not be sustainable in the long run.

In this section we show that the effects of trade on growth are usually positive in empirical studies, but they may be more elusive whenever the nature of trade and the forces it activates, its channels, the theoretical framework, and the subset of countries are considered. Furthermore, whenever one takes account of the short and long effects of trade on inequality, the latter may exert a negative impact on growth, acting as a diverting rather than a supporting force. We reach the conclusion that, when choosing a growth-supporting policy (and trade integration is often a good candidate), it is crucial to assess not only its direct growth-effects, but also its distributional impacts, because these may become a constraint on growth in the long run.

16.2.1. The Growth Impact of Trade: the Consensus Estimate

The usual size of the welfare gains from trade liberalization generates disappointing static estimates, seldom larger than 1 per cent of GDP. Why is trade liberalization often a priority policy for closing the 1 per cent difference in per capita incomes between developed and less developed economies? The solution lies in the often invoked dynamic relationships, with the widely accepted conclusion that international trade and economic growth are positively and significantly related.

The large body of empirical literature on the subject has emphasized the statistical significance of the trade–growth relationship more than its economic significance. The trade–growth nexus is quantitatively important because the 'consensus' estimate shows an average coefficient of 0.22 across a very large number of cross-section and time-series sources of growth regression models testing the relationships between trade and growth.[6] Many researchers have found evidence that this coefficient is different between groups of countries, suggesting that trade generates more growth in developed economies than in developing ones. Some problems and questions arise when conducting balanced evaluation of this empirical result.

First, opinions differ on the empirical evidence

Because of trade indicators. The empirical literature is still affected by serious limitations, and it is unable to generate satisfactory indexes of trade policy orientation. Indicators of trade policy have been inappropriately used, and they have been selected in order to systematically bias the results so that they show a statistically and quantitatively significant link between trade liberalization and growth (Rodriguez and Rodrik, 2000).

Because of multicollinearity, simultaneity, omitted variables bias, measurement errors, and so on, or the well known problems that may give rise to inconsistent results in many of the empirical studies of trade and growth, studies often motivated by the newer econometric methods and better data sets that promised to correct the statistical problems of earlier studies. At the same time, recent work shows that growth and trade have a common driver: institutions.[7] Researchers have sought good instruments

(Frankel and Romer, 1999) for actual trade/GDP ratios as well as for institutional quality (Acemoglu et al., 2001). The simultaneous use of these instruments is intended to identify the respective contributions of institutions, geography and trade to cross-country income levels (Rodrik et al. 2002), focusing on discrimination among competing stories in explanation of striking variations in cross-national incomes around the world: the conclusion being that the institutional explanation outweighs the other stories.

Because there is a gap between the results at the firm level and at the macro level. The proponents of trade liberalization have argued that at the firm level, it will force firms to produce closer to the production possibility frontier, and that the frontier will move out more rapidly. Empirical results (Bernard and Jensen, 1999a and 1999b) provide little evidence that firms derive technological or other benefits from exporting *per se*; the more common pattern is that efficient producers tend to self-select in export markets. This finding can be explained by a self-selection model (only the most productive firms engage in exporting) rather than by a learning by exporting model. In other words, the causality runs from productivity to exports, not vice versa. These results hold across countries at different levels of development[8] and indicate a potential gap between firms' absorption of exporting experience and country absorption of trade benefits: in other words, between the positive impact of trade at the macro level (Harrison, 1996) and the debatable impact of learning by exporting at the micro level (Bernard and Jensen, 2004; Tybout, 2003; Pavcnik, 2002).

Second, the question is whether or not there is any positive and economically significant link between trade and growth

Or in other words, whether or not the many econometric improvements have substantially aided understanding of why the statistically significant relationship is robust across many samples, data sets and regression models, corrected for omitted variables and simultaneity, and so on. Indeed, the 'missing discrimination' problem is still unsolved: Rodriguez and Rodrik's criticism that trade merely serves as a proxy for other important policy variables cannot be easily rejected. The channels through which trade influences growth are still undervalued.[9]

An answer to this question requires two viewpoints to be adopted: the first looks at the nature of trade flows because they may have level effects on output and consumption or growth effects or both;[10] the second looks at the theoretical framework utilized: since endogenous growth models are often thought to have provided the missing theoretical link between trade openness and long-run growth, it is useful to examine why such models may provide ambiguous rather than definite answers.

With respect to the first viewpoint, trade flows may be either in goods or in ideas and knowledge, or in both. Opening to trade in goods only will clearly raise the value of output at world prices at each point in time if there is no change in the path of factor accumulation. This is the so-called level effect. Whether there will be a growth effect (that is whether there will be any change in the economy's steady-state growth rate), and if there is, whether it will be transitory or permanent depends on the response of factor accumulation to the increase in income levels, and on whether the marginal returns to factor accumulation eventually diminish to zero.[11]

Ideas and knowledge are conduits for technological progress, and international spillovers are made possible by openness to international trade: ideas flow rapidly and machines incorporating better technologies can be imported. But even in this case, in the endogenous framework, a subset of countries may indeed experience trade-diminished growth depending on their initial factor endowments and levels of technological development. The ambiguous explanations of the growth effects of trade vary according to whether the forces of comparative advantage push the economy's resources in the direction of activities that generate long-run growth (through externalities in R&D, expanding product variety, upgrading product quality and so on) or whether they divert them from such activities. For example, a country lagging behind in technological development may be induced by trade to specialize in traditional goods and thus experiences a reduction in its long-run rate of growth (Grossman and Helpman, 1991; Matsuyama, 1992, 1996; Young, 1991).[12] Even in the presence of complete technological spillovers, and in the absence of any barriers to technology transfers, there may be problems of technological 'appropriateness' (or technology–skill mismatch) when technological trade flows are biased toward the needs of the richer economies, giving rise to productivity differences and to large output gaps between trading countries (Acemoglu and Zilibotti, 2001).[13]

These two viewpoints are useful for identification of when and why there are theoretical presumptions in favour of an unambiguous relationship between trade and growth rates in the types of cross-national data sets typically utilized. But they also show that it is relatively easy to construct a well specified model which generates the conclusions that many opponents of trade openness have reached – namely that free trade may be detrimental to some countries' economic prospects when these countries are lagging behind in technological development and have an initial comparative advantage in 'non-dynamic' sectors, or when technological–skill mismatches arise. More broadly, they illustrate that there is no *general* theoretical link between trade protection and growth once real world phenomena such as learning, technological change and market imperfections (captured by a learning by doing externality) are taken into consideration.

For these reasons, the openness–growth link cannot be an empirical matter alone. Researchers have not yet reached consensus on the effects of trade on growth at a theoretical level, and there are endless discussions on the empirical side. The debate is even greater if the distributional effects of trade openness and their impact on future growth are considered: in this case trade-induced inequalities matter for the sustainability of the growth process. We discuss this issue in the following sections, where we first look at the impact of trade on within-country inequality and then the impact of the latter on economic growth.

16.2.2. The Inequality Impact of Trade

The effects of trade on inequality can be examined in terms of its effects on wage inequality via the factor income share or via outsourcing. In both cases the effects concern the so-called skill-gap and trade may be a source of within-country inequality in one or all countries.

Establishing a link between trade and inequality is particularly difficult because internal and external liberalization may be at work,[14] and because different concepts of inequality may be used, together with a variety of methods and measurement techniques (see Kanbur, 2000; Goldberg and Pavcnik, 2004).

If we read the inequality impact of trade through the specific lens of the factor income implications[15] suggested by traditional or less traditional trade theories, we encounter at least one problem: the long-run Stolper–Samuelson's predictions are in contrast with the wage inequality impact of international trade and especially with the widening skill gap in less developed countries (LDCs).[16] Standard theory predicts that trade liberalization will increase *within*-country inequality in advanced countries (ACs) and decrease it in LDCs, thus implying an improvement in global inequality. The reason for this result is that trade widens the skill gap in ACs, and reduces it in LDCs.

As is well known, within-country inequality has indeed increased in ACs, and candidate explanations for this phenomenon have been trade with LDCs and skill-biased technical change.[17] However, as predicted by Stolper and Samuelson, trade integration could explain rising inequality within ACs if, at the same time, it is associated with a reduction of inequality in LDCs. But this prediction is in contrast with the growing body of empirical evidence that within-wage inequality has also increased in some LDCs since the early 1980s as well as in transitional economies. This is the case of East-Asian exporters, of six out of seven Latin American countries, of the Philippines, Eastern Europe and Central Asia.[18] Therefore, the observed evolution of inequality in both developed and developing countries cannot be traced back to trade liberalization, at least in the form examined in standard theory. If skill bias and skill premium matter for widening skill–gaps in LDCs, the mechanism must operate not via trade (in goods) but via, for example, foreign direct investment or the many forms of technology transfers. In this case, trade cannot be the main culprit, but is simply a conduit for North–South skill-upgrading transfer, an important aspect of technological change.

This role of trade as the conduit of technological transfer is emphasizsed by the case which envisages trade as a conduit for skill-upgrading between North and South. This literature integrates two strands: the sorting mechanism whereby product cycles, outsourcing or organizational change (Aghion et al., 1999) during the 1990s generated wage inequality in LDCs (Feenstra and Hanson, 1996a, 1996b; Harrison and Hanson, 1999); and product innovation and technology transfer as emphasised in both endogenous growth theory (Grossman and Helpman, 1991) and empirical work on international technology spillovers (Coe et al., 1997). In these cases, skill–gaps increase in both ACs and LDCs,[19] with a consequent deterioration in global inequality. Therefore, trade is a source of inequality within both countries and thus a source of deterioration in global inequality.

Left unresolved by all the interpretations of trade-induced wage inequalities are the questions as to which area, and who within the area, is gaining from trade reforms, especially at the empirical level.[20] On the one hand, various examples of increasing wage inequalities in the presence of trade reforms can be shown to be consistent with the many forms of trade as conduits for skill-upgrading transfers. On the other, internal factors as well as specific sector-bias or factor-bias, patterns of protection before trade reforms, political economy models of protection, FDI, distribution of human capital, all play a role in explaining wage distributional impacts.

Two examples may prove useful: Mexico, where wage inequality had been declining prior to the 1985 reform, and which experienced overall increasing inequality after the reform; and Colombia, where the overall effects of trade reforms on the wage distribution were modest compared to Mexico. The difference between the two cases warrants exploration.

Mexico is a particularly interesting case because wage inequality had been declining in the decades prior to reform in 1985. Following the trade reform, the ratio of skilled to unskilled wages increased dramatically, with the costs of adjustment falling disproportionately on unskilled workers. Examination of the Mexican case shows that much of the adjustment occurred through falling wages in previously protected sectors. Explanations for increasing wage inequality include outsourcing, foreign direct investment, or skill-biased technological change. Empirical results (Hanson and Harrison, 1999) show that the skills gap is wider in plants which upgrade intensively through licensing arrangements, that foreign investment locates in sectors with more income inequality, that foreign firms pay higher premiums to skilled workers, and that the share of royalty payments is positively and significantly correlated with relative wages. All these results point to the importance of labour demand by incoming foreign firms skewed towards skilled workers (Feenstra and Hanson, 1997). Mexican trade liberalization was undoubtedly disappointing for those who had hoped that globalization would benefit the poor in the developing countries.

The other empirical investigation, on Colombia's gradual trade liberalization beginning in 1985, focuses on the effects of drastic tariff reductions in the 1980s and the 1990s. The use of detailed micro data from the National Household Survey combined with data on trade policy changes enabled thorough examination (Attanasio et al., 2002) of how wage distribution was affected by different channels: through increasing returns to education (the skill premium), changes in industry wages, and inter-sectoral reallocation of the labour force. Overall, although inequality gradually increased over the period, the effects of trade reforms on wage distribution were small, and the increase was by no means as pronounced as in Mexico.[21]

The difference between Colombia and Mexico warrants further examination because it can show the conditions under which efficiency-promoting policies like trade reforms may have a significant (or relatively) small impact on wage distribution. The role of FDI in the case of Mexico (Cragg and Epelbaum, 1996; Feenstra and Hanson, 1997) and the active role of the Colombian government in improving social conditions may have magnified (or offset) the negative impact of trade reforms on wage distribution.

These two examples suggest that the effects uncovered by empirical work can be very different, and that governments may play an important role in offsetting the negative impact of trade reforms on inequality.[22] As already mentioned, understanding and balancing the distributional consequences of trade policies is crucial if we are interested in the sustainability of the growth process. Indeed, if trade worsens within-country inequality, trade-supported growth will not be sustainable in the long run because high inequality may be harmful for growth. In order to clarify this point, the next section discusses the main theoretical and empirical findings on the effects of inequality on growth.

16.2.3. The Growth Impact of Inequality

According to the traditional view, inequality was good for growth, and growth would either reduce inequality and poverty in the long run (Kuznets,

1955) or it would at worst be distribution-neutral, with the poor benefiting from growth like everyone else. Interest then turned to the possibility that inequality has a negative effect on growth, after Lucas (1993) raised the famous puzzle about the different growth performances achieved by South Korea and the Philippines between the early 1960s and the late 1980s.[23] In order to collect evidence on this relationship, a large number of empirical studies in the first half of the 1990s ran cross-country regressions of GDP growth on a vector of control variables and various measures of income and/or land inequality (as a proxy for wealth inequality). Unfortunately, the empirical literature is far from reaching a consensus on the sign of the relationship between inequality and growth. Results differ according to the data-set, the time interval and the inequality measure considered.[24] However, we can highlight the following points:

- studies that find a positive partial correlation between inequality and growth typically focus on the *short-run* relationship, usually considering five year intervals;
- studies that use data over a longer time span – generally 25 years – tend to find a negative partial correlation between inequality and growth, with the exception of Barro (2000) who finds little overall relation between inequality and rates of growth, but uses panel data for ten-year periods;
- a recent study by Knowles (2001) shows that when consistently measured data are used, there is no evidence of a significant correlation between *gross income* inequality and economic growth, whereas there is evidence of a significant negative correlation between net income (or expenditure) inequality and growth.

We can therefore agree with Ravallion (2001), who notes that ‘the existing evidence using cross-country growth regressions appears to offer more support for the view that inequality is harmful to growth than the opposite view’, but some qualifications are necessary. First, this relationship seems to emerge over long time periods and it seems to involve net rather than gross income. Second, at the empirical level it is not clear whether this relationship is different for countries at different stages of development.[25] More generally, results appear highly sensitive to the sample of countries included: in some studies, inequality becomes statistically insignificant when continental dummies are included or country fixed effects are allowed for (Ravallion, 2001). Third, it is not clear how inequality is related to other known growth determinants. In some cases, the coefficient on inequality is reduced (but remains significant) after the inclusion of variables that are typically negatively correlated with inequality, such as enrolment in and stocks of secondary education or regional dummies (Benabou, 1996); in other cases controlling for inequality reduces the coefficient associated with other factors, such as trade openness (Mbabazi et al., 2001).

Since results are sensitive to the countries included in the sample, and since specifications allowing for the proper interactions may be difficult to implement, one should turn to theoretical analyses for insights. Theory itself is not conclusive, because there are arguments in favour of both a positive and a negative effect of inequality on growth. However, these arguments provide guidance for judging whether in a particular context the forces underlying positive or negative arguments are more likely to emerge. Moreover, they help us identify which factors are likely to act in the short run and in the long run.

The main channels that link inequality to growth in the long run are physical and human capital investments. Traditionally, inequality was considered good for growth because the affluent have a greater propensity to save, and because of the indivisibility of investments.[26] However, in a world with decreasing returns, the marginal product of the poor is higher, and therefore, when *credit constraints* prevent the poor from undertaking the efficient amount of investment, the aggregate level of output, and in an endogenous growth model also its rate of growth, will be lower. A similar result is derived when investment involves a *minimum project size*, thus generating a threshold level of wealth below which agents do not invest, or do not leave enough to their offspring for them to invest. Unfortunately, the empirical evidence on this issue is rather limited.[27]

A similar process operates for human capital accumulation, especially when educational choices are modelled jointly with *fertility*. For example, Galor and Zang (1997) show that, for a given distribution of income, a higher (exogenously given) fertility rate means that fewer resources are available within each family to finance the education of each child; with fixed costs of education and borrowing constraints, fewer children will be able to attend school. In a more recent paper, de la Croix and Doepke (2003) highlight the role of the fertility *differential* between the rich and the poor in explaining why countries with greater inequality will accumulate less human capital, and therefore grow more slowly.[28] At the empirical level, Perotti (1996) finds that an increase in the share of the middle class is associated with a fall in fertility and with an increase in the female secondary school enrolment ratio, and that both these phenomena lead to higher growth.

Unequal access to investment opportunities across individuals, and a high degree of capital market imperfections may also generate persistent credit cycles,[29] leading to *macroeconomic fluctuations* and lower growth (Aghion et al., 1997, 1999). Indeed, empirical studies find that income inequality is positively correlated with volatility measured by the standard deviation of the annual rate of growth of GDP (Hausmann and Gavin, 1996; Breen and Garcia-Penalosa, 1999). Cross-country regressions also find that greater volatility of the growth rate consistently reduces the average rate of growth during that period, and that this is partly due to its deterring effect on physical and human capital investment (Ramey and Ramey, 1995; IDB, 1995).

These arguments suggest that inequality may be good for growth in those situations in which credit availability is low and large initial investments are necessary. However, when decreasing returns start to emerge, and when the average level of human capital of the population becomes important, a high level of inequality may have a negative effect on growth. Therefore, the implementation of a policy that boosts growth but at the same time increases inequality within a given country, may lead to positive results in terms of economic growth in the short run, but it may generate problems for the process of growth in the future.

It is worth noting that inequality may have a direct negative effect on growth even in the short run when it generates *sociopolitical instability*.[30] Indeed, a highly unequal and polarized distribution of resources creates strong incentives for organized individuals to pursue their interests outside normal market activities or the usual channels of political representation, engaging in rent-seeking behaviour or other manifestations of sociopolitical instability, such as violent protests, assassinations and coups. In turn, sociopolitical instability discourages investment by creating uncertainty over the political and legal environment, and by disrupting market activities and labour relations. Perotti (1996) shows that a larger share of the middle class is indeed associated with lower sociopolitical instability, and that this in turn is associated with higher growth; however, the estimated relationship is much stronger in rich countries.

Generally, the models mentioned above examine the effect of an unequal distribution of wealth or resources, without considering either the exact form of this distribution or its composition in terms of wealth sources of the different percentiles. These factors are particularly important in developing countries, in which particular forms of redistribution like land reforms are often invoked in order to support economic growth. Indeed, empirical evidence on the effects of these reforms is mixed. Aghion et al. (1999) stress that 'redistribution in the form of land or education reform has played an important role in fostering economic growth', whereas De Janvry et al. (2001) find that access to land is not a sufficient condition to secure higher household incomes. The explicit consideration of the composition in terms of wealth sources of the different percentiles is crucial not only because different forms of redistribution may have different consequences in terms of subsequent economic performances, but also because this composition is likely to affect the structure of growth, that is its sectoral and geographical composition, and this may be important for determining both the growth rate and its distributional effects.

Another aspect that the theoretical and empirical debate summarized above tends to ignore is the interaction between inequality and growth determinants other than physical and human capital accumulation. In particular, it is not clear whether the inequality effect that these models identify is simply an additive one which works independently on the level of other variables (for example the distribution of schooling or the degree of trade openness), or whether there are links between initial inequality and these other variables that may give rise to an overall effect that differs from the simple sum of the effects associated with each factor. As a consequence, it is rather difficult to determine whether policies aimed at reducing inequality have impacts on growth which differ according to the level of other factors, such as education or trade openness.[31] Secondly, it is even more difficult to gain an idea of the overall result of policies that affect different growth determinants simultaneously.

Summarizing, at a theoretical level the link between inequality and growth emerges along two directions: on the one hand inequality may affect the process of physical and human capital accumulation when it is combined with imperfections in asset markets, and when educational choices are modelled jointly with fertility. On the other hand, inequality may have important consequences for sociopolitical instability, and through this it may affect economic activities at all levels and time horizons. Taking into account the different situations and time horizons over which these two different effects are likely to emerge should help in distinguishing among different results from empirical work and to guide policy choices. For example, inequality may be good for growth in those situations in which credit availability is low and large initial investments are necessary. However, when decreasing returns start to emerge, and when the average level of human capital of the population becomes important, a high level of inequality may have a negative effect on growth.

All the empirical and theoretical studies summarized above suggest that ignoring the distributional consequences of a given growth pattern may generate problems for the process of growth in the future. In other words, the implementation of a policy that boosts growth but at the same time increases inequality within a given country, may lead to positive results in terms of economic growth in the short run, but it may generate problems in the long run. Furthermore, in assessing these effects, one should keep in mind that they may differ according to the initial structure of inequality and other specific conditions of the country, such as the distribution of schooling or the degree of trade openness.

We now turn to the poverty reduction perspective, first examining the link between trade and poverty via growth and then discussing the short-run direct impact of trade on poverty via factor markets. We will present evidence of the wide range of poverty responses to growth and of the way in which the initial level of inequality may influence the poverty/growth elasticity. Although the section is mainly motivated by poverty concerns, it is worth noting that, since many of the theoretical channels through which inequality may affect growth appear more closely related to the proportion of people below a certain threshold level of wealth than to inequality *per se*, the effects of trade on poverty may again become a crucial issue in defining whether the process of trade-induced growth will be sustainable in the long run.

16.3. THE PERSPECTIVE OF REDUCING POVERTY

The literature on the trade impact on poverty does not show well-established results (Winters et al., 2004). Trade liberalization programmes[32] have brought important economic and social changes in countries where they have been implemented. As such, they are likely to affect the absolute level of poverty within and between households, the chances of a household falling into and/or escaping poverty, as well as the amount of time each household spends in poverty.

However, empirical identification of the relationship between trade liberalization and poverty poses a significant challenge. The most important long-run channel through which poverty is likely to be affected is growth, but the elusive relationship between trade and growth identified in Section 16.2 leaves little hope of establishing a further link to poverty. Although growth generally reduces poverty, the magnitude of this effect may differ substantially across countries and over different time periods. In particular, a certain policy choice that pushes the structure of growth in a given direction may have different consequences on poverty according to specific conditions of the country at hand, such as the underlying evolution of the supply of skilled labour, the distribution of schooling, the level of inequality and so on. The poverty-growth problem highlights different empirical reactions of poverty to growth, across countries and over time horizons (Section 16.3.1). Again, even in the long-run perspective, specific conditions of the country concerned (the structure of poverty and other initial conditions) affect the reaction of poverty to growth.

What seems promising for explaining the trade/poverty nexus is a focus on factor markets as they have emerged as the most important linkage between trade and poverty.[33] This recognition ensued from the Conference on Poverty held in Stockholm in October 2000, which brought together economists working with household surveys and researchers taking a more macroeconomic approach, as well as researchers using computable general equilibrium models with a poverty focus (Gurgel et al., 2003; Harrison et al., 2003, Decaluwé et al., 1999). However, there is still much disagreement on the right approach to use in analysis of the problem. Despite strong interest at the political level and the close attention paid to the matter in recent years, research into the impact of trade policy on low-income households has been relatively scant.[34]

Although factor markets are critical for the trade–poverty linkage, they have been relatively neglected by much of the poverty research conducted to date.[35] The section devoted to these aspects (16.3.2) will focus only on specific labour market and consumption effects that may fall disproportionately on the poor, and reports the results on poverty in developing countries generated by a simulation of trade liberalisation policies (Hertel et al., 2003b). The results show the crucial role played by the structure of poverty in each country in framing the differential impact (not always positive) of identical trade liberalization policies on poverty.

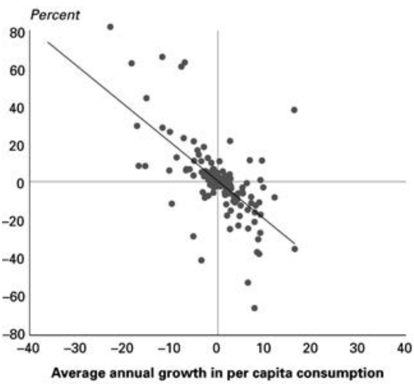
16.3.1 The Poverty–Growth Elasticity

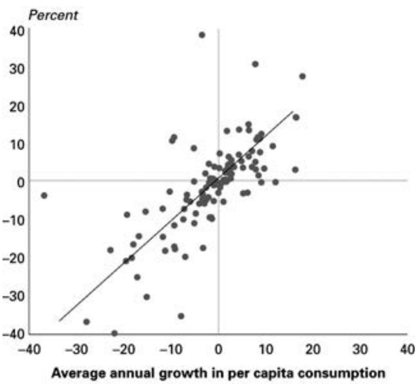
There is substantial consensus among researchers in the trade/growth tradition that 'on average' the incidence of absolute poverty in developing countries tends to fall with growth. The link between trade liberalization and economic performance within developing countries has been often summarized as 'Globalization is good for the Poor', with evidence from individual cases and cross-country analyses supporting the view that open trade regimes lead to faster growth and poverty reduction in poor countries (Dollar and Kraay, 2001, 2002). The core argument is that openness is associated with higher growth, but with no systematic tendency for inequality to increase (on average). The acceleration in growth rates that accompanies expanded trade usually translates into proportionate increases in the incomes of the poor.

Some points about this 'double' relationship are worth noting. First, as discussed in Section 16.2, there is no theoretical and empirical consensus on the view that open trade regimes give rise to faster growth.[36] Second, we have seen various situations in which trade liberalization can lead to increasing inequality. Finally, while it seems quite well established that on average the incidence of absolute poverty in developing countries tends to fall with growth (Figure 16.1), various authors stress that the experience is diverse when one looks behind averages (see Ravallion, 2001 and 2004; Lustig et al., 2002; Bourguignon, 2002).

With respect to the average relationship between poverty and growth, Ravallion (2001) reports estimates of the poverty/growth elasticity that vary from –2.5 to –1.96 according to whether survey data or national account data are used in the estimation. Moreover, as noted by Dollar and Kraay (2001), the elasticity of the average income of the lowest quintile of the population is not significantly different from –1 (Figure 16.2). Regressing the (log) income share of the bottom quintile on GDP per capita, and several other variables, such as trade openness, monetary and fiscal policy, financial development, and the rule of law, Bourguignon (2002) finds that none of these variables is significant. This result would imply that the income share of the bottom 20 per cent is insensitive to growth and known growth determinants, that is growth would affect the income of the poor in the same proportion as that of other groups, whatever the nature of the policy variables behind it and its sectoral structure.

However, as already mentioned, there is a significant dispersion around the average relationship: that is, in some countries and over some periods, there is a significant decrease in poverty as the economy grows; in others the response is much less appreciable. The 95 per cent confidence interval of Ravallion's estimate implies that a 1 per cent rate of growth in average household income or consumption will give rise to anything from a modest





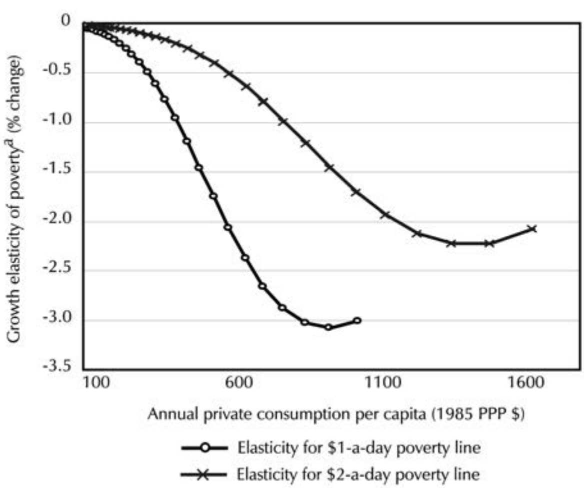
Note: The data, drawn from Chen and Ravallion (2000), span 65 developing countries in the 1980s and 1990s.

Source: World Development Report 2000/2001, *Attacking Poverty*, Washington, DC: World Bank,

Figure 16.1 Average annual growth in share of population living on less than 1$ a day

Figure 16.2 Average annual growth in per capita consumption of poorest fifth of population drop in the poverty rate of 0.6 per cent to a more dramatic 3.5 per cent annual decline. Moreover, within the lowest quintile, growth has the least effect on the income of the poorest. Using changes in the 'generalized mean' as dependent variable, Foster and Székely (2001) find that in Latin American and Caribbean countries, as more weight is given to the income of the poorest, the elasticity falls to zero; that is those living in extreme poverty benefit very little from growth. This finding is confirmed by Karshenas (2001), who shows that the poverty/growth elasticity in absolute value is an increasing non-linear function of average private consumption per capita (Figure 16.3): for the $1 poverty line, the growth elasticities of poverty may range from –0.5 to about –3.0; and countries with average annual consumption per capita of less than $300 (1985 PPP) have an elasticity lower than 1 in absolute value.

Which factors are able to explain this variation have not yet been clearly identified. *At a theoretical level*, Kakwani et al. (2003) show that total poverty elasticity, which determines the extent of a country's poverty reduction, depends on three factors: the growth elasticity of poverty, the inequality elasticity of poverty and the inequality elasticity of growth. The former two elasticities depend on the country's initial level of economic



Note: The growth elasticity of poverty is the percentage change in the proportion of the population living below the poverty line following a 1 per cent increase in average annual per capita private cunsumption

Source: Karshenas (2001)

Figure 16.3 The relationship between the growth elasticity of poverty, the poverty line and the average level of private consumption development and inequality.[37] By contrast, as regards the inequality elasticity of growth, it is not possible to say *a priori* what its sign and magnitude will be.[38] However, if a given growth pattern leads to higher inequality – as in some cases with trade openness policies – it will reduce the poverty impact of growth.

At *the empirical level*, two sets of factors have been found to play an important role in reducing the poverty/growth elasticity: the initial level of inequality and the way in which inequality changes over time. Ravallion (2004) shows that the elasticity of poverty to growth declines appreciably as the extent of initial inequality rises: it diminishes from –4.3 to –0.6 as the Gini index rises from 0.22 to 0.6. Ravallion (2001) also shows that, although on average poverty is falling even in countries in which inequality is rising with growth, it typically falls at a much slower rate than in countries experiencing more equitable growth:

> the median rate of decline in the proportion of the population living below $1/day among countries with both rising average income and rising inequality was 1.3 per cent per year; by contrast, the median rate of poverty reduction was seven times higher, at about 10 per cent per year, among the countries that combined growth in average living standards and falling inequality.

Besides initial inequality and the change of income distribution over time, country studies have shown that the response of poverty to growth depends on some specific conditions in the economy under analysis. For example, by comparing rates of poverty reduction across states of India, Ravallion and Datt (2002) showed that the response of poverty to non-farm output growth varied significantly among states and that this difference reflected *systematic differences in initial conditions*: low farm productivity, low rural living standards relative to urban areas and poor basic education all inhibited the chances of the poor to participate in growth of the non-farm sector. Bourguignon (2002) shows that a crucial role is played by the underlying evolution of the *supply of skilled labour*: if the latter lags behind growth, then the rate of return to skill is bound to increase, resulting in more inequality. In addition, the demand for skilled labour is affected by the *structure* of growth which may result from *policy choices*.[39] Two other variables that appear particularly important in explaining different growth–inequality relationships are the *distribution* of schooling within the population of working age and the *convexity of earning profiles* with respect to education.[40]

Following these studies, Kakwani et al. (2003) have proposed that, in order to achieve a rapid reduction in poverty, a poverty equivalent growth rate should be maximized, rather than the growth rate itself. The former is basically an index of pro-poor growth which takes account of both the magnitude of growth and the benefits of growth that the poor receive. This index will be equal to the growth rate itself when growth is distribution-neutral, that is when everyone in the society receives the same proportional benefits of growth. Therefore, in order to assess the consequences of trade liberalization on poverty, we should take account of both its effects on the growth rate and its distributional consequences, keeping in mind that they may be different according to the initial distribution of income and wealth and to other initial conditions.

When analysing the poverty/growth elasticity we showed that, while on average the incidence of absolute poverty in developing countries tends to fall with growth, the absolute value of the poverty–growth elasticity may differ substantially across countries and over different time periods. However, which factors are able to explain this variation have not yet been clearly identified. Two factors that appear particularly important are the initial level of inequality and the way in which the income distribution changes over time. Growth enhancing policies that yield a worsening of the income distribution will be much less effective in reducing poverty. Therefore, since trade liberalization is likely to lead to increasing inequality, there appears to be no general presumption in favour of the view that ‘globalization is good for the poor’, unless specific conditions of each single country are considered.

The empirical identification of the relationship between trade liberalization and poverty poses a significant challenge. The most important channel through which poverty is likely to be affected by trade openness is growth, but both the relationships between trade and growth, and between growth and poverty are elusive. This suggests that a focus on factor markets could be a promising way forward. Factor markets have emerged as the most important linkage between trade and poverty also in the short run: labour market and consumption effects may fall disproportionately on the poor according to the structure of poverty in each country.

16.3.2. The Poverty Impact of Trade: An Example

In spite of methodological diversities and very mixed results,[41] there appears to be increasing agreement that factor market effects constitute the crucial linkage among trade, trade policy and poverty, for at least three reasons:

- the ‘magnification effect’ (changes in commodity prices due to trade liberalization ‘magnify’ the resulting changes in factor prices: see Jones, 1965);
- households appear to be more specialized in factor earnings than they are with respect to consumption;[42]

- the combination of complete reliance on one income source together with the magnified change in returns may easily dominate the impact of food prices on the farm household.

This is reinforced by four empirical results:

1. the patterns of the structure of poverty (Hertel et al., 2003a) with systematic clear disadvantages of specialized income earners[43] in 14 developing countries, showing that:
   i) there is a strong negative correlation between per capita GDP and the share of the population specialized in agriculture and a positive correlation between the non-agricultural stratum's concentration (wage and salary specialization) and per capita GDP;
   ii) the poor are more specialized than the population at large;
   iii) the poverty rate tends to decrease with increasing income, but considering the intensity of poverty across the earning strata, specialized strata are poorer than average while diversified strata are less poor than average.
2. Income effects accounted for more than two-thirds of poverty alleviation in the Philippines when there was a rise in agricultural productivity (Coxhead and Warr, 1995).
3. The welfare and distributional effects of the proposed rice export tax in Thailand show that the negative income effects (on unskilled labour in the Thai rice industry) outweigh the consumption benefits, so that both the rural and urban poor are harmed by the export tax (Warr, 2001).
4. The adverse impact of trade liberalization on the average household in Turkey has been driven by the source of income rather than the pattern of expenditure (Harrison et al., 2003).

Hertel et al. (2003a) report the results of an exercise linking income and expenditure profiles based on household surveys from 14 developing countries (Malawi, Zambia, Uganda, Bangladesh, Vietnam, Indonesia, Philippines, Peru, Venezuela, Colombia, Thailand, Brazil, Mexico and Chile), with prescriptions of country-specific or multilateral trade policies. They assess the impact of trade liberalization, by country, by strata, and within the strata, proceeding through the following steps: i) utilization of detailed earnings data from household surveys; ii) an econometrically estimated demand system reflecting the change in consumption patterns across the income spectrum and providing a natural vehicle for analysis of household welfare and poverty; iii) a globally consistent framework for projecting the price impacts of trade liberalization. The method used tracks commodity price shocks resulting from trade policy through factor prices to poor households, embedding the household disaggregation within a computable general equilibrium (CGE) model.[44] The global trade model[45] can be used to generate the price change to be fed into the microsimulation analysis.[46]

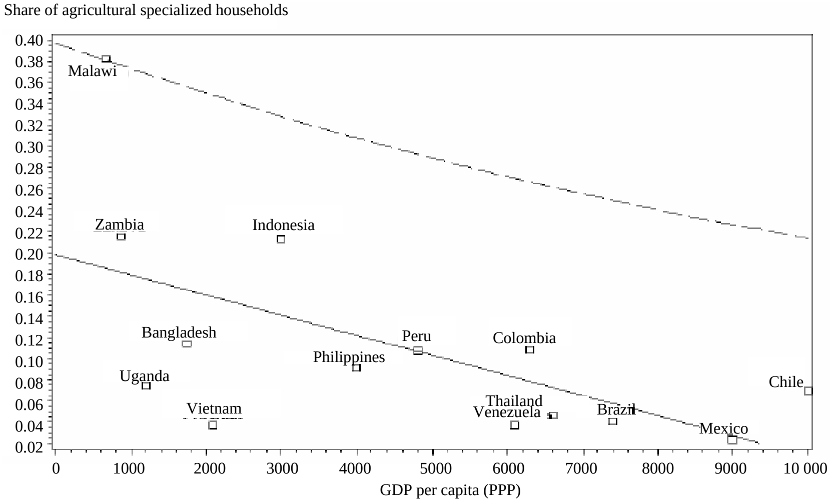
Given the systematic pattern of earning specialization identified in the previous empirical results, households are stratified in five types according to the primary source of income (at least 95 per cent of their income) from:

- agriculture (specialized households where the poor are overrepresented)
- non-agricultural business (households specialized as self-employed in non-agricultural sectors)
- labour (specialized households specialized in wages and salaries)
- transfer payment – specialized households
- diversified income, the fifth complementary type.

On average, they account for 56 per cent of the poor in the 14 countries, who have specialized income patterns (the same is not true for the non-poor) and are more vulnerable to price shocks due to their earnings and expenditure patterns. They are therefore likely to be disproportionally affected by trade liberalization, especially in the short run.

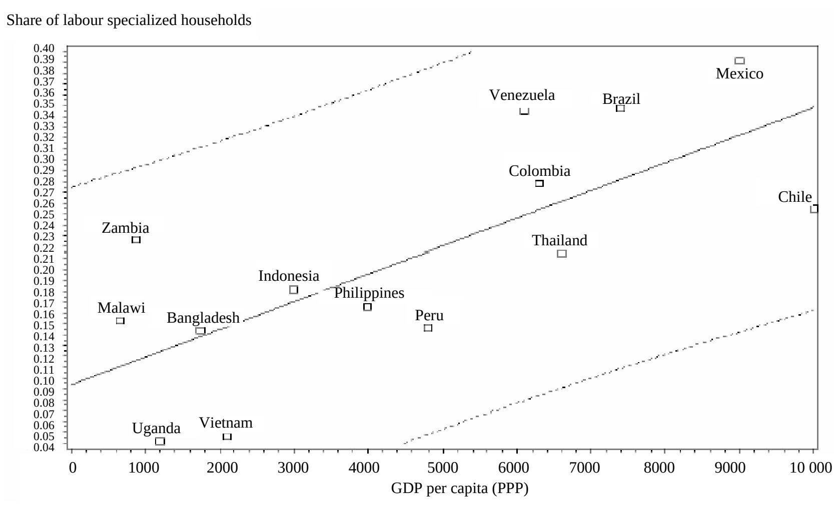
The simulation experiment assumed the elimination of all import barriers; this assumption, together with the information on the structure of protection in the developing countries was expected to have a significant impact, especially on the trade in clothing and agricultural products (both highly protected and relatively unskilled labour intensive).[47] The results emphasized the differential short-run[48] impacts of multilateral trade liberalization on poverty across countries, across and within strata, thereby highlighting the links between the structure of poverty and the national impacts of trade liberalization. The results can be summarized as follows:

i) *Results across countries*. These capture per capita real income effects due to changes in per capita earnings and the price change that consumers must pay for goods and services. Most of the 14 countries examined displayed a modest positive (less than 1 per cent) per capita gain from trade liberalization. This finding is quite consistent with per capita results of most studies on multilateral trade liberalization, which typically show that most, but often not all, developing countries gain from trade liberalization, and that these gains are rather small (Martin and Winters, 1996). Poverty is reduced in all countries when per capita welfare rises.[49] But earning and spending effects differ: when account is



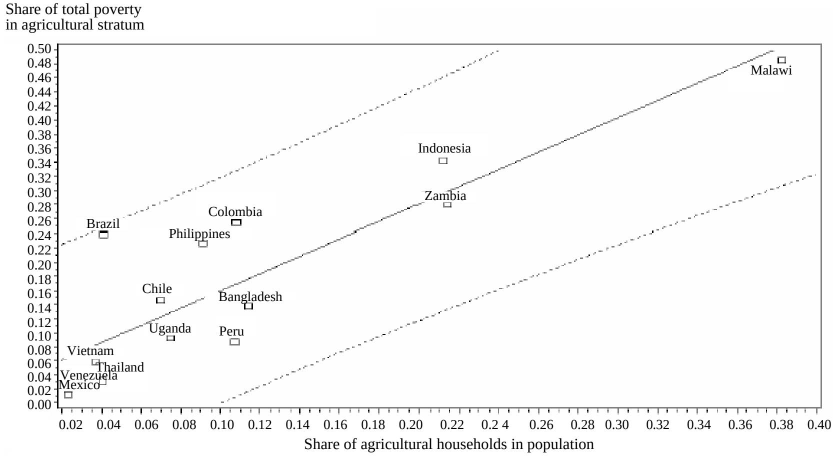
Source: Hertel et al. (2003a)

Figure 16.4 Correlation between per capita GDP and the share of agricultural specialized households in the total



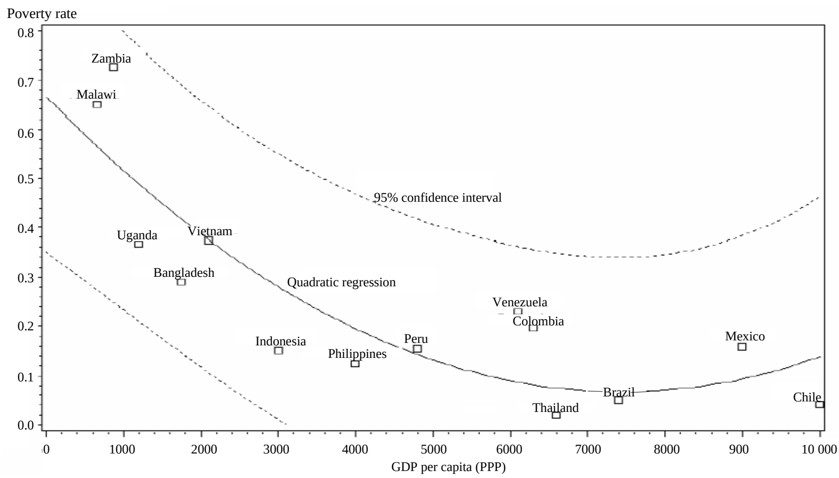
Source: Hertel et al. (2003a)

Figure 16.5 Correlation between per capita GDP and the share of labour specialized households



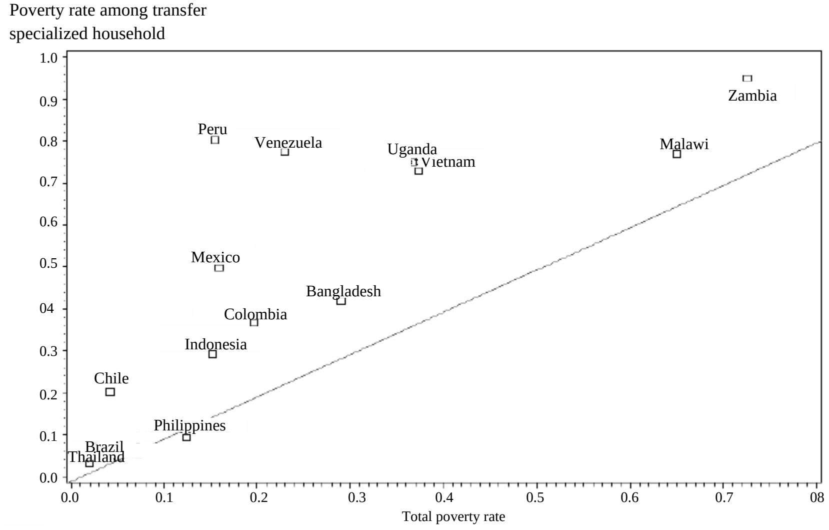
Source: Hertel et al. (2003a)

Figure 16.6 Correlation between per capita GDP and the share of agricultural specialized households in the population and their share in the poor



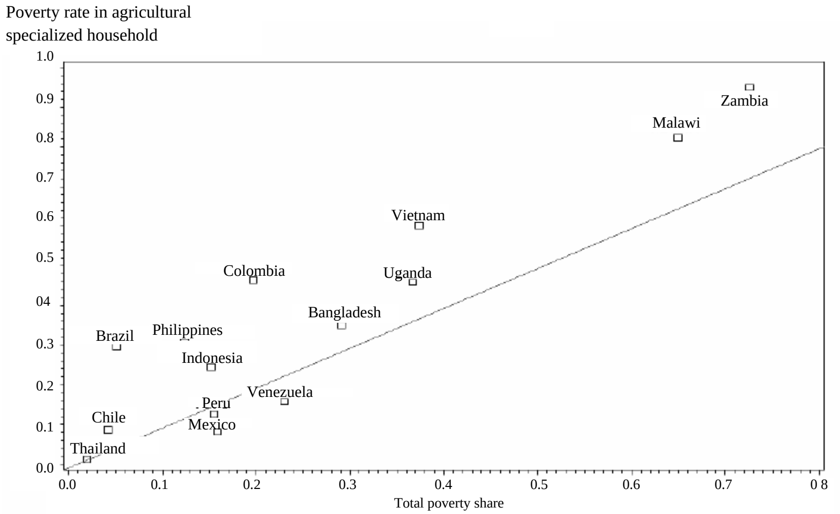
Source: Hertel et al. (2003a)

Figure 16.7 Correlation between the overall poverty rate GDP/per capita



Source: Hertel et al. (2003a)

Figure 16.8 Total poverty rate vs poverty rate among transfer specialized households (line denotes locus of points with equal poverty rates)



Source: Hertel et al. (2003a)

Figure 16.9 Total poverty rate vs poverty rate among agricultural specialized households (line denotes locus of points with equal poverty rates) taken of differential spending patterns across households, poverty rates may rise or fall in a number of countries. The fact that the poor are more exposed to food price increases works to their disadvantage in the context of global trade liberalization.

ii) *Results across strata*. When cross-stratum differences are introduced, poverty rates do not fall (increase) uniformly within each country. Poverty reduction amongst agriculture-specialized households in some countries where agricultural profits rise strongly (the case of Brazil, Chile and Thailand) are associated with a deteriorating position of the non-agriculture specialized stratum and especially the labour-specialized stratum. For example, in Chile, Indonesia and Thailand, this group experiences the largest increased poverty as households are hit by the combination of declining income and higher food prices.

iii) *Results within the strata*. Decomposition of the determinants of stratum-specific poverty changes requires the introduction of within-stratum variation in income sources. For example, within the wage–labour specialized stratum, low income households rely on unskilled as opposed to skilled wages. With unskilled wages rising relative to skilled wages in 12 out of 14 countries, the poorest experience a somewhat larger gain which results in a larger reduction in poverty rates for this stratum. These intra-stratum earnings effects are strong in some countries and are able to turn a poverty increase into a reduction or vice-versa.

iv) *Results on national poverty* (see Table 16.1). Weighting the poverty changes with stratum shares of total poverty gives us the national change in poverty. According to Table 16.1 poverty falls in 11 out of 14 countries. Poverty rises in non-agriculture-specialized strata in both Brazil and Indonesia, but it falls in both cases because the effects on these strata (accounting respectively for 45 per cent and 14 per cent of the poor) are more than compensated by the much larger poverty reduction of agriculture-specialized households in Brazil and by the small incidence of the poorer population in Indonesia. This stands in contrast to Venezuela, where the percentage reduction in agricultural poverty is quite large, but this stratum accounts for only 2.5 per cent of the poor. Not surprisingly, poverty increases in Venezuela. But a different weighting system yields different results: when the interaction between the structure of the poverty and trade liberalization is weighted by applying the average weights instead of the true population weights to the stratum-specific poverty changes (as if the share of poverty amongst the stratum was as large as the sample average), overall poverty may fall rather than increase (the case of Venezuela).

Table 16.1 Approximating national poverty changes using true vs. sample average population weights for each stratum

Country	True weights	Average weights	True average
Bangladesh	–0.10	–0.13	1.25
Brazil	–2.39	–2.41	1.01
Chile	–3.87	–4.54	1.17
Colombia	–2.23	–2.27	1.02
Indonesia	–1.47	–1.11	0.76
Malawi	–2.03	–1.61	0.79
Mexico	–0.21	0.27	–1.27
Peru	1.43	2.00	1.39
Philippines	–3.14	–2.23	0.71
Thailand	5.68	5.93	1.04
Uganda	–0.49	–0.50	1.02
Venezuela	0.34	–1.16	–3.38
Vietnam	–5.60	–4.92	0.88
Zambia	–0.01	–0.03	4.20

In conclusion, although trade reforms are not directly a poverty-reduction strategy, they may have a significant impact on poverty which is either positive or negative according to the structure of poverty in each country. The impact of trade liberalization on different households groups is quite varied and not always positive because the poor tend to be more specialized in their earning sources, and this makes them more vulnerable to trade policy changes which favour one sector at the expense of another. The sectoral composition of the overall poverty picture as well as the weighting system for each stratum (average weights or true population weights) are crucial for isolating the role of the inter-stratum poverty composition in determining the overall poverty changes. This result is even more important if one considers that a lack of major labour reallocation across sectors following large tariff reductions in the 1980s and 1990s has been consistently documented by empirical work (see Goldberg and Pavcnik, 2004).

In this section we have shown that trade reform may have potentially significant effects on poverty through changes in relative prices, which affect both earnings and consumption, and that they may be of paramount importance for households living on less than one dollar a day. In turn, these effects crucially depend on the initial poverty structure, that is on the initial pattern of households' earnings specialization, and their sectoral and occupational mobility. These conditions, and particularly the latter, may also be important for the dynamic link between trade and poverty which operates via growth. Whatever the case may be, trade policies inevitably involve redistribution, with the consequence that there is a crucial need to provide safety nets, which are particularly important for the poor.

16.4. CONCLUSIONS

Establishing a clear link between trade liberalization and poverty is a significant challenge:

- because of the interplay between static and dynamic links over different time horizons
- because of many measurement and identification problems at the empirical level
- because it is difficult to identify common patterns across countries in the presence of trade liberalisation episodes
- because of conflicting empirical evidence on some important issues.

Despite the difficulties and uncertain results, this chapter has started from the separated agenda for growth-supporting and poverty-reducing strategies and analysed their linkages rather than their separation from two different perspectives.

From the perspective of sustaining economic growth, the results can be summarized as follows:

1. The effects of trade on growth are not univocally positive at theoretical level. Usually positive in empirical studies, they are affected by serious shortcomings. Thus, there is no *general* evidence in favour of the view that trade liberalization 'is always good for growth'.
2. Trade may worsen within-country inequalities and trade liberalization may contribute to the rise in the skill premium
3. Since the initial structure of income distribution as well as its changes induced by specific policy interventions matter for future growth, the effect of trade openness on economic growth may be reduced or even reversed in the long run.

In other words, the theoretical debate on the effects of trade on growth (1) and the endless discussions on its empirics are extended if we consider the distributional effects of trade openness (2) because of their effects on future growth (3). Indeed, initial as well as trade-induced inequalities matter for the sustainability of the growth process. These conclusions show that it is necessary to focus on the conditions under which trade may generate more growth and less inequality, and more generally on the distributional effects of different trade-driven growth paths.

From the perspective of poverty reduction, the results can be summarized as follows:

1. the direct impact of trade on poverty, that is a short-run effect, may be potentially significant by operating through changes in relative prices (which affect both earnings and consumption), and not necessarily positive.
2. as regards the link between trade and poverty via growth, that is a long-run effect, trade may reduce poverty if it generates more rapid growth and this, in turn, increases the income of the poor. Although the sign of the link between growth and poverty is generally negative, the absolute value of the poverty/growth elasticity may differ substantially across countries and over different time periods. Two factors that appear particularly important in explaining this variation are the initial level of inequality and the way in which the income distribution changes over time. Growth policies that yield a worsening of the income distribution will be much less effective in reducing poverty. Therefore, since trade liberalization is likely to lead to increasing inequality, there appears to be no general presumption in favour of the view that 'globalization is good for the poor', unless specific conditions of each single country are considered.

In the absence of any general evidence in favour of the views that trade liberalization 'is always good for growth' and that 'growth is good for the poor' we are left with some important common findings:

Initial as well as trade-induced inequalities matter for both the sustainability of growth processes and the speed at which poverty is reduced.

- As a consequence, the problem of supporting economic growth and the problem of reducing poverty are closely linked: we cannot address the former while ignoring the distributional consequences of the policies we seek to implement, and we cannot address the latter while ignoring the precise way in which growth is achieved.
- Specific conditions (sectoral composition, wealth and land distribution, distribution of schooling, convexity of earning profiles, specialization of income sources, and so on) can explain why, at the macro level, similar growth rates may have such different impacts on poverty and why the same policies may have different effects on growth performance. Thus, the problem of poverty reduction cannot be separated from the context in which trade is liberalized.

Assessing the social, political and economic interdependencies among growth, inequality and policy instruments in the context of globalization, and measuring poverty reduction, is one avenue to pursue in defining and evaluating the quality of growth. It is also a means to gain an understanding of why globalization has produced winners and losers, and marginalized those unable to gain access to it.

NOTES

* Preliminary versions of this chapter have been presented not only at the Conference but also at the International Conference on 'Poverty Inequality and the Quality of Growth', held in Trento, Italy, December 5–6, 2003. We would like to thank the participants and the discussants at these Conferences, the members of the GRADE research group for their many helpful comments and suggestions on previous drafts, and Francesca Modena for excellent research assistance. We thank two anonymous referees for their helpful comments. Financial support provided by the University of Trento is gratefully acknowledged. Responsibility for this study is the authors' alone.

1. Igniting economic growth and sustaining it are different enterprises as Rodrik (2003) outlines.
2. This should be clear even from the two broad definitions of pro-poor growth that have emerged in the literature, which focus on inequality outcomes or, instead, on poverty outcomes (see Ravallion, 2004).
3. We leave aside an important stream of the literature which incorporates demand-side considerations into the analysis of trade and growth, focussing on the demand channels through which inequality affects international trade patterns. For example see the role of non-homothetic preferences in the literature on the 'Linder hypothesis' and the income effects in Matsuyama (2000).
4. See stylized fact no. 6 'Growth on output and growth in the volume of international trade are closely related' in Jones (2002, p. 15).
5. As implied by Stolper and Samuelson or by the effects of international reorganization of production.
6. Lewer and van der Berg (2003). For every percentage point increase in the growth of trade, the rate of economic growth, defined as either an increase in real GDP or real per capita GDP, rises by slightly more than one-fifth of a percentage point (East Asia during the 1980s). These results suggest that a country whose exports grow by 12 per cent a year will grow about 2.5 percentage points faster than a country whose trade grows by 2 per cent a year (Sub-Saharan Africa during the 1980s).
7. A growing body of evidence indicates that institutions, and especially their quality, are the driving force behind differential growth rates: the size and density of social networks and institutions (Temple, 1998), social capital (Dasgupta, 2000), the 'social infrastructure' (Hall and Jones, 1999). For a broader perspective see Rodrik et al. (2002) and Kraay (2004). On why and how inadequate institutions can act as a hidden tax on trade flows and underestimate home-bias see Anderson and Marcouiller (1999). The excellent survey by Anderson and van Wincoop (2004) shows why trade costs vary widely across countries and products.

8. Roberts and Tybout (1996). In the case of transition economies (Russia, Ukraine and Belarus) there is some evidence that partially supports the learning by exporting model, where the causality goes from exports to productivity.
9. See Frankel and Romer (1999). Wacziarg (2001) indicates that the most important channel through which trade influences growth is investment, which accounts for 63 per cent of trade's total growth effect, with technology (22.5 per cent) and stabilizing macroeconomic policy (18 per cent) accounting for the remainder. This obviously reinforces Levine and Renelt's (1992) findings that trade acts through investment.
10. The discrimination between whether opening trade in goods, or trade in both goods and ideas, may have different effects is very important. Excessive focus on the long-run effects may well distract from the welfare effects of a given parameter change, giving too much emphasis to the growth effect over the level effect. For instance, if the prime concern is the level of welfare, then it is the sequence of level effects supporting growth processes that is more important to investigate than any long-run growth. See the discussion in Temple (2003).
11. On static losses in the presence of market distortions see Bhagwati (1971); on immiserizing growth see Bhagwati (1958). Samuelson (2004) shows by Ricardo–Mill analysis that 'sometimes a productivity gain in one country can benefit that country alone, while permanently hurting the other country by reducing the gains from trade that are possible between the two countries' (ibid., p. 142).
12. Such models/examples are formalizations of traditional arguments on infant industries and the need for temporary protection to enable 'catching-up' with more advanced countries. These issues have been clarified with the help of two-country models of trade and endogenous growth in the presence of 'learning by doing', where the role of historical advantages is decisive.
13. These difficulties in adapting advanced technologies to the needs of less developed countries, or the importance of the 'appropriateness' of technology, are well reflected in the empirical analysis showing that the largest total factor productivity gaps between advanced countries and less developed countries are in the least skill-intensive sectors, rather than in the skill-intensive sectors.
14. See Cornia (2003) for the role of internal and external liberalization in explaining distributive impacts of trade reforms as well as for an overview of between-countries inequalities.
15. The focus on wage inequality prevents consideration of employment responses to changes in either trade policies or trade flows in developing countries. This exclusive focus may be misleading but in general the price (wage) response to trade liberalization is more pronounced than the quantity response. For an analysis of the transmission channels see Winters in WTO (2000)
16. The similar debate on the role of openness in driving wage inequality in advanced countries (AC) is left to the excellent surveys available.
17. There is no evidence that trade liberalization had major effects on wage distribution. This is often due to the high level of aggregation usually used in household surveys – too high to detect worker reallocations across firms within the same sector in response to trade reallocation. Evidence in favour of technological change does not rule out trade policy effects on wage distribution, as technological change is often seen as a response to intensified competition from abroad. See Wood (1995) and Acemoglu (2002).
18. The case of Taiwan, where higher growth was associated with greater equity, is associated with many other cases in Latin America, Africa and some countries in South-East Asia where those to benefit from opening up are quite small in number: see Kanbur (2000) and Winters et al. (2004).

19. Trade could still contribute to the rise in skill premium without being the main culprit, if technological change had itself been an endogenous response to 'openness' (Acemoglu, 2002).
20. A well known puzzle in LDCs is that their comparative advantage sectors were the most protected before trade reforms: see Mexico in Hanson and Harrison (1999) and Marocco in Currie and Harrison (1997).
21. Compare the results in Attanasio et al. (2002) with the results in Cragg and Epelbaum (1996).
22. On the role of market rigidities in this perspective see Goldberg and Pavcnik (2004).
23. In trying to explain this difference, researchers noted that although the two countries were similar as regards all major economic aggregates, 'the initial distribution of income was considerably more unequal in the Philippines' (Benabou, 1996).
24. Some studies have focused on a short-run relationship (5-year interval; Li and Zou, 1998, Forbes, 2000; Deininger and Olinto, 2000), while others have used data over a longer time span (25 years). Some papers combine data on net income, gross income and expenditure, and also combined data based on households and individuals (Alesina and Rodrick, 1994; Rodrick, 1999; Easterly, 2000; Keefer and Knack, 2000; Sylwester, 2000), whereas others consider transformations of the data to make them more comparable (Perotti, 1996; Deininger and Squire, 1996; Li and Zou, 1998; Forbes, 2000; Barro 2000). Some authors argue that these transformations have little consequence for the estimated effects of inequality on growth and investment (Deininger and Squire, 1998; Barro, 2000); others stress the importance of using inequality measures that are defined consistently (Atkinson and Brandolini, 2001; Knowles, 2001). For studies predating the release of the Deininger and Squire (1996) dataset, problems of data quality are more relevant (Persson and Tabellini, 1994; Alesina and Rodrick, 1994; Perotti, 1996).
25. Barro (2000) finds evidence of a negative relationship for poor countries, but a positive relationship for rich countries. In contrast, Perotti (1996) finds that the negative relationship between inequality and growth becomes much stronger if the poorest countries in the sample are dropped. This may be due to the fact that, although the data have improved, international comparisons of distributional statistics are still plagued by both conceptual and practical problems (measures derived from household surveys or other sources; differences in the measures of living standards used or in the ways in which income from or consumption of non-market goods is evaluated, and so on. See Fields (1994), Chen et al. (1994), Ravallion et al. (1991), Ravallion and Chen (1997), Milanovic (2002).
26. A more recent argument is based on incentive problems in situations involving moral hazard (see Aghion et al., 1999).
27. We know from Perotti (1992) that greater credit availability measured by the loan-to-value ratio for domestic mortgages has a positive and significant effect on the growth rate, and that this effect increases as the income share of the lowest two quintiles decreases. Perotti's (1996) finding that the relationship between inequality and growth becomes much stronger if the poorest countries in the sample are dropped, appears more difficult to rationalize in the context of the borrowing constraint approach. However, one can think of other explanations as well.
28. On their argument, the fertility differential matters because it affects the accumulation of human capital: since poor parents tend to have numerous children and provide little education, future average education will be low. By means of calibration, de la Croix and Doepke (2003) show that the effect of the fertility differentials is quantitatively important and accounts for most of the empirical relationship between inequality and growth.

29. Aghion et al. (1999) suppose that only a fraction of the active population has access to high yield investment opportunities and that investors can borrow only a limited amount of funds.
30. Among others, Alesina and Perotti (1996), Benhabib and Rustichini (1996), Gupta (1990). Another class of models developed at the beginning of the 1990s encompassed models of political economy (or endogenous fiscal policy models) where distributional effects arose through the balance of power in the political system. Empirical tests of this theory, however, have found that the relationship between inequality and transfers is rarely significant (Perotti, 1996).
31. For example, it would be interesting to know whether the effect of a given income redistribution on fertility differs among countries with different distributions of schooling.
32. One recent IMF review of seven Poverty Reduction and Growth Facility programmes found that each loan came with seven trade policy conditions attached.
33. The role of trade and effects of trade reforms can be analysed in a wider context (WTO, 2000) via the various transmitting channels, of which product and labour markets are only two among many others. The product channel may be critical, especially for highly commodities-dependent countries (the protracted crisis in commodity markets does not even register on the global agenda), but the focus here is on the factor channel. For comprehensive papers on the many linkages see Cirera et al. (2001), the contribution by Winters in WTO (2000), Bannister and Thugge (2001), Winters (2001), Winters et al. (2004).
34. Furthermore, most studies focus on a single country, which makes it difficult to disentangle findings driven by methodological assumptions from characteristics of the country in question.
35. A point emphasized in the ground-breaking work by Adelman and Robinson (1978) as well as in the more recent work by Bourguignon et al. (2003).
36. Countries like China, Thailand and Vietnam have strong records of economic growth (and poverty reduction), but they have liberalized imports very slowly and still have relatively restrictive trade barriers. Conversely, countries like Brazil, Haiti, Mexico, Peru and Zambia have been world-beaters when it comes to import liberalization, but they have weak records on growth (and poverty reduction). In short, many first-rate globalizers have fifth-rate records on poverty reduction (Winters, 2001; Winters et al., 2002).
37. Kakwani and Son (2002) demonstrate analytically that the growth elasticity of poverty is a decreasing function of the initial level of mean income and an increasing function of the initial level of inequality; and that the inequality elasticity of poverty is an increasing function of the initial level of mean income and a decreasing function of the initial level of inequality.
38. Bourguignon (2002) provides four country stories (Mexico, Taiwan, Indonesia and Brazil) in which growth was responsible for significant changes in the distribution of income, but with very different overall effects: slow growth was potentially unequalizing in Brazil, whereas fast growth was also unequalizing in Taiwan, and neutral in Indonesia.
39. For instance, 'the reason why demand for skilled labor grew so much in Taiwan may have to do with the openness of the economy and the strong changes it caused in the structure of production toward sectors more intensive in both physical and human capital' (Bourguignon, 2002).
40. Bourguignon (2002) shows that schooling increased in all the four countries, but it had very different effects on the distribution of income: for example schooling expansion increased inequality in Mexico and in Indonesia but reduced it in Taiwan and Brazil. 'The reason for this difference is mostly that earning profiles with respect to education are less convex in Taiwan and Brazil than in Mexico and the expansion of education in the former countries

may have been stronger, in absolute value, at the bottom than at the top of the schooling range'.
41. Papers are categoriszed by Reimer (2002)
42. This implies that households tend to be more specialized with respect to income than with respect to consumption. Historically most poverty research has focused on the consumption side of the question, since it is easier to measure, more reliable and less volatile than income (Lipton and Ravaillon, 1995). Mitra and Trindade (2003) show that consumption rather that income specialization plays a crucial role in developed countries, and they explain trade as driven by consumption specialisation.
43. As will be clear from the empirical example, 'specialization' refers to households that earn 95 per cent or more of their incomes from, for instance, agricultural profits.
44. With CGE it is possible to link household types (specialized by factor incomes) with prescriptions of country-specific or multilateral trade policies. Utilization of CGE is almost the only tool with which to predict the effects of future trade policy changes, but care must be taken with the parameters and functions assumed: they identify predictions and are complementary rather than substitutes for genuine empirical work on ex-post data. For a survey see Reimer (2002).
45. In this case the GTAP model and data base are utilized and are interesting especially because of the regional disaggregation (78 regions in version 6).
46. Details of the micro simulation model behind this exercise cannot be given here. See Hertel et al. (2003b) for the analytical work behind the exercise on Indonesia.
47. A previous study (Hertel et al. 2000) on how global trade liberalization affects poverty in each of seven different developing countries showed that multilateral trade liberalization reduces overall poverty in Indonesia, Philippines, Uganda and Zambia but increases overall poverty in Brazil, Chile and Thailand. Within regions, the results vary considerably by household group. The largest poverty reduction occurs among agriculture-specialized households in Brazil, while the largest increase occurs among non agricultural, self-employed and wage labour households in Brazil, Chile and Thailand.
48. Long-run results may be different: see the analysis of the Indonesian case in Hertel et al. (2003b).
49. The poverty level of utility (the utility of the household at the poverty line) in each country (before and) after liberalization is calculated by recomputing income as well as consumption and utility level for each percentile in each stratum with post liberalization prices.

REFERENCES

Acemoglu, D. (2002), 'Technical change, inequality and the labor market', *Journal of Economic Literature*, **40**(1):7–72.

Acemoglu, D. and F. Zilibotti (2001), 'Productivity differences', *Quarterly Journal of Economics*, **116**(2):563–606.

Acemoglu, D., S. Johnson and J.A. Robinson (2001), 'The colonial origins of comparative development: an empirical investigation', *American Economic Review*, **91**(5):1369–401.

Adelman, I. and S. Robinson (1978), 'Income distribution policy: a computable general equilibrium model of South Korea', Stanford: Stanford University Press.

Aghion, P., A. Banerjee and T. Piketty (1997), 'Dualism and macroeconomic volatility', *CEPREMAP Working Paper 9720*.

Aghion, P., E. Caroli and C. Garcia-Penalosa (1999), 'Inequality and economic growth: the perspective of the new growth theories', *Journal of Economic Literature*, **37**:1615–60.

Alesina, A. and D. Rodrik (1994), 'Distributive politics and economic growth', *Quarterly Journal of Economics*, **109**:465–90.

Alesina, A. and R. Perotti (1996), 'Income distribution, political instability and investment', *European Economic Review*, **40**(6):1203–28.

Anderson, J.E. and D. Marcouiller (1999), 'Trade, insecurity, and home bias: an empirical investigation', *NBER Working Paper 7000*.

Anderson, J.E. and E. van Wincoop (2004), 'Trade costs', *Journal of Economic Literature*, **XLII**(September):691–751.

Atkinson, A.B. and A. Brandolini (2001), 'Promise and pitfalls in the use of "secondary" data-sets: income inequality in OECD countries as a case study', *Journal of Economic Literature*, **39**(3):771–99.

Attanasio, O., P.K. Goldberg and N. Pavcnik (2002), 'Trade reforms and income inequality in Colombia', Paper presented at 2002 IMF Conference on Macroeconomic Policies and Poverty Reduction, Washington, DC, March 14–15.

Bannister, G.J. and K. Thugge (2001), 'International trade and poverty alleviation', *Finance and Development*, **38**(4):48–51.

Barro, R. (2000), 'Inequality and growth in a panel of countries', *Journal of Economic Growth*, **5**(1):5–32.

Benabou, R. (1996), 'Inequality and growth', *NBER Macroeconomics Annual*, **11**:11–74.

Benhabib, J. and A. Rustichini (1996), 'Social conflict and growth', *Journal of Economic Growth*, **1**:129–46.

Bernard, A.B. and J.B. Jensen (1999a), 'Exceptional exporter performance: cause, effects or both?', *Journal of International Economics*, **47**(1):1–25.

Bernard, A.B. and J.B. Jensen (1999b), 'Why some firms export', *The Review of Economics and Statistics*, **86**(2):561–69.

Bernard, A.B. and J.B. Jensen, J.B. (2004), 'Exceptional exporter performance: cause, effect, or both?", *NBER Working Paper 6272*.

Besley, T. and R. Burgess (2003), 'Halving global poverty', *Journal of Economic Perspectives*, **17**(3):3–22.

Bhagwati, J. (1958), 'Immiserizing growth: a geometrical note', *Review of Economic Studies*, **25**(June):201–5.

Bhagwati, J.N. (1971), 'Trade-diverting customs unions and welfare improvement: a clarification', *Economic Journal*, **81**(323):580–87.

Bourguignon, F. (2002), 'The distributional effects of growth: case studies vs. cross-country regressions', *DELTA Working Paper 2002-23*.

Bourguignon, F., A.S. Robilliard and S. Robinson (2003), 'Representative versus real households in the macro-economic modeling of inequality', *DELTA Working Paper 2003-05*.

Breen, R. and C. Garcia-Penalosa (1999), 'Income inequality and macroeconomic volatility: an empirical investigation', *GREQAM Working Paper 99b11*, Universite Aix-Marseille III.

Chen, S. and M. Ravallion (2000), 'How did the world's poorest fare in the 1990s?', *Policy Research Working Paper*, Washington, DC: World Bank.
Chen, S. et al. (1994), 'Is poverty increasing or decreasing in the developing world?', *Review of Income and Wealth*, **40**:359–76.
Cirera, X., N. McCulloch and L.A. Winters (2001), *Trade, Liberalization and Poverty: A Handbook*, London: CEPR.
Coe, D.T., E. Helpman and A.W. Hoffmaister (1997), 'North–South R&D spillovers', *Economic Journal*, **107**(440):134–49.
Cornia, G.A. (2003), 'The distribution of income over the last twenty years in the developing and transitional economies', paper presented at SIE 44th Meeting, Salerno, 24–25 October.
Coxhead, I. and P. Warr (1995), 'Does technical progress in agriculture alleviate poverty? A Philippine case study', *Australian Journal of Agricultural Economics*, **39**(1):25–54.
Cragg, M.I. and M. Epelbaum (1996), 'Why has wage dispersion grown in Mexico? Is it the incidence of reforms or the growing demand for skills?', *Journal of Development Economics*, **51**(1):99–116.
Currie, J. and A.E. Harrison (1997), 'Sharing the costs: the impact of trade reform on capital and labor in Marocco', *Journal of Labor Economics*, **15**(3):44–71.
Dasgupta, P. (2000), 'Economic progress and the idea of social capital', in P. Dasgupta and I. Serageldin (eds), *Social Capital: A Multifaceted Perspective*, Washington, DC: World Bank, pp. 325–424.
De Janvry, A., E. Sadoulet and W. Wolford (2001), 'The changing role of the state in Latin American land reforms', in A. de Janvry, G. Gordillo, E. Sadoulet and J.-Ph. Platteau (eds), *Access to Land, Rural Poverty, and Public Action*, Oxford: Oxford University Press, pp. 279–303.
de la Croix, D. and M. Doepke (2003), 'Inequality and growth: why differential fertility matters', *American Economic Review*, **93**(4):1091–113.
Decaluwé, B., A. Patry, L. Savard and E. Thorbecke, E. (1999), 'Poverty analysis within a general equilibrium framework', *CREFA Working Paper 9909*, Université Laval.
Deininger, K. and P. Olinto (2000), 'Asset distribution, inequality, and growth', *World Bank Policy Research Working Paper 2375*, Washington, DC: World Bank.
Deininger, K. and L. Squire (1996), 'A new data set measuring income inequality', *World Bank Economic Review*, **10**:565–91.
Deininger, K. and L. Squire (1998), 'New ways of looking at old issues', *Journal of Development Economics*, **57**:259–87.
Dollar, D. and A. Kraay (2001), 'Growth is good for the poor', *World Bank Policy Research Department Working Paper 2587*, Washington, DC: World Bank.
Dollar, D. and A. Kraay (2002), 'Trade, growth, and poverty', *Journal of Economic Growth*, **7**:195–225.
Easterly, W. (2000), 'The middle class consensus and economic development', *Development Research Group Working Paper 2346*, Washington, DC: World Bank.
Feenstra, R.C. and G.H. Hanson (1996a), 'Globalization, outsourcing, and wage inequality', *American Economic Review Papers and Proceedings*, **86**(2):240–45.

Feenstra, R.C. and G.H. Hanson (1996b), 'Foreign investment, outsourcing, and relative wages', in R.C. Feenstra, G.M. Grossman and D.A. Irwin (eds), *Political Economy of Trade Policy: Essays in Honor of Jagdish Bhagwati*, Cambridge, MA: MIT Press, pp. 89–127.

Feenstra, R.C. and G.H. Hanson (1997), 'Foreign direct investment and relative wages: evidence from Mexico's maquiladoras', *Journal of International Economics*, **42**:371–93.

Fields, G. (1994), 'Data for measuring poverty and inequality changes in the developing countries', *Journal of Development Economics*, **44**(June):87–102.

Forbes, K. (2000) 'A reassessment of the relationship between inequality and growth', *American Economic Review*, **90**(4):869–87.

Foster, J. and M. Szekely (2001), 'Is economic growth good for the poor? Tracking low incomes using general means', Inter-American Development Bank, mimeo.

Frankel, J. A. and D. Romer (1999), 'Does trade cause growth?', *American Economic Review*, **89**:379–99.

Galor, O. and H. Zang (1997), 'Fertility, income distribution and growth: theory and cross country evidence', *Japan and The World Economy*, **9**:197–229.

Goldberg, P.K.D. and N. Pavcnik (2004), 'Trade inequality and poverty: what do we know? Evidence from recent trade liberalization episodes in developing countries', *NBER Working Paper 10593*.

Grossman, G. and E. Helpman (1991), *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press.

Gupta, D. (1990), *The Economics of Political Violence*, New York: Praeger.

Gurgel A., G.W. Harrison, T.F. Rutherford and D.G. Tarr (2003), 'Regional, multilateral and unilateral trade policies of MERCOSUR for growth and poverty reduction in Brazil', *World Bank Working Paper 3051*.

Hall, R.E. and C.I. Jones (1999), 'Why do some countries produce so much more output per worker than others?', *Quarterly Journal of Economics*, **114**(1):83–116.

Hanson, G.H., and A. Harrison (1999), 'Trade liberalization and wage inequality in Mexico', *Industrial and Labor Relations Review*, **52**(2):271–88.

Harrison, A. (1996), 'Openess and growth: a time series, cross country analysis for developing countries', *Journal of Development Economics*, **48**:419–47.

Harrison, A. and G. Hanson (1999), 'Who gains from trade reform? Some remaining puzzles', *Journal of Development Economics*, **59**(1):125–54.

Harrison, G.W., T.F. Rutherford and D.G. Tarr (2003), 'Trade liberalization, poverty, and efficient equity', *Journal of Development Economics*, **71**(1):97–128.

Hausmann, R. and M. Gavin (1996), 'Securing stability and growth in a shock-prone region: the policy challenges for Latin America', in R. Hausmann and H. Reisen (eds), *Securing Stability and Growth in Latin America*, Paris: OECD, pp. 23–64.

Hertel, T.H., M. Ivanic, P.V. Preckel and J.A.L. Cranfield (2000), 'Multilateral trade liberalization and poverty reduction', paper presented at Conference on Trade and Poverty, Stockholm, Sweden, October 20–21.

Hertel, T.H., M. Ivanic, P.V. Preckel and J.A.L. Cranfield (2003a), 'Trade liberalisation and the structure of poverty in developing countries', *GTAP Working Paper 25*, Purdue University.

Hertel, T.H., Ivanic, M., Preckel, P.V and Cranfield, J.A.L. (2003b), 'The earnings effects of multilateral trade liberalization: Implications for poverty in developing countries', *GTAP Working Paper 16*, Purdue University.

IADB (Inter-American Development Bank) (1995), *Report on Economic and Social Progress in Latin America: 1995*, Washington, DC: Inter-American Development Bank.

Jones, R.W. (1965), 'The structure of simple general equilibrium models', *Journal of Political Economy*, **73**(December):557–72.

Jones, C.I. (2002), *Introduction to Economic Growth*, 2nd edn, New York: W.W. Norton & Co.

Kakwani, N. and H. Son (2002), 'Pro-poor growth and poverty reduction: the Asian experience', The Poverty Center, Office of Executive Secretary, ESCAP, Bangkok.

Kakwani, N., S. Khandker and H. Son (2003), 'Poverty equivalent growth rate: with applications to Korea and Thailand', paper presented at the WIDER Conference on Inequality, Poverty and Human Well-Being, Helsinky, Finland, 30–31 May. Available at: http://www.wider.unu.edu/conference/conference-2003-2/conference2003-2.htm.

Kanbur, R. (2000), 'Income distribution and development', in A.B. Atkinson and F. Bourguignon (eds), *Handbook of Income Distribution*, Amsterdam: Elsevier, pp. 791–842.

Karshenas, M. (2001), 'Measurement and nature of absolute poverty in least developed countries', background report for *The Least Developed Countries Report 2002,* Geneva: UNCTAD.

Keefer P. and S. Knack (2000), 'Polarization, politics, and property rights: links between inequality and growth', *World Bank Policy Research Working Paper 2418*, Washington, DC: World Bank,.

Knowles, S. (2001), 'Inequality and economic growth: the empirical relationship reconsidered in the light of comparable data', Paper presented at the WIDER Conference on Growth and Poverty, Helsinki.

Kraay A. (2004), 'When is growth pro-poor? Cross-country evidence', *IMF Working Paper 47*.

Kuznets, S. (1955), 'Economic growth and income inequality', *American Economic Review*, **45**:1–28.

Levine, R. and D. Renelt (1992), 'A sensitivity analysis of cross-country growth regressions', *American Economic Review*, **82**:942–63.

Lewer, J.J. and H. Van den Berg (2003), 'How large is international trade's effect on economic growth?', *Journal of Economic Surveys*, **17**(3):363–96.

Li, H. and H. Zou (1998), 'Income inequality is not harmful for growth: theory and evidence', *Review of Development Economics*, **2**(3):318–34.

Lipton, M. and M. Ravallion (1995), 'Poverty and policy', in J. Behrman and T.N. Srinivasan (eds), *Handbook of Development Economics*, Amsterdam: North-Holland, vol. III, pp. 2551–657.

Lucas, R. (1993), 'Making a miracle', *Econometrica*, **61**:251–72.

Lustig, N., O. Arias and J. Rigolini (2002), 'Poverty reduction and economic growth: a two-way causality', *Inter-American Development Bank, Sustainable*

Development Department Technical Papers Series POV-111, Washington, DC, March.

Martin, W. and L.A. Winters (1996), *The Uruguay Round and the Developing Countries*, Cambridge: Cambridge University Press.

Matsuyama, K. (1992), 'Agricultural productivity, comparative advantage, and economic growth', *Journal of Economic Theory*, **58**(2):317–34.

Matsuyama, K. (1996), 'Why are there rich and poor countries?' Symmetry-breaking in the world economy', *Journal of the Japanese and International Economies*, **10**(4):419–39.

Matsuyama, K. (2000), 'A Ricardian model with a continuum of goods under nonhomothetic preferences: demand complementarities, income distribution and North–South trade', *Journal of Political Economy*, **108**(6):1093–120.

Mbabazi, J., O. Morrissey and C. Milner (2001), 'Are inequality and trade liberalization influences on growth and poverty?', *WIDER Discussion Paper 132*, Helsinki, November.

Milanovic, B. (2002), 'True world income distribution, 1988 and 1993: first calculations based on household surveys alone', *Economic Journal*, **112**(January):51–92.

Mitra, D. and V. Trindade (2003), 'Inequality and trade', *NBER Working Paper 10087*.

Pavcnik, N. (2002) 'Trade liberalisation, exit and productivity improvement', *Review of Economic Studies*, **69**:245–76.

Perotti, R. (1992), 'Fiscal policy, income distribution and growth', *Columbia University Working Paper 636*, November.

Perotti, R. (1996), 'Growth, income distribution, and democracy: what the data say', *Journal of Economic Growth*, **1**(2):149–87.

Persson, T. and G. Tabellini (1994), 'Is inequality harmful for growth?', *American Economic Review*, **84**(3):600–621.

Ramey, G. and V.A. Ramey (1995), 'Cross-country evidence on the link between volatility and growth', *American Economic Review*, **85**(5):1138–51.

Ravallion, M. (2001), 'Growth, inequality and poverty: looking beyond averages', *World Development*, **29**(11):1803–15.

Ravaillon, M. (2004), 'Pro-poor growth: a primer', *Policy Research Working Paper 3242*, World Bank (available at http://poverty.worldbank.org/library/view/15174/).

Ravaillon, M. and S. Chen (1997), 'What can new survey data tell us about recent changes in distribution and poverty?', *World Bank Economic Review*, **11**(2):357–82.

Ravallion, M. and G. Datt (2002), 'Why has economic growth been more pro-poor in some states of India than others?', *Journal of Development Economics*, **68**:381–400.

Ravallion, M., G. Datt and D. van de Walle (1991), 'Quantifying absolute poverty in the developing world', *Review of Income and Wealth*, **37**(December):1803–15.

Reimer, J.J. (2002), 'Estimating the poverty impacts of trade liberalization', *GTAP Working Paper 20*, Purdue University.

Roberts, M.J. and J.R. Tybout (1996), *Industrial Evolution in Developing Countries: Micro Patterns of Turnover, Productivity, and Market Structure*, New York: Oxford University Press.
Rodrik, D. (1999), 'Where did all the growth go? External shocks, social conflict, and growth collapses', *Journal of Economic Growth*, **4**(4):385–412.
Rodrik, D. (2003), 'Growth strategies', *NBER Working Paper 10050*.
Rodriguez, F. and D. Rodrik (2000), 'Trade policy and economic growth: a skeptic's guide to the cross-national evidence', in B.S. Bernanke and K. Rogoff (eds), *NBER Macroeconomics Annual*, Cambridge, MA, and London: MIT Press.
Rodrik, D., A. Subramanian and F. Trebbi (2002), 'Institutions rule: the primacy of institutions over geography and integration in economic development', *NBER Working Paper 930*5, November.
Samuleson P.A. (2004), 'Where Ricardo and Mill rebut and confirm arguments of mainstream economists supporting globalization', *Journal of Economic Perspectives*, **18**(Summer):135–46.
Sylwester, K. (2000), 'Income inequality, education expenditure and growth', *Journal of Development Economics*, **63**:379–98.
Temple, J. (1998), 'Initial conditions, social capital and growth in Africa', *Journal of African Economics*, **73**(3):309–47.
Temple, J. (2003), 'The long run implications of growth theories', *Journal of Economic Surveys*, **17**(3):497–514.
Tybout, J.R. (2003), 'Plant and firm-level evidence on new trade theories', in E.K. Choi and J. Harrigan (eds), *Handbook of International Trade*, Oxford: Basil Blackwell, pp. 388–415.
Wacziarg, R. (2001), 'Measuring the dynamic gains from trade', *World Bank Economic Review*, **15**(3):393–429.
Warr, P.G. (2001), 'Welfare effects of an export tax: Thailand's rice premium', *American Journal of Agricultural Economics*, **83**(4):903–20.
Winters, L.A. (2001), 'Trade, trade policy and poverty: what are the links?' *CEPR, Working Paper 2382*, London.
Winters, L.A., N. McCulloch and A. McKay (2004), 'Trade liberalization and poverty: the evidence so far', *Journal of Economic Literature*, **62**(1):72–115.
Wood, A. (1995), 'How trade hurt unskilled workers', *Journal of Economic Perspectives*, **9**(3):57–80.
World Bank (2001), *World Development Report 2000/2001: Attacking Poverty*, New York: Oxford University Press.
WTO (2000), 'Income disparity, trade and poverty', *Special Study 5*, Geneva: World Trade Organization.
Young, A. (1991), 'Learning by doing and the dynamic effects of international trade', *Quarterly Journal of Economics*, **106**(2):369–405.