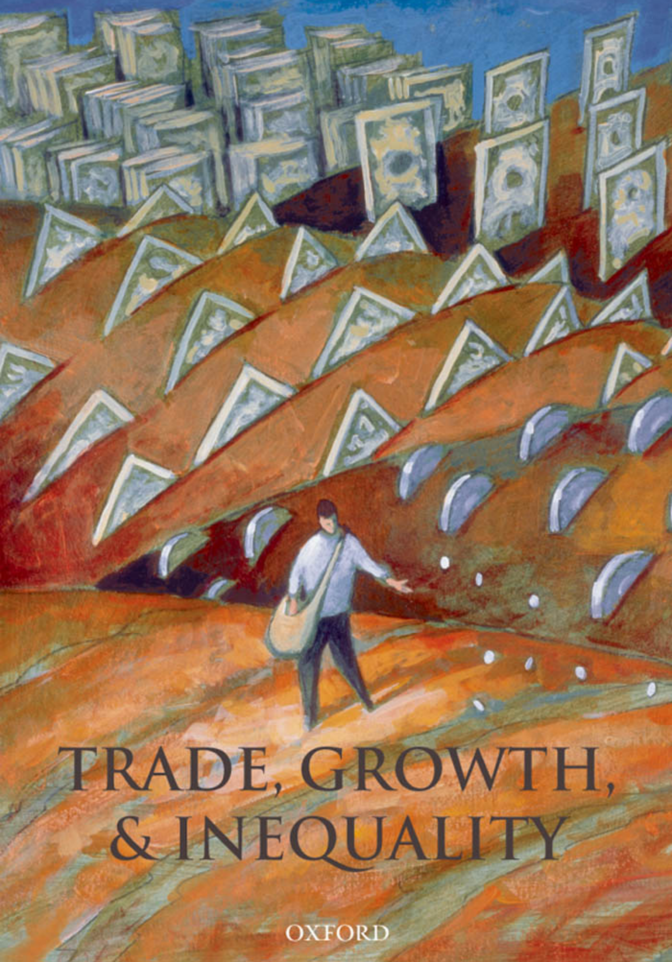
CHRISTOPHER BLISS



# TRADE, GROWTH, & INEQUALITY

OXFORD

Trade, Growth, and Inequality

*This page intentionally left blank*

# Trade, Growth, and Inequality

Christopher Bliss

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam



Published in the United States
by Oxford University Press Inc., New York



First published 2007



British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Data available

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
Biddles Ltd., King's Lynn, Norfolk

ISBN 978–0–19–920464–9

1 3 5 7 9 10 8 6 4 2

# Contents

# Preface

This volume brings together a large part of my research and teaching over the last eight years. The term research suggests a formal academic programme, one that leads to discussion papers, seminar presentations, and publications. There is all of that, certainly. However over these years I have done a great deal of thinking about the problems of world economic development, the results of which are sometimes too informal to stand alone as publications. A great benefit of writing a book is that it gives its author the space to roam in speculative areas, and to offer ideas that would never get past the gate-keepers of professional journals. I have availed myself of that freedom. The reader will find informal material in this book, but also much formal economic analysis. This is not a textbook as such. It contains material, however, that can supplement taught courses for graduates; courses in international trade theory, in economic development, and in economic growth. The chapters do not divide rigidly between those fields. However, Chapters 5, 6, 9, and 11 are particularly concerned with trade. Chapters 3, 4, 7, 8, and 10 offer material especially to do with development. Economic growth recurs frequently as a theme, so that it is harder to pick out particular chapters as featuring it. That said, growth is strongly involved in the arguments of Chapters 3, 4, 12, and 13. Another word of my title is inequality. No single chapter treats inequality by itself. Instead, the issue of inequality, and its changes via policy and over time, is one that is repeatedly visited throughout the volume.

While I refer freely and frequently to the work of others, this is not a volume of surveys. In most cases there are already good surveys of my major fields, and I refer the reader to these. Some of my published work is reflected in the book, some new unpublished results will be found here also. Generally I have avoided the lengthy exposition of established theory; although Chapter 5, on classical trade theory, is original only where it finds a new angle, or a different emphasis, for what is otherwise textbook territory. Originality is hard to define precisely, and nowhere is this more true than in the area of endogenous growth and its relation to trade. In Chapter 12 the reader will find several new models, and may yet think that they have a 'variations on a theme' flavour. If so, I do not apologize. The literature is solid and impressive, and to presume to sweep it aside and start again would indicate *folie de grandeur*.

One reason why I have been thinking about questions of trade and development during recent years, is that this is a time both of exciting changes, and also of anxieties concerning the future of the world economy. The excitement results from the dramatic rise in world trade, involving notably the take-offs into rapid economic growth to be seen in Brazil, China, India, and elsewhere. The anxieties come from an insidious growth in protectionist sentiments in rich countries, and also, and possibily even more worrying, a marked failure by large regions of the world to participate in the globalization experiment, with consequent stagnation and rising poverty. Many of the critics of globalization would consider these anxieties to be far too limited in scope. I am not worried that increased trade will be a massive cause of poverty, or that it will destroy the world environment. It is not that I am complacent about the environmental problem. Rather, I believe that poverty is more likely than trade to make the problem worse. And, in any case, the restriction of trade is a stupid and ineffective way of addressing the environmental problem. On trade and the environment see Frankel (2005).

During the years in which this volume was under preparation, I have accumulated countless debts; too many to attempt to list inclusively, without running the risk that I will omit a name more deserving acknowledgement than one included. Among many others I am indebted to: Anthony Atkinson, Alan Beggs, Paul Collier, David Cox, Ernesto Dal Bo', Rafael di Tella, Avinash Dixit, Cecilia Garcia Penelosa, Macartan Humphreys, Godfey Keller, Margaret Meyer, Danny Quah, John Quah, Joseph Stiglitz, Jon Temple, David Vines, Adrian Wood, and Martin Wolf. All of these have influenced this work, sometimes in ways of which they may not be aware. Thanks are due also to participants in seminars at Oxford and elsewhere, to anonymous referees, and to my family, who have had to suffer living with a frequently distracted member.

I am indebted above all to graduate students at Oxford. These have suffered the indignity of acting as experimental consumers of my drafts, always with patience. And they have frequently provided helpful comment and criticism. Most university teachers feel that we have to teach too much. Nevertheless, a great benefit of teaching is that one always

thinks that one understands something clearly until one has to teach it. And even when the preparation of material for classroom use has forced one to straighten out ideas, sharp graduate students demonstrate that it could all be more transparent and precise. Any imperfections that remain are entirely my fault, but there would have been many more but for my students.

The translation of the work of many years into typescripts, and thence into a printed volume, represents a huge effort. In this I have been wonderfully aided by my secretary, Elaine Herman, and, at the Oxford University Press, by Carol Bestley, Sarah Caro, and Jess Smith.

# 1

# Introduction

## 1.1 Focal Points

The title of this work indicates both the breadth of the field which it explores, and also some particular points of focus within that field. The first word of the title, *Trade*, pinpoints the starting point of the analysis. International trade is of growing importance in the contemporary world, just as it has been important ever since the Second World War. It has been a major topic of interest to economists since the beginning of the subject in its modern form, in the late eighteenth century. Also it has always been a strongly contested area. At the present time anti-trade positions are growing in impact, both in the arena of journalism and political debate, and again on the streets of any city where a major intergovernmental conference concerned with trade or finance is taking place—and now in every city where a major international conference takes place. Many of the 'anti-trade' parties define themselves as opposed to 'globalization'. Globalization represents free trade, certainly, but much more besides. It includes in particular international mobility of capital and technology. It is an indication of how globalized our world has become that this volume has been typeset in Pondicherry, India.

One might say that globalization embraces in addition cultural dissemination. Street demonstrators frequently target McDonald's hamburger outlets. Why is this so? It is easy to mount a case that cheap hamburgers are poor-quality cuisine. Yet the mobs are not energetic defenders of culinary standards. They have never been known to target local low-standard restaurants. McDonald's stands as a symbol for the international migration of technology, of remote power and of alien life-style (perhaps seen as more offensive because of its ready popularity with local customers).

Economic theory and evidence can only contribute to such a grand debate, never arbitrate the issues by itself. Yet its contribution is central and important. And international economics is an essential tool in the argument which follows. It is to be applied rather as a fishing boat might be used to transport travellers—as a vehicle usable for that task, but needing considerable adaptation to make it effective.

Two major respects in which international economics needs recasting and development are advertised by the other two main words of this book's title: *Growth* and *Inequality*. It is no exaggeration to say that received international economic theory is 90 per cent static. For that reason trade theorists have largely ignored growth. In the past, and again recently, attempts have been made to marry trade theory and growth theory. And empirical studies have examined the connection between these two variables. The results of this research, however, are far from satisfactory, and one of the large tasks to be undertaken below is to show why that is the case, and to suggest ways forward to improve the position. Where inequality is concerned the problem is not so much that international economic theory has ignored the issue. Rather the most influential model of international trade has produced a particular angle on inequality, which is at best quite partial and at worst misleading. For that reason another task which has to be addressed below is the reformulation of some basic trade theory to make it more helpful and pertinent to the great issues of trade and inequality in the modern world. It turns out that one has to strike a balance between conviction and simplicity. A useful model has to be too simple to wholly convince, but rich enough to reflect large aspects of reality.

Like growth, inequality can be considered at many levels: inequality between nations; between regions; between cultural subgroups within nations; and between individuals. Much modern thinking questions the centrality of the nation state in terms of power and cultural definition. Similar doubts have always voiced themselves. Thus people have asked why inequality between nations should monopolize our concern when rich countries contain some extremely poor individuals, and poor countries contain some extremely rich individuals. The cynical view which says that international aid is a system under which poor people in rich countries give money to rich people in poor countries reflects such doubts.

With all due recognition for scepticism concerning the centrality of the nation, it has to be said that this is a line which is easy to overdo. Modern nation states remain powerful. When they are weak it is usually for internal reasons. This is far more often the explanation for weakness than it is the case that powerful international corporations have chained the nation's hands. Of course all nations are constrained by the facts of the world in which they find themselves, just as are individuals. Still, one seldom hears individuals described as weak on that account. And a further reason why the nation is of central importance is that nations to a great extent define the economic environments within which individual economic actors—persons, families, and companies—play out their parts. It is indeed one of the leading weaknesses of mainstream trade theory that it fails to model this notion of economic environment, and the formalization of what is involved will take up space in Chapter 8 below and elsewhere.

## 1.2 The Motivation of this Book

If any of the three topics of this book is taken alone, the concerned literature is enormous. The theory of international trade has been a lively area since the beginning of economics, and large blossomings of the theory were seen from the 1930s to the 1970s, and later. Even more, the theory of economic growth has developed even faster than the economies that it studies. Finally, inequality has always captured the attention of economists. They are aware of the problem of measuring inequality, and find its persistence a feature always in need of analysis. Any two of these three topics are connected in the existing literature. Yet these connections are often unsatisfactory; they frequently reveal inadequacies in the single component theories on which any cross-connections must be founded.

Thus, to illustrate the point, trade theories too often attribute the inequalities between nations to differences in costs of production, without explaining why these arise. Or, where cost differences between nations are attributed to differences in factor supplies, theory concludes that trade should take care of differences in factor prices, at least for somewhat similar countries; where observation says that this does not happen. And in any case, our world is full of starkly dissimilar countries; so what happens in that case? A vision of the world that can be called *neoclassical*, without stretching that term absurdly, says that knowledge is easily communicable across the world; when everything that atomistic agents will choose to do depends upon the prices that they face. While prices are important, they are only a small, and a declining, part of the total picture. Different countries provide different economic environments for their producers, and these differences account for massive variations in economic performance. Everyone knows this, and the new institutional economics aims to formalize these effects. However the existing literature fails to expose how institutions make their influence felt at the micro level; and how precisely they impact upon patterns of trade and growth. By considering the three topics; trade, growth, and inequality together, this volume is able to tighten up some of the cross-connections between the three.

The relationship between trade and growth is an area on which this book attempts to shed fresh light. Empirical studies show trade and growth to be positively correlated. In particular openness to trade and growth are associated. As with any cross-section studies, there are conceptual problems, and in any case the direction of causal relations is difficult to ascertain. Where empirical connections are hard to establish firmly, the theoretical palate allows for the endless creation of desired conclusions. That this should be the case reflects the fact that there is no evident rock-solid connection between the two variables. Even if only recent contributions are considered, the literature on growth is extensive. Contemporary work concentrates on endogenous growth, where the long-run rate of growth is determined within the model, and not as an exogenous value. Some of the models that connect trade and growth are what this author describes as sticky-tape models. The term indicates that a growth-enhancing effect is tagged onto what would otherwise be standard economic activities; as an incidental add-on, not taken into account by the economic actors. The trouble with that type of approach is that it makes almost any conclusion possible. Stick growth onto an activity that trade expands; and trade is growth-enhancing. Stick growth onto an activity that contracts with more trade; and trade is growth-diminishing.

A more rigorous approach follows that pioneered by Grossman and Helpman (1991). These authors treat growth as driven by R & D activity. This means that growth is the result of profit-maximizing decisions that balance cost against revenue, just as do any economic activities, such as baking bread. There is an important difference. R & D activity produces knowledge, and knowledge is by its nature a public good. Chapters 12 and 14 circumvent that problem by computing the socially optimal expenditures on R & D, which allows of definite results. Then we know that while actual private expenditures will be below the socially optimal levels, they will move in the same direction if private agents capture a fixed share of social benefits. This work develops a strangely neglected idea: that growth is influenced by the relative standings of different trading nations; that catching-up in technology plays a large role in the growth of many nations. If a certain country engages in productivity-improving research, how is the success that it can expect affected by the level of its own technology relative to those of the other countries in the world? If the relatively backward country enjoys an advantage, because it can learn from the leaders, we can talk of catching-up. Catching-up can apply to research technology, in which case countries will converge in growth rates, but not in levels. Or it may apply to the technological level itself, in which case countries will converge in levels. The important and surprising conclusion is that catching-up connections can be so weak, as to be almost negligible, and yet the asymptotic results still hold. Of course these are asymptotic results, and it may take so much time for them to determine what happens as to make them practically irrelevant.

In a much-cited study, Sachs and Warner (1997*a*) claim a negative connection between resource abundance and economic growth; the so-called resource curse. They offer a few off-hand suggestions as to why this relationship may exist, but little formal theoretical analysis. Chapter 13 attempts to repair this lacuna. It offers two models of growth and the resource curse. They differ according to the way in which resource abundance is introduced into a simple trade model. In the first model, resource abundance takes the form of a positive addition to the balance of trade, such as when the country sells the rights to its newly discovered oil rights for an annual payment of dollars. There is no great effect on domestic relative prices: resource abundance is essentially an income effect. In the second model, resource abundance takes the form of a large improvement in the terms of trade, such as would happen with a boom in coffee prices, as in the 1970s. That was a temporary episode. The model, however, examines the consequences for growth of a permanent shift in the terms of trade. In this case there are changes in domestic prices, as with a change in the terms of trade in the HOS model, and it is the implications of these changes for growth that deliver the results.

Martin Wolf, one of the most perceptive and stimulating economic journalists of our time, once remarked to me that the single most important variable for the prediction of a newborn baby's life prospects is the country into which that baby is born. It would follow that a prudent baby choosing where to be born would opt for the USA. I commented that there are huge differences in life prospects between a baby born in Harlem, New York, and a baby born in White Plains, not twenty miles away. Even as I voiced that argument I knew that this is not a strong rebuff to the original point. On a world scale of life prospects, the differences between Harlem babies, even those of the old Harlem before the recent move towards gentrification, and White Plains babies are significant, but not enormous. Martin Wolf's observation would produce no surprise from the typical educated individual, uncorrupted by an economics education, and the same individual would be amazed to know that economists spend a huge amount of time on models in which place of residence has no influence on personal prosperity and welfare. Naturally economists do not neglect place of birth because they are stupid. The problem rather is that received theories of trade and welfare allow no place for international differences in economic efficiency as such; that is, those not attributable to factor prices or scale. The final answer to the issues raised by this discussion will be found neither in the following pages, nor for the most part elsewhere. However this book makes a start by examining what would be required to model precisely an economic environment. That demands more than just a broad institutional approach, according to which this or that feature is associated with economic success. We need to know not just that institutions matter, but also how, and how much, they matter, and this for particular activities and groups. One point can illustrate the insights that a fresh look offers here. Some contemporary studies conclude that big government is bad for economic performance and growth, probably because it is too intrusive. Similarly, it is argued that corruption is bad for economic performance and growth, for obvious reasons. These effects cannot be considered separately. Corruption is typically a sign of weak government, whether it is intrusive or not. Where government is strong it will always attack corruption, because corrupt systems are wasteful and inefficient, and what strong government would want that? Successful anti-corruption drives have come from strong governments, such as La Guardia's New York administration from 1934, or the government of Hong Kong in the 1970s.

Where the nation is the focus of many discussions of inequality and different rates of economic growth, one should never wholly neglect smaller units. One needs to ask in particular how within-nation inequality is associated with national performance. Also it is an embarrassing fact that too many current economic models treat nations as if they were super-individuals. In truth a nation is more than an arbitrary aggregation of individual micro units. But while the small components that make up the nation are not free-floating units, they act autonomously to great extent; for which reason economic theory is better at modelling action at a micro level than at building grand models of national dynamics. The key to solving this tension between the large and the small is the idea of an economic environment. This aims to capture the way in which institutional features of a nation or region constrain and influence the actions of small and local actors.

## 1.3 China and the World

Some of the points at issue here are well illustrated by the most dramatic development in the world economy at the present time: the explosive economic growth of China. Naturally it is not strictly China that grows, but rather various Chinese cities and regions. These, however, are all influenced by the same fundamental changes in economic policy that have been implemented by the Chinese Communist Party, but with variable results in different places. It is certainly the case that economic growth in China has increased inequality within that country. This is an inevitable consequence of sudden rapid economic growth, where that growth has not been experienced to the same extent everywhere. Sudden bursts of rapid growth are never experienced equally across different regions. And the same can be said of the incomes of different types of individuals, whether these differ by education, place of residence, or in any other respect.

Where the regions of China are concerned, there are two possibilities for the future here. It may be that the slower-growing regions, particularly inland regions away from the major coastal cities, will catch up in due course. In that case the present inequality will be the type analysed long ago by Kuznets, in an approach to be examined in detail in Chapter 3 below. The Kuznets story depicts eventual convergence, as units slow to develop catch up in due course. That idea, however, may not reflect a second possible future for China. It could be that the inequalities of Chinese economic growth in the early twenty-first century will be seen to reflect fundamental structural features of China's geography and society. At a superficial level at least, Chinese economic development may resemble the experience of Italy, starting from the Italian 'economic miracle' of the 1950s and 1960s. There it was the North that experienced rapid growth. The South did not simply take longer to follow the same path; it never did so; and as a result Italian regional inequality increased.

The recent experience of China illustrates the manner in which the three key words of our title capture features that coexist and interact in the contemporary world. Trade has been crucial to China's economic growth, and the variable participation of its citizens and regions in that trade and growth plays a large role in explaining the increasing inequality in that nation. And it is not only variable participation that lies behind these changes. Trade, to employ the terminology of economic theory, alters factor prices. The wages earned by unskilled workers in China have risen, along with a huge increase in the employment opportunities for those workers. As a result, millions have been lifted out of poverty, if poverty is defined as falling below a basic subsistence minimum.

In many respects China represents the world. There has been a massive increase of trade worldwide. Even more than within China there has been, across the world, greatly variable participation in that revolution. And on a world scale there have been changes in factor prices and big shifts in income distributions. On a world view the question of participation is even more pressing than it is within China. While many countries have joined the movement towards greater trading, and fuller exploitation of comparative advantage, some countries and regions have barely participated. Notable among the non-joiners are sub-Saharan Africa and the Arab world. This volume is concerned with how far economics can throw light on contemporary developments and problems. From that perspective the non-joiners are of at least as much interest as the participators, and they will receive special attention, particularly in Chapter 7.

## 1.4 Poverty Traps

The foregoing discussion has brought to the foreground a leading question of this book. Does increased trade and the growth that it makes possible produce convergence? Even if they are slow to get underway, will poorer countries, regions, or families eventually catch up? Imagine for example choosing one disadvantaged family and asking what barriers stand in the way of that family improving its position over time and moving from destitution to some comfort and dignity. The issues involved are multiple and complex. Some of the barriers are to be found 'within' the family (poor education, health, etc.). Other barriers are located 'outside' the family concerned. These include such factors as crime and poor social support, as well as discrimination generated by limited information. Where employment is at issue a prejudice against the family's background may have some justification, and if it is costly to obtain detailed personal data, our subject may suffer unfair but rational discrimination. If we turn now to the case of a poor country, it is notable how the problems of explanation are sometimes surprisingly similar. Again barriers are sometimes 'within' the nation and sometimes 'outside'. In the former category come resources and cultural and economic environment problems. In the latter category are to be found the institutions of international economic relations, such as protectionism, managed trade, and international economic regulation as administered by the World Trade Organization. For these reasons both national and individual poverty will be examined in Chapter 3, which is entitled, significantly, *Why the Poor Stay Poor*.

Imagine a notional poor family. This may suffer what is sometimes called 'zip-code discrimination'. In the UK post-code discrimination has the same meaning. The term refers to the practice of some insurance companies in the US and elsewhere that charge different rates for motor vehicle or household insurance according to exactly where the insuree lives. These rates are then based on the average experience of policies in the same small locality (zip-code district). Consider now a similar argument for the case of employment, where an applicant (call him Mr Smith) gives an address of a district which on average provides poor employees. Mr Smith may come up against zip-code discrimination, which seems to be unjust. Yet the employer might argue that he cannot know or assess in detail the true ability and prospects of a particular individual, or that to discover this information would be too costly. Therefore, with or without sympathy and regret, the employer cuts corners and discriminates by zip-code. If this is the situation, there is a further knock-on implication that can be of great importance. Mr Smith cannot afford to move to a more favourable address. However a local government offers free employment-orientated education. Should Mr Smith enrol in a course under this scheme? The question is trivial if there are no costs associated with joining a course. It is reasonable to assume that serious costs will be involved. They may consist of loss of leisure, the pain of studying, or lost opportunities to search for work. Mr Smith has to balance these costs against the probable benefits of being better trained. Suppose that some education will mitigate the zip-code discrimination faced by Mr Smith, but will leave a substantial amount of it in place. That may well tip the balance against enrolling in the course.

The point of the above story is that Mr Smith is caught in a *poverty trap*. Because he is poor he lives in a poor area. Because he lives in a poor area he suffers negative discrimination when he seeks employment. Because he will suffer negative discrimination when he seeks employment it is not worth his while to improve his educational level. And because he will not improve his educational level he will remain poor. If instead Mr Smith comes from a district known to be largely inhabited by blacks, the cause of the discrimination against him could well not be a zip-code antipathy, but might instead take the form of simple racism. In that case there is just as great a possibility that Mr Smith will be discouraged from pursuing educational improvement. The same point applies. Discrimination lowers the return to educational investment and tends for that reason to lock in poverty. However this is not a poverty trap. Mr Smith's difficulty in climbing out of poverty arises because he is black in a racially discriminating society.

The crude way of defining a poverty trap is to say that it is a situation in which a major part of the explanation for poverty is the very poverty of the poor itself. The idea of poverty traps is an old one in economics but it has never lost its relevance. Indeed I will argue below that some kind of poverty-trap model is required to explain the long-term persistence of poverty. While the concept has been explained for an individual, it has as much, if not more, relevance for the nation. If it is not obvious why poor nations do not progress rapidly to improved situations—and I will argue that it is certainly not as obvious as might be supposed—one has to look for traps which hold the poor nations back. There are many possibilities for the modelling of such national poverty traps and they form leading themes of Chapters 3, 8, and 12. Poverty traps and their modelling are a main theme of Chapter 3. It may not seem obvious at this stage what is the connection between the example of the poverty trap based upon zip-code discrimination against a family, and the type of poverty trap that may afflict a nation. There are great differences between the two cases. Things that unite both examples, however, are the importance of imperfect information, and also the crucial role of the inhibiting effects of large start-up costs; and these two are connected. A worthwhile expenditure on education for our Mr Smith is huge; anything less is a waste. Similarly for a nation marginal in world trade to break into competitive international markets may require a big push. Nothing less may suffice. For a wide-ranging review of poverty traps in the neoclassical growth model, see Azariadis (1996).

## 1.5 The Kuznets Model

Simon Kuznets (see Kuznets 1955, 1963, 1971, and 1973), posited that economic development would proceed unevenly across individual families and across different regions and groups within a country. Imagine that in the initial state there is inequality but that its scope is limited because all agents are fairly poor. Economic development is conceived as a step jump from traditional to modern practice, which brings with it a large upward jump in income. Then the unevenness of development translates to an initial increase in inequality, just as China has experienced recently. Later, as all agents adopt modern practices, income inequality declines. Thus the path of inequality over time as development happens is an increase followed by a decline. This inverted U-curve is known as the *Kuznets Curve*. This is a brilliant model, beautifully simple and direct. As is often the case with very simple models it fails the test of realism. Time series of inequality for countries experiencing economic development do not usually exhibit an inverted U-pattern. Some empirical investigators have placed another interpretation on Kuznets's analysis. They examine income levels and inequality in cross-section, looking to see whether middle-income countries will show the greatest inequality, with poor and rich countries relatively more equal. Again the findings are negative. A chief problem with this approach is located in the manner in which the development story is choreographed. All developing countries start out about 1950 in a state of similar economic backwardness, in which there is little inequality. In fact, many societies in a 'primitive' pre-industrial condition exhibit great inequality. In this connection China is exceptional, because forty years of communism led to it starting from a position of unusual equality.

A deeper issue concerns the manner in which the increase in inequality at the start of the development process is described. The picture is stereotyped and simplistic. All agents in the model are supposedly the same. And the benefits of adaptation, the pay-off from grabbing modernization with both hands, are considerable. Despite this some agents move quickly, others more sluggishly. Why is this? Differential rates of adjustment should ideally be modelled, not simply assumed. The question of whether the poor can rightly be characterized as slow and non-optimal adjustors is one of the important questions to be addressed in Chapter 3. These questions, posed for the case of individuals within a country, can be presented with equal force to the case of countries within a developing world. With regard to income distribution the time pattern of development which the Kuznets model predicts takes the form of increased inequality followed by convergence to much greater equality. In a simple version of the model the eventual outcome is complete equality. In a sense there is economic convergence throughout the story, because all agents are moving towards the same destination, each at a pace appropriate to that particular agent. When they have all completed their journeys they have converged to equality.

This is an optimistic view of development. The poor, such a view says, are those who have been left behind, or started late. Their time will come, when they have traversed the same road as the rich passed over earlier. Put like that the description sounds far-fetched. And it should be said in fairness to the economists who have constructed models of economic convergence that they do recognize that there are major cases in which their models seem not to fit the facts well. Important work has been done to try to capture what kind of features need to be added to explain why convergence is not always observed. Even given these major extensions and qualifications, the view most commonly held by economists may well be that theory says that economies (and other economic units) should converge, but in practice that often does not happen. A leading contribution of this volume will be the subversion of the idea that economic theory shows that convergence is to be expected. In the main body of the literature convergence is given a local interpretation, later to be identified as $\beta$-convergence. It is not because the core assumptions of the convergence models are probably inaccurate that these models should be questioned. Even given these core assumptions, convergence, particularly of incomes, is uncertain, and in any case less simply structured than much received theory indicates. These issues are considered in detail in Chapter 4.

## 1.6 The Stiglitz Model

In Chapter 3 the model of Stiglitz (1969) is adapted to produce poverty traps. In its original unadapted form the Stiglitz model pinpoints perfectly the reason for which the question of why the poor stay poor is a serious non-trivial issue. Stiglitz takes the Solow (1956) growth model and disaggregates it in the following manner. The population which provides the input of labour into the aggregate production function consists of individuals all alike in that they each provide labour at the same flow rate. They own together the aggregate capital stock, but this is distributed unequally, so that some individuals are poor; others rich. All factor inputs are paid their marginal products. So all earn the same wage income, given at any time by the then marginal product of labour; and all earn the same return to any unit of capital owned at the time. All individuals save the same proportion of their total income, regardless of source. Stiglitz shows that this disaggregated Solow model converges to the same steady state as does the standard Solow model with the same saving rate. And in that asymptotic steady state all agents are alike. They all converge to owning the same capital.

In the Stiglitz model the poor do not stay poor. Their wealth grows faster than does the wealth of richer individuals. The reason for this is simple. For agents the rate of growth of their personal capital at any time is their saving divided by the capital that they own. That is the same as saying that the rate of growth of their personal capital is proportional to the ratio of income to capital. This ratio is larger for the poor and smaller for the rich, simply because all enjoy the same wage income, and this accounts for a larger share of income for the poor.

It goes without saying that this model has to be subjected to some forensic analysis. However little may be left standing after the model has been criticized, it makes a powerful point, that always has to be taken into account in any analysis of poverty and the problems of getting out of poverty. It is often, if not always, the case that the first step up the ladder is the easiest. If your game of tennis is really bad, an hour's coaching, even quite indifferent coaching, can do wonders. If you start as a really fine tennis player it can only be difficult, if not impossible, to make your game better still. Similarly, if you start with \$1 million, you will need to find \$100,000 to increase your wealth by 10 per cent. But if you start with \$1 you need only find or save 10c to increase your wealth by 10 per cent. You might find that 10c in the street; but even if you save it from your wage income, the effort may not be impossibly burdensome.

The chief reason why the first step up the ladder may not be so easy after all is the poverty trap. That is why the development of a poverty-trap version of the Stiglitz model is of considerable interest. A key to the modelling of poverty traps is found in one term: *non-convexity.* In economic models with convexity small-scale operations are highly effective. That means that the poor can pull themselves up by investing on a small scale and allowing that small wealth to grow. In fact serious non-convexities are everywhere. The problem is not so much to find one; rather it is to determine which of the many possible causes of non-convexity will be most significant in practice.

The idea of the poverty trap is not original, and the demonstration that non-convexities can lead to multiple steady-state equilibria is unsurprising. What the extended Stiglitz model can show, however, is rich and intriguing dynamics. For example, general capital accumulation can produce the result that an intermediate medium-level steady state disappears: it, or anything close to it, ceases to be an equilibrium. This is another anti-convergence finding.

## 1.7 The Diamond Capital Model

Another model is the Diamond capital model. This is an overlapping generations model with capital, where the capital stock represents the saving of the young to finance their retirement. It has long been known that there can be more than one steady state in this model. This is because low capital implies a low wage rate, which in turn implies low saving, which implies low capital. The circle is closed and we have the perfect theoretical realization of a poverty trap. It is surprising that this attractive possibility has not been given much attention. While many textbooks mention it, it is usually dismissed as a theoretical curiosum. The reason why this has been the typical reaction is illuminating. The attractive multiple-equilibrium pattern cannot be obtained in examples specified by simple functional forms. In particular there is no good multiple-equilibrium case when the elasticity of intertemporal substitution (the EIS) is constant. Yet the assumption of a constant EIS is not only uncomfortably special. It is also particularly *unrealistic*.

The case that $\beta$-convergence is far less theoretically plausible than is commonly maintained depends to a great extent on allowing the EIS to vary in a reasonable manner. With the Diamond capital model the effect of introducing a variable EIS is dramatic. A variable EIS is a feature of a specific well-behaved class of utility functions, and if a member of this class is suitably sculpted we can have a continuum of steady-state solutions to the Diamond model. Then everything depends, at least locally, on initial conditions, and the poverty trap describes the world completely.

## 1.8 Growth Empirics and Growth Theory

Chapter 4 looks at the substantial empirical analysis of cross-section growth data, in particular research directed to testing the hypothesis of $\beta$-convergence (that poor countries grow faster). That literature has been reviewed extensively elsewhere; see for instance Durlauf and Quah (1999). The focus of the argument of Chapter 4 is directed somewhat differently from that of most parallel discussions. Attention is on placing the problem of making inferences, of the type at issue, within the context of standard applied economics, and classical and non-classical regression analysis. Then, contrary to much received opinion, cross-section growth studies are not particularly disreputable on account of the fact that variables selected to be 'independent' or right-hand-side variables are significantly intercorrelated. This is a problem, yes, but one nearly always encountered in multivariable applied economics.

A deeper issue concerns what inferences can be drawn from any multivariable regression analysis, whether or not individual coefficients are shown to be significant. These studies do not reveal the results of counterfactual experiments. All the information that is the input in the study is in the variance-covariance matrix and the sample size. Time-series data sets allow richer possibilities, as they can be truncated and sectioned as desired. Even so, knowing how the various variables in the data set are related is far from knowing why they are related, as they are related, in the world. To put it another way, many applied analysts see themselves as revealing how one or more particular variables are causally affected by the various factors measured by the right-hand-side variables. Causality is a slippery and dangerous notion, in economics as in physics. The latter dispenses with the concept except for informal expository purposes. Economists find it irresistable. By comparing regressions of growth on standard explanatory variables with a regression including latitude as an independent variable, the argument will make clear some of the leading difficulties that confront any attempt to reach firm and reliable conclusions.

The theoretical sections of Chapter 4 contain one of the most distinctive contributions of this volume. As has already been noted above, in Chapter 3 the Stiglitz model predicts that unequal agents will converge in the limit to equality. Its crucial assumptions can be questioned. However, the model does challenge the unthinking claim that if one starts far down the income distribution, it will be impossible to climb up, however slowly. Poverty, says the model, may confer tangible advantages. It may be relatively easier to improve gradually on a very low position than to maintain, or improve upon, a high position. That feature powers the neoclassical convergence models.

The convergence hypothesis may be taken as an intuitive proposition to be tested on the data without the need for elaborate theoretical justification. Thus Baumol (1986) in a pioneering paper offers little theory to motivate his test of the convergence of post-war OECD per capita incomes. Baumol's discussion hints at a catching-up story. The idea of convergence as catching up deserves development, and that is attempted in Chapter 12. There a surprising conclusion emerges, Catching-up effects can make a spectacular difference, at least asymptotically, even when their strength is extremely weak. In particular the standard conclusion that convergence will not be a feature of an endogenous-growth world is not necessarily correct when endogenous growth is affected by catching up.

However, the leading theoretical argument which is used to justify the hypothesis of $\beta$-convergence, which says that the incomes of poor units grow faster than those of rich units, is the neoclassical growth model. We have seen above how the Stiglitz disaggregated-growth model leads to $\beta$-convergence. Robert Barro has promoted an alternative approach which treats different units of observation (these would be countries in the cross-country studies) as on different points of a Ramsey optimal-growth path converging to a common stationary state. See Barro and Sala-i-Martin (1995).

The model has an orthodox Ramsey-Solow production function, common to all units. All agents have the same utility discount rate. If the discount rates of agents are constant but differ, there will be no convergence, and in the limit all capital will be owned by those agents with the lowest discount rate. That points to another possible reason for enduring poverty. The poor stay poor because they have high discount rates. An economist may assume a discount rate to be a constant parameter, but it is in principle an endogenous variable. Even so, there is no reason, as this chapter will demonstrate, why the poor should not be less willing to save, in a sense that will be made precise. The neglect of this possibility is a shameful lacuna in the literature. It will be seen that willingness to save depends upon the utility discount rate, of course, but also on how sensitive marginal utility is to a change in the level of consumption (the elasticity of marginal utility). This last value is simply the inverse of the EIS already discussed above.

A notable theorem is demonstrated in Chapter 4. Take any time path of capital which:

- is monotonically increasing and asymptotic to the Ramsey steady-state level;
- implies a monotonic increase in consumption given a particular production function.

Then there exists a concave utility function such that the above path is the optimal Ramsey solution starting from the initial level of capital and maximizing the said utility function. This result blows away any notion that the behaviour of different units at different points on their Ramsey paths to the common steady state can be shown to conform to any simple pattern. How can this be? Do not Barro and Sala-i-Martin prove $\beta$-convergence for a standard Ramsey model? Indeed they do, but they slip in a little assumption. They assume the elasticity of marginal utility (and therefore the EIS) to be constant. For the arbitrary increasing path of capital in the theorem just stated to be optimal, the EIS as a function of consumption might have to behave in quite a strange manner.

Yet we do not need some bizarre relation between consumption and the EIS to tell a story which may well throw as much light on the failure of $\beta$-convergence in broad cross-country studies as the other explanations available. If the poor have a low EIS their capital holdings will grow slowly. And this idea is highly plausible. When one is very poor the intertemporal substitution of consumption which saving requires is painful. It is not necessarily painful because utility is discounted strongly. Because the rate of interest will be high in a low-capital state, the rate of fall of marginal utility will be rapid. But this need not be because consumption is falling rapidly. It could be just because marginal utility is highly sensitive to slowly growing consumption.

In the original convergence models, countries were treated as if they were perfectly integrated with regard to the transfer of technology—hence the common production function—and perfectly isolated when capital mobility is concerned. The strong contrast between the way in which the mobility of capital and the mobility of technology are specified is hard to accept either for a nation or for a world of nations. Later developments have included capital mobility in various forms. With perfect capital mobility there is immediate convergence of output per head and of wage rates. However, uneven ownership of capital implies that incomes will remain unequal, for a long time at least.

This finding is not fundamentally altered if adjustment costs of capital accumulation are included, or if part of capital accumulation takes the form of human capital, always provided that the production function is concave, although rates of convergence are affected. Barro, Mankiw, and Sala-i-Martin (1995) consider a model (called the BMS model for convenience) which includes a credit constraint. The accumulation of human capital must be financed by own saving, because human capital cannot be used as collateral for outside borrowing. This constraint stops the immediate convergence of output per head and may increase the eventual extent of income convergence. The BMS model is examined in detail in Chapter 4. These authors consider a special case. It is shown that somewhat more general cases yield different conclusions and possibly more relevant insights.

## 1.9 Trade and the HOS Model

In a model of the Heckscher-Ohlin-Samuelson (HOS) type, or its extensions, trade in goods substitutes for factor mobility. It is no surprise therefore that it does not need perfect capital mobility to undermine the tendency to convergence advertised by Barro and Sala-i-Martin. In some contexts free trade in goods will be equally fatal to $\beta$-convergence. This implication of the HOS model was first noted by Ventura (1997). Ventura's model, together with other models mentioned above, shows how some theoretical conclusions can mock what intuition would seem to tell us. It is common to hear politicians or economists declaring that poor countries can grow rapidly and catch up with richer countries if they are given free access to international capital markets, and also allowed free trade in goods. Now we have models, which are far from absurd, that arrive at exactly the opposite conclusion.

Free capital movement and free trade can increase national welfare, at least for one-agent countries with no interpersonal conflicts of interest. However, national welfare and asymptotic convergence to identical steady states are not at all the same thing. That point in turn exposes a serious problem with an undue emphasis on convergence. A convergence argument says to a poor country: 'Do the right things and eventually you can be rich; forget about the transition between where you are now and where you will tend.' A counter-response from the poor country might be: 'I do not care about where I am going to asymptote; I am hungry now, and why do you not help me?' Incidentally this discussion may point up the perils of identifying economic growth and welfare too closely. They are not the same thing, a point that is revisited in a different context in Chapter 12.

The first and enduring insight was originally Ricardo's. The principle of comparative advantage says that countries can gain from trade if their relative costs or different demand structures lead to their having different relative prices before trade. An economics training is liable to leave its subject with the impression that this result, however important it may be, is somewhat obvious. In fact for the world at large it is strongly counter-intuitive. The influential idea according to which beneficial trade requires a 'level playing field' almost amounts to a denial of the result. It can be no part of any respectable argument to attempt to undermine the principle of comparative advantage. It does need to be interpreted correctly. In particular, the effects of non-tradeables when they are present, and also the whole issue of the economic environment, must be taken into account. These issues are examined in depth in Chapters 7 and 8.

Within the same family of competitive general-equilibrium trade models is to be found the most popular and influential of all trade models: the HOS model. It was originally designed by Heckscher to explain the pattern and effects of trans-Atlantic trade in the nineteenth century. In that application the two factors are land and labour. The model was further formalized by Ohlin, and later still rigorously analysed by Samuelson. Samuelson in particular, in tandem with Stolper, used the model to analyse the effect on the returns to capital and labour of a freeing of trade. Stolper and Samuelson (1941–2) proved a 'magnification' result. They showed that factor prices are changed more proportionately than a change in goods prices. That is the reason why labour suffers from a cut in the tariff on a labour-intensive product, because the real wage falls even if measured in the labour-intensive good. In the context of the Stolper-Samuelson model the two factors have become capital and labour. More recently Krugman in several papers and Wood (1994) have applied a version of the model to the analysis of globalization. Now the two factors are skilled and unskilled labour. The evident uncertainty as to which are the two factors, and a parallel uncertainty as to which are the two goods, defines one of the first criticisms that has to be advanced against the HOS model: that it is too low dimension for even a stylized depiction of reality. There are further issues which have to be explored. The HOS model assumes the best technology of production to be freely available to all countries, whereas evidence indicates that productivity, even for similar techniques, such as production-line motor-car assembly, varies greatly from one country to another.

For all its problems, the HOS model is effective in highlighting some central questions for the analysis of international trade. The most important of these questions have to do with income distribution and inequality. The model does not depict inequality explicitly. Rather it focuses on the prices of goods and factors. Those prices can be translated into an income distribution once the distribution of factor ownership is known. Many theoretical treatments choose an extremely simple assignment of factor ownership. Thus Stolper and Samuelson treat capital and labour as class and lobbying interests, not just as factors that have market prices. The Krugman-Wood redefinition of the HOS model, with skilled and unskilled labour as the factors, seems to make a simple identification of factor prices and inequality inescapable. That model can be interpreted, however, to allow individuals to own varying proportions of skilled and unskilled labour (corresponding to different levels of skill training). Individuals can be allowed to own a mix of factors in any multi-factor model, and once that is done account can be taken of the point that factor ownership alters over time according to the dynamics of change given by a model of factor accumulation, such as the Stiglitz model. The HOS model says that trade tends to make factor prices more equal between different countries. As Wood shows, equalization in that direction may be accompanied by either an increase in equality or a decrease in a particular country.

Chapter 6 looks at models which recast the simple classical trade theory framework to make it richer and able to yield new results, and to answer questions which elude the basic model. The developments concerned involve models which are higher dimension than the two-factor two-good HOS model. Some classical results, for example a version of the Rybczynski theorem, do generalize to a large high-dimension version of the model. However other implications of the HOS model, in particular factor-price equalization, do not generalize beyond the simple case. A particular two-good three-factor model is examined. The key feature of this model is that one factor, called without loss of generality skilled labour, is only used in one of the two production sectors. The other two factors are used in both sectors. In the most tractable case the cost of the third factor (the one used in only one sector) enters *separably* into total costs for that sector. That means that the price of this factor has no effect on the relative use of the other two factors. Then the following results may be derived, and compared directly with a parallel list of results for the HOS model:

- Factor-price equalization. Can come about only if two countries have the same price of skilled labour. With different prices of skilled labour, two countries can support different rates of profit and different unskilled wages, even if they produce both goods and share the same technology.
- Magnification. Is observed but now its force is weakened.
- Rybczynski's result. When capital moves into a country which faces fixed world output prices, and produces both goods, and which remains diversified after the capital inflow, there can be a fall in the return to capital. Suppose that the capital inflow tends to expand the high-tech sector. This will push up the wage rate of skilled labour, and that may depress the return to capital.

The results above are more satisfactory, for the intuition at least, than those derived from the HOS model. For example, any model which predicts factor-price equalization causes some embarrassment, as it is plainly at odds with reality. Similarly, the Rybczynski result goes somewhat against reasonable intuition, and to have it softened can only be welcome. On the other hand, while it must be recognized that other changes apart from trade liberalization have influenced inequality in developed and underdeveloped countries, the Krugman-Wood account of trade liberalization and inequality is attractive. For that reason it is reassuring that this new model preserves the essentials of that account.

## 1.10 Unequal Access to Trade Opportunities

Chapter 7 re-examines the old concept of comparative advantage. It appears, and this case has been argued vigorously, that every country must have a comparative advantage in something. This is because comparative advantage is concerned with relative prices prior to trade, and it is argued that even in the most desperately poor and inefficient country some good must be relatively cheap. This relative cheapness, so the argument has it, provides the scope for the advantageous export of that commodity. Without wishing to reject the evident rightness of this comparative-advantage argument for many circumstances, we propose the concept of *participation*. The idea is that only a subset of those goods which we would normally think of as tradeable, can in practice move in international trade. To be traded goods have to pass tests of quality and reliability, and in the limit a country may be unable to produce any of those goods. In that case the relative prices of those goods are given by their international prices, and the country concerned will of necessity have no comparative advantage in any of the goods that are in fact exchanged in international trade. This way of looking at things is not dissimilar to that provided by Macartan Humphreys, one of the most able Oxford economics graduate students of recent years, in his M.Phil. thesis, see Humphreys (2000). He noted that when a poor country is opened up to trade it will not favour the producers of labour-intensive goods if their produce is undercut by capital-intensive substitutes. An instance would arise when local buckets, produced using labour-intensive methods from wood and basic iron, have to compete with plastic buckets sold on the international markets by rich-country corporations.

These ideas may throw light on the situation of countries that are to a great extent outsiders where world trade is concerned: including the vast region that goes by the name of sub-Saharan Africa (SSA). Another case is the Arab world, the merchandise trade of which, as has been noted by the UNDIP report, is less in total than that of Denmark. These regions include many failing states, and poor trade performance is but one aspect of dreadful economic performance, which in turn is only one aspect of social and governmental failure. In Chapter 7 a theory of participation is developed. It makes monopolistic-competition-style modelling the norm and shows how the ideas of Macartan Humphreys can be generalized and formalized.

Chapter 8 moves away from the basic neoclassical view of producers and their decision structure. In neoclassical theory the producer's decision problem is completely defined by technology (the production set), and prices. Such a specification can never model the consequences of government, failing states, and corruption; to name but a few features of the economic environment that are certainly not technology or prices. These features are examined, departing from traditional neoclassical analysis, yet keeping as close as possible to the micro-foundations of economic decisions, that are the strength of neoclassical analysis. Rent-seeking is a good and insightful concept of received theory. The chapter presents a new concept: rent-looting (basically, insecurity of property rights); it underlines its importance and discusses how it affects individual decisions. The chapter also looks at corruption, and how it affects producer decisions. It argues that corruption is a symptom, not a disease, and that it usually indicates weak government. This view is in contrast to the opinion that bad government is always overintrusive government. Corruption flourishes when governments are weak, and even when they are autocratic, yet still weak.

A concept developed and examined in Chapter 8 is *endogenous corruption*. The idea is that corruption is not simply the product of particular cultures, prone to corruption; although the cultural influence on corruption should not be dismissed. In addition, corruption is not unlike a communicable disease. When prevalent, corruption creates environments that nurture corruption. This happens for many reasons. There is the 'everyone is doing it' effect; corrupt societies punish uncorrupt behaviour. There is also a shift in economic incentives. In deeply corrupt societies the returns to honest behaviour are reduced; as are the returns to corrupt behaviour (reallocation as it is called here). In less corrupt societies, both returns increase. This allows low and high levels of corruption to be possible solutions to the same model. A simple mathematical model makes all this precise.

Chapter 9 examines the real exchange rate. After discussing definitions, it notes that in theory there are as many real exchange rates as there are non-tradeable goods. That point applies with particular force to the housing market, where house-price bubbles can lead to the real-housing exchange rate departing significantly from other real exchange-rate measures. Evidence is presented to show that different non-tradeable prices converted to dollars are weakly correlated. It is argued in addition that the misalignment of an individual non-tradeable price need not result in that particular market not clearing. Rather when the balance of trade constraint is weak, because liberal capital markets are willing to finance deficits, the balance of trade may absorb the disequilibrium. If overvaluation of the real exchange rate, meaning that prices of tradeables are low relative to non-tradeables, is a common occurence, that calls for an explanation. The chapter finds that explanation in two key features. First the political economy of the real exchange rate implies that overvaluation often favours powerful and elite groups, notably the urbanized and the westernized. Secondly, overvaluation is frequently the result of anti-inflation policies that tie the domestic currency to a hard currency, such as the US dollar. These policies often succeed in limiting inflation, but only after they have produced an overvalued real exchange rate.

## 1.11 Mobility and Urbanization

Chapter 10 looks at mobile factors and urbanization. The simple HOS model, in the fully diversified case, with factor-price equalization, produces the strange and counter-intuitive conclusion that factor migration is unnecessary, because factor prices are equal everywhere. Should migration happen, perhaps on account of political reasons, it is entirely innocuous. The Rybczynski result says that factor migration will be accommodated by suitable changes in output levels. So in-migration of labour, for example, will be accompanied by an expansion of the labour-intensive sector; all at the same factor prices. Migration affects growth but never inequality. The simple HOS model is unrealistic. When we move to more realistic models, we encounter many reasons why migration of all factors can be well motivated. Many commentators on the contemporary international economy emphasize the increased facility of international factor migration. However, barriers to the international migration of factors deserve at least as much attention. These are various and numerous. They range from formal barriers, as with much labour migration, to cultural blocks, language, imperfect information, and non-transferable skills. Most real-life labour migration does not accord with the Harris-Todaro model of isolated individualistic migration into labour markets that are blind to anything apart from true productivity and skill. Similarly with capital markets, in-migration highways, founded in imperfect information, play a large role. Contrary to the simple Rybczynski model, in-migration of labour can harm existing residents, although it confers considerable benefits as well.

The huge emphasis on migration of factors across international borders reflects the politics of mainly rich industrial countries far more than the facts of factor movements. This can be illustrated by the case of labour migration. Much, probably most, labour migration is internal: it takes place within national boundaries, or at least between adjacent poor countries. These movements include the massive migration from country to city that is happening over most of the world. Urbanization is the great economic revolution of the transition between the twentieth and the twenty-first centuries. Shortly after this volume appears a majority of the world's peoples will live in cities, be that in comfortable housing, or in urban slums. Much economic analysis of the consequences of this revolution largely remains to be done. Chapter 10 makes a start by analysing the trade between city and country, using the same techniques as have proved to be effective in the analysis of international trade.

## 1.12 Globalization and International Governance

Chapter 11 looks at international trade rules and the institutions that govern international trade. For various reasons unilateral trade liberalization is often not the rational best strategy for an individual nation. This can be because protection gives national benefits, as with the optimal tariff case, or for political-economic reasons. Then tariff reduction requires coordination. Also, even if the unilateral reduction of tariffs would be in the national interest, it might pay to hold onto protection as a bargaining counter. The GATT system, later the WTO, was founded on the most-favoured nation (MFN) principle. This has been regularly undermined by the formation of customs unions, and by the ever-increasing use of bilateral trade agreements. It is argued that these bilateral agreements cause serious trade diversion, and that they inhibit successful multilateral reduction in protection.

Some political actors have been called anti-globalization activists. Being against globalization is somewhat like being against earthquakes; it is nearly inconceivable that the world will halt globalization. And were that to happen the consequences would be catastrophic, as they were in the 1930s, when such globalization as was present at that time was largely rolled back. Leaving aside the lunatic fringe, who would abolish capitalism if they had their way, anti-globalization positions embrace quite a range. A leading position holds that international trade is good in itself, but that it is dreadfully managed in today's world. Criticism is directed at the institutions that govern the world, notably the IMF and the WTO.

Chapter 14 examines the case against the IMF and its structural adjustment programmes. The discussion is set against the background of the so-called Washington consensus, the components of which are argued to be valid as long-term objectives, yet sometimes a poor guide to the design of structural adjustment programmes. That said, some of the criticisms of IMF-style adjustment are shown to be incoherent, especially where the control of inflation is required. If the claim that globalization has greatly increased inequality, within and between countries, is correct, we should see it in the data. It is shown that inequality is a slippery concept, as it can be measured in several ways with different implications. Also global inequality is an integral concept that can never reflect fully disastrous outcomes in small subsets of nations. Noting that point, nearly all measures show a reduction in inequality in the last ten or fifteen years, or no rise at worst. The reason is mainly the rapid growth of incomes in China and India, countries whose huge size makes their changes felt at the level of world measures.

## 1.13 How to Use this Volume

This book is not a textbook of the usual kind. First it reaches across fields that are usually kept separate in standard course designs and allocations. Secondly, it includes numerous original approaches and results, which supposedly the textbook is not meant to do. That said, there is much material here that I hope will prove to be of value as supplementary reading for courses on economic development, international economics, and economic growth. I have made little effort to keep the accessibility of the material constant throughout the book, but have allowed the prior level required for each chapter to 'choose itself', as seems appropriate in each case. Even when the argument inescapably moves towards difficult mathematics, the reader is always given intuition and a simple account to aid understanding. Important components of the arguments that make up this volume are inescapably mathematical in character. I have carefully avoided glossing over the mathematical nature of the analysis when to do so would be to obscure the essential nature of a case. That said, there are instances when mathematical detail might intrude on the underlying points at issue. In some such cases the mathematics is removed to a mathematical appendix at the end of the relevant chapter. References are collected together at the end of the volume.

# 2

# Trade in an Unequal World

## 2.1 Globalization

Because this book has been several years in the writing, I have been asked by friends and colleagues on many occasions: what is your book about? My usual reply has been: you can have the short answer, that takes a few seconds; or, if you have an hour, the long answer. It was always the short answer that was requested. In that case I said: 'My book is about globalization, although I do not like that term.'

My dislike of the term globalization is not simply a snobbish disdain for a word that is currently on many lips. Neither is my aversion to the word to be explained by the opinion that it means nothing. On the contrary, globalization means too much. It is an all-embracing term for several things that are happening in the contempory world. As Bertrand Russell realized a long time ago, words are perilous devices. We only need to utter them, or write them down, and we come to feel that there must be something in reality that corresponds to our verbal creations. '*N* is the largest integer.' is a well-formed sentence, and sounds convincing. But it is also pure nonsense. When we talk about globalization, there must be something that corresponds to what we are discussing. Perhaps there may even be just one thing that corresponds to what we are discussing. In fact many components make up the scene that encourages people to talk about globalization. Some are new; others less so. Globalization should be a global phenomenon; but its characteristic features are not to be found everywhere. Capital is more mobile than in recent decades, but certainly not completely mobile. Capitalism is triumphant, except where it is not. Trade is often free, but also frequently constrained. Even when trade is not constrained by formal barriers, distance still matters hugely. Distance here means physical distance, as where Beijing is a long way from New York.

But distance can also mean separations created by cultural differences, or by gaps of information.

For a simple analogy that captures some of the points at issue, consider the difference between shopping for groceries by visiting a local retailer, or a local market; or, as is now possible in several countries, shopping for groceries on the Internet. The Internet has the advantage of convenience, and the additional cost is small. Its disadvantage is the loss of the ability to inspect products offered for sale. With completely standardized products, that is no problem. Many fresh products, however, cannot be standardized. Meat or fruit, for example, can vary from excellent to awful without the simple description changing. The internet shopper has to be familiar with the seller, and to trust that seller. The costs of establishing that link of familiarity and trust are a barrier to trade, similar to those that exist across the world. A leading theme of the following chapters will be the vital importance of the 'wall-flower' countries and cultures; those that have not gone onto the globalization dance floor to join in. These are important because they are likely to generate problems: unacceptable poverty, or violence and division. They are also important because they test our theoretical ideas about the contemporary world.

When globalization is taken to embrace the universal dissemination of knowledge and ideas, then its validity becomes even more questionable. It is far from clear that over the last thirty years basic ideas about what is a good society and how to run one, have converged across the world. On the contrary, in certain areas, there have been increasing signs of division. One might think that the collapse of Soviet Communism represented a striking convergence of ideology, but this may be misleading. The fact is that the communist method of running societies never commanded deep widespread support. The Russian people never voted for it; it was imposed on Eastern Europe; and developing countries supposedly allied to the Soviet Union frequently rejected it.

On the other hand, the worldwide resurgence of religion has increased the influence of strongly divergent intellectual tendencies. Religion was always there, but after the Second World War it was pushed into the background. Its changing role is evidenced by the growing popularity of creationism in the USA; by the Iranian Revolution; and in an opposite direction by huge increases in secularization in Ireland, Italy, and Poland. Beyond religion as such, the status of science, that was extremely high from the 1940s, has declined. It is questioned by so-called 'fundamentalist' religious believers; but also by faddish sociologists; by the secularly superstitious; and even by some who do not question its intellectual foundations, but who depict it as an instrument of evil. Worse still, fewer and fewer students in industrial countries want to study science, which is rejected in favour of 'softer' subjects.

It is notable that the Internet, the World Wide Web, that was meant to unite humanity with a common stock of knowledge, has done nothing of that kind. It has become a massive depository of invaluable data, an archive of erodite facts, and a world-size shopping mall of nutty opinion. Of course, all this existed previously. What the Internet has done is to make the sorry picture available at the click of a mouse. In the age of the human genome 25 per cent of Americans believe that Elvis Presley is still alive, and at least that proportion hold to the view that the world is 4600 years old. In the case of the Internet it does not follow that globalization leads to convergence.

It follows that the features that will be depicted in this volume are those of a messy world. Fortunately there will be more economic modelling than depiction, and gross oversimplification is basic to the manner in which economic modellers go about their task. Nonetheless, economic models are worthless if they do not reflect some essential features of a tangled reality. Leading aspects of the world as it has developed over the last twenty-five years, and which explain the growing employment of the word globalization are:

1. A great increase in world trade relative to world income. See Irwin (2005). Over the period 1960 to 2000 the ratio of world exports to GDP increased from 12.5 per cent to 25 per cent (quoted by Glyn 2006: 97).
2. A large increase in the sophistication of the use of international trade, as when trade in components grows more rapidly than trade in final products.
3. A substantial rise in the scale of capital movements relative to past levels, and relative to world income. Over the last three decades the ratio of foreign direct investment (FDI) to total investment has increased rapidly. FDI increased eightfold during the 1990s (see Calomiris 2005), and now exceeds, but only just, its pre-1914 level (see Glyn 2006: 100–1).
4. A large decrease in the costs of long-distance transport, especially by sea, but also by air. Much of this fall in costs can be attributed to the use of containers, as these make possible door-to-door delivery of bulk consignments, without the need for dockside sorting.

5. A huge migration of manufacturing to newly developing countries. Notable in this connection is China. That country is now responsible for 50 per cent of world production of clothing, and 75 per cent of world production of childrens' toys.
6. A rise in the mobility of people, both within and between nations. Ours is an age of mass migration and massive urbanization.
7. The international transmission of technical knowledge is ancient. Today however it has become rapid to an extent unimaginable in the past. Sometimes knowledge cannot move by itself; it needs a vehicle, such as foreign direct investment. Such vehicles abound in the globalized world.
8. Great changes in levels of income, in the form of rapid income growth for certain countries and for particular groups. And great variability in the extent to which various groups participate in the growth of incomes.
9. Some signs of increasing instability. While fluctuations in output have moderated, if anything, and inflation has fallen across the world, exchange-rate crises have become more frequent. Often this reflects a failure to design good policies to combine capital mobility with stability.

Not included in the above list is the near-instant transmission of news worldwide, with supporting TV pictures. A large accident anywhere is quickly seen everywhere. That this is the case is undeniable. Its implications, however, are not clear. First it is not obvious that it makes a radical difference that we know so much so quickly. Secondly, the dramatic accident, inescapably newsworthy worldwide, may be a somewhat misleading case. Generally speaking, news channels in different countries, or aimed at different groups, are highly selective in the 'news' that they choose to present, and in the manner in which it is presented. This can be seen by comparing the treatment of the Israel-Palestine conflict in the US media with the treatment of the same conflict in the European media. The Arab TV channel *Al Jazeera*, provides an even more striking illustration of the fact that the same news is different according to who transmits it.

How well does traditional economic modelling accord with the above description? The most popular model of international trade, the HOS model described in detail in Chapter 5, fits the specification perfectly in some respects, and not at all in others. In the simple version of the model there are no tariffs and no transport costs; though these can be added, and the model shows how to treat them. Technical knowledge is a worldwide public good; no country can enjoy an advantage over another in that field. Climate and natural resources are irrelevant for comparative advantage. This is a world far more globalized even than the one we inhabit today.

By way of striking contrast, factors in the basic HOS model are completely immobile, as in fact they never were, not even when mobility was highly restricted. This last feature, like the absence of tariffs, can be dispensed with, and the model offers a picture of the consequences of factor migration, as largely unnecessary, but entirely harmless. Factor migration in the contemporary globalized world is at an absolute level unprecedented in the past. And it is certainly not without effect. For that reason, the construction of models that provide a more useful account of migration will be an important concern in the following pages.

## 2.2 Problems with Globalization

Imagine that we choose two of the developments that have defined globalization listed above: increased trade, and the liberalization of capital movements. We ask a representative sample of economists to state which of these developments has been of more definite benefit; and which has been the more ambiguous blessing. Without doubt the majority choice would be that increased trade has brought clearer and more widespread benefits than has capital liberalization. Why is this? The answer is a mixture of experience and theory, with the former playing the leading role. There are bad trade stories, and the protectionists have done their best to collect and advertise them. The fact remains that increased trade has, generally speaking, done well for the countries that have opted for it. Looking at the issues more deeply, shows that free trade is not a simple single policy. Opening up to trade is a radical shift in policy, and as such needs to be designed with subtlety and care. When that has been done well, the results are impressive and reassuring. There are losers, there always are with any large change. Without going into the various details of loss and gain for numerous countries, the general picture is that the losers could usually be compensated, and, because of the typical political economy of protection, the poor are more often gainers than losers. This reflects the fact that protection is often put into place to protect the interests of the rich and powerful.

When we turn to capital mobility, the picture is different. The crucial point is that poorly designed schemes for freeing the movement of capital allow of the possibility of modest transitory gains, and serious dangers of instability. This is the story of the bad histories of the freeing of capital movements, as in the case of currency crises, in Asia, and elsewhere. The complexity here is that, in not a few cases, two policy programmes were combined and intertwined. One policy objective is to free up capital movements, allowing capital to seek high returns, and ameliorating the shortage of capital in the home country. Another policy is the control of inflation, to which end the home currency is pegged to a hard currency, usually the US dollar. Now notice a subtle implication of currency pegging. It will always lack complete credibility, if it does not permit some short-term capital transactions across the exchanges.

Suppose that the national authorities say that you can convert the national currency to US dollars, provided that you give twelve months' notice. Currency holders will ask why the authorities do this. If it is because they expect the home currency to be worth fewer US dollars in a year's time, then the exchange-rate pegging is without credibility. If instead the national authorities say that you can convert the national currency to US dollars in twelve months' time, at today's exchange rate, then people will ask what benefit this postponement provides for the government. At best it could help with a short-term liquidity problem. However, governments with reliable and credible policies do not suffer from short-term liquidity problems. So it comes about that the two policies; the free flow of capital, and inflation control, interact to encourage capital inflows, because the lenders believe that they are insured against currency risk. But when that belief evaporates there is a currency crisis. This is badly designed policy, and the Asian currency crises show how it all falls apart when the US dollar itself becomes overvalued, and the currency peg becomes unsustainable and incredible.

These are the experiences that have convinced many economists that capital mobility may be harmful. This is once again an example of a handy term convincing the careless that there must be one thing that corresponds to it: that capital mobility is a unity; you either have it all, or none at all. That is not the case. Consider the case of government borrowing denominated in US dollars. The government can borrow long or short, where short here means short relative to the period over which the returns that will service the debt will arrive. If the government borrows short it will need to renew its borrowing in the future, when it may find that the cost of borrowing is higher, or even that the funds required are unavailable. This is an example of imprudent policy; the prudent course involves the matching of the time structure of returns and liabilities. Similarly it can be argued that pegging the exchange rate to the US dollar is imprudent, because that particular currency may move substantially relative to a trade-weighted basket of world currencies.

As Stiglitz (2005: 254) writes:

> Some critics of the critics of globalization have argued that globalization is inevitable: to resist it is futile. That, I would argue, is the wrong perspective. Globalization may be inevitable, but how countries respond to it is not. They do not, for instance, have to open up their capital markets to short-term capital flows. Different countries have, in fact, responded to globalization in different ways, and some have managed globalization better than others.

Stiglitz does not detail precisely what not opening up capital markets to short-term capital flows entails. It cannot mean, as I have argued, not having any kind of currency peg and convertability. It should mean the avoidance of gross imbalance between returns, and payment, term structures; and that rule should apply to both public and private borrowers. It should also entail choosing a currency peg close to the trade-weighted exposure of the home currency to foreign currencies; as Singapore does, and Thailand failed to do. All these policy strictures are implications of financial prudence. Heavy short-term borrowing because short rates are low is manifestly imprudent.

How do these ideas translate to stock markets? It is sometimes claimed to be an undesirable feature of stock markets that they encourage short-termism, because the investor can buy a stock today and sell it next week (or be evaluated on how it has performed by next week). Keynes employed this type of argument to criticize the use of stock markets for the finance of investment. The truth is considerably more complex than the argument implies. Most stock offerings are close to being qualified perpetuities. Perpetuities means that they will pay (if they do) forever. Qualified means that the sums payable are not, as with a government bond, fixed; rather they will depend upon the performance of the firm. It follows that in most cases the firm is borrowing forever. The lenders, those who buy the stock, may think that it is a liquid investment that they can sell at any time. For a sale, excepting the case where the firm buys its own stock, there must be a buyer. That buyer may see his commitment as transitory; but he too can only sell if there is a buyer. It is almost as if the firm is selling a perpetuity to the market, and the market will hold it forever.

How this all works out depends upon how expectations operate, and how sharp market revaluations affect confidence, perceived risk, and liquidity constraints. In a dramatic instance there is a huge negative confidence shock affecting a certain country. To highlight the issues assume that the national stock market quotes and transacts in the local currency. All stock holders want to sell their stock, there are no buyers, and the market crashes. Individual investors think that they can take money out of the market, but for the collective there is no money to be taken out; any taken out by individuals comes from buyers who are putting money in. So any notion of money fleeing the stock market and going abroad is ill-founded. There may however be asymmetries between the buyers and sellers; perhaps the sellers want to get out of the local currency and into dollars. Why this asymmetry should be large is not clear, but it might exist. In that case the pressure translates to the foreign exchange reserves, as domestic currency taken out of the stock market is sold for dollars, and in this case the other side of the market is the central bank.

The lengthy argument laid out above might be interpreted to imply that a stock market is a potential source of instability, and that wisely designed globalization should avoid establishing stock markets. Unfortunately, the implied potential threat to the foreign exchange reserves is there with any liquid market for non-tradeable assets. It might be land, housing, unquoted companies, cars, or cattle; though the last two might just be tradeable. The true conclusion is that it is impractical in a market economy to maintain a fixed exchange rate after a huge negative confidence shock. If the nation has signalled that the currency peg will always be maintained come what may, then that itself is imprudent policy.

## 2.3 Globalization and the Future

When we are living through history it is extraordinarily difficult to take our bearings. It is too easy to suppose that the current is the eternal; that history has ended. As that view was always grossly incorrect in the past, it may be assumed safely that it will be wrong again. But precisely how it will be wrong is impossible to know. The forces that have powered the move towards globalization described above will not suddenly disappear, but they may generate counter-forces that will make the future take on a different character from the present, and not just be more of the same. We see now signs on a small scale of the kind of effect that may become seriously important in the future. I refer to the increase in the influence of protectionism, both as policy and as sentiment. For the time being these developments are more sad lost opportunities than disasters, and with good fortune so it may remain.

To see how bad things can be, we need only to cast our minds back to 1914. Prior to that year the world had enjoyed over forty years of peace and prosperity. It was, as many commentators note, a world more globalized in many respects than anything seen subsequently, until very recently at least. It is true that the early years of the twentieth century were not without signs of trouble. Anarchists were active, in Russia and elsewhere. The Austro-Hungarian Empire was showing signs of strain, with Serbia particularly troublesome, and Czarist Russia more than willing to meddle in that mess for its own ends. Turkey was the 'sick man of Europe', drowning in its own debt. Rivalry for economic, and also naval, pre-eminence, between Britain and Germany was evident, but could surely not lead to war. The whole of Europe was bound together by economic links, and treaties that promised mutual aid in the case of conflict. How ironic that those same treaties are what tipped Europe into a World War, when one assassination, with Russian involvement suspected, set Austro-Hungary to war with Russia. The War destroyed not only the possibility of optimism for a generation at least. It also killed globalization, led directly to the Russian Revolution, and set the stage for the interwar Depression. The cruel lesson of this history is that if it is 1913, one knows nothing of 1914.

## 2.4 Capital and Labour

Glyn (2006) has the term 'Globalization' in its subtitle. A valuable feature of this book, however, is its lengthy historical perspective, taking in the whole of the post-1945 years. A narrative with that reach serves well to underline the changing picture of the world capitalist economy. Glyn's argument has a Marxist flavour in that he places his emphasis on the relative power of capital against labour. This relation depends simply upon the relative abundance of the two factors. Marx provides a particularly simple version of this balance, based on the assumption of fixed input coefficients for the two factors. The capitalist economy oscillates between two distinct states. If investment is insufficient to employ all the labour supply, then wages crash to some minimal level; profits are high; and investment is high. This is the outcome designated by the term 'the reserve army of the unemployed'.

Eventually the amount of capital invested is so large that capitalists cannot find workers to man their machines; and wages increase to such a high level that profit disappears. This is the classic 'crisis of capitalism'. We can represent the Marxian story by a simple game. At the beginning of any period there are $N$ workers, with $N$ constant, and $M$ investors, who collectively own $K$ units of capacity. One unit of capacity employs one worker, and produces $X$ units of product. At the start of the period each investor chooses how much new capacity to order, which fixes the capacity he will own during the period. For simplicity assume that capacity costs nothing. Choices of capacity levels are uncoordinated, so the investors cannot know what will be the total capacity, and thence the total demand for labour.

The story that the above game represents is Marx with one big difference. For Marx capital accumulation was a mechanical process, driven by the profit available to finance accumulation, and by an almost obsessive need on the part of capitalists to accumulate as far as possible. To the game we can apply the standard rational calculation methodology of received game theory. That is revealing, not because it solves the problem, but because it shows how problematic is any proposed solution. Nash equilibrium is easy to determine. Any choices of capacity levels that in sum employ the total labour force are Nash-equilibrium strategies. So Nash solves one of Marx's problems; there is no need for cycles between profitable capital accumulation and crises.

This solution only creates another problem for the Marxian view. Fundamental to Marx's thinking is his observation that capitalism is not slavery. Workers voluntarily accept employment contracts; so how can they be exploited? If capital and labour are both necessary for production, why can workers not exploit capitalists? When there is a reserve army of unemployed, the question is easily answered. Exploitation of labour flows from its disadvantage on account of its excess abundance. Capital and labour both compete in markets, but with labour in excess, the playing field is not even: labour will always be on the losing side. A crisis of capitalism redresses the balance in labour's favour, but only as a temporary episode.

So much for Nash equilibrium. That solution concept, however, is of uncertain validity in the present context. The game is of the type known as a *coordination game*. There are many Nash solutions, all of which dominate bad outcomes, in which in this instance investors collectively create too much capacity. How are uncoordinated investors to fix upon individual investment levels, without negotiation to a mutually agreed plan?

Schelling (1960) suggests that a natural focus solution will be chosen. In his example, two friends who are to meet in Manhattan, lacking an agreed meeting point, both go to Grand Central Station, where presumably they have met before. How such intuitive solutions could work in a complex changing economy is far from clear. It seems more likely that collective under- and overinvestment will alternate, as with Marx and Glyn.

The fixed coefficient model is special, and unnatural. It is not needed for an account of history in terms of the relative strengths of the positions of capital and labour. A neoclassical model can do the same. When labour is abundant its marginal product is depressed, and when it is scarce the marginal product is high. In world economic history, the 1950s and 1960s represent a period when labour was scarce. The early years of the twenty-first century, given the consequences of opening up world markets to huge supplies of labour, from China, India, and elsewhere, are a time of labour abundance. That would point to low wages and high profitability, and there are indications that this has been happening.

The account just offered is too simple. It has only two factors: capital and labour. A full picture needs more: skilled and unskilled labour, and natural resources. Also, the balancing of the demand and supply for saving is poorly represented in a static production-function model. While some of these deficiencies are taken care of in models presented in following chapters, no economic-model-builder can hope to fully represent the huge complexities of the world. What the reader will find instead are several snapshots that may help to provide an indication of the developing scene.

## 2.5 Capitalism and Globalization

Many of those who voice a dislike of globalization are really expressing a distaste for capitalism. The arguments that are heard are standard items in the prosecution case against the capitalist system. In particular, the accused in the dock, known by various names—capitalism, free markets, liberal economics, and globalization—stands accused of numerous crimes. These include: harming the poor to fatten the rich; despoiling the environment; spiritual emptiness; and gross immorality.

This position is correct, at least to the following extent. Capitalism can never commend itself as a beautiful and noble way for humanity to organize its affairs. It is an ugly system based on selfishness, and it can, and often does, lead to great inequality. The question is not: is this a lovely idea? Rather the issue must be: is there any better way for humanity to organize its economic life? Of course many idealistic alternatives to capitalism have been proposed. So far experience seems to indicate that they do not deliver the goods; using that last phrase in a literal, not a metaphorical, sense.

If capitalism, yes or no, is not a sensible way of addressing our economic predicament, a better idea comes from the realization that capitalism is not one thing. It comes in many varieties, and much benefit can flow from working for the best kind of capitalist system, because the worst kind is truly awful. The same point applies with equal force to the design of the governance of the international economy. To the question free trade, yes or no, the answer will surely be yes. Free trade, however, comes in various forms, and the choice of the best variety can bring substantial rewards.

## 2.6 World Inequality

Long before globalization as we now know it was present in the world, there was great inequality between nations and within nations. In the weakly globalized world that followed the Second World War the inequalities between nations increased to a great extent. There was some convergence among rich nations, but this was more than offset by the gross failure of the poor nations to share in post-war economic growth. Within the rich nations inequality did not alter dramatically; if anything it increased.

The decades after 1980 have seen the world picture change, as described above. Some poor countries are now growing rapidly. Inequality within many nations, poor and rich, is rising; but this increase is not great enough to offset the fall in inequality on a worldwide scale. Dollar (2005), in a generally optimistic account of what has been happening, notes that 'wage inequality is rising world wide'. He means by this that wages have been becoming more unequal between the same category of workers in different nations, and between the same category of worker within nations. One reason for the last effect is the increasing market determination of wage levels. In the past football players mainly received standardized wages, depending upon the division of the player's club, but not apart from that upon the ability of the individual. Now each player has his payments decided by an international auction, with the consequence that a David Beckham or a Zinedine Zidane can command an income ten times at least the level of some of his fellow players. A similar change is to be observed with lawyers, academics in some countries, and in the financial sector. Specialities that always featured great inequality, such as acting and writing, still show at least as much inequality as in the past.

Looking forward to the following chapters, wage inequality provides a good opportunity to take note of the strengths and the weaknesses of economic theory. Increasing inequality between the wages paid for the same work in different countries is an effect on which theory may hope to throw some light, although the problem is certainly complex. The increase in marketization of individual wage levels is to a considerable extent an institutional shift. It could always have happened in the past, but it did not happen, and perhaps it would have been felt to be distasteful and unjust. Now it is with us, and we take it for granted. At the outset, then, we are reminded that economic theory can do much but not everything.

# 3

# Why the Poor Stay Poor

## 3.1 Explaining the Persistence of Poverty

Consider an economic unit, large or small. Examples would include an individual, a family, a region, or a nation. Suppose that at some moment in time the unit is poor. In this context poverty will mean relative poverty; being poor in comparison with comparable units; although we leave open the possibility that absolute poverty may play a role in determining subsequent developments. Allow a suitable period of time to elapse. We expect that a poor unit is likely to remain poor. Beggars may win lotteries, or poor regions discover oil. Yet typically poverty gives rise to poverty. Even so, if the inheritability of poverty is less than perfect, the poor unit may tend to be closer to average wealth or income than was the case with the initial observation.

Friedman (1992) notes that a tendency for poor units to improve their state relative to rich units, as measured say by the growth rate of wealth, does not imply that the variance of wealth tends to decline. Serially uncorrelated random individual shocks would alone explain a tendency for the very poor to improve their positions. Given that sharp variations in wealth within a short time span are rare, notwithstanding the National Lottery, it is certain that most variation in wealth is not the result of transitory shocks.[1] For poor nations the evidence shows little tendency to *convergence* towards the incomes of rich nations. In that sense the poor tend to stay poor. On the variation and stability of earnings, see Atkinson, Bourguignon and Morrisson (1992). The same is probably true of individual family lines. See in this connection Erikson and Goldthorpe (2002).

[1] In fact Friedman looked at income, repeating his familiar argument according to which transitory income varies relatively to permanent income—which is similar to wealth.

Why is this? If a tendency to convergence does not assert itself, is that because there is no reason to expect it; or are there forces pushing in the direction of convergence which somehow get blocked by other important influences?

These are enormous questions, and the complexity of the influences that might need to be taken into account is daunting. Economic theory might help. However, many models can be constructed, sometimes leading to different conclusions. Also, when the interaction between different agents is taken into account, or the consequences of random shocks, especially with the possibility that randomness affects agents' decisions, the models are extremely complicated—sometimes practically insoluble. Economic theory will be useful in pointing out the kind of factors which it is important not to neglect. It is less likely to provide definite answers. For these reasons this chapter is as much a review of modelling techniques and modelling issues as a compendium of completed models, although some of these are included.

Where theory is reticent, empirical investigation may be informative. Unfortunately, data which can be used to investigate questions of long-run convergence is quite scarce. For countries there is data from Maddison, and the Summers-Heston data set. The latter provides comprehensive data prepared with attention to accuracy and comparability for most countries, but only for post-World War II years. For regions, still more for families, accurate intertemporal data, especially panel data, is usually unobtainable. We do know a good deal about the shape of actual income distributions and something of their development over time.[2] Note however that decreasing income inequality, even when that happy outcome applies, is consistent with either weak or strong convergence.

Consider two extreme cases. The incomes of individual agents are drawn from a common distribution, independently by time and agent. The variance of the distribution declines through time, so that inequality tends to decrease. Convergence however is stronger than the decline in overall inequality, as there is regression to the mean. This is Friedman's point again. An agent with a very low drawing at $t$ is highly likely to do better at $t+1$. Alternatively, assume that position in the distribution is fixed once and for all by the initial drawing. Inequality falls over time but there is little convergence.

[2] See, for instance, Atkinson, Rainwater, and Smeeding (1995).

## 3.2 Economic and Non-Economic Influences

There is a question which at first seems to be clear, if not easy to answer. On closer examination it turns out to be not at all precise. To what extent is economic inequality caused by economic factors? The trouble with the question is that the boundary between the economic and the non-economic is not easy to define. In particular, anything which tends to influence economic outcomes is of interest to economists. Again, even if a factor is certainly non-economic in origin, as with race, it is still the task of economics to elucidate how it expresses itself in economic outcomes.[3]

The general point can be illustrated by the consideration of a controversial book. In *The Bell Curve*, Herrnstein and Murray (1994) argue that the fundamental explanation of economic inequality in the US is differences in IQ, and that the transmission of low IQ within the black minority and across generations is the major explanation of the enduring economic disadvantage of blacks and other subgroups in the US population. The arguments and evidence deployed by Herrnstein and Murray will not be examined here. This is a popular book written by authors whose professional standing in the field is quite marginal. Formidable and mainly hostile reviews are provided by Goldberger and Manski (1995) and Kamin (1995). For present purposes, however, the argument serves well as an example of a view directly antithetical to what might be called the 'enlightenment' view. According to that view all human beings, and by extension any countries they inhabit, have the same potential.

Herrnstein and Murray take it for granted that high IQ favours earnings, other things being equal. Certainly other influences are important—top athletes may be no fools, but it is not their brains that account for their exceptionally high earnings. Even focusing on IQ, we are not told what determines the shape of the function that relates IQ to earnings, although the authors argue that the function has been changing shape over time. Notably they claim that technological change has been such that the distance between low-IQ earnings and the mean has been rising over time. Herrnstein and Murray are not economists, and it would be inappropriate to add to the heap of criticism which their psychometric writing has attracted the additional charge of not doing economics. All the same, an economist would be interested in the issues of how labour inputs of various IQ levels cooperate in income generation and how market processes determine the valuation of different IQs; how technological

[3] On the economics of race, see Becker (1971).

design, education, and training can ameliorate a tendency of technical change to be biased against the slow-witted; and what public interventions might be useful. In short, Herrnstein and Murray's economics is far too mechanistic.

It has been seen that a Bell Curve model needs to incorporate some serious economic analysis. Even so, that model is not of the type most commonly analysed by economists. It emphasizes the effects of more-or-less immovable non-economic differences between agents. Agents have to differ somehow, or there can be no economic inequality. However economists have usually preferred to view different agents as fundamentally the same, and to ask how and whether their situations might come closer to each other by way of natural processes of economic adjustment.

If human beings come into the world naked, without the inhibitions of low IQ, or similar disadvantages, what about nations? Do they also enjoy equal potential, or are there countries burdened by low PP—prosperity potential? If there are such innately disadvantaged nations, the explanation cannot be low IQ, as that should average out across a population. In nineteenth-century colonial times many argued shamelessly that some races or cultures are innately inferior. This ran contrary to an earlier eighteenth-century view that saw all humanity as fundamentally the same.

This view of the world is reflected, for instance, in Adam Smith's idea that other nations would eventually catch up with Britain's technological lead. Even so, Smith saw international economic convergence as conditional on the social and political convergence which he hoped would be the result of the spread of enlightened thought. As he puts it:

> Little else is requisite to carry a state to the highest degree of opulence from the lowest barbarism but peace, easy taxes, and a tolerable administration of justice: all the rest being brought about by the natural course of things.[4]
>
> (Lecture 1775)

It is interesting to note that Smith's belief that the 'lowest barbarism' could graduate to opulence, is qualified. It is conditional on what may be broadly described as good government. That view has an optimistic aspect, and a pessimistic one too. It is like being told by a doctor that your disease is curable, but that a cure will require sticking to a difficult regime. That balance of optimism and pessimism did not survive the arrival on

[4] Quoted in Stewart (1793).

the scene of the wildly optimistic Karl Marx. For him economic progress was the unconditional outcome.

> The country that is more developed industrially only shows to the less developed the image of its own future.[5]

Over 200 years later modern approaches echo Smith's view of things by locating the barriers to economic convergence mainly in the sphere of institutions, in particular government. Even so, there are some interesting shifts of emphasis. Thus Barro and McCleary (2003) and Myrdal (1968) take seriously a possible influence of religion on economic development. Adam Smith probably thought that enlightened progress would relegate the influence of religion to the merely marginal. We know now that global history did not develop in that way.

Religion is part of the knotty issue of culture. Even if all human beings come into the world equally naked and essentially alike, they grow up 'clothed' by different cultures. And these cultures may have powerful effects on economic behaviour. Could it be that some cultures induce the existence and replication of poverty? An examination of these questions from a long-term historical perspective is provided by Landes (1998).

Regardless of whether one accepts that all economic units are fundamentally the same, the question remains: will those units which are basically the same eventually come together? Is inequality of similar units simply a matter of differential degrees of adjustment towards a common equilibrium point?

## 3.3 The Kuznets Model

The *locus classicus* of the idea that inequality is related to different rates of economic adjustment is found in the writings of Simon Kuznets. He was concerned with changes in income distribution during the process of economic development. See Kuznets (1955) and (1963). That the Kuznets model is almost absurdly simple is an advantage for present purposes, as it allows us to consider a particular type of approach through a clear and stark example. At the start of development the population share in common the use of pre-modern production techniques. These generate an income distribution with a relatively mild degree of inequality. As development proceeds, individuals or families switch to high-output modern

[5] Quoted by Myrdal (1968: 674).

techniques. This causes inequality to rise, but it falls later as adoption of modern techniques becomes universal. The inequality behind the Kuznets curve is the result of uneven economic development. On the question of why the poor stay poor the Kuznets model is optimistic. The poor do not stay poor, they just take longer to raise their incomes to a high level.

When the Kuznets model is extended to incorporate continuing technical change, it epitomizes one possible view of the poor and why they stay poor. The poor on this view are the slow movers; the last to see the chances provided by change; those whose economic adaptation is sluggish. Such a philosophy raises several questions. Is it true that poverty is particularly associated with conservatism and lack of adaption to economic circumstances? One might think that their straightened circumstances would encourage the poor devote themselves energetically to self-improvement and make them less choosy and fastidious about means than the better off.[6] A leading example of the view that the poor are bad at adapting is provided by demography, where the poor have often been criticized for failing to adapt their reproduction to modern conditions. Yet research lends little support to this idea. A large number of children is frequently in the individual interest of the poor family. There may well be serious problems caused by externalities. If so, the problem is less a failure to adapt on the part of the poor, and more a failure on the part of governments to address externalities and to adapt public policy to modern conditions. See in this connection Dasgupta (1993: chapter 12). In addition, if it is true that poverty is caused by sluggish adaptation to changed circumstances, what makes particular groups bad at adapting? Might it even be poverty itself? Finally, if slow-footedness characterizes some agents in a world repeatedly changing, and frequently experiencing technological shocks, will they eventually catch up the quick-movers, as in the Kuznets account, or might they fall further and further behind?

If the subjects of economic development were all exactly alike in the initial situation, they would develop in the same way. Therefore, as Kuznets recognized (see Kuznets 1973), the roots of uneven economic development lie in some initial differentiation of the population. What are these differences? If they are unchangeable factors there is no guarantee that there will be convergence in the limit, hence no guarantee

[6] In Shaw's play *Pygmalion* Professor Higgins is shocked to discover that Mr. Dolittle would not object to his daughter sharing Higgins's house, in what he, Dolittle, assumes to be immoral circumstances if he were to receive some money. 'Have you no morals?', asks Higgins. 'No guv, can't afford them', is the reply.

that the Kuznets curve will eventually slope downwards. One cannot model uneven economic development, as opposed to simply assuming it, without opening up the possibility that the limit of uneven development will be developed inequality.

The shape of the Kuznets curve and the issue of convergence have to be considered together. They are really different ways of looking at a general Kuznets curve. Convergence focuses on the limit of the curve for high time values; whether that limit is or is not complete equality. Classic Kuznets curve analysis focuses on the middle of the curve; on whether it rises and then falls.

To illustrate the point, suppose that improvement in economic condition is associated with migration. In a state of transition inequality will increase as non-migrants fall behind migrants. Why, in that case, do some migrate and others not do so? In the Harris-Todaro model (Harris and Todaro 1970), migrants and non-migrants enjoy the same expected income, although the outcome for migrants is variable. So the Harris-Todaro model provides another simple and focused example. Agents are all the same, yet purely random effects (who gets a city job and who does not) generate inequality. This is a special case of a wider family of models in which inequality is the result of random shocks. To put it simply, the poor are poor because they are unlucky. Regardless of whether bad luck is the sole cause of poverty, the question of how random shocks affect long-run income distributions is one that must not be neglected, and it will be considered again below.

## 3.4 The Stiglitz Model

In an early and important paper, Stiglitz (1969) considered what would happen in a neoclassical growth model in which the ownership of capital is initially unevenly distributed, all units supply the same amount of labour at any time, and all save the same fraction $s$ of all income. The model is essentially the Solow-Swan model; see Solow (1956 and 1970). The difference is that agents are disaggregated. All agents supply labour at the rate of $e^{\gamma t}$ units per period. An agent might be a family and $\gamma$ the rate of continuous growth of its numbers. There is no human capital in this model. The other component of income is income derived from capital owned. There is a fully efficient capital market, so that all agents earn the same rate of return on capital owned. Then total capital income for any agent is simply proportional to capital owned.

In the Solow-Swan model, aggregate normalized capital will converge to a unique steady-state value $k^*$ which satisfies:

$$s \cdot F\,[k^*, 1] - \gamma \cdot k^* = 0 \tag{1}$$

The factor prices $r$ and $w$ will converge to the corresponding marginal products.

**Theorem 3.1:** *(Stiglitz) Regardless of where they start, all agents will converge to holding capital $k^*$ per head.*

In Stiglitz's model the poor do not stay poor. They converge in the limit to the same level of per capita wealth as every other agent in the model. Consider, for instance, the case in which initial aggregate capital per head of the population is below its long-run steady-state value. In that case only the right-hand side of the Kuznets curve applies. There is an unbroken movement from more inequality to less. The convergence result implies that inequalities in holdings of capital decline over time. If some agents start with more capital than $k^*$ their capital will fall over time, and this again implies convergence.

The distribution of income is always more equal than the distribution of capital because all agents receive the same wage. On this count, the higher the wage rate, the more equal will be the distribution of income. In the case where aggregate capital per head is rising over time, the wage rises, and makes an increasingly important contribution to income, and therefore to equality. The contribution of capital to income distribution, although declining in importance, is also in the direction of equality. If it seems strange to assume that all agents receive the same wage, we may note that part of the capital accumulated by an agent may take the form of human capital. Then total capital accumulation is allocated between human and non-human forms, so as to equalize returns at the margin. In that case all capital can be considered together, as a production-increasing and income-generating unity. And a high wage income which reflects high human capital (as in the case of a doctor's salary) is still income in the Stiglitz model, and still generates saving at the same proportional rate.

## 3.5 Altering Stiglitz's Assumptions

Stiglitz adopted the specification of the Solow-Swan model, relatively new and popular at the time, and examined the consequences of disaggregation with respect to the ownership of capital. How far can we alter those assumptions? One route is to allow random shocks to influence the histories of individual agents. That case will be considered later. Then there is convergence to a distribution, independent of starting point, which is a generalization of the simple notion of convergence. Another radical redefinition of the problem will have agents saving optimally, rather than using an arbitrary fixed saving share. This is considered in Chapter 4 below. Without going that far, the next result slightly weakens a Stiglitz assumption.

**Definition:** *An agent is regular if s is a continous non-decreasing function of k, and:*

$$\frac{\gamma}{s\,(k)} > r \tag{2}$$

Note that if the production function is Cobb-Douglas and initial total capital is small, when $r$ could be arbitrarily large, it must be possible to violate (2). In that case the rate of growth of capital for the agent will be large and the agent will soon become regular. Normally (2) is satisfied in the standard Solow-Swan model and thus in the Stiglitz derivative of that model. There $s(k)$ is a constant independent of $k$. Although $\gamma$ is normally less than $r$, $\frac{\gamma}{s}$ will be considerably greater than $\gamma$, and (2) is a perfectly reasonable requirement.

In the Mathematical Appendix it is shown that for a model where all agents are regular, the system is convergent in aggregate. Individual agents converge together as in the basic Stiglitz model, and all converge to the unique Solow-Swan steady state that applies when the saving rate is endogenous. The full import of this last result will only be fully appreciated when it is seen against an argument of Chapter 4. There it emerges that optimizing agents with correct foresight in a fully integrated markets model will not converge together.

Then the question arises: why does the Stiglitz model give a different result from a model with optimization? Our latest result indicates that the reason is not that Stiglitz assumes constant saving rates. We can obtain the same result with a variable saving rate that responds to the agent's wealth. The crucial feature is that the saving rate in the Stiglitz model is not responsive to factor prices. This is the feature that is inconsistent with optimization.

## 3.6 Convergence and the Discount Rate

To many non-economists the question of why the poor stay poor will seem to admit of an obvious answer. If one starts far down the income distribution, it will be claimed, it is evidently difficult to climb up. Later in this section, and in this chapter and others, an argument will be developed according to which the unwillingness of the poor to save will be shown to be a rational response to the intertemporal substitution possibilities available. That conclusion would be unsurprising if the difference between the poor and richer agents consisted in the returns to saving available to them, with that comparison of returns favouring the rich. The opposite case will be demonstrated: the poor can be rationally unwilling to save themselves rapidly out of poverty, even when they enjoy returns more favourable than those available to the rich. And even if they do save more than the rich, they may not do so to a sufficient extent to give the $\beta$-convergence which is a leading idea of received growth theory.

Take again the claim that if one starts far down the income distribution it is difficult to climb up. One point that presumption neglects to consider is that where self-improvement is concerned, poverty may confer tangible advantages. It may be relatively easier to improve gradually on a very low position than to maintain, leave alone improve on, a high position. That is the feature which powers the neoclassical convergence models.

The convergence theorists look at the history of diverse nations, but the method has more general application. The underlying model has an orthodox Ramsey-Solow production function, common to all units. All agents have the same discount rate. If the discount rates of agents are constant but differ, there will be no convergence, and in the limit all capital will be owned by those agents with the lowest discount rate. That points to another possible reason for enduring poverty. The poor stay poor because they have high discount rates.

While an economist may assume a discount rate to be a constant parameter, it is in principle an endogenous variable. It cannot be trivially obvious why the poor should have a high discount rate. Firstly one has to distinguish between the *utility discount rate*, which is pure impatience, and consumption discounting, which is affected by the rate of growth of consumption over time.

Because the poor are poor the needs of present consumption are pressing. But so too are the needs of future consumption, unless the agent expects to be better off. Therefore one easy explanation of why the poor might have a high rate of discount works only by assuming that they will become less poor in future. That argument works well for the temporarily poor, those visited by low transitory income. In that case the high discount rate cannot be used to explain the persistence of poverty. This argument is concerned with the consumption discount rate, while it is the rate at which utility is discounted which the convergence theorists assume to be common to all agents. So far we have only looked at the possibility that the utility discount rate might vary between individuals, with poor units being those with a high discount rate.

Another possibility is that a high utility discount rate might be endogenous to poverty with the rate depending on current consumption. If so, for convergence with the poor improving their position rapidly one requires that the discount rate be *low* for agents with *low consumption*. Surely those who would wish to make the discount rate depend upon the level of consumption would not envisage the relationship operating in that direction.

## 3.7 Dynamic Inconsistency

There is another issue which is thrown up by the idea of an endogenous discount rate. This will not be consistent with full rationality of intertemporal choices. How can that be? There is no dictatorship where tastes are concerned. However much one may disapprove of a high discount rate, if it is consistently reflected in choices, it as good from the point of view of economic rationality as a preference for pinball over Pushkin.

A deep problem is exposed in a great paper by Strotz (1955). Intertemporal preferences are dynamic in the sense that they have to order not only consumption streams, but also parts of consumption streams. It is in that context that an endogenous discount rate depending upon current consumption leads to difficulties. It violates time consistency, a concept now more familiar in the field of game theory. See in this connection Selten (1975). A rational individual should not regret a decision just because time has passed, with nothing else in the structure of the choice altered. An endogenous discount rate creates just such an outcome, as the following example illustrates.

Table 3.1 shows three series of utility values available to the agent according to the choices made. In effect the agent makes one choice at time 1, whether or not to save a little. If saving is chosen, utility in period 1 is reduced by 5 per cent. The saving choice makes available higher utility

**Table 3.1.** Dynamic Inconsistency

| | 1 | 2 | 3 |
|---|---|---|---|
| I | 95 | 20 | 25 |
| II | 95 | 22 | 22 |
| III | 100 | 18 | 18 |

levels in periods 2 and 3. Exactly what those levels are is governed by a further period 2 saving decision. If the agent saves again in period 2 much of the benefit of the saving only arrives in period 3, as a utility of 25. Otherwise the gain from period 1 saving is a 22 per cent increase in utility above that offered by choice III in each of the periods 2 and 3.

Note an obvious feature of this example. The economic situation of the agent is deteriorating over time. This is important to the argument. Were the agent always poor there would be no possibility of distinguishing between a high discount rate with poverty, and a high discount rate endogenous to poverty. Now suppose that with consumption in the range 99 to 100 the discount factor (the weighting of future utility against current) is approximately .9 per period, while with consumption in the range 20 to 22 it is .5 per period. Then viewed from time 1 the present value of utility for respectively I, II, and III is 133.25, 132.65, and 130.78. In each case these totals are arrived at using weights (1, .9, .81). So the optimal choice for the agent at time 1, and assuming the ability to *precommit* later decisions at the start, is invest and then invest again. I is the best choice from that perspective, which is the viewpoint of high consumption.

The choice to invest in period 1 excludes the history III, but leaves open the choice between I and II. When the agent reaches period 2 the decision whether to invest again remains open. This is shown in Table 3.2. Now poverty makes itself felt in the shape of a far higher discount rate. The agent is now faced with these choices:

**Table 3.2.** The Second-Stage Decision

| | 2 | 3 |
|---|---|---|
| I | 20 | 25 |
| II | 22 | 22 |

Now the discount factor is ·5, hence the present values of the part sequences I and II are respectively 32.5 and 33. So in poverty, II is more attractive and the agent fails to invest at the second opportunity. This is the particular sense in which endogenous discount rates are irrational. Poverty, to borrow a wonderful term used by Strotz, makes the consumer 'spendthrifty'.

## 3.8 The Elasticity of Intertemporal Substitution

There is another variable which has to be brought into the argument to make it complete. This is the *elasticity of intertemporal substitution (the EIS)*, denoted $\sigma$. This variable, and it must be insisted that it is a variable, is defined as:

$$\sigma = -\frac{\frac{dU}{dc}}{c\frac{d^2U}{dc^2}} \tag{3}$$

Note from (3) that because $\frac{d^2U}{dc^2} < 0$ for a concave utility function, $\sigma$ is defined to be a positive number. The size of $\sigma$ measures how powerfully a rise in consumption pushes down the marginal utility of consumption. The larger $\sigma$ is, the more attractive at the margin is intertemporal consumption substitution in the form of consumption postponement. For more on $\sigma$, see the Mathematical Appendix to this chapter.

Clearly then, $\sigma$ is a most important value, and a central question will be: how will $\sigma$ vary with the level of consumption? This is much the same as asking how $\sigma$ will differ between rich and poor agents. Should low values of $\sigma$ be associated with poverty, there will be a strong effect deterring the poor from saving to pull themselves out of their poverty.

This possibility is no mere curiosum. Consider the condition of a poor unit (an individual or a family). Imagine the said unit going to bed hungry each night and struggling to maintain a modicum of dignity in the way it lives. Even in this stark situation provision can be made for the future. If there is no expectation that income will be much better in the future, the only way to generate a small rise in consumption over time is to postpone current consumption and use the small resources released by that action to gain a return by means of which consumption later may be raised.

Is this intertemporal consumption substitution unattractive for the poor because they expect consumption to be higher later, and for this reason marginal utility lower later than it is today? As has been remarked, that cannot be an explanation for the persistence of poverty, as it assumes that poverty will not persist. Might intertemporal consumption substitution be unattractive for the poor because they discount future utility more strongly than do the better off? That is taken here to mean an endogenous discount rate higher for the poor. Frank Ramsey called the discounting of utility 'a failure of the imagination'. So are we to end by saying that the poor suffer from a special kind of feeble-wittedness which weakens particularly their imaginations concerning the future? That type of argument is all too familiar. It involves blaming the poor, however politely, for their poverty. They are too indolent, dim, drunk, whichever feature is at hand to lay on the poor the blame for their own condition.

To reiterate the fundamental point, an unwillingness to save when poor requires no kind of irrationality. While time discounting at variable rates can involve problems of rationality, there is no similar possibility where the curvature of the utility function is at issue. One may dislike others' preferences but they cannot be refuted. So should agents happen to have low values of $\sigma$ at low levels of consumption, that is their business. Given the apparent importance of this question, it is natural to ask what the current growth literature has to say about it. The answer is simple: this issue is nearly always assumed away. Constant elasticity utility functions are commonly employed, so that $\sigma$ becomes a *constant* no matter where it is evaluated. See Barro and Sala-i-Martin (1995: 64).

It is right to feel uncomfortable when economists make simplifying assumptions for analytical convenience. Yet this goes with the territory: everyone does it. The important question is whether too much rides on a particular simplification.

Basing the analysis of growth, convergence, and the explanation of poverty on the assumption that $\sigma$ is a constant involves a huge restriction of the richness of the analysis, and may be seriously misleading. Barro and Sala-i-Martin have argued that their empirical investigations into $\beta$-convergence are based on a firm foundation, in that this is a feature predicted by optimal growth theory. On closer examination, the optimal growth theory concerned turns out to be a house of cards. If $\sigma$ is assumed to be a constant, the result follows. If $\sigma$ is allowed to vary, and to vary in an entirely reasonable manner, there is no similar implication. Note that optimal growth based on capital accumulation within a common technology is not the only way of giving $\beta$-convergence a theoretical basis. A catching-up model, similar to those displayed in Chapter 12, can do the job.

## 3.9 A Variable Elasticity $\sigma$ and the Structure of Preferences

One thing which may have deterred economic theorists from investigating the consequences of a variable level of $\sigma$ is that such a feature should ideally emerge from a general overarching utility function. Then the different values of $\sigma$ would simply express themselves when various budget constraints are presented to the agent. A tight budget constraint (poverty) would yield a low value of $\sigma$ locally; a slack budget constraint (prosperity) would yield a high value of $\sigma$. All cases would represent the same agent with the same underlying preferences, but in different situations. Deriving functional forms that yield such an outcome, and working with them is certainly not impractical. That is shown in the Mathematical Appendix to this chapter. In the Mathematical Appendix to Chapter 4 another result is shown, of great import to the present argument. Once one abandons a constant $\sigma$, a strong structure, such as $\beta$-convergence, cannot be sustained any more. Indeed the opposite picture applies. If $\sigma$ can vary, then more or less anything goes: numerous patterns of growth and wealth are possible, given suitable well-behaved preferences.

## 3.10 Poverty Traps

It sometimes happens that an economic model is useful, not so much because one can believe it, but because it makes clear a set of assumptions at least one of which has to be altered if the dubious implications of the model are to be avoided. The Stiglitz model applied to show economic convergence may be such an example. What do we need to change in the model to get away from its perhaps overoptimistic implication that the poor do not stay poor, but only take longer to become rich?

There are many possibilities. In Chapter 4 below we examine at length the consequence of making saving rates endogenous when, in combination with perfect foresight, one gets non-convergence, for incomes at least. But that model is patently unrealistic. One might also assume different discount rates, but the rationale for that is not clear. In any case, if that is the route taken, one might as well go all the way and claim that the poor stay poor because they have low productivity, or for any other reason. Then the message is disturbing for social reformers, or reassuring for the rich: the poor stay poor because they are of poor quality.

Even if one accepted that idea, a more optimistic gloss could be placed on it. One of the things that may cause the poor to be of poor quality is their poverty itself. If poverty leads to malnutrition, which in turn lowers labour productivity;[7] if poverty makes education inaccessible to the poor;[8] or if the poor receive a low rate of return on their saving;[9] all of these are examples of *poverty traps*. Yet an examination of the Stiglitz model shows that if the poor have a high value for $\frac{w}{k}$, they may have capital per head growing faster than is the case for the rich, even if either $r$ or $w$, or both, are low in their case. It depends on the exact values involved.

Many models which yield a poverty trap generate it by introducing three innovations relative to the disaggregated Solow-Swan model:

- Each agent has an individual production function which describes all or part of the production possibilities available to that agent.
- The individual production functions are non-concave.
- Some kind of capital market imperfection stops individual agents from filling in the non-concave segment of the production function.

Thus the nutrition-productivity relation is necessarily about the individual and is naturally non-concave. Many education and human capital models assume a non-concavity. Bliss (1995*b*) explicitly assumes a non-concave relation; Garcia-Penelosa (1995) posits that a fixed minimum amount of education must be acquired if it is to have market value. Barham *et al.* (1995) assume that the function relating expenditure on education to effective units of labour provided is concave. However their assumption that agents must either spend a whole period on education or receive none at all produces a non-concave return function. These authors also illustrate how capital market imperfections play a role. They assume that in order to finance their education the young can only borrow from their parents.

How satisfactory is it to have non-convergence and poverty traps depending so strongly on non-concave specifications? Some of them look rather ad hoc. As educationalists we may have a vested interest in claiming that a minimum level of education is required to deliver anything of

[7] See Mirrlees (1975); Bliss and Stern (1978); and Dasgupta (1993: chs. 14–16).

[8] See Barham *et al.* (1995); Bliss (1995*b*); and Garcia-Penelosa (1995).

[9] Most authors have stressed the point that the poor pay high rates of interest when they borrow. See Dasgupta (1993: chs. 9 and 9*) and Stiglitz and Weiss (1981). Oddly this carries the implication that the poor should experience a higher rate of return to saving if that saving displaces costly borrowing.

value: but is this really plausible? Even the productivity-nutrition relationship is not always non-concave in the relevant region. These models mostly apply to individuals. The convergence theorists have tended to treat nations as if they were huge individuals, but as a literal specification it cannot be taken seriously. Non-concavities for nations are even more elusive than for individuals. When they specialize appropriately, small nations have sometimes been notably successful.

## 3.11 A Model with Human Capital

The idea that capital might take two forms, standard and human, can be applied to the Stiglitz model considered above. Then the assumption that all agents earn the same wage might not be so absurd as first appears. If that wage is the market return to base unimproved labour; then human capital augments the total wage. It is simpler in that case not to treat those additional labour earnings as wage income, but rather to consolidate them with capital income. That makes sense especially because the constant saving coefficient applies to the aggregate accumulation of capital, standard and human, regardless of the composition of that total.

This last approach is convenient but is only possible under special assumptions. These assumptions do not hold in the model that follows. The key issue is whether the return to capital of all kinds can be expressed as the product of a market return to capital $r$ and the total capital accumulated, whether as standard or human capital. That is not possible when human capital is a personal non-marketable asset and where non-convexity is important.

While many convergence models can easily be treated mathematically, the next model lends itself to diagrammatic treatment. Now allow agents to invest to augment the market value of their labour. To keep things really simple, suppose that saving can be directed at choice to either the accumulation of physical capital or to the accumulation of human capital. Assume also that an individual agent at any time can sell either type of capital held and buy the other type of capital of an equal value without cost. If human and physical capital were just two types of capital, even if one augmented the value of labour proportionately, it would make little difference to the analysis of the previous section. What causes the addition of human capital in the present model to make a large difference is the following assumption.

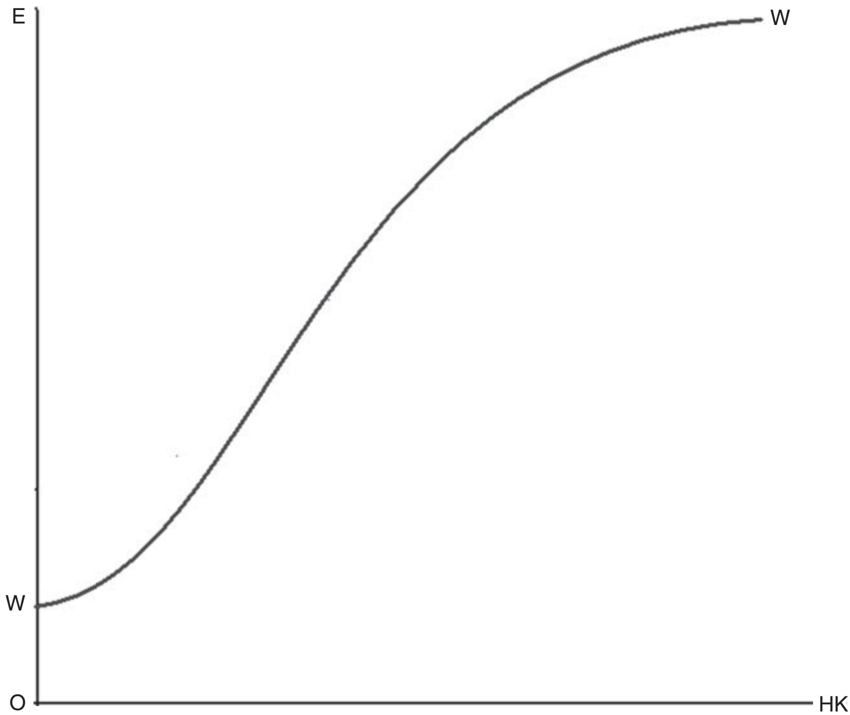



**Figure 3.1.** Human Capital Augments Productivity

**Assumption:** The function which relates that part of an agent's wealth allocated to human capital to the quantity of effective labour supplied is increasing but non-concave.

Figure 3.1 illustrates the type of relation supposed for the present model. The horizontal distance from O measures the quantity of human capital, denoted HK. The vertical distance from O measures the number of efficiency units of labour which the agent can supply, denoted E. The curve WW shows how increasing human capital generates extra efficiency units of labour, above the basic unit which even agents without human capital can supply, shown by the distance OW. Allocating capital in this way is something internal to the agent or the family. One could imagine spending capital on education which results in a non-transferable ability to generate units of labour supply which is sold on the labour market. The production function for this process is non-concave, as the figure shows. For small allocations of capital to augmenting labour supply the return is low. A little education is not a dangerous thing, as the saying has it, but its value is modest. For larger allocations of capital to augmenting labour supply the return improves—this is the sharply rising portion of the curve—then the return falls away again on account of eventual diminishing returns.

This account should be clear where only one individual is concerned. Where single agents are members of families the case may be substantially altered. Should the family own too little capital to send four sons to school, it could educate just one son in return for his undertaking to transfer income to his brothers in future, or to take off his brothers' hands the responsibility of caring for the parents in their old age. That is not to say that such arrangements are simple and proof against defection from his part of the bargain by the favoured son. Surely, however, it is reasonable to assume that intertemporal agreements within a family are more reliable than would be the parallel arrangements through a capital market that we are assuming to be infeasible.

Figure 3.1 shows efficiency units of labour as a function of capital allocated to the human capital category. Given a wage rate for one efficiency unit of labour, a curve differing from Figure 3.1 only by vertical scale shows labour income for the agent. A high wage rate stretches the figure vertically. A low wage rate shrinks it vertically. The positive intercept is explained by the fact that an agent who allocates no part of wealth to accumulating human capital still supplies some quantity of effective labour, taken to be one unit. Even that agent benefits from a higher wage, so the stretching of the vertical axis raises the intercept.

Assume factor prices, given for the present moment of time. Consider a particular agent with total wealth W. This wealth can be translated costlessly between physical and human capital.[10] For this reason the agent has to solve an allocation problem consisting of the decision where to put the wealth. The solution is demonstrated in Figure 3.2.

The curve WW is shaped as in Figure 3.1. However it shows the income flow from human capital, so the vertical height is E multiplied by the wage rate for efficiency labour. The vertical axis in Figure 3.2 measures income as a flow. Physical capital earns a yield of income at a fixed rate $r$. This is shown on Figure 3.2 by a straight line of slope $r$, WA, which again shows income as a function of wealth. Starting from zero wealth, when

[10] This assumption may seem implausible. However during an actual history of accumulation the adjustment will mainly be made at the margin. An exception, as will be seen, is when a switch is made from holding only physical capital to holding entirely human capital. That switch corresponds to something like buying a large amount of education all at once. This is not absurd given the simplified nature of the underlying model.

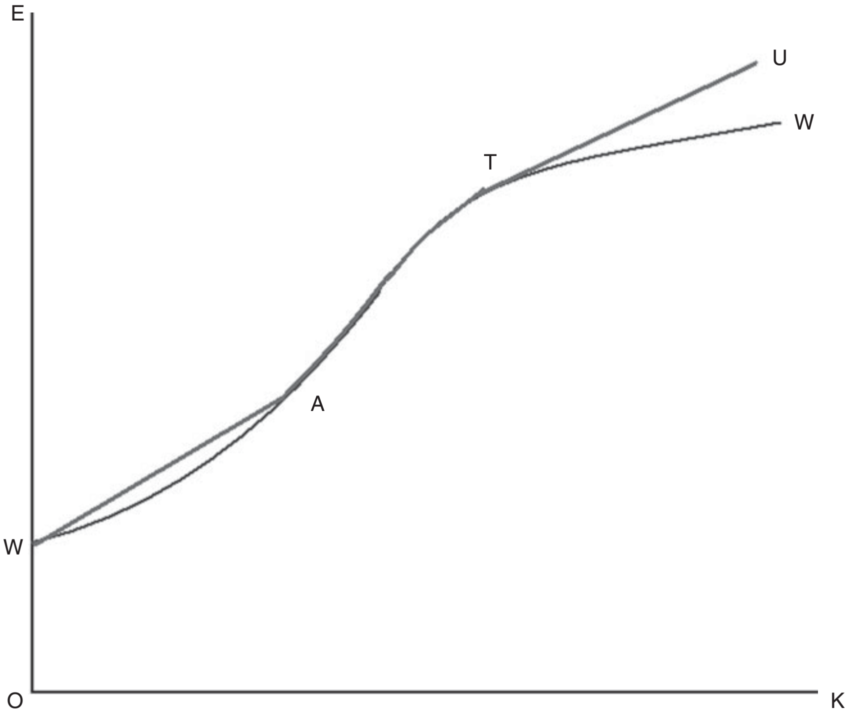



**Figure 3.2.** Optimal Allocation of Human and Physical Capital

the agent's income is OW, income increases most rapidly with wealth when wealth is allocated to physical capital. Income rises with wealth along the line WA. This line is thick in the figure because thickness is used to show the highest income obtainable by allocating wealth optimally.

Immediately beyond A income is higher when all wealth is allocated to human capital. So A indicates a threshold level of wealth at which it becomes optimal to invest entirely in human capital. That continues to be the case along the curve until its slope falls to $r$. That happens at T. Beyond T investment in physical capital does better than further investment in human capital. Therefore income against wealth follows the line TU, where the slope of TU is again $r$. To summarize, the part-linear locus of income against wealth is WATU. Its three sections correspond to different states of specialization and diversification in the distribution of wealth. On WA the agent holds only physical capital. On AT the agent holds only human capital. On TU the agent is diversified, holding both physical and human capital.

The shape of the curve WATU depends critically upon the assumption that there exists no capital market in which the agent can go short, that is borrow wealth against future income. Neither are agents able to save via financial intermediaries. Those could show a return equal to the steepest line connecting W and any point on WATU, which is evidently larger than $r$. The model assumes that the agent is a stand-alone investor who can invest savings in a common capital market investing in physical capital and giving a rate of return $r$. The agent cannot borrow to invest in its own or anyone else's human capital. This same assumption is used by Barro, Mankiw, and Sala-i-Martin (1995) to explain why convergence is slower than a perfect capital market model would predict.

Figure 3.3 shows a snapshot of the dynamic evolution of per capita wealth for a particular agent. It is called a snapshot because it depends on the particular factor prices ruling at the moment to which the figure applies. These themselves are changing with time. If $r$ is falling and $w$ rising, the curve WATU will be stretched vertically while its linear sections become flatter.[11] That can have radical effects, as will be seen shortly.

The non-linear curve shows maximized income $Y$. The variable $k$ on the horizontal axis is per capita wealth. The rate of change of per capita wealth is given by the standard, Solow, formula:

$$\frac{dk}{dt} = sY - \gamma k \tag{4}$$

The direction in which $k$ is changing can be read from the figure by comparing the level of $Y$ and the staight line $OK$. This line passes through the origin and has slope $\frac{\gamma}{s}$. The figure only shows one possibility, but it is an interesting one, as the figure suggests by calling this case the splitting of the middle class.

As with Kuznets's original analysis, we describe economic development in terms of initial conditions which are disturbed by the introduction of new opportunities to augment income. Hence suppose that initially per capita wealth is distributed unequally, such that some agents have per capita wealth corresponding to values everywhere along the horizontal axis. Then accumulation opportunities represented by the curve WATU are introduced. The very poor start to accumulate towards per capita income corresponding to the intersection between WA and OK, marked L on the Figure. Here convergence works much as in the disaggregated

[11] The exact changes are even more complicated. To see exactly what happens one has to go back to Figure 3.2, adjust the curve and lines, then maximize again.

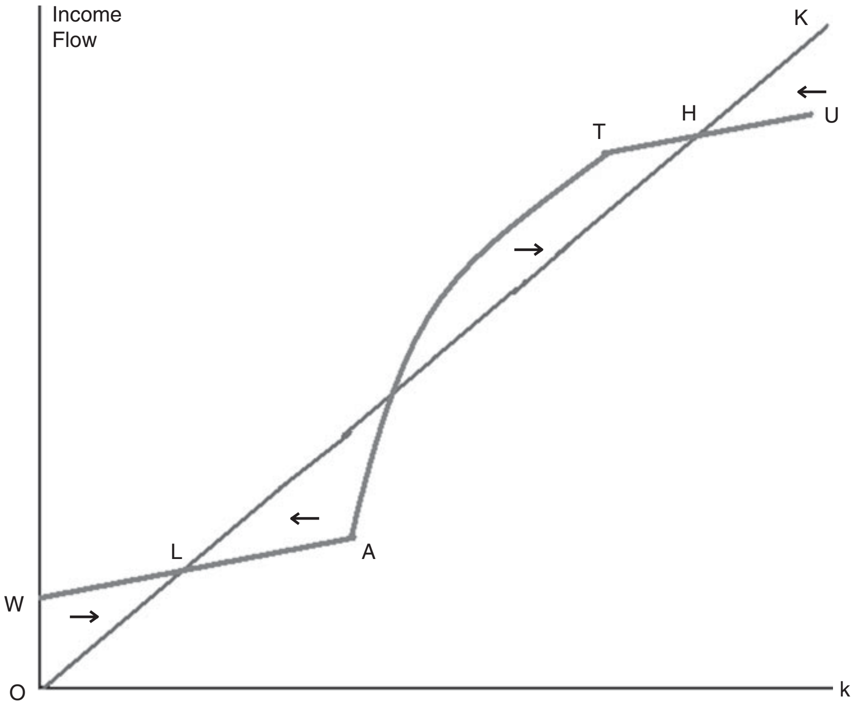



**Figure 3.3.** Dynamics with Human Capital (The Splitting of the Middle Classes)

Solow-Swan model. Those with per capita capital in the range above the low intersection save. Yet they are unable to keep up with growth in numbers and per capita wealth falls away towards L. These agents are middle class in the following sense. They are above the poorest in terms of wealth. They may or may not have enough wealth to accumulate some human capital. Even if they have passed the threshold of wealth to make human capital accumulation optimal, however, (and even A can be associated with a positive holding of human capital) they lack enough wealth to keep up with growth in numbers and they are involved in an inescapable process of decline in per capita wealth. If factor prices did not change, then, in the limit, the declining middle class would meet the rising poor at L. The description of this outcome provides excellent scope for the novelist. From the analytical point of view it is clear.

Agents with per capita wealth to the right of M—the upper middle class—are in a completely different situation. They have enough wealth to finance positive per capita accumulation with the same savings share as everyone else. Notice that these individuals must have human capital. Their wealth will grow towards H, where, if factor prices did not change they would join the declining super-wealthy who started with wealth to the right of H but are unable to keep up in per capita terms. The striking feature of this very stylized model is that over a substantial range, the rich get richer and the poor get poorer. However the very poor benefit from economic development. And for them trickle-down works: they benefit particularly from a wage rate which economic development causes to rise.

## 3.12 Initial Conditions and Subsequent Developments

As has been made clear, Figure 3.3 strictly cannot be used to determine the long-run development of wealth holdings. This is because over time factor prices, and hence the curve WATU, change. Even so, some points are evident from the figure if we imagine that factor prices do not change much, or from other considerations. If initially average wealth is low, and its frequency distribution shows a low dispersion of wealth holdings, as assumed in both cases by Kuznets, then all the action takes place around a point like L in the figure. We have convergence. Saving will carry hardly anyone up to wealth levels where they can accumulate human capital. If however it happens that aggregate capital is accumulating considerably, perhaps because in the initial position most were very poor, the effect of that accumulation on the structure of Figure 3.3 may be dramatic. As OW rises, and WA and TU flatten, the low-level equilibrium at L may disappear. Subsequently all will converge to a point like H. It will be the Stiglitz story again.

In detail this account can support a Kuznets curve with eventual convergence. Yet consider what happens during the dynamic process of development. Agents switch from holding no human capital to holding a complete minimal quantity. At the precise moment when that happens there is no discontinuity in either income or wealth, because agents switch at precisely that point at which the two forms of wealth holding show equal returns. Even so, as may be seen from Figure 3.2, the switch in the composition of capital causes a discontinuous shift in the shape of the production function local to the agent concerned. And that may cause divergences in the subsequent accumulation patterns of different agents. Specifically, rich agents, who switch capital composition early, may start to accumulate wealth more rapidly following the switch. This will result in marked but transitory divergences of per capita income. That is consistent with a Kuznets curve, rising inequality followed by falling inequality.

While positive saving guarantees that wealth accumulation will always be positive, this does not entail that per capita wealth will rise; neither in aggregate, nor for particular agent types. Even for total physical capital, stationarity if not decumulation is in principle possible. This is shown by examining a case in which a large mass of agents has reached wealth holdings close to or below the point A in Figure 3.2. They all switch into human capital, and thus create a large volume of sales for physical capital. These sales soak up the demands for physical capital generated by less wealthy agents who are still saving into physical capital. As a result there is no aggregate accumulation for physical capital. In that case, per capita physical capital must be falling.

### 3.12.1 *Long-run Dynamics*

Although the discussion of Figure 3.3 above has trespassed into the consideration of development of wealth into the long run, strictly one should never forget that the figure shows no more than a snapshot, good only for given factor prices. The figure indicates in which direction wealth accumulation is taking agents from various starting points given factor prices. What those factor prices will be will depend upon the frequency distribution of wealth. That is given by history. Those factor prices in turn will develop over time. How that will happen is discussed further below. However if economic development is to imply an aggregate accumulation of per capita wealth, it must be the case that the predominate weight of wealth endowments along the horizontal axis of Figure 3.3 is in regions corresponding to rightward movement.

As the argument of the previous section has made clear, there is no guarantee that aggregate per capita wealth accumulation will correspond to the aggregate per capita accumulation of physical capital. Hence there is no certainty of a rising wage rate and of a falling rate of return to physical capital. Note however that a rising wage rate constitutes a powerful force for convergence. In the Stiglitz model of Section 3.4, the existence of a positive wage rate is most helpful to convergence. When the wage rate rises through time, that accelerates the tendency to convergence. When the wage rate can fall, convergence is less certain, especially at particular times.

### 3.12.2 *The Diamond Capital Model*

Diamond (1965) put capital into the overlapping generations model (OLG model) invented by Allais and Samuelson. A fine exposition and analysis of the model in its many ramifications can be found in De La Croix and Michel (2002). In the basic case the consumer lives for two periods. In the first period of her life she supplies 1 unit of labour inelastically and earns the wage rate corresponding to the marginal product of the capital which the previous generation saved for its retirement. She may also save part of her wage and this becomes the capital saved until the next period. Population grows at rate $\alpha$, so that each generation is $(1 + \alpha)$ the size of the previous generation; $0 < \alpha$. The consumer maximizes lifetime utility. In line with nearly all treatments of this model, we assume utility to be additively separable. Thus the utility of a consumer born in period *t* is:

$$U\left(c_t^t\right) + \delta U\left(c_{t+1}^t\right) \tag{5}$$

with $0 < \delta < 1$.

The production function gives gross output as a function of gross capital. The budget constraint of the consumer born at *t* is:

$$c_t^t + \frac{1}{1 + r_t} c_{t+1}^t \leq w_t \tag{6}$$

where, for consumption, superscripts denote the generation which consumes the quantity concerned, and subscripts denote the period when it is consumed. The factor prices $r_t$ and $w_t$ in (6) are equal to the respective marginal products.

$$w_t = F_2\left[k_{t-1}, 1\right] \tag{7}$$

$$1 + r_t = F_1\left[k_t, 1 + \alpha\right] \tag{8}$$

For capital, subscripts denote the generation which invested that capital. Note that the marginal product of capital gross corresponds to 1 plus the rate of interest. Also, when maximizing, an atomistic individual saver treats $r_t$ as a constant and this fact is incorporated in the demand function for first-period consumption. Subscripts denote partial differentiation.

The demand function for first-period consumption is:

$$c_t^t = c^1\left(w_t, \frac{1}{1 + r_t}\right) \tag{9}$$

which must be in conformity with the budget-constraint condition:

$$w_t(k_{t-1}) = c^1\left(w_t(k_{t-1}), \frac{1}{1+r_t(k_t)}\right) + k_t. \tag{10}$$

In equation (10) the dependence of $w_t$ on $k_{t-1}$, and $r_t$ on $k_t$, is shown by placing those variables in brackets after these factor prices. Equation (10) defines a non-linear first-order difference equation in $k_t$. First-period consumption is assumed to be a normal good; i.e.

$$0 < \frac{\partial c^1}{\partial w_t}\left(w_t, \frac{1}{1+r_t}\right) < 1 \tag{11}$$

which says that if the wage income of the consumer increases, the proportion of that additional income spent on first period consumption is positive and less than 100 per cent.

A fundamental theorem for the Diamond Model says that $k_t$ increases with $k_{t-1}$. Figure 3.4 illustrates. Multiple equilibria have always fascinated economists. They bring sophistication and subtlety to economic analysis and allow the theorist to tease the intuitions of simple-minded colleagues. In that guise they may represent no more than technical sleight-of-hand. Often the truth is that economic theory is either straightforward and simple or it is not much use. Multiple equilibrium in the Diamond capital model is not of this character. It is not just a technical tease. Rather it embodies the theoretical realization of a vital and simple idea.

Consider two stable steady-state equilibria of the model, one with a low level of capital, the other with a far higher level of capital. The model can support poverty or prosperity, each as long-run equilibria. How can that happen? Plainly the model specification is identical in each case; these are two solutions to the same model. Look at the individuals in the low-capital equilibrium: why are they and their successors poor? It is not their tastes or technology that explains their state, for these influences are common to all cases. The answer is clear. In the poverty solution capital is low, for which reason the wage rate is low. In the Diamond capital model the wage rate decides everything; it fully determines the level of saving and hence capital next period, from which the wage-rate net period follows.

Thus the poverty in a low-capital equilibrium is a consequence of low capital. We have one simple example of a powerful and widespread effect. *The cause of poverty is poverty itself*. Or, to express the point in language which has already been employed freely above, the poor are caught in

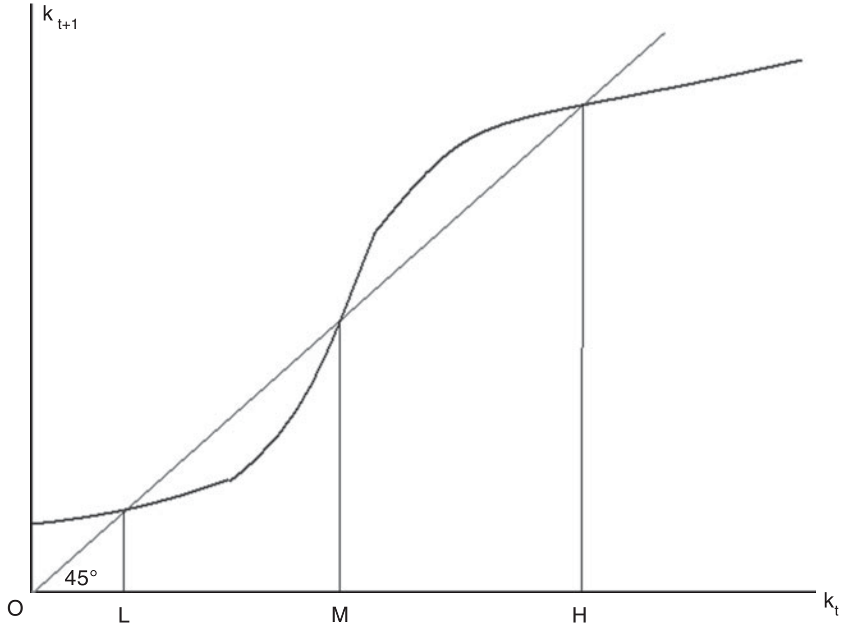



**Figure 3.4.** Multiple Equilibria in the Diamond Model

a poverty trap. The difference here is that the poverty-trap outcome does not depend upon any non-convexity or externality. Neither does it depend upon imperfect information, as in the Harlem resident's job-seeking problem explained in Chapter 1. The idea that the poor are poor because they are poor is one that will come as a surprise to few visitors to the world's poor regions. In particular a gross waste of human resources is one of the most depressing features of many poor countries. One meets frequently people whom one knows would flourish in a society which provided better opportunities, including the employment openings, education, and health care, who now live out a low-level existence and return low productivity.

It seems then that the Diamond capital model could offer the perfect theoretical realization of the poverty-trap concept. In fact, however, the multiple equilibrium possibility has never attracted much theoretical interest. It is explained in every textbook that treats the model; yet this case is never developed. It has become a footnote point. Why is this? Probably a major explanation for the unpopularity of the multiple solution case is that it is seldom realized in combination with two highly appealing features:

- stability of steady-state solutions of interest
- simple standard functional forms

It is most straightforward to produce multiple steady states when one of these is the so-called *corner steady state* (see De La Croix and Michel (2002: 28). This is the case in which the economy has a zero-capital no-activity equilibrium which is locally stable. Usually this case is judged to be uninteresting as there are no examples of zero capital economies in the world, or any such have disappeared and lost their populations. Perhaps the Empty Quarter of Saudi Arabia is a degenerate corner solution economy. If so, the theory of its non-activity is not challenging, and in fact it cannot conceivably be seen as a particular solution to a general case embracing otherwise like economies. Rather it is an instance in which climate certainly accounts for poverty and low growth. More interesting is to note that a small modification of the Diamond model can convert any stable corner steady-state solution to a low-level positive-income stable steady-state trap. Simply modify the budget-constraint equation (6) to make it:

$$c_t^t + \frac{1}{1+r_t} c_{t+1}^t \leq w_t + w_0 \tag{12}$$

where $w_0$ is the income which an agent can obtain from marginal activities outside the formal economy. Thus $w_0$ might be the value of vegetables grown at home with no capital input. Then $k_{t-1} = k_t = 0$ is no longer a solution to the model, and $k_{t-1} = 0$ implies $k_t > 0$. It is possible with simple functional forms to obtain multiple stable steady-state solutions with simple functional forms, but these cases are a minority of all the cases concerned.

Probably the lack of multiple solutions in the Diamond model with simple functional forms explains why this possibility has not received a great deal of attention from economic theorists. Every textbook notes the case concerned and then passes on. It seems that the position here is exactly parallel to the problem with $\beta$-convergence examined above, in the following sense. Even the CES functions examined by De La Croix and Michel are quite special. If we assume a Cobb-Douglas production function, often believed to guarantee a unique solution to the model, but allow a complete freedom for the choice of the utility function, subject only to standard continuity and convexity properties, then what is possible?

The following result is shown in Bliss (2008). If one allows the computer to draw the utility function; that is to solve a differential equation that generates a function with the required property, then a Diamond model, with a Cobb-Douglas production, can be created that has not one, not two, but an uncountable infinity of solutions. In fact every level of $k$ on an interval defines a steady-state equilibrium. Once one has that result, then as many isolated stable equilibria as required can be obtained by perturbing the model. This is achieved with a utility function for which the elasticity $\sigma$ increases continuously with the level of consumption. So the Diamond model, quite contrary to the impression left by De La Croix and Michel (2002), is well able to model poverty traps via multiple stable solutions.

## 3.13 Which Model: Diamond or Ramsey?

The argument of the present chapter has taken us down some highly theoretical paths. Theory, however, is not the ultimate concern of the investigations. For explaining poverty and its changes over time the crucial issue is which model comes closest to reality, in the sense that it best mirrors the major factors that make themselves felt in actual cases. Of course any simple model can only offer a faint image of the many-sided features of a complex world. So we are forced with only primitive tools in our hands to design models that get the essentials right.

The Diamond capital model and the Ramsey optimal saving model give strongly different results. For instance, the former can generate a steady-state poverty trap, while the latter cannot do that. For this reason it is important to ask which of the two models is essentially correct, in the sense that it gets to the heart of the economic growth story. In particular we have to know which model fits best the condition of poor countries, or poor individual agents, because it is for these that the poverty-trap model is most apposite.

It might seem that the Diamond model is most appropriate for a rich country with a well-developed system of private pensions. Then first-period saving could be interpreted as the purchase of units in a national capital trust which invests its funds in the real capital accumulated by private firms. This is not the only interpretation available. Thus in many poor countries parents can invest in human or physical capital for the use of their children. Buying land for sons or educating any children would be cases in point. Often such saving will be part of an implicit contract which requires children to care for their parents in their old age. Of course it is far from clear that the return on this intergenerational transfer would be the same as a market-determined rate. Equally the exact quantitative level of childrens' 'repayment' of their parents' favours may not be critical for the type of features shown above.

## 3.14 Implications for Policy

A Solow-Swan-style model does not support the Kuznets view of the development of inequality over time. There is an inevitable and monotonic decrease in inequality as all agents converge to the same limiting income per head. Then with integrated capital markets it is shown in Chapter 4 below that the agents cannot be fully optimal rational-expectations savers. Yet real-life agents are surely not rational-expectations optimal savers. When we introduce the non-concave specifications discussed above, there is no necessary tendency to convergence, not even for mechanically saving agents. There can be a low-wealth trap from which agents cannot save enough to escape. Also, and independently, even when all agents converge to the same limiting state, income distribution may not behave monotonically.

The idea that convergence is a natural tendency represents quite a complacent view of development and distribution. It tells the poor that time and the right policy regime will solve their problems—eventually. No income transfers are required. Yet even in a rational-expectations model the consequences of a low income may endure significantly forever. See Chapter 4 below for further discussion. Even when that is not the case, convergence could take a long time, in which case the inequality that persists while one waits for convergence to complete itself will surely cause concern. And if there are important non-concavities, some units may never reach a high level of wealth. Policy might well attempt to help individuals to surmount some of the non-convexities, and certainly avoid creating more of them. For instance, educational systems should be designed to be flexible and accessible to the poor. And, ideally, social-security systems should not create non-concave return functions which trap the poor in their poverty; although designing fully acceptable systems to avoid that problem is well-nigh impossible.

Perhaps the most important message would be that there is no economic immune system which fights back poverty by itself. Convergence is not guaranteed and the poor may well stay poor indefinitely. They may suffer forever from the consequences of starting with low wealth. They may get caught in poverty traps caused by non-concavities, including those created by policy. And even if convergence works it may take too long to dispose of the acute inequalities which characterize our world of nations or our nation of families.

Most of the argument of this chapter has been conducted as a discourse concerning the poverty, or the escape from poverty, of micro units—individuals or families. It is then a small step, or perhaps a huge unwarranted leap, to treat larger units, such as regions or nations, as being much the same as micro units writ large. The nation is then seen as a vehicle steered by a giant representative agent, who optimizes against constraints just as any person or family.

There are two leading problems with this manner of modelling macro poverty. First if the representative agent is an aggregation of many individuals, how is that aggregation to be effected and can it be justified? An aggregation exercise either depends upon impossibly restrictive conditions, which typically require the units aggregated to be essentially the same, or it is a serious oversimplification. A second difficulty arises because assumptions that are reasonable for the micro case may be less so for a macro application. For example a non-convexity that is huge for a small agent may be insignificant for a nation. On the other hand nations may face large problems, such as environmental externalities, that are beyond the control of the small agent, and hence outside his planning and decisions.

The truth is that we have not yet attempted to construct a serious model of the nation. The nearest to that concept would be the Stiglitz model, which does not aggregate at all, and has no representative agent. That model only yields useful results because all its agents are precisely the same, except that they start with different levels of wealth. Also the field in that case is a set of agents sharing a unified capital market. That could be more or less than a nation. Below, especially in Chapters 7 and 8, more progress will be made towards depicting something that really resembles a nation.

## Mathematical Appendix 3.1

### A Generalization of the Stiglitz Result

Recall that an agent is *regular* if $s$ is a continous non-decreasing function of $k$, and:

$$\frac{\gamma}{s(k)} > r \qquad \text{(A3.1)}$$

Note that if the production function is Cobb-Douglas and initial total capital is small, when $r$ could be arbitrarily large, it must be possible to violate (A3.1). In that case the rate of growth of capital for the agent will be large and the agent will soon become regular.

**Theorem 3.2:** *If all agents are regular, factor prices and per capita capital holdings converge to their steady-state Solow-Swan values, $(w^*, r^*)$ for factor prices and $k^*$ for per capita capital holding.*

**Proof:** The value of $\frac{dk}{dt}$ for an agent is given by:

$$\frac{dk}{dt} = s(k)\,[w + rk] - \gamma k \tag{A3.2}$$

where $k$ is the per capita capital holding of the agent. The sign of $\frac{dk}{dt}$ in (A3.2) is the same as the sign of:

$$[w + rk] - \gamma \frac{k}{s(k)} \tag{A3.3}$$

Define $\widetilde{k}$ at any time as the solution to:

$$\left[w + r\widetilde{k}\right] = \gamma \frac{\widetilde{k}}{s(\widetilde{k})} \tag{A3.4}$$

Note that $\widetilde{k}$ defined by (A3.4) is unique. The left-hand side of (A3.4) increases at rate $r$. The right-hand side is an increasing concave function of $\widetilde{k}$. Its slope with respect to $\widetilde{k}$ is $\frac{\gamma}{s(\widetilde{k})}$ which is greater than $r$ by the regularity property. Therefore the left-hand side of (A3.4) is positive at $\widetilde{k} = 0$ while the right-hand side is zero at $\widetilde{k} = 0$. The right-hand side increases more rapidly, and only one intersection is possible.

The solution to (A3.4) varies with time because $w$ and $r$ vary with time. Then $\frac{dk}{dt}$ is a function of $k$ for the particular factor prices that rule at a moment of time. Over time this relation will shift around as factor prices change in response to alterations in aggregate capital. All such movements in the curve are continuous.

Let $D$ be defined as:

$$D = \left(k^{\max} - k^{\min}\right)^2 \tag{A3.5}$$

where the two $k$ values in (A3.5) are respectively the max of $k$ across all agents, and the min of $k$ across all agents. The identity of these two agents never changes, as it is not possible for an agent to catch up another in finite time.

Then:

$$\frac{dD}{dt} = 2\left(k^{\max} - k^{\min}\right)\left(\frac{dk^{\max}}{dt} - \frac{dk^{\min}}{dt}\right) \tag{A3.6}$$

The signs of the two terms in curly brackets will be opposite. Therefore $D$ is non-negative and $\frac{dD}{dt} < 0$ whenever the system of many agents has not converged as required by the theorem. $D$ is a Lyapunov norm and all the $k$s $\longrightarrow k^*$. This implies $(w, r) \longrightarrow (w^*, r^*)$ and the proof is complete. □

# 4

# Convergence in Theory and Practice

## 4.1 Introduction

In Chapter 3 we have already encountered the question of whether the incomes of distinct infinitely lived agents will converge together. This issue can be considered from either a theoretical or an empirical angle. For example, the Stiglitz model that was examined in Chapter 3 represents the theoretical approach. In fact the boundary between theory and the empirical is never absolute and not always obvious. So the striking theoretical findings of Stiglitz force us to view the empirical evidence in a new light. If persistent inequality is obvious and self-evident, then there is no need to seek out its major causes. But if that same persistent inequality is contrary to what theory would predict, as the simple Stiglitz model indicates, then the empirical evidence has to be viewed and analysed in a different way. This leads on to a major idea that is visited in the present chapter, *conditional convergence*. Conditional convergence means that while there is an underlying tendency to convergence, this is moderated by the influence of other important factors.

## 4.2 Convergence Empirics

### 4.2.1 *An Overview of the Findings*

Cross-section convergence empirics began with the work of Baumol (1986). He looked at what is now called $\beta$-convergence for the OECD countries for the period 1870–1979. De Long (1988) advanced the criticism that because membership of the OECD is endogenous to successful economic growth, the Baumol study is subject to sampling bias. To illustrate this point imagine that one had to select using information available in 1960 the countries likely to be rich in 1979. It is far from obvious that Japan would be included in the sample in that case. This argument begins a long line of writing critical of cross-section studies of economic growth. Today cross-section empirics are often viewed with considerable scepticism by the profession. I will argue that some of this criticism is misplaced. It is true that these studies are subject to many and serious problems. Yet the largest difficulties at issue are encountered in almost any multivariable regression analysis. It is not fair to single out the cross-section growth analysts for disdain or ridicule, when their problems are those of the greater part of applied economics.

Since Baumol, growth empirics have moved on. In particular the sampling bias issue has been largely subvented by the use of near-comprehensive data sets. The work of Barro and Sala-i-Martin, see Barro and Sala-i-Martin (1992) and the 1995 book by these authors is the most thorough exposition. The data employed is from Summers-Heston data set augmented by additional variables. See also Mankiw, Romer, and Weil (1992) where the cross-section method is used to assess how far the Solow-Swan model can account for differences in growth rates. Durlauf and Quah (1999) provides a comprehensive review. See also Easterley (2002) who examines attempts to explain economic growth rates from the standpoint of a sceptical but engaged observer.

The results of large-sample cross-section growth studies will be called for convenience the Barro regressions. This convenient term is not intended to detract from the great contributions of other researchers, Barro's co-worker Sala-i-Martin, or Quah, to name but two. Barro shows that in a sample of 117 countries for the years 1960–85 initial per capita income and the growth rate of per capita income over that period are essentially uncorrelated. However the analysis is extended to embrace an equation of the form:

$$y = \alpha + \beta x + \sum_i \gamma^i z^i + \epsilon \tag{1}$$

where $y$ is per capita economic growth, $x$ is log initial income, the $z$ values are additional likely explanations of economic growth, called here conditioning variables, $\epsilon$ is a random error, and $\alpha$, $\beta$, and the $\gamma$ values are constants.[1] In this extended regression analysis it is found that $\beta$ is significantly negative. This is $\beta$-convergence conditional on the $z$ variables being constant.

[1] Lagged values of conditioning variables are used to moderate the problem of right-hand-variable endogeneity. This important point need not concern us here.

### 4.2.2 *Interpreting Correlation*

Start with the single-variable regression of growth rate of per capita income 1960 to 1985 on the logarithm of per capita income in 1960. The former variable is denominated *growth*, the latter *LnPCI60*. This is also the starting point for Barro and Sala-i-Martin (1995), who note that the two variables are only just significantly correlated, and that correlation is positive. The correlation coefficient between *growth* and *LnPCI60* for 117 countries is .227. The regression for these countries is shown in Table 4.1.

**Table 4.1.** A Simple Growth Regression

| Variable | Coefficient | t-value |
|---|---|---|
| *constant* | −.0135 | −.998 |
| *LnPCI60* | .0046 | 2.50 |
| $N = 117$ | $F = 6.245$ | |

This is a good moment to pause and consider what a regression exercise like that reported in Table 4.1 means. A non-zero coefficient for a single regression variable depends upon that variable being correlated with the dependent variable. We remind our graduate students that correlation is no proof of causation. In fact the relationship between correlation and causation is even more difficult and perilous than this simple slogan indicates. One can equally say that the absence of correlation is no proof of the absence of causation. It will be seen shortly how looking inside growth regressions perfectly illustrates that last point.

A correlation between two variables usually indicates something significant, especially if the sample is large. Imagine that one assembled one thousand ordered data sets, each comprising one hundred observations, and these numbers consisted of strange and seemingly unconnected values. They might be three digits taken from a given position in randomly selected telephone numbers; or the last two digits of stock-market quotations; or the running times of 100 horses selected from various races. We then have nearly half a million pairs of these variables and between one pair at least there is bound to be a correlation which viewed by itself would seem to be of overwhelming statistical significance. This weird illustration shows what ought to be meant by the much-abused term *spurious correlation*.

Most correlations encountered in practical empirical work are not spurious in the sense that they have arisen entirely by haphazard chance. But they may well not be due to a simple causal connection between the two variables involved. Many correlations arise wholly or in part because the variables related are not stationary. Beginning econometrics, students are delighted by the fine result they obtain when they regress consumption on income. The high correlation between the two series guarantees a high $R^2$ and a hugely significant regression coefficient. Of course it is reasonable to believe that income is a major causal determinant of consumption. Other correlations between non-stationary variables however may lack any causal element. It could be that a regression of arrests for drunkenness on teachers' salaries will show a positive coefficient and a high $R^2$. If so there is plainly no proof that paying teachers more causes drunkenness. Equally the correlation is not spurious: it does not arise simply by chance.

Examples of non-stationary series fall under the more general heading of the missing variable. For non-stationary time series that variable is time itself. Because both series are trended they are correlated. Missing variables are a major headache for the applied economist. A missing variable here is a variable that plays a part in the data-generating process (DGP) but which has not been included in the regression analysis. Missing variables cause problems when they are correlated with included variables. In that case the regression coefficients on included variables are biased estimates of the true linear relation in the DGP. We encounter just such a case with the cross-section growth regression. A simple regression of growth on log initial income shows a near zero coefficient. Nevertheless multivariable $\beta$-convergence estimation depends on the presumption that the simple estimate is biased. It is biased by the omission of the conditioning variables $z$.

A regression for 73 countries from the Barro-Lee data set follows. The right-hand variables and their means and standard deviations, together with the same values for the growth rate, are given in Table 4.2.

A regression of these variables on the growth rate gives the numbers shown in Table 4.3.

The inclusion of five extra variables in the regression reported in Table 4.3 transforms the position with regard to the relationship between $LnPCI60$ and $growth$, whether its basis be causal or not. Now the coefficient on $LnPCI60$ is significantly negative. For that reason the hypothesis of a tendency to $\beta$-convergence is not rejected in this sample. All the words in this statement count. The multivariable regression does not show that $\beta$-convergence is what happens generally. The simple regression of $LnPCI60$ on growth, reported in Table 4.1, shows that it does not. Rather the multivariable regression does not reject the hypothesis that

**Table 4.2.** Regression Variables

| Variable | Definition | Mean | Standard deviation |
|---|---|---|---|
| growth* | Growth rate PCI 1960–85 | .0226 | .0161 |
| lnPCI60* | Log of PCI in 1960 | 7.5201 | .8930 |
| bmp** | Forex black market premium | .1188 | .1675 |
| govsh4** | Government 'consumption' over GDP | .1571 | .0656 |
| geerec** | Public expenditure on education over GDP | .0245 | .0103 |
| I/Y* | Investment to GDP ratio | .0968 | .1893 |
| pinstab** | Measure of political instability | .1916 | .0859 |

* *Source*: Barro and Sala-i-Martin (1995).
** *Source*: Barro-Lee data set.

other things being equal (though they never are), a low initial income is associated with rapid growth.

For present purposes the multiplicity of the right-hand variables is unimportant. What is significant is to understand how the inclusion of any additional variable or variables can transform a coefficient from a value close to zero to a significantly negative value. It assists the following discussion to consider a regression in which to $LnPCI60$ only one further variable is added, The variable selected is the most successful in the multiple-variable regression, $I/Y$. It is true that the investment share is suspect as a root cause of economic growth, as it is likely to be jointly determined with economic growth by other factors. At this point in the argument that issue is not material.

The regression results are shown in Table 4.4 below.

Even the addition of one extra right-hand variable, in this case $I/Y$, changes the coefficient on $LnPCI60$ from a significant positive value ·0046 (Table 4.1) to an equally significant negative value −.0048.

**Table 4.3.** Regression with many Variables

| Variable | Coefficient | t-value | Partial R-squared |
|---|---|---|---|
| constant | .0698 | 3.83 | .1821 |
| LnPCI60 | −.01133 | −3.89 | .1863 |
| bmp | .0035 | .345 | .0018 |
| govsh4 | −.0419 | −1.66 | .0400 |
| geerec | .4922 | 2.71 | .0999 |
| pinstab | .0003 | .029 | .000 |
| I/Y | .1673 | 6.02 | .3545 |
| $N = 73$ | $F = 8.326$ | $R^2 = .4308$ | |

**Table 4.4.** Regression with Two Right-Hand Variables

| Variable | Coefficient | t-value | Partial $R^2$ |
|---|---|---|---|
| Constant | .0281 | 2.17 | .0403 |
| LnPCI60 | −.0048 | −2.33 | .0463 |
| I/Y | .1502 | 7.08 | .3092 |
| $F$ = 29.57 | $N$ = 115 | | |

### 4.2.3 *Looking Inside a Growth Regression*

Let $g$ be economic growth; $lY$ log initial income; and $z$ another variable of interest, such as $I/Y$, which is itself positively correlated with growth. All these variables are measured from their means. We are interested in a case in which the regression coefficient of $g$ on $lY$ is near zero or positive. So we have:

$$v_{glY} \geq 0 \tag{2}$$

where $v_{ij}$ is the product of $N$, the sample size, and the covariance between two variables labelled $i$ and $j$. Thus $v_{glY}$ is $N$ times the covariance between $g$ and $lY$. Now consider the multiple regression:

$$g_i = \beta lY_i + \gamma z_i + \epsilon_i \tag{3}$$

The least squares estimators satisfy:

$$\begin{bmatrix} v_{glY} \\ v_{gz} \end{bmatrix} = [\beta, \gamma] \begin{bmatrix} v_{gg} & v_{zg} \\ v_{gz} & v_{zz} \end{bmatrix} \tag{4}$$

So that:

$$v_{glY} = (\beta)\left(v_{gg}\right) + (\gamma)\left(v_{gz}\right) \tag{5}$$

Then if $v_{glY} \geq 0$ and $v_{gz} > 0$, (5) requires that either $\beta$ or $\gamma$, but not both, be negative. If $v_{glY} > 0$ then $\beta$ and $\gamma$ may both be positive, but they cannot both be negative.

One way of explaining that conclusion is to say that a finding of $\beta$-convergence with an augmented regression, despite growth and log initial income not being negatively correlated, can happen because the additional variable (or variables on balance) are positively correlated with initial income.

Further inspection of (5) shows that if $z$ is any variable sufficiently positively correlated with initial income, then the inclusion of that variable in the regression will force $\beta$ to be negative. With $z = I/Y$ the variance-covariance matrix between the three variables $g$, $lY$, and $I/Y$ is shown in Table 4.5.

**Table 4.5.** Variance-Covariance Matrix

| | $g$ | $lY$ | $I/Y$ |
|---|---|---|---|
| $g$ | .00034 | | |
| $lY$ | .003838 | .823248 | |
| $I/Y$ | .000921 | .052161 | .007801 |

Then from (4) we have:

$$.003838 = (\beta)\,(.823248) + (\gamma)\,(.052161) \tag{6}$$

The algebra just completed implies that adding any variable positively correlated with initial income will have the effect of pushing the estimate of $\beta$ from close to zero or positive towards a negative value. Thus consider the following regression.

$$g_i = \beta lY_i + \gamma L_i + \epsilon_i \tag{7}$$

where $g$ and $lY$ have the same definition as above, and $L$ is latitude, defined as the angle subtended with the equator by the capital city of the country concerned, this angle being positive whether the said capital city should be north or south of the equator. It is no surprise that $L$ is highly correlated with log initial income. The correlation coefficient is ·7337. The variable $L$ has several appealing features. It varies greatly across the sample, from near zero for Ecuador or Gabon to a huge 64 for Iceland. And unlike some of the conditioning variables employed in cross-section growth regressions it cannot conceivably be argued that it is simultaneously generated by the same data-generating process that yields the initial income values.

The correlation matrix for the three variables growth, $LnPCI60$ and $L$ is shown in Table 4.6.

**Table 4.6.** Correlation Matrix with Latitude

| | growth | LnPCI60 | L |
|---|---|---|---|
| growth | 1 | | |
| LnPCI60 | .17480 | 1 | |
| L | .32184 | .73373 | 1 |

**Table 4.7.** Correlation Matrix for the Three Variables

| Variable | Coefficient | t-value | Partial $R^2$ |
|---|---|---|---|
| Constant | .016141 | 1.03 | .0092 |
| LnPCI60 | −.000831 | −.348 | .0011 |
| Latitude | .000435 | 3.34 | .0893 |
| $F$ = 8.99 | $N$ = 117 | | |

The regression (7) gives the following result:

If only the signs of variables are noted, introducing latitude into the growth equation gives a $\beta$-convergence finding similar to the Barro estimate. However relative to Table 4.4 the detailed result is very different. The negative coefficient on $LnPCI60$ is now insignificantly different from zero. The observation that latitude might be positively associated with economic growth is not new; see Masters and McMillan (2000).

### 4.2.4 *Correlation and Cause*

Some people will dislike the comparison of equation (7) with a standard growth-regression exercise. The difference it will be asserted has to do with causality. The Barro equation is founded in economic theory, or at least in intuitive causal relations between right-hand-side variables and economic growth. It is absurd to imagine that high latitude could cause economic growth. In fact it may not be completely ridiculous to posit that such a relation is a causal connection. It would be the relationship between climate and economic growth already mentioned in Chapter 1. Be that as it may, this case serves well as a stepping-off point for an examination of the connection between causation and correlation.

If Barro-style growth theory is valid, an equation like (1), repeated here for convenience as equation (8), describes exactly the true data-generating process.

$$y = \alpha + \beta x + \sum_i \gamma^i z^i + \epsilon \tag{8}$$

Then the regression:

$$y = \alpha + \beta x + \epsilon \tag{9}$$

gives a biased estimate of the coefficient $\beta$ in (8). The bias is caused by the omission of the important $z$ variables, and this can only happen because the $z$ values are correlated with $x$. The only plausible manner in which

(8) can be the data-generating process with $\beta$ negative and the $\gamma$ values positive, is if larger $z$ values cause higher economic growth on average ($\epsilon = 0$) and larger levels of $x$ retard it on average. Then low $x$ causes economic growth although its simple correlation with economic growth is near to zero. The absence of correlation is no proof of the absence of causation. Indeed a positive correlation would be no proof that a negative causal connection between the two variables does not exist.

Applied economists dream of classical regression analysis where all variables of concern are joint normally distributed, and all the variables save one (the dependent variable) are uncorrelated. These are the independent variables. Then omitting one or more of the independent variables causes no problem, except that the proportion of the variance in the dependent variable explained by variance in the independent variables is reduced. In life variables of interest are usually correlated. They are the product of complex and sometimes obscure data-generating processes. And hidden within those unseen data-generating processes may lie causal connections of various kinds.

We can see how difficult it is to establish cause and to test for it if we return to equations (3) and (7) above. Equation (3) is a standard growth regression with conditioning variables aggregated for convenience. Equation (7) is similar but includes latitude in place of the usual conditioning variables. The growth theorists will insist that the conditioning variables are direct causes of economic growth. They will say that high latitude happens to be correlated with economic growth but cannot conceivably be its cause. Who is to say whether these views are correct? The experimental method can hardly be applied. We cannot tow Iceland to the equator and observe how its economic growth develops over a few following decades. We cannot even vary conditioning variables such as religion or democracy a great deal. Nature or history has provided the experimental variation in these variables, but it has done so in a messy, highly correlated manner that would disgust a laboratory experimenter.

Granger-Sims causality tests (see Geweke 1984) require time-series data, so they cannot be applied to simple cross-section growth regressions. The root idea of Granger causality is that causes precede effects in time, so that shocks to causal variables come first in time. In truth the method allows sophisticated analysis of the relations between variation in different time-series variables, to the results of which analysis the term causality is a convenient label. In general the temporal association between cause and effect can take many forms. If I set my alarm clock to wake me at sunrise, then whether the sun will rise first or my alarm clock ring will depend upon how accurate is my estimate of the time of sunrise, how carefully I have set the clock, and on how precisely my clock keeps time. I may always set the alarm to go off promptly, so as never to be in bed at sunrise. Then time-series analysis may seem to indicate that my alarm clock causes the sun to rise. Granger causality which depends upon the sequential relationship between shocks to time series would not support that conclusion. My clock ringing before the sun rises is not causality in the sense of making things happen. The truth is that the sun plus my reaction to it is what makes the alarm clock ring.

Galileo said that Physics should concern itself not with cause and effect but with what he called 'the geometry of motion'. His view was in opposition to Aristotle's account of motion in terms of its causes. The principle of inertia sees motion as the natural state of objects, requiring no cause to account for it. With a geometry of economic growth, being a complete mathematical description of the world, including policy variables, we might be equally disdainful of cause and effect. But we are never even close to such a complete economic model. So we have to worry whether the correlations we observe indicate simple causal connections, or whether these correlations arise for other reasons. It is observed that corruption and low economic growth are strongly correlated. That may or may not entail that a successful anti-corruption drive will raise economic growth. Yet that is the kind of policy question to which economists need answers.

The conclusion must be that growth regressions do not reliably establish the causes of economic growth. Certainly the latitude equation will never convince anyone not already of that opinion that high latitude causes economic growth. Whatever the conclusion of theorizing, the fact remains that growth rates are hugely variable, and that globalization has affected them. See Sachs (2005).

## 4.3 Convergence Theory

### 4.3.1 *The Solow-Swan Model*

As Chapter 3 has explained already, the dynamics of the Solow-Swan model imply convergence for all agents with identical production functions. This is true for two different important cases:

- there are several isolated economies, each one a version of the same Solow-Swan model, with the same saving share; only the level of capital attained at a particular time distinguishes one economy from another.

- there is one economy with a single integrated capital market and numerous agents each with the same saving rate; only the level of capital attained at a particular time distinguishes one agent from another. This is the MASS model.

For the issues raised for theory the first case is most pertinent here. Given convergence for identical economies, which is straightforward, a different question deserves the attention that it has not received in the literature. If convergence is conditional on various additional variables, how precisely do these variables make their effects felt? Barro and Sala-i-Martin refer to the influence of the extra variables as an alteration to the steady state to which the country concerned is converging. Fair enough; but more than one shift in the production function will influence the steady-state solution. The shift that is posited is supposed to lead to equation (1), repeated here for convenience as (10):

$$y = \alpha + \beta x + \sum_i \gamma^i z^i + \epsilon \tag{10}$$

How do we arrive at this equation, taking into account that it may be a linear approximation?

We work in continuous time, just for convenience. For country $i$ income at time $t$ is:

$$A^i F\left[K^i(t), L^i(t)\right] \tag{11}$$

where the countries are assumed to have the same underlying constant-returns production function, $K$ and $L$ are respectively labour and capital, and $A$ is a measure of national efficiency. As variations in $A$ correspond to differences in total factor productivity, $A$ will be referred to as the TFP coefficient. It is assumed that it is through the value of the TFP coefficient that the conditioning variables make their direct effects felt. They may have an indirect effect via the saving rate. Let the growth rate of employed labour be $n$.

If country $i$ has a saving share $s^i$ the rate of change of per capita capital $k(t)$ in that country satisfies the standard Solow-Swan equation:

$$\frac{dk^i(t)}{dt} = s^i A^i f\left[k^i(t), 1\right] - nk^i(t) \tag{12}$$

where $f\left[k^i(t), 1\right]$ is the per capita production function.

Then the rate of change of per capita product is:

$$\frac{dy^i(t)}{dt} = A^i f_1\left[k^i(t), 1\right] \frac{dk^i(t)}{dt} \tag{13}$$

where subscripts to $f$ denote partial differentiation. Then:

$$\frac{1}{y^i(t)} \frac{dy^i(t)}{dt} = \frac{A^i f_1\left[k^i(t), 1\right] \frac{dk^i(t)}{dt}}{A^i f\left[k^i(t), 1\right]} \tag{14}$$

Rearranging (14) taking into account (12) gives:

$$\frac{\frac{dy^i(t)}{dt}}{y^i(t)} = \frac{k^i(t)\, f_1\left[k^i(t), 1\right]}{f\left[k^i(t), 1\right]} \frac{\frac{dk^i(t)}{dt}}{k^i(t)} = \zeta^i(t)\left[s^i A^i \frac{f\left[k^i(t), 1\right]}{k^i(t)} - n\right] \tag{15}$$

where $\zeta^i(t)$ is the share of capital in country $i$ at time $t$.

Equation (15) says that the growth rate of per capita income in a Solow-Swan model, with a variable TFP coefficient, satisfies the following properties, each to be read in an other-things-being-equal sense.

The growth rate is larger:

1. The larger is capital's share;
2. The larger is the saving share;
3. The larger is the TFP coefficient;
4. The smaller is capital per head;
5. The smaller is the rate of population growth.

The first relation has usually been ignored in the literature. In so far as it matters, it would tend to support the hypothesis of $\beta$-convergence. Arrow, Solow, Chenery, and Minhas (1961) arrived at the idea of a CES production function from the observation that the share of capital is typically larger in poor countries than in rich countries. They showed that with an elasticity of substitution less than unity this is exactly what is to be expected.

Effects 3 to 5 are in the spirit of standard growth theory. The chief problem for growth empirics is to disentangle effects 3 and 4. These jointly decide initial income, with $A$ and $k$ pulling in opposite directions. Mankiw, Romer, and Weil (1992) show that 80 per cent of cross-section differences in growth rates can be accounted for by effects 2 and 5 by themselves. As regards effect 5, the coefficient on $n$ should be capital's share.

The influence of capital is through the level of $\frac{f\left[k^i(t), 1\right]}{k^i(t)}$, the output-capital ratio for an economy with $A = 1$. With a Cobb-Douglas function, for instance:

$$\frac{f\left[k^{i}(t), 1\right]}{k^{i}(t)} = k^{\alpha-1} \tag{16}$$

For equal values of $A$ this value decreases with income and the relation is non-linear. Thus:

$$\frac{dk^{\alpha-1}}{dy} = \frac{dk^{\alpha-1}}{dk}\frac{dk}{dy} = (\alpha - 1)\,k^{\alpha-2}\frac{1}{\alpha k^{\alpha-1}} = \frac{\alpha-1}{\alpha}\frac{1}{k} \tag{17}$$

and the slope of the relation is inversely proportional to $k$.

## 4.4 The Ramsey Model

Robert Barro, as the most prominent proponent of a theory-based convergence model, has made the optimal growth model of Ramsey (1928) the central support of his account of convergent economic growth. Ramsey himself considered a version of his optimal saving problem which will be called here the many-agent Ramsey model (MARM). He looked at steady states, and noted a paradoxical feature of many-consumer steady states. If agents discount future utility, and use different constant discount rates, then in any steady state all the capital will be owned by agents with the lowest discount rate.[2] One way round having all capital end up owned by one agent type would be to have the discount rate depend on consumption per head. For this to help, however, the discount rate would have to be low for the poor, which is the opposite of what intuition may suggest. There is also the issue of time inconsistency; see Chapter 3 above. The optimal growth problem with many consumers is examined by Lucas and Stokey (1984).

Just as with the Solow-Swan model there are two different cases to consider:

- isolated economies, each one a version of the same Ramsey model, with the same utility function and the same utility discount rate; the level of capital attained at a particular time distinguishes one such economy from another.

[2] Barro and Sala-i-Martin (1995: 100–1) discuss the implications of differences in discount rates. However these authors do not provide a full discussion of what happens out of steady state when all agents have the same discount rate.

- one economy with a single integrated capital market and numerous agents each with the same utility function and the same utility discount rate; again only the level of capital held distinguishes agents.

It is useful to have a short title for the second of the above cases. We call it the MARM—the many-agent Ramsey model. Barro, Mankiw, and Sala-i-Martin (1995) build a model which is a hybrid of the above two cases. By distinguishing two kinds of capital they allow capital markets to be perfectly integrated in one case, and isolated in another. This is similar to the extended Stiglitz model exposited in Chapter 3, but without the non-convexity that played an essential role in that case. Discussion of that model is postponed to a later section. Again a short term comes in handy, so the last model will go by the title the BMS model.

## 4.5 Optimal Growth with Isolated Economies

For isolated economies the argument of Chapter 3 has already made clear that there is no general connection between the level of $k$ and $\frac{1}{c}\frac{dc}{dt}$, the rate of growth of consumption. In a Ramsey model growth solves:

$$\text{Max} \int_0^\infty U\,[c(t)]\,e^{-rt}dt \tag{18}$$

subject to:

$$\frac{dk}{dt} = F\,[k(t), 1] - c(t) \tag{19}$$

and the initial condition $k\,(0) = k_0$.

The necessary condition for optimal growth derived in the Mathematical Appendix can be written in the form:

$$-\frac{c\frac{du}{dc}}{u}\left[\frac{1}{c}\frac{dc}{dt}\right] = F_1\,[k(t), 1] - r \tag{20}$$

where $u$ is $U_1\,[c\,(t)]$, the marginal utility of income. When $k(t)$ takes a low value the right-hand side of (20) is relatively large. If the growth rate of consumption is not to be relatively large, the elasticity of marginal utility $-\frac{c}{u}\frac{du}{dc}$ must be large. That could well be the case. Therefore the idea that $\beta$-convergence follows from optimal growth theory is somewhat suspect.

## 4.6 Growth in the MARM

It is known that in the MARM equal-discount-rate unequal-income steady states can be observed. Agent $i$ solves:

$$\text{Max} \int_0^\infty U\left[c^i(t)\right] e^{-rt} dt \tag{21}$$

subject to:

$$\frac{dk^i(t)}{dt} = F_1\left[\sum_j k^j(t), 1\right] k^i(t) + w(t) - c^i(t) \tag{22}$$

where the factor prices $F_1$ and $w$ in (22) are treated as constants, notwithstanding the fact that an individual influences $\sum^k$ and thence $w$. In a stationary state factor prices are constant over time, as is capital held by any type of agent. No agent wishes to alter consumption so as to transfer it marginally between periods. That requires that the analogue of (20) for this case be satisfied:

$$-\frac{\frac{dU_1\left[c^i(t)\right]}{dt}}{U_1\left[c^i(t)\right]} = F_1\left[\sum_j k^j(t), 1\right] - r \tag{23}$$

for all $i$.

In steady state the left-hand side of (23) is zero. It follows that:

$$F_1\left[\sum_j k^j(t), 1\right] = r \tag{24}$$

is a necessary condition for a steady state.

As all agents have the same discount rate $r$, there is only one possible steady-state value for aggregate capital $\sum k$, that which satisfies (24). The wage rate will be the one corresponding to aggregate capital defined by (24). Although only one value of aggregate capital is consistent with steady state, it can be distributed between agents in any way.

Assume that steady state holds. This is equivalent to supposing local stability of the system. We vary the capital holding of agents of significant weight, then we have to vary the capital holdings of other agents so as to keep total capital in the economy constant. If one large group of agents decides, for whatever reason, to accumulate more capital, they will drive down the rate of return, and other agents through their optimizing responses will eventually end up holding less capital. In any case, if one starts in steady state with total capital equal to its long-run equilibrium level, there will be no tendency at all for agents' incomes to converge.

## 4.7 A Non-Convergence Result

So far we have only shown that unequal-income steady states exist. This by itself points up a contrast with the Stiglitz model, as in that case an unequal-income steady state is not an equilibrium. That point granted, it is not immediately obvious that solutions starting off the steady state will not converge.

In fact starting from arbitrary unequal initial conditions, and analysing optimal developments in the MARM, convergence to income equality for otherwise like agents never happens. It is surprising that an optimal general-equilibrium solution to the MARM is always inconsistent with what follows from Stiglitz's fixed saving coefficients. But such is evidently the case, as different convergence conclusions are implied in the two instances. In showing a tendency to convergence, Barro and Sala-i-Martin (1995) assume that capital markets are highly segmented, so that low-wealth countries enjoy high rates of return to saving. For the basic MARM model it will be shown that:

1. Non-converging steady states are possible;
2. Strict asymptotic convergence can never occur;
3. Partial convergence or divergence (clubs) are possible depending on the sign of the third derivative of the utility function.

In the Mathematical Appendix it is shown that any MARM equilibrium solution solves a program of the form:

$$\text{Max} \sum_{i=1}^{N} \alpha^i \int_0^{\infty} U\left[c^i(t)\right] e^{-rt} dt \tag{25}$$

where the weights $\alpha$ are constants independent of time. Simple inspection of (25) shows immediately that convergence of consumptions can never be the outcome. Imagine two agents, indexed $m$ and $n$. Agent $m$ starts with less capital, so in an equilibrium must have a lower weight $\alpha$. Then that lower weight attaches to agent $m$ forever. Were it to happen that the consumption levels of agents $m$ and $n$ converged, then the respective marginal utilities of consumption would not converge, because the marginal utility of consumption would be lower for agent $m$. So full convergence could never be the solution.

This conclusion comes from a model which incorporates strong simplifications. Some of these might be regarded as particularly favourable to convergence. In particular:

- All agent types have the same tastes and the same discount rates;
- All supply the same quantity of labour in all periods and earn the same wage;
- All have access on exactly equal terms to the same capital market, where they all earn the same rate of return;
- All have perfect foresight and there are no stochastic effects in the model to upset convergence.

## 4.8 Convergence Clubs

We have been able to rule out strict asymptotic convergence: the unequal cannot become completely equal, not even in the limit. This result has little relevance to empirical studies of convergence. These only examine convergence over quite short periods of time. The present theory can throw light on partial convergence. However the findings are ambiguous. Unequal agents may come closer together (meaning here that the ratio of their consumption levels moves closer to unity), or they may move further apart (meaning here that the ratio of their consumption levels moves away from unity). Just as both these outcomes may be observed, so both may occur simultaneously in different areas of the global income distribution across agents. The possibilities are rich. This is not inconsistent with the findings of concrete empirical studies, which similarly seem to suggest a variety of possibilities. Too much should not be made of that fact. The model is highly stylized and unrealistic.

To see what happens to the distribution of consumption over time, consider the maximization of the objective function:

$$\sum_{i=1}^{N} \alpha^i \int_0^{\infty} U\left[c^i(t)\right] e^{-rt} dt \qquad (26)$$

In the mathematical appendix it is shown that if the maximization of (26) provides total consumption $C(\tau)$ in period $\tau$, then the $C^i(\tau)$ values must maximize:

$$\sum_{i=1}^{N} \alpha^i U\left[c^i(\tau)\right] \qquad (27)$$

Then the growth rate of consumption increases with its level if and only if the elasticity of marginal utility:

$$\xi = -\frac{U_{11}[c]}{U_1[c]} \cdot c \tag{28}$$

decreases with the level of consumption.

How $\xi$ varies with the level of consumption is decided by the third derivative of the utility function, which cannot be determined with certainty.

Now is the time to collect together the implications for the comparison of growing economies of the non-convergence result and the calculations just completed. Even with no random shocks, $\beta$-convergence need not imply the asymptotic convergence of incomes. Suppose, for instance that the time paths of the logarithm of per capita income $y$ in respectively Country I and Country II are:

$$\ln y^I = a^I - \frac{b}{t+2} \tag{29}$$

and:

$$\ln y^{II} = a^{II} - \frac{b}{t+1} \tag{30}$$

with $a^I > a^{II}$.

Then:

$$\ln y^I - \ln y^{II} = a^I - a^{II} + b\left[\frac{1}{t+1} - \frac{1}{t+2}\right] > 0 \tag{31}$$

So Country I has the higher per capita income at all $t$. The growth rates of $y$ for countries I and II respectively are:

$$b(t+2)^{-2} \text{ and } b(t+1)^{-2} \tag{32}$$

So Country II has the lower income and is always growing faster. Yet Country I's $\ln y$ asymptotes to $a^I$, and Country II's to $a^{II}$, and there is no asymptotic convergence.

We have seen that empirical investigation, at least for diverse samples, lends only powerfully qualified support even to the hypothesis of $\beta$-convergence, whether or not that hypothesis implies asymptotic convergence. Yet those empirical investigations are presented as tests or measurements of a theoretical model that supposedly predicts $\beta$-convergence. In fact, as has been shown above, there is more than one theoretical model at issue. And it is misleading to suggest that it makes little difference which model is used.

Broadly speaking, identical isolated economies converge asymptotically, even if they start from different initial positions. This is only to say that the models presently under consideration have unique stable steady states. That asymptotic convergence neither implies, nor does it require, $\beta$-convergence, or any other simple local property. All this ignores the effect of random shocks, although these are surely important in the growth histories of individual countries. If random shocks of variable size arrive each period, there is no strict asymptotic convergence. If the shocks are of large size on average there is a limiting probability density function that measures the probability that any particular country will find its per capita income within any interval at a time far in the future ($t$ large). Bliss (2002) shows how the shape of that distribution is influenced both by the form of the differential equation for per capita income, particularly its non-linearity, and also the probability density of shocks.

Two intuitive cases illustrate the type of results available:

- Suppose that low-income countries grow slowly; middle-income countries rapidly; and high-income countries converge slowly to the steady state. Then if shocks are large enough to throw many countries far away from the steady state, the low and high-income range will form basins of attraction in which many countries will be found, while the middle range will be sparsely populated. This would make the Quah (1997) empirical observation of twin peaks in the current world income distribution a possible long-run equilibrium feature.
- If shocks are highly asymmetric this will affect the probability distribution of income levels even if the differential equation for income is linear. Imagine for example that the shock each period is the sum of relatively small-scale noise and the possible arrival of a huge 'earthquake' shock that reduces income by 25 per cent. Then the density of incomes will show a mass around the steady state, formed of countries that have largely converged. Below the steady state there will be a mass of countries that have experienced an 'earthquake' in the more or less recent past. Above the steady state density will be low because only low-level random noise will take any country there.

With integrated capital markets it is no longer the case that asymptotic convergence is the norm, regardless of the model chosen. With a MASS model there is asymptotic convergence. With the many-agent Ramsey model (the MARM), there is a sharp contrast between per capita production and per capita income. The former converges immediately, the latter never, as the non-convergence result shows.

It is plain now that integrated capital markets are not good for income convergence. It is also evident that world capital markets are far from being fully integrated. Capital does not flow freely from where it is abundant to where it is scarce. On why this should be the case see chapter 2 of Lucas (2002) and further arguments below, particularly in Chapter 8. One model that directly addresses this problem is the BMS model, that is examined immediately in the next section.

## 4.9 The BMS Model

To explain why capital does not migrate instantly between different agents to equalize the marginal product of capital, Barro, Mankiw, and Sala-i-Martin (1995), here (BMS), propose an extended Solow-Swan model. Human capital is introduced into the production function. By itself this does not make a great difference, as human capital is simply accumulated optimally to combine with physical capital. These authors, however, add an extra assumption. The accumulation of human capital cannot be financed by borrowing.

The idea is that imperfect capital mobility by impeding output convergence will assist income convergence. That is what happens in the case considered by the authors. This case is quite special. One small low-wealth country converges to a steady state which the rest of the world occupies from the start. The converging country is borrowing-constrained all the way to steady state. The large size of the rich country fixes the world rate of return to physical capital at the common utility discount rate of all countries. This must be the case when the rich country is in steady state.

The inclusion of human capital by itself is fairly innocuous and uninteresting. The assumption that human capital cannot be used as colleteral for international loans, on the other hand, makes a huge difference. The point is that it is a *value constraint*. To check whether a particular agent is satisfying the constraint at a certain time, one compares physical capital owned by that agent with physical capital employed by the agent. The constraint says that physical capital owned cannot take a negative value. That means that a country cannot borrow more than it uses as physical capital and use the proceeds to finance the accumulation of human capital. Problems arise because ownership depends upon prices via budget constraints. So one cannot proceed by mapping simply from prices to demands.

In the examination of the Ramsey MARM we saw that a weighted sum of individual utility integrals is maximized. If value constraints constrain the histories which can be experienced we have to proceed in a different way and we will arrive at qualitatively different types of equilibrium. We will no longer be able to say in any simple sense that equilibrium maximizes a weighted sum of individual utility integrals. More generally it is known that existence of general equilibrium with quantity constraints is more than a little problematic.

A poor country asymptotes to a steady-state equilibrium in which the return to its human capital is equal to that common utility discount rate. All initially poor countries are identical in the limit. They own the same level of human capital per capita, and they all own no physical capital. All the physical capital they use is leased from abroad. There is convergence of product per capita in the limit, while per capita incomes never converge. Thus in its limiting properties the BMS model replicates and exaggerates the outcome predicted by a model in which capital is always perfectly mobile. We get productivity convergence without income convergence. With perfect capital mobility a country which starts with more physical capital than another will always be better off. In the BMS model there is no asymptotic advantage to a small country from starting with more capital.

If one has a value-constrained equilibrium, which need not be unique, some of the qualitative features noted by BMS will apply. For instance, at the start convergence will be accelerated for low-initial-human-capital agents. Equally some of the results derived above will also hold. Convergence clubs will be possible, indeed they may be more probable. Now, however, they may not be so simply associated with the sign of the third derivative of the utility function. It is not obvious that strict asymptotic convergence can never occur but it seems hard to describe precisely. If one agent is always in the interior of unconstrained states, another agent can never converge to that first agent. For once the second enters the unconstrained region we can apply earlier results to rule out strict convergence. If one agent is right on the boundary of the constrained region, could another agent converge to the same point? I am not certain but it must be quite a special case.

Of greater interest to the present concerns is a simple generalization of the BMS model, as follows. There is again one large country (or mass of countries) in the steady state. Now a distinct but large country is added.

This is a poor country with both per capita physical and human capital below the steady-state level. Now the world as a whole has less capital of both kinds than the steady state. Physical capital moves from the rich country to the poor country, while the poor country divests itself of the ownership of all physical capital. While it is working on the accumulation of human capital, the poor country wastes none of its saving on buying, as opposed to leasing, physical capital.

One might think that the poor country need not necessarily have a higher rate of return to human capital than a rich country. Because it has less human capital, the poor country will attract less physical capital. Less human capital implies a higher rate of return; but less physical capital for the human capital to cooperate with should lower the rate of return to human capital.

There is no genuine ambiguity, as is shown in the mathematical appendix. So long as $F$ [.] is a strictly concave function, the marginal product of human capital in the poor country:

$$F_2\,[k^p, h^p, 1] = \rho^p \tag{33}$$

where the superscript $p$ indicates poor, and $k^p$ varies so as to keep $F_1$ [.] equal to an externally given return to physical capital is a decreasing function of $h^p$.

Then the optimal accumulation of human capital in the poor country is governed by a necessary condition analagous to (23), viz:

$$-\frac{\frac{dU_1[c^i(t)]}{dt}}{U_1\,[c^i(t)]} = F_2\,[k^p(t), h^p(t), 1] - r \tag{34}$$

The lower is human capital, the higher is $F_2$ [.]. Then given a constant elasticity of intertemporal substitution, consumption grows faster in the poorer country. Again, as with the basic Barro story derived from the Ramsey model, the constant elasticity assumption is seriously suspect.

A further simple generalization of the BMS model highlights the point that the conclusions of that paper lean heavily on the particular construction used by the authors. There is again one large country (or mass of countries) in the steady state. Now there is one rich country initially in steady state. Two different poor countries are added, both with per capita physical and human capital below the steady-state level. Again the world as a whole has less capital of both kinds than the steady state. Physical capital moves from the rich country to the poor countries, and these both accumulate human capital on optimal growth paths.

In the model now proposed, with two or more large poor countries, many of the simple findings of the BMS model go out of the window. Now the large mass of poor countries pushes the world return on physical capital above the utility discount rate. Then the return to human capital for a borrowing-constrained country must be even higher. As a consequence a poor country, not the poorest, may break through the borrowing constraint in finite time, when it will start to accumulate both kinds of capital optimally. Per capita product may converge for some countries before it converges for all. Not all poor countries will be equal in the limit. The BMS model is far richer in its possible implications than the simple example exhibited by these authors indicates.

## 4.10 Concluding Remarks

Most readers of the empirical and theoretical literature concerned with growth and convergence will take away three impressions:

1. There is no simple statistical association between initial income and subsequent growth, hence no support for $\beta$-convergence from a basic two-variable analysis.
2. With multivariate analysis there is good support for a causal connection between growth and initial income, on an other-things-equal basis.
3. Theoretical models with common technology and preferences strongly confirm the $\beta$-convergence hypothesis.

Of these three propositions only the first can be allowed to pass without the addition of a note of qualification or caution. With regard to the second proposition, the extensive discussion above has made it clear that it is nearly impossible to establish a causal connection from a limited cross-section sample. Indeed it is not even plain what a causal connection means. Physics, at least in its more formal manifestations, manages well without the notion of a cause. Economists for the most part find the concept nearly irresistible. That may be a result of the economist's inescapable engagement with policy questions, which typically take the form of 'what if?'. That said, the idea of a causal connection is slippery and perilous. All the information we have is in the variance-covariance matrix of our variables, plus the sample size. This is an approximation to the limiting variance-covariance values defined by the data-generating process. Any regression we compute from the data we have embodies a view that a particular variable is dependent, in the sense of being generated by a linear combination of the other variables, plus an uncorrelated random variable. When, as is always the case, the 'right-hand-side' variables are highly cross-correlated, their influence is complex and the coefficient on any one of them can be powerfully influenced by the absence or presence of the others. Cross-section growth regressions should be seen as estimates of the coefficients of a model proposed and not directly tested by the regression analysis.

Finally with regard to the third proposition, it is true that there are theoretical models that predict $\beta$-convergence, on average at least. The Solow-Swan model meets that description. One version of the Ramsey model exhibits the same feature. Yet this model is far more special than the standard presentations suggest. A truly general Ramsey model carries no implications for comparative growth. Particular functional forms allow stronger results. However a constant elasticity of intertemporal substitution is not a convincing assumption.

With optimal saving and unified capital markets there is immediate convergence of product per head, but wealth inequality persists, and it never disappears. It persists because it is not optimal for individual agents to remove it by their own saving. They may partially remove it within a convergence club but there is no guarantee of this, and there may indeed be anti–convergence clubs, within which agents' wealth holdings tend to diverge.

The assumption of perfect capital mobility is so seriously inaccurate that it is natural that theorists should try to model its failure. The most prominent attempt to do that is the BMS model. The idea is a good one. Yet the authors only exhibit one special case, and it conceals as much as it displays. Earlier analysis has shown how assuming a constant elasticity of intertemporal substitution limits the ability of models to capture possibly important aspects of reality.

There is a further respect in which the theoretical models are unsatisfactory. They mostly treat of a non-stochastic world, which is not the world that we inhabit. The econometric analysis on the other hand has to take account of random shocks. And many writers, including Friedman, stress that random shocks result in the difference between $\beta$- and $\sigma$-convergence. It is not enough strictly to build a non-stochastic model and then throw in a bit of randomness as an afterthought. Randomness affects optimal growth planning, and it may affect it differentially at different income levels.

The excessive stability of income distributions in the MARM model is due to the fact that intergenerational transmission of wealth is perfect, which makes the system extremely conservative. Less perfect wealth transmission would help there but we need to avoid its opposite. We may not want the only possible long-run state to be one of convergence to a unique equilibrium. In this case stochastic shocks are helpful. They spread wealth out and avoid unrealistic convergence.

So our arguments show that there are important deficiencies in the received growth theory that supposedly underlies economic convergence. Given the importance that its proponents attach to this theory, their use and exposition of it is extraordinarily casual.

Finally it is worth remarking that the huge body of literature concerned with economic convergence has given relatively little attention to catching-up as a driving force of convergence. This is surprising given that imitation and catching-up seem to be important factors in, for example, post-war European growth. A formal catching-up model is developed in Chapter 12.

The type of theory considered in this paper is fundamentally long-run, so that it is not easy to distinguish the realistic from the unrealistic. We know that global income distributions can shift significantly within individual lifetimes, due to macroeconomic developments, policy shifts, or technical changes. Yet other studies show a remarkable stability of inequality. Even partial convergence in the models examined will require many generations. Qualitative findings, however, may be suggestive. If there is no tendency to convergence, as can happen within plausible models, the shortness of our line of view will matter less.

## Mathematical Appendix 4.1

*Strange Accumulation Paths can be Optimal*

Given the model:

$$\text{Max} \int_0^\infty U\,[c(t)]\, e^{-rt} dt \tag{A.1}$$

subject to:

$$\frac{dk}{dt} = F\,[k(t), 1] - c(t) \tag{A.2}$$

and the initial condition $k\,(0) = k_0$.

The Hamiltonian is:

$$p_0 \cdot U\,[c(t)]\,e^{-rt} + p_1 \left\{F\,[k(t), 1] - c(t)\right\} \tag{A.3}$$

and we have:

$$\frac{dp_1}{dt} = -p_1 F_1\,[k(t), 1] \tag{A.4}$$

With $p_0 = 1$, the maximization of the Hamiltonian with respect to $c$ gives:

$$U_1\,[c(t)]\,e^{-rt} - p_1 = 0 \tag{A.5}$$

Differentiating (A.5) with respect to $t$ gives:

$$\frac{dU_1\,[c(t)]}{dt} e^{-rt} - rU_1\,[c(t)]\,e^{-rt} - \frac{dp_1}{dt} = 0 \tag{A.6}$$

Thence from (A.4)–(A.6):

$$-\frac{\frac{dU_1[c(t)]}{dt}}{U_1\,[c(t)]} = F_1\,[k(t), 1] - r \tag{A.7}$$

**Theorem 4.1:** *Given a standard production function $F\,[k, 1]$, and a monotonic increasing differentiable function of $t$, on $[0, \infty]$, denoted $k(t)$, with $lim_{t\to\infty} k(t) = k*$, where $r = F_1\,[k*, 1]$, and such that:*

$$F\,[k(t), 1] - \frac{dk(t)}{dt} \tag{A.8}$$

*is monotonically increasing, there exists a concave utility function $U\,[c]$ such that $k(t)$ is the Ramsey-optimal path for capital.*

**Proof:** The differential equation:

$$\frac{dx}{dt} = F_1\,[k(t), 1] - r \tag{A.9}$$

has its right-hand side decreasing with $t$, because $F\,[k, 1]$ is a concave function. Now:

$$x(t) = \int_0^t \left\{F_1\,[k(\theta), 1] - r\right\} d\theta \tag{A.10}$$

is strictly increasing in $t$ so long as $F_1\,[k(t), 1] > r$, which only requires $k < k*$.

Now define $u\,[c(t)]$ to satisfy:

$$ln\,[u\,[c(t)]] = -x(t) \tag{A.11}$$

which implies:

$$u\,[c(t)] = e^{-x(t)} \tag{A.12}$$

Differentiating both sides of (A.11) with respect to $t$ and taking into account (A.9) gives:

$$-\frac{\frac{du[c(t)]}{dt}}{u\,[c(t)]} = F_1\,[k(t), 1] - r \tag{A.13}$$

Thus $u\,[c(t)]$ is a marginal utility of consumption function consistent with $k(t)$ being an optimal accumulation path for the problem (A.1) and (A.2).

It remains to show that $u\,[c(t)]$ integrates up to make a well-behaved concave utility function. Note that from (A.12) and (A.13) $u\,[c(t)]$ is strictly decreasing in $t$, hence strictly decreasing as $c$ increases. Now define $U\,[c(t)]$ to satisfy:

$$U\,[c(t)] = \int_0^t e^{-x(\theta)} d\theta \tag{A.14}$$

As $e^{-x(\theta)} > 0$ everywhere, $U\,[c(t)]$ increases with $t$ , hence with $c(t)$. □

## Non-Convergence in the MARM

The non-convergence result follows immediately from the observation that any $N$-agent MARM equilibrium is a solution to a program of the form:

$$\text{Max} \sum_{i=1}^{N} \alpha^{\mathrm{I}} \int_0^{\infty} U\left[c^{\mathrm{I}}(t)\right] e^{-rt} dt \tag{A.15}$$

where the maximization is constrained by the initial total capital endowment $K_0$ and the aggregate production function. A detailed proof that the equivalent of (A.15) is maximized, for the case of a discrete model with Koopmans recursive preferences, is provided by Bliss (2004*a*). Here a sketch of the parallel argument for the present case follows.

The principle of optimality says that any section of an efficient growth path from $t^{\mathrm{I}}$ to $t^{\mathrm{II}}$ must be efficient subject to the levels of $K$—aggregate capital—attained at respectively $t^{\mathrm{I}}$ and $t^{\mathrm{II}}$. Denote the utility accumulated by agent $i$ over that interval by $V^{\mathrm{I}}$. Then:

$$V^{\mathrm{I}} = \int_{t^{\mathrm{I}}}^{t^{\mathrm{II}}} U\left[c^{\mathrm{I}}(t)\right] e^{-rt} dt \tag{A.16}$$

The set of feasible $V$ levels is a convex set, and an efficient point of that set must lie on its upper boundary. There it maximizes a weighted sum of the form:

$$\sum_{i=1}^{N} \alpha^{\mathrm{I}} V^{\mathrm{I}} \tag{A.17}$$

This means that the equivalent of (A.15) is maximized, but with the integration over the range $t^{\mathrm{I}}$ to $t^{\mathrm{II}}$, rather than 0 to $\infty$.

In a fully differentiable model the weights $\alpha$ will be unique, and that is assumed here to keep the argument simple. It only remains to show that the $\alpha$ weights in (A.17) are invariant to the choice of the section of the growth path. Let there be two sections, $t^{\mathrm{I}}$ to $t^{\mathrm{II}}$ and $t^{\mathrm{III}}$ to $t^{\mathrm{IV}}$ for which the $\alpha$ weights are unequal in the sense that normalized to sum to one they take different values. Denote the utility accumulated by agent $i$ over the interval $t^{\mathrm{I}}$ to $t^{\mathrm{II}}$ by $V_{\mathrm{I}}^{\mathrm{I}}$, and the utility accumulated by agent $i$ over the interval $t^{\mathrm{III}}$ to $t^{\mathrm{IV}}$ by $V_{\mathrm{III}}^{\mathrm{I}}$. Then the sets of feasible levels of respectively $V_{\mathrm{I}}^{\mathrm{I}}$ and $V_{\mathrm{III}}^{\mathrm{I}}$, and also $V_{\mathrm{I}}^{\mathrm{I}} + V_{\mathrm{III}}^{\mathrm{I}}$ are convex sets. The frontier of this last set is only attained when the support hyperplanes to the feasible sets of $V_{\mathrm{I}}^{\mathrm{I}}$ and $V_{\mathrm{III}}^{\mathrm{I}}$ are identical. It is similar to international trade between two countries each with its own production-possibility set. Trade equalizes marginal rates of substitution between outputs in the two countries, when joint production is efficient.

Because it is infinite integrals of discounted utility that matter, efficiency requires that the frontier of feasible levels of $V_{\mathrm{I}}^{\mathrm{I}} + V_{\mathrm{III}}^{\mathrm{I}}$ be attained. Hence the conclusion that a function of the form of (A.15) be maximized, with the $\alpha$ values constants independent of time.

## Convergence Clubs

Consider the maximization of the objective function:

$$\sum_{i=1}^{N} \alpha^{\mathrm{I}} \int_{0}^{\infty} U\left[c^{\mathrm{I}}(t)\right] e^{-rt} dt \tag{A.18}$$

- The solution which maximizes (A.18) also maximizes:

$$\sum_{i=1}^{N} \alpha^{\mathrm{I}} \int_{\tau}^{\infty} U\left[c^{\mathrm{I}}(t)\right] e^{-rt} dt \tag{A.19}$$

for any value of $\tau$ subject to the total capital available at time $\tau$ determined by the solution to (A.18).

- Given that the solution to the maximization of (A.18) provides total consumption $C(\tau)$ in period $\tau$, then the $C^{\mathrm{I}}(\tau)$ values must maximize:

$$\sum_{i=1}^{N} \alpha^{\mathrm{I}} U\left[c^{\mathrm{I}}(\tau)\right] \tag{A.20}$$

on any measurable subset of values of $\tau$. The maximization is subject to:

$$\sum_{i=1}^{N} c^{\mathrm{I}}(\tau) \leq C(\tau) \tag{A.21}$$

Now (A.20) and (A.21) imply:

$$\alpha^{\mathrm{I}} U_1\left[c^{\mathrm{I}}(\tau)\right] - \lambda(\tau) = 0 \tag{A.22}$$

for all $\tau$, where $\lambda(\tau)$ is the value of the Lagrange multiplier on the constraint (A.22) at $\tau$. Taking any two agents, say $i$ and $j$, we have, from (A.22):

$$\frac{U_1\left[c^{\mathrm{I}}(\tau)\right]}{U_1\left[c^{j}(\tau)\right]} = \frac{\alpha i}{\alpha^{j}} = B \tag{A.23}$$

where the right-hand side of (A.23) is a constant, denoted $B$. Then:

$$\ln U_1\left[c^{\mathrm{I}}(\tau)\right] = \ln U_1\left[c^{j}(\tau)\right] + \ln B \tag{A.24}$$

Differentiating (A.24) with respect to $t$ and rearranging gives:

$$-\frac{U_{11}\left[c^{\mathrm{I}}(\tau)\right]}{U_1\left[c^{\mathrm{I}}(\tau)\right]} \cdot c^{\mathrm{I}}(\tau) \cdot \frac{\frac{dc^{\mathrm{I}}(\tau)}{dt}}{c^{\mathrm{I}}(\tau)} = -\frac{U_{11}\left[c^{j}(\tau)\right]}{U_1\left[c^{j}(\tau)\right]} \cdot c^{j}(\tau) \cdot \frac{\frac{dc^{j}(\tau)}{dt}}{c^{j}(\tau)} \tag{A.25}$$

Then (A.25) implies the growth rate of consumption increases with its level if and only if the elasticity of marginal utility:

$$\xi = -\frac{U_{11}[c]}{U_1[c]} \cdot c \tag{A.26}$$

decreases with the level of consumption.

## Growth of the Poorer Country in the BMS Model

Let $\nu$ be the world marginal product of perfectly mobile physical capital. Then:

$$F_1\left[k^p, h^p, 1\right] = \nu \tag{A.27}$$

where $F[.]$ is the production function, subscripts denote partial differentiation and the superscripts $p$ indicate that this is the poor country. Differentiating (A.27) totally with respect to $h^p$ gives:

$$F_{11}\left[k^p, h^p, 1\right] \frac{dk^p}{dh^p} + F_{12}\left[k^p, h^p, 1\right] = 0 \tag{A.28}$$

Or,

$$\frac{dk^p}{dh^p} = -\frac{F_{12}\left[k^p, h^p, 1\right]}{F_{11}\left[k^p, h^p, 1\right]} \tag{A.29}$$

As $F_{12} > 0$, $\frac{dk^p}{dh^p} > 0$. Now define $\rho^p$ to be the marginal product of human capital in the poor country, so that:

$$F_2\left[k^p, h^p, 1\right] = \rho^p \tag{A.30}$$

Differentiating (A.30) totally with respect to $h^p$ gives:

$$\frac{d\rho^p}{dh^p} = F_{21}\left[k^p, h^p, 1\right] \frac{dk^p}{dh^p} + F_{22}\left[k^p, h^p, 1\right] \tag{A.31}$$

Taking into account (A.29), (A.31) may be written:

$$\frac{d\rho^p}{dh^p} = -\frac{F_{12}\left[k^p, h^p, 1\right] F_{21}\left[k^p, h^p, 1\right]}{F_{11}\left[k^p, h^p, 1\right]} + F_{22}\left[k^p, h^p, 1\right] \tag{A.32}$$

Assuming $F\,[.]$ to be a strictly concave function:

$$F_{11}\,[.]\,F_{22}\,[.] - F_{12}\,[.]\,.F_{21}\,[.] > 0 \tag{A.33}$$

Then from (A.31) $\frac{d\rho^p}{dh^p} < 0$.

# 5

# Competitive Trade Theory

## 5.1 Introduction

Competitive trade theory is the application of the competitive general-equilibrium model (the GE model), sometimes called the Arrow-Debreu model, to an international economy. That definition fails to note that the GE model is so general as to incorporate only those assumptions essential to prove the existence of an equilibrium, and as such is to a great extent devoid of specific results. For instance, outside of special cases there are no general GE comparative static results. Those that can be derived depend upon particular restrictive assumptions, such as gross substitutability. Trade theory has followed a very different course. Its leading competitive equilibrium models are so special as to make a GE theorist laugh. Yet such mockery might disguise some envy, for the particular lends to trade theory numerous definite results. Then a critical question is whether special models—they will be called 'toy models' later—can give insights concerning a world which in its complexity and elaborate structure resembles more the GE model than a simple classroom trade model. We will return to that difficult question more than once below but without reaching any final or definite conclusion.

This chapter may be taken as a review of generally well-known theory. Indeed major sources for its results are the fine surveys of trade theory provided by Chipman (1966), Dixit and Norman (1980), and Feenstra (2004), as well as other standard sources. Arguably the chapter is too dry and technical to provide much pleasure to my readers. A reader with a basic training in trade theory could well skip it. It is always there to be referred to as required.

New value however may be found here because the argument is tailored to the particular focus of the book. Also, surprisingly, to the author at least, there is some meat to be picked from these old bones. While indirect functions, or duality, have been widely employed in expositing trade theory, the nice insights offered by the mixed price-quantity revenue functions have not been fully exploited. I hope to convince the reader of this point below. The concept of a mixed price-quantity revenue function is the same as the normalized restricted profit function of Lau (1976). That writer noted that such a function permits the easy derivation of factor shadow prices.

This is a chapter with a large, almost an unwieldy, mathematical appendix. That reflects the fact that this part of economic theory is highly technical in certain respects. However all the technical detail has been banished to the appendix because the most important need is to give a picture of the essentials. Given that picture, the interested reader may pursue the details in the appendix, which it is hoped provides an accurate and more complete account.

Diewert (1982) provides a wide-ranging review of duality theory, with many references and some application to trade theory. See also Woodland (1974). The indirect functions that will be used here are the *Revenue Function*:

$$R\,[\mathbf{p}] \tag{1}$$

which is the maximum net profit when prices are the vector $\mathbf{p}$. $R\,[\mathbf{p}]$ is homogeneous of degree one in $\mathbf{p}$, and weakly convex. As a convex function it is continuous and is differentiable almost everywhere. The most important property of $R\,[\mathbf{p}]$ is this. Its partial derivatives are profit-maximizing outputs. This is an envelope property. If outputs are unaffected by a small change in a price, then it is evident that a small increase in the price of slippers increases profit at a rate equal to the output of slippers. Any substitution of production in favour of more highly valued slippers is a second-order effect and may be neglected.

Another indirect function is the *Expenditure Function*:

$$E\,[\mathbf{p},U] \tag{2}$$

which gives the minimum cost of buying a basket of goods with utility at least as large as $U$ when prices are $\mathbf{p}$. The function $E\,[\mathbf{p},U]$ is increasing in $U$ and is concave in $\mathbf{p}$. The partial derivatives of $E\,[\mathbf{p},U]$ are denoted $E_{\mathbf{p}}\,[\mathbf{p},U]$ and are demands when prices are $\mathbf{p}$. This again is an envelope result and is valid because any substitution resulting from the price change is a second-order change.

## 5.2 Two Examples with Results

To illustrate how powerful is the indirect function approach this section offers two examples of the method at work. The first is a straightforward generalization of a familiar textbook result for the one-consumer economy. It is followed by a demonstration of the projection of that result to the many-consumer economy, where it is seen that a far weaker conclusion holds.

*1. Gains from Trade* The proposition that trade is gainful is one that is demonstrated, usually with a simple diagram, in almost any undergraduate textbook. For the sake of diagrammatic simplicity, but also because the result as stated depends upon that assumption, this is a theorem for a one-consumer economy. Figure 5.1 shows one consumer gaining from trade at fixed prices shown by the slope of the straight line through B. To allow for a diagram it is a two-good example for which the result is shown. Here we show a result not of gain as such, but of minimum gain against a hostile price-setter, which the same figure can also demonstrate.

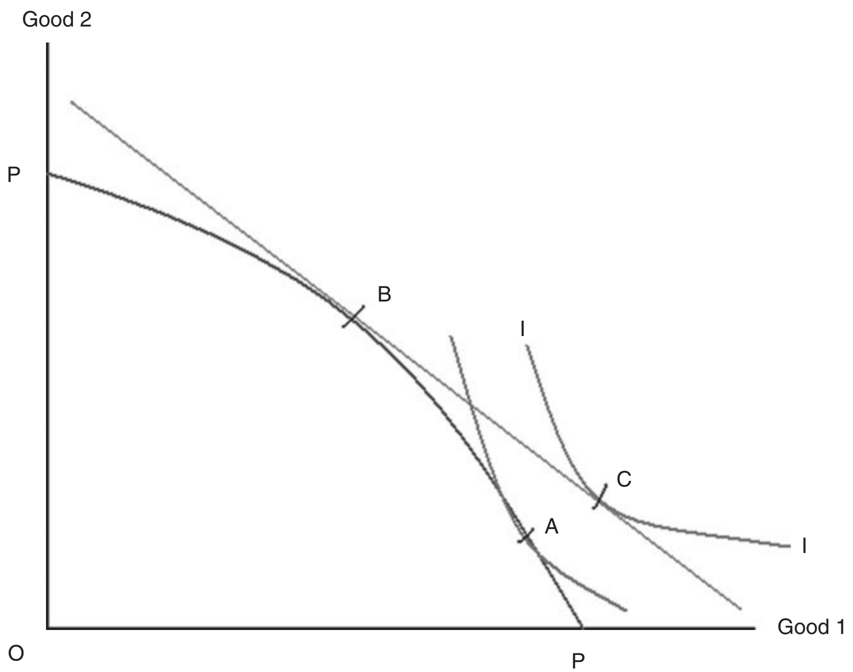



**Figure 5.1.** The Gains from Trade

The production possibility curve is the curve PP. The worst possibility for this consumer is relative international prices shown by the slope of the tangent to the PP curve at A. With those bad prices the consumer consumes at A, and the welfare level is that of the lower indifference curve. With trade the consumer consumes at C on a higher indifference curve.

By mentally adjusting the position of the price line the reader may confirm that the consumer can never end up at a lower indifference level than that delivered at A. However at the worst prices the consumer consumes at A. In other words, rather than showing that trade is a welfare gain, it is shown that autarky is a welfare minimum. So imagine that a demon, who hates our consumer, can choose international prices to make him as badly off as possible. In minimizing the consumer's utility the demon is constrained by the fact that the consumer can optimize at whatever prices obtain. Also the demon cannot loot national resources, for which reason the consumer's expenditure can be as large as the value of national production. The programme is:

Maximize:

$$-U \tag{3}$$

subject to:

$$E(\mathbf{p}, U) - R(\mathbf{p}) \geq 0 \tag{4}$$

Having the demon maximize $-U$ is the same as writing for him a role in which he minimizes $U$. The use of the indirect functions automatically incorporates maximization on the part of both the consumer and producers in aggregate. The demon is like a chess player planning a move designed to cause maximum harm to his opponent's position. However the choice of that move must take into account that the opponent will select (or must be assumed to select) the best possible reply from his point of view. The situation is that of a game in which the demon must make Nash-optimal choice of the variables $(U, \mathbf{p})$. He selects those variables in the knowledge that given his choice the agent will choose optimal consumption and production levels given $\mathbf{p}$ and $U$. Those optimal agent replies are already written into the indirect functions. This is because the best reply on the part of the agent to any prices he may face is to maximize the value of production at those prices, and to minimize the cost at those same prices of attaining the given level of utility $U$.

Let the Lagrange multiplier for the constraint (4) be $\lambda$. Then the Lagrangean to be maximized is:

$$-U + \lambda \cdot [E(\mathbf{p}, U) - R(\mathbf{p})] \tag{5}$$

and the first-order conditions are for respectively $U$ and $\mathbf{p}$:

$$-1 + \lambda \cdot \frac{\partial E(\mathbf{p}, U)}{\partial U} = 0 \tag{6}$$

$$\lambda \cdot \left[E_{\boldsymbol{p}}(\mathbf{p}, U) - R_{\boldsymbol{p}}(\mathbf{p})\right] = 0 \tag{7}$$

The equation (6) just defines the multiplier $\lambda$ as the inverse of the marginal international money cost of additional utility. Equation (7) is the key. It says that the two vectors $E_{\boldsymbol{p}}(\mathbf{p}, U)$ and $R_{\boldsymbol{p}}(\mathbf{p})$ are exactly equal component by component. That is the same as stating that there is national self-sufficiency (i.e. autarky) with regard to each and every good. The demon does his worst by choosing international prices exactly equal to national autarky prices. The best definition of comparative advantage defines it as a difference between world equilibrium prices when the country of interest engages in trade and the autarky prices of the same country. Notice that this is not a small country definition. Thus the demon in doing his worst to a country denies it any comparative advantage whatsoever.

To extend the analysis to many agents, let individual agents $i = 1, \ldots, N$ be characterized by individual profit and expenditure functions $R^i(\mathbf{p})$ and $E^i(\mathbf{p}, U^i)$. Before trade prices are $\mathbf{p}$ and give an autarkic general equilibrium.

$$\sum_{i=1}^{N} \left[R^i_{\boldsymbol{p}}(\mathbf{p}) - E^i_{\boldsymbol{p}}(\mathbf{p}, U^i)\right] = 0 \tag{8}$$

where the $U^i$ values are those of the autarkic equilibrium. After trade world prices are $\mathbf{p}^0$ and:

$$\mathbf{p}^0 \cdot \sum_{i=1}^{N} \left[R^i_{\boldsymbol{p}}(\mathbf{p}^0) - E^i_{\boldsymbol{p}}(\mathbf{p}^0, U^{Fi})\right] = 0 \tag{9}$$

where the $U^{Fi}$ values are those of the free-trade equilibrium.

The following theorem is proved in the Mathematical Appendix. The proof is much the same as a standard proof of the Pareto efficiency of a competitive equilibrium for a closed economy. While not difficult, the proof takes up too much space to allow it to interrupt the flow of the argument here.

**Theorem 5.1:** *Relative to autarky free trade cannot be Pareto inferior.*

Theorem 5.1 contrasts sharply with the earlier demonstration that trade is gainful in a one-agent economy. In that case there is only gain, for the unique agent who must gain, or to be precise cannot lose. Here at least one agent among many, but it could be only one, may gain. That conclusion might come as a shock to one who has been exposed to simple pro-trade advocacy. That kind of argument often makes freerer trade sound like penicillin, a benign innovation that helps many greatly and harms no-one. Some have talked of trade as a win-win policy; a benefit to all and any. Whatever the facts of particular examples, economic theory does not teach that trade by itself is win-win. Probably no real-life change of any complexity benefits all without exception.

In excluding Pareto inferiority Theorem 5.1 is consistent with a situation in which the benefits from trade are concentrated and the costs are widespread. Yet it is also perfectly consistent with the opposite case in which a few lose but many gain. That was the picture of the abolition of the Corn Laws in mid-nineteenth-century Britain as painted by the abolition lobby, including notably David Ricardo. A similar case is made today by the more sophisticated advocates of increased trade. Economic theory says only that this too is possible. It might seem that the only general conclusion is the weak non-dominance result of Theorem 5.1; not everyone can lose. For a similar reason economic theory also fails to support the antithesis of win-win optimism: 'the rich get richer and the poor poorer' pessimism. General theory cannot adjudicate on the costs and benefits of trade. That requires careful empirical analysis of the particular example.

An easy point to make against the model of trade encapsulated in equations (8) and (9) is that it is static, while the world is dynamic. As it stands this argument is incorrect. The many goods and agents of the model could be dated, which would lend to the model a dynamic (strictly an intertemporal) aspect. The dating of goods is a familiar device of general equilibrium theory. Bread becomes bread delivered, or consumed, at various dates. The dating of agents might bring into account the yet unborn agent. For trade, just as much as many other economic policy choices, affects the as yet unborn. The key issue here is not what can be done formally, but rather what is realistic and useful. To take the unborn first, they cannot represent themselves directly in current markets, not even if those markets are present markets for future goods. So their interests inescapably can only be felt through the altruistic actions of the presently living.

Among the many problems posed by dated goods is uncertainty and how to deal with it. There is much evidence that typical consumers do not behave rationally under uncertainty. So the model of the rational agent comes under serious stress when uncertainty is involved. In addition futures markets can only cover the known and the standardized, and that must exclude the newly invented. Such considerations make it clear that the competitive model in its simple form has serious limitations. A more realistic treatment requires the inclusion of missing markets which makes for greater realism but also great complications. For a demonstration of how missing markets can change the gains-from-trade analysis, see Newbery and Stiglitz (1984). In conclusion two points may be noted. First, pointing out problems with the competitive model is easier than building a useful alternative. Secondly, the main problems are inherent in the competitive model in any form; they are not difficulties with its extension to international trade.

*2. Reform sequencing* Another example of the application of the indirect function approach is provided by the question of whether capital mobility is a good idea for a distorted economy. The distortion here will be the presence of tariffs, presumably arbitrary tariffs not justified by any second-best consideration. The problem goes back to Johnson (1967) whose diagrammatic treatment is generalized here. The model focuses the issue sharply, because what is analysed is a gift of capital, which must be employed in the home country. Obviously if a free gift of capital can be harmful, as will be shown to be the case, an inflow of return-seeking commercial capital would be even less favourable. This analysis also introduces for the first time a simple case of the mixed-price-quantity revenue function, which will be used extensively below.

The economy is a single-consumer economy, which makes welfare evaluation easy. World prices are fixed, so this is a small country, and their value is the vector $\mathbf{p}$. Domestic prices are $\mathbf{p} + \mathbf{t}$, where the tariff vector $\mathbf{t}$ is not proportional to $\mathbf{p}$, when it would be without effect. The key innovation here is the use of the revenue function:

$$R\,[\mathbf{p} + \mathbf{t}, K] \tag{10}$$

where $K$ in the national capital stock, and the dependence of maximized value on the quantity of capital is made explicit. Of course other factors may be as important as capital. However as these quantities do not vary they are not shown explicitly as arguments of (10). The trade balance is always zero in this model, so that:

$$\mathbf{p} \cdot \left[R_{\mathbf{p}}(\mathbf{p} + \mathbf{t}, K) - E_{\mathbf{p}}(\mathbf{p} + \mathbf{t}, U)\right] = 0 \tag{11}$$

where the subscripts $p$ denote partial derivatives with respect to the first arguments of these functions evaluated at $\mathbf{p} + \mathbf{t}$. Differentiating (11) totally with respect to $K$ gives:

$$\mathbf{p} \cdot \left[R_{\mathbf{p}K}(\mathbf{p} + \mathbf{t}, K) - E_{\mathbf{p}U}(\mathbf{p} + \mathbf{t}, U)\frac{dU}{dK}\right] = 0 \tag{12}$$

The term $\frac{dU}{dK}$ can be negative if either but not both of the other terms within the square brackets are negative when multiplied by world prices. The first such term $\mathbf{p} \cdot R_{\mathbf{p}K}(\mathbf{p} + \mathbf{t}, K)$ can be negative if additional capital expands capital-intensive sectors already overexpanded by tariff protection, so that the value of production at world prices is lowered by a gift of extra capital. The second term $\mathbf{p} \cdot E_{\mathbf{p}U}(\mathbf{p} + \mathbf{t}, U)$ can be negative if the value of the increase in demand associated with a higher level of utility, which must be positive at domestic prices, is negative at world prices.

The argument just completed indicates that the problem of how to sequence reform is difficult, and that wrong choices can cause harm. The model shows that liberalizing capital movements first when the economy is distorted by tariffs may not be a good idea. Suppose instead that trade is liberalized first with capital movements restricted. Must that be a good thing? The answer is: not necessarily. Capital inflows are a kind of trade, trade with an inescapable intertemporal aspect. So asking whether tariff liberalization with capital restriction must be good is formally similar to asking whether removing some tariffs while others remain in place is always a good thing. There the well-known conclusion is that it all depends. Imagine that the restriction of capital imports, say by policy hostile to foreign capital, has caused the labour-intensive sector to be overexpanded and the capital-intensive sector to be too small. Then it is possible that tariff reductions would further expand the labour-intensive sector and worsen the existing distortion.

Some commentators on economic reform have proposed a simple way round the difficulties exposed above. They advocate a 'big-bang' solution to the problem, Everything should be liberalized immediately. Would that solve the problem? Plainly it would if it implied an instantaneous switch to the first-best equilibrium. Sadly things are seldom that simple. In real life adjustment is sluggish. When prices jump to new values, agents behave as if prices had only adjusted partially. That inescapably entails that the problem of reform sequencing reasserts itself. Suppose, for instance, that capital is 'quicker off the mark' once reform is implemented, so that perfect capital mobility becomes a fact soon after reform, while production and consumption decisions take some time to respond. Then we are back with a situation somewhat similar to the original difficulty rehearsed at the beginning of this section. Capital mobility is not necessarily a good thing when the economy is distorted by tariff-affected prices. We just substitute a formally equivalent conclusion. Capital mobility is not necessarily a good thing when the economy is distorted by partial adjustment.

The exercises above are illustrative of method more than final policy conclusions. Let the discussion conclude therefore on a more positive note. If the analysis undermines some simple rules for reform sequencing, it does not thereby imply the conclusion that anything is as wrong as anything else. In any particular case it may well be possible to work out a good approximation to an optimal transit from distortion to liberalization. Our arguments suggest that a somewhat gradual adjustment on all fronts will often be the best choice. The detailed solution must be left to the policy-maker faced with the facts of a particular case. If that policy-maker understands that simple rules are suspect it is more likely that his decisions will be good.

## 5.3 The Simple HOS Model

The Hecksher-Ohlin-Samuelson (HOS) model is to the international economist what a sharp knife is to a chef. It is a tool used all the time because nothing else does so much so well. In a skilled hand the chef's knife cuts, chops, trims, and shapes. In a clumsy hand it causes injury. Similarly, used well the HOS model is an endlessly flexible device for depicting international trade between countries whose comparative advantage differences are modelled, rather than simply assumed. Used clumsily it is as bad as any other model.

It reflects well on the model that its application has tracked long-term developments in the world economy according to how its basic formal structure is filled out. At its birth it was directed to trade between the New World and the Old World, and the factor endowments that differentiated the trading regions were relative supplies of labour and land. Later when cross trade in manufactures became of central importance, the factors became capital and labour. Leontief's famous test of the model, on which see Chipman (1966) and Feenstra (2004), was designed to see whether US exports are capital intensive in comparison with its imports. The point is that the US is taken to be well endowed with capital. Capital has become something with which a nation is endowed, much like land. In the Leontief exercise land, and also climate, is pushed off stage, although in fact the importance of these two actors for US trade cannot be ignored. The consideration of US exports of timber and imports of coffee makes that clear.

The Atlantic trade which the original model was designed to explain was accompanied by a high level of labour migration to the labour-scarce New World. A good part of this migration had political causes (Jews fleeing Czarist oppression), or economic catastrophe more than economic migration (famine in Ireland and elsewhere). The model depends upon some barriers to factor migration, for without these, differences between relative factor endowments would be obliterated. If factor migration is partial the HOS model is still relevant. In that case, as claimed by Heckscher at the beginning, trade substitutes for migration. Thus, rather than migrating to the Americas, European labour produces labour-intensive manufactures and sends those across the sea. In return the Americas produce land-intensive food which is shipped in the opposite direction. Factor prices in the two regions are brought closer together and the economic motivation for migration is weakened.

Thus incomplete economic migration of factors plays an essential part in the HOS model. Where the transnational migration of labour is concerned this specification is reasonable. What about the transnational migration of capital? For the greater part of the post-war years capital mobility was highly restricted. The restrictions encompassed formal legal restrictions. These have relaxed over time but have never wholly disappeared. They have certainly declined in significance in relation to the non-formal barriers to capital mobility. The greatest of these is undoubtedly imperfect information. It is always easier to know the local than to know the distant. And typically agency is not the answer because the missing knowledge that the agent might provide is needed to select the agent. Of course economies of scale apply to information as to other areas. So the large international firm can afford to expend the resources needed to cut through the undergrowth of imperfect knowledge and to take advantage of a high return in a strange and distant country. It was large companies, such as motor-vehicle builders, or electronics producers, that originally opened up the major possibilities for profitable international investment.

For reasons sketched above, the HOS model with factors capital and labour has retained great relevance over many years. However capital has increased its international mobility over time, and this has forced the contemporary economic theorist to choose between two options. One choice is to posit that capital is as mobile as desired, that this effectively equalizes the rate of return between countries, and that international differences in technology are what explain variations in real wage rates across countries. In solving one problem that creates another. Why is technology not as mobile, or even more mobile, than capital? Lucas (2002) provides an excellent discussion of the issues. An alternative choice is to posit again perfect capital mobility but to let its role be taken on by another immobile factor. This can be done by having two kinds of immobile labour, skilled and unskilled. That route is taken by the Krugman-Wood model, which is examined below.

## 5.4 Equilibrium in the Simple HOS Model

The model has two sectors each of which has its own constant-returns production function with two inputs. The sectors are called food and machines. The factors are capital and labour. Plainly it makes no difference to the formal analysis what are the inputs and outputs, or how they are labelled. Let the fixed factor supplies of capital and labour be respectively $K_0$ and $L_0$. Let output prices for respectively food and machines be $p^f$ and $p^m$. For such a simple model it is easy to write down the revenue function.

$$R\left(p^f, p^m\right) = \text{Max}_{k,l}\left[p^f F^f(k,l) + p^m F^m(K_0 - k, L_0 - l)\right] \tag{13}$$

where the superscripts on the production functions $F$ show the sector, and $k$ and $l$ are the factors employed in the food sector. Equation (13) shows a feature which has already been exploited in the treatment of reform sequencing above, is general, and will be made use of further below. The maximum output value given product prices depends upon total factor supplies. Thus the left-hand side of (13) can be written:

$$R(p^f, p^m, K_0, L_0) \tag{14}$$

More properties of the function (14) will be elucidated below. We may note immediately that it is homogeneous of degree one in the goods prices, holding factor supplies constant, and in the factor supplies, holding goods prices constant. Then when partial derivatives are single-valued we can apply Euler's theorem and the basic properties of indirect functions to deduce:

$$R(p^f, p^m, K_0, L_0) = p^f y_f + p^m y_m = w_K K_0 + w_L L_0 \tag{15}$$

where the $y$ values are optimal outputs of the two sectors, and the $w$ values are shadow prices of the factors indicated by the respective subscripts. Equation (15) shows an ideal accounting balance. Maximized revenue is the sum at given output prices of the value of both optimal output levels. This sum in turn is equal to the value at optimal factor shadow prices of the two given factor supplies.

Figure 5.2 illustrates the form of the revenue function in the ideal case for the HOS model, with factors capital and labour, and when there are no capital-intensity reversals. The space shown in the figure is a Cartesian space of factor quantities. The curves are isoquants for capital-intensive

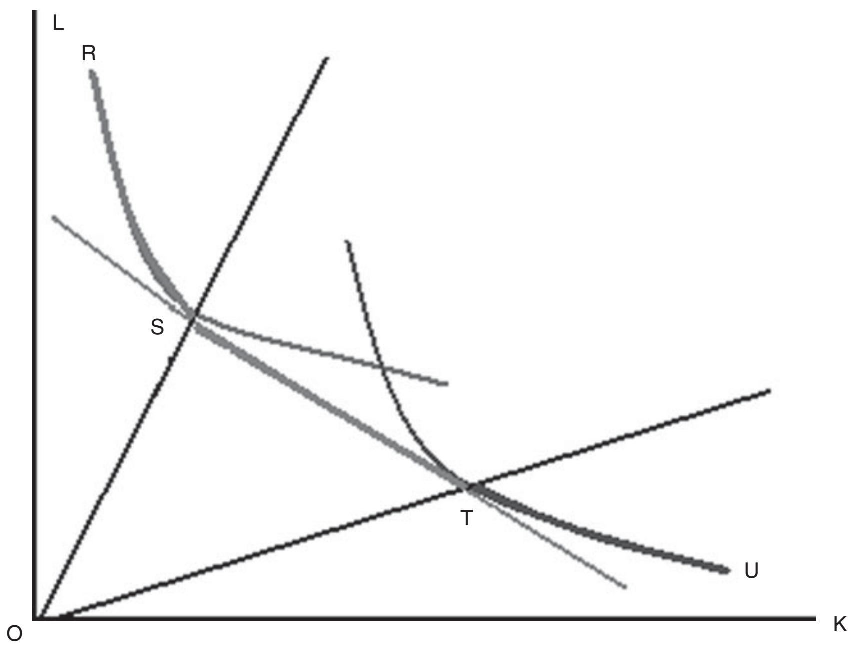



**Figure 5.2.** The HOS Revenue Function

machines (TU and its extension) and labour-intensive food (RS and its extension). The isoquants drawn correspond to \$1m worth of total output for the given output prices. All factor pairs on the part-non-linear, and part-linear, curve RSTU, or its extensions beyond the end points R and U, can produce \$1m worth of output. With factor supplies anywhere on the line ST an appropriate average of the supply points S and T can be chosen to fully employ those factors by mixing the output levels of food and machines as required. When the ratio of factor supplies lies outside the cone shown by the two rays from O, the factors can still produce \$1m worth of output, but now this is achieved by employing all the factors in one sector only. This is the case of the specialization of production in one product only.

Notice an important property of the locus RSTU: it is everywhere differentiable. This follows from three of its properties. First, the section ST is linear, and therefore evidently differentiable. Secondly, the sections RS and TU are made up of parts of isoquants, assumed to be differentiable. Finally, and critically, at the points of connection S and T the components of RSTU are *smooth pasted*. This means that the slope of RS at S is the same as the slope of the straight line ST. This follows from the fact that ST is the common tangent to the two isoquants. Similarly the slope of TU at T is the same as the slope of the straight line ST.

Maximizing outputs are shown by the intersection of a ray showing the relative factor endowments of the economy with the locus RSTU. Because RSTU is differentiable it follows that: *value-maximizing outputs are differentiable functions of factor supplies*. It is natural to call these changes *Rybczynski effects* as they are just the changes analysed in the famous paper by Rybczynski (1955).

A basic symmetry condition is explored in the mathematical appendix. This is similar to the familar symmetry property of consumer-demand functions, according to which $\frac{\partial x_i}{\partial p_j} = \frac{\partial x_j}{\partial p_i}$ where the $x$s are demands and the $p$s are prices. The response of outputs to factor supplies is obtained by differentiating $R(p^f, p^m, K_0, L_0)$ first with respect to prices and then with respect to factor quantities. Differentiating the same function first with respect to factor quantities and then with respect to prices gives the same quantitative result. Now however the resulting term measures the response of a factor shadow price to an alteration in an output price. This is a *Stolper-Samuelson effect*, so called after Stolper and Samuelson (1941–2).

## 5.5 Cross Partial Derivatives

The differentiability of outputs with respect to factor supplies implies that a second-order partial derivative such as:

$$\frac{\partial^2 R(p^f, p^m, K_0, L_0)}{\partial p^f \partial K_0} \tag{16}$$

is uniquely defined. And in the case of the basic HOS model it will equal:

$$\frac{\partial^2 R(p^f, p^m, K_0, L_0)}{\partial K_0 \partial p^f} \tag{17}$$

This symmetry property, which appears puzzling at first sight, has a clear intuition. If the price of food increases slightly, with all other prices constant, the only effect on the shadow price of capital is the rate at which extra capital leads to more food (or less food as the case may be) being produced. Below in the section of this chapter entitled Generalizations it will be seen that this last result is perfectly general, and always equally intuitive. Here is an instance of the point.

**Proposition:** *The additional quantity of pianos which the economy produces when there is a small increase in the supply of french-polishers is equal to the increase in the shadow price of french-polishers when there is a small increase in the price of pianos.*

The intuition is the same as before. The shadow price of french-polishers is the market value of the marginal change in all outputs when there is a marginal increase in their supply. The price of pianos affects only the piano component of this vector. Hence the result.

We can use the equality of (16) and (17) to show conveniently a crucial point in the original Stolper-Samuelson analysis. This is the famous *magnification* property. When protection is removed from the labour-intensive good the price of that same good falls and the wage rate falls even more proportionately than the fall in the output price. Let the price of the labour-intensive good in terms of the other good be $p$. The price of the other (capital-intensive) good is 1. The effect on the wage of a small change in $p$ is then:

$$\frac{\partial^2 R(p, 1, K_0, L_0)}{\partial L_0 \partial p} \tag{18}$$

and this is equal to:

$$\frac{\partial^2 R(p, 1, K_0, L_0)}{\partial p \partial L_0} \tag{19}$$

which measures the effect on the output of the labour-intensive good of an increase in the supply of labour. Provided only that both outputs are produced, (19) is positive and the increase in output is more than proportional to the increase in labour supply. The last property follows from the fact that when more labour expands one sector the other sector must contract to provide the expanding sector with the additional capital that it requires. Then (18), which takes the same value, implies that a rise in $p$ raises the wage rate more than proportionately, and that is magnification.

What happens to factor prices when factor supplies change? That involves evaluating terms such as:

$$\frac{\partial^2 R(p_f, p_m, K_0, L_0)}{\partial L_0^2} \tag{20}$$

In that particular case the term (20) is less than or equal to zero. This follows immediately from $R(\cdot)$ being concave in the factor quantities. Note however that this same term may be zero. That will be the case when factor supplies are on the line ST.

## 5.6 Key Conclusions

The time has come to gather together some key conclusions concerning the simple HOS model. Its implications for inequality are examined below in a section devoted to that issue. Here we focus on what the model implies for convergence of prices through trade, especially the prices of factors. There are no large new insights here. The model is old and well used and anyone who thinks that he has a big new result from the model has surely made a mistake. That said, the emphasis here is different from that which is sometimes encountered in the literature. There has been a tendency to concentrate on the canonical version of the 2X2 model, and then to suggest that matters are much more complex and difficult in broader models. In the canonical model there are no factor-intensity reversals and all countries are diversified in production. Then there is factor-price equality between all such countries. In Figure 5.2 all countries are found somewhere on the linear segment ST.

One can characterize these same conclusions in such a way that broader generalization is immediately possible.

- If two countries are sufficiently similar with regard to their factor supplies, and share a common technology, they will either have exactly the same factor prices, or closely similar factor prices. In particular every country diversified in production is in a cone which contains other possible countries with different factor supplies but the same factor prices. The closely similar qualification above allows for two countries on different sides of a point like S or T in Figure 5.2.
- If the gain-loss evaluation of a change in output prices is calculated from the point of view of pure factor interests, then such a computation always shows that price changes, and therefore trade liberalization, are conflictual. That means that there are both gainers and losers. This is stronger than the demonstration above that trade is a Pareto improvement. That is consistent with all gaining. Now we see that if individuals are associated with the ownership of particular factors in the simple HOS model, one factor or other will lose.
- Changes in factor supplies cause changes in outputs. In the simple HOS model with diversified production, the change of outputs takes place at constant factor prices. In other cases factor prices change.

## 5.7 The Krugman-Wood Model

It has been noted already that the designation of the two factors in the simple HOS model has varied over time. Where once the two factors were land and labour, they later became capital and labour. More recently, in an attempt to pinpoint a crucial force operating on contemporary international trade, the two factors have become skilled and unskilled labour. The idea is that the most important distinction between the 'North' (the rich industrial countries) and the 'South' (the less developed countries) is their relative endowments of highly skilled educated labour on the one hand, and basic labour with few formal skills on the other. Adrian Wood promoted this approach from the late 1980s. See Wood (1994). Paul Krugman has also used the approach. See for instance Krugman and Lawrence (1993). Certainly other writers have made use of the model, but the short title Krugman-Wood model is convenient.

While most would agree that skill differences are of great importance in today's world, what about other essential differences between trading nations? In particular differences in national endowments of capital and variations of technology come to mind, Wood has argued vigorously that capital is fully mobile in the modern world, and that technology is equally footloose. A major problem with supposing perfect capital mobility is that it seems to depend on perfect information, which is a most questionable assumption. It is possible also that some countries have better technological knowledge, or better delivery of technology, than others. Again variations in economic environments, a concept that will be developed below, inhibit the free mobility of the application of technology, if not of pure technological knowledge. These questions will receive detailed examination in later chapters. Chapter 6 will establish a basic framework which will permit trade with skilled and unskilled labour to be analysed together with imperfect capital mobility.

Notwithstanding the doubts just expressed, a version of the simple HOS model with just two immobile factors, skilled and unskilled labour, is attractive. Not only does it permit new insights to be obtained from a model which in its formal aspect is old and familiar; it also produces conclusions that make sense and which indicate directions for policy. In a standard application of the model, two countries (actually groups of countries), North and South, are differently endowed with skilled and unskilled labour, the North being relatively skilled-labour-abundant. There are two products, called for convenience bicycles and computers. The production of bicycles is unskilled labour-intensive. The production of computers is skilled labour-intensive.

Now trade between the two countries becomes easier, so that their relative domestic goods prices move closer together. If the two countries are diversified and producing both the products, their respective ratios of skilled to unskilled wage rates will move closer together. That ratio will rise in the North, as bicycles become cheap relative to computers. The same ratio will fall in the South, as the improved trading opportunities raises the relative price of bicycles to computers. Note that under the maintained assumption of perfect capital mobility, the international flow of capital could even be from South to North. That would happen if the computer sector were to be capital-intensive as well as skilled-labour intensive. This analysis holds out the hope of explaining the growing inequality of incomes that has characterized the North over recent decades, as well as, in an ambitious extension, the so-called 'reverse capital-flow paradox'.

## 5.8 Trade and Inequality in the Competitive Model

A brief examination of the Krugman-Wood model has brought the dry technical analysis of this chapter face to face with the question of income inequality. Inequality is a central issue for this volume. So now is a good moment to look back and ask what competitive trade theory has to say about inequality. To return to a point with which this chapter opened, there are two most different approaches to trade theory. General equilibrium theory is, as its name implies, exceedingly general. Even so it delivers a clear and strong message where inequality is concerned. It says that what individuals get from a market economy is a function simply of the market value of the resources that they own. Those who come armed with valuable resources, be those physical assets or marketable skills, do well. Those who come with little get little. In line with the rich generality of general equilibrium theory the equilibrium may not be unique. So what individuals get from a market economy may depend upon happenstance.

In Chapter 4 we have already seen how the rule that what the agent gets springs from what the agent has applies even to a dynamic growing economy. The market will not iron out initial inequalities in starting positions, not even in the long run. From these ideas there follows a clear corollary: the effect of any change on inequality is determined solely by how that change affects the market value of the resources commanded by specific individuals. As the opening up of trade for whatever reason, reduced protection or lower transport costs, is an instance of economic change, it follows the same rule. Trade can increase inequality if it raises the market value of resources owned by the already well-off (the skilled in the North in the Krugman-Wood model). Trade can reduce inequality if it raises the market value of resources owned by the badly-off (the unskilled in the South in the Krugman-Wood model).

It is plain that there can be no general mathematical demonstration that freer trade is good for either the rich or the poor. Yet many students of particularly less developed countries will feel that this argument is too abstract and general to capture an important feature of many actual situations. Very often, it will be claimed, less developed countries (LDCs) have adopted policies that restrict trade in a manner that is especially costly to the poor. For this reason the position in the South as depicted by the Wood model is singularly appropriate. Why might that be the case?

The answer would come from the political economy of protection. In most LDCs mechanisms of political accountability, leave alone formal democracy, are weak if not absent. That leaves policy formation in the control of powerful and to a great extent non-accountable groups, and these are typically the rich. It is no surprise then if the type of protection and trade restriction chosen is not too onerous for the rich and falls with a heavy weight on the poor. If that is the case then it follows that a reversal of restrictive policy will impose a cost on the rich and come as a relief to the poor.

That is a neat argument, but is it not met by a telling objection? Suppose that political economy drives trade policy. If restraints on trade in the North harm the rich particularly, as the other side of the Krugman-Wood model indicates, then why do we see some powerful protection in the North, notably for agriculture and fibres? Possibly this asymmetry, that the rich choose policy to suit themselves in the South, but do not do so in the North, may be explained by political accountability again. In the North democracy, in the messy imperfect form that it takes everywhere, is the norm. For that reason protection is often a response to populist pressure or vote seeking. It may then be aimed, or be supposed to be aimed, to benefit certain poor interests. Agricultural protection in the North is the perfect example of this story. It is meant to benefit ‘poor’ farmers, and by protecting these it necessarily anti-protects other activities, and in doing that it impacts negatively on the interest of rich groups. If ‘big business’, meaning here the interests of rich capitalists, really did run things in industrial countries, we would not see heavy agricultural protection.

That said, note that agricultural protection in the North, while it is not good for the rich, is strongly regressive in its effects, so that the cost that it entails is paid particularly by the poor. The single mother in the North, struggling to feed her family, feels the weight of agricultural protection far more sharply than does the rich company executive deciding what to order for his dinner in a fashionable restaurant.

## 5.9 Generalizations

Many of the results shown above do generalize. Some of these generalizations are explored in the appendix. Two prominent instances are listed here.

1. It is generally true that for a full-rank case any national equilibrium will have close neighbours (whether occupied by any concrete country or not) which share techniques and factor prices, though not of course production levels.

2. It is generally true that any small change in factor supplies is almost always associated with a unique vector of output changes (with positive and negative elements).

## Mathematical Appendix 5.1

*Indirect Functions*

We have already met the *Revenue Function*:

$$R\,[\mathbf{p}] \tag{A.1}$$

which is the maximum net profit when prices are the vector $\mathbf{p}$. $R\,[\mathbf{p}]$ is homogeneous of degree one in $\mathbf{p}$, and weakly convex. As a convex function it is continous and everywhere has directional derivatives. Directional derivatives means that a convex function has left-hand and right-hand derivatives everywhere. See Eggleston (1963). Then a concave function, which is the negative of a convex function, also has directional derivatives. For a convex function, naturally, the left-hand derivative is less than or equal to the right-hand derivative.

This result allows us to talk of the derivatives of a convex function, recognizing that such derivatives may be ranges of values rather than single numbers. That permits a simple statement of the most important property of $R\,[\mathbf{p}]$. Its partial derivatives are profit-maximizing outputs. Then of course when left-hand and right-hand derivatives are unequal, profit-maximizing outputs are non-unique. The partial derivatives of $R\,[\mathbf{p}]$ are denoted $R_{\mathbf{p}}\,[\mathbf{p}]$ which is a vector of the same dimension as the number of commodities, with a positive element for a net output and a negative element for a net input.

Another indirect function is the *Expenditure Function*:

$$E\,[\mathbf{p},U] \tag{A.2}$$

which gives the minimum cost of buying a basket of goods with utility at least as large as $U$ when prices are $\mathbf{p}$. The function $E\,[\mathbf{p},U]$ is increasing in $U$ and is concave in $\mathbf{p}$. The partial derivatives of $E\,[\mathbf{p},U]$ are denoted $E_{\mathbf{p}}\,[\mathbf{p},U]$ and are demands when prices are $\mathbf{p}$.

Where the question of the second-order partial derivatives of a convex function is concerned there is no general result for an arbitrary convex function. While a convex function always has first-order directional derivatives, it may not have second-order partial derivatives, not even with the weak directional definition. Nevertheless, such second-order directional derivatives are well defined for trade models of interest, and in those cases they satisfy the Young's theorem property so that the order of partial differentiation makes no difference to the result. All this is most readily appreciated through the basic 2X2 HOS model as the treatment above has shown.

### *Gains from Trade*

**Theorem 5.1:** *Relative to autarky free trade cannot be Pareto inferior.*

**Proof:** Let $U_i^A$ be the utility level of agent $i$ under autarky, and $U_i^F$ the utility of the same agent under free trade. If the theorem is false there must be a case in which:

$$U_i^F \leq U_i^A \tag{A.3}$$

with strict inequality for at least one agent. Then (A.3) implies:

$$E^i(\mathbf{p}^0, U_i^A) \geq E^i(\mathbf{p}^0, U_i^F) \tag{A.4}$$

and because of (A.4) with at least one strict inequality:

$$\sum_i E^i(\mathbf{p}^0, U_i^A) > \sum_i E^i(\mathbf{p}^0, U_i^F) \tag{A.5}$$

Now (A.5) and homogeneity of an expenditure function imply:

$$\sum_i^{0} \mathbf{p} \cdot E^i_{\mathbf{p}}(\mathbf{p}^0, U_i^A) > \sum_i^{0} \mathbf{p} \cdot E^i_{\mathbf{p}}(\mathbf{p}^0, U_i^F) \tag{A.6}$$

With free trade the balance of payments condition is:

$$R(\mathbf{p}^0) - \mathbf{p}^0 \cdot \sum_i E^i_p(\mathbf{p}^0, U_i^F) = 0 \tag{A.7}$$

Therefore from (A.6) and (A.7):

$$\sum_i^{0} \mathbf{p} \cdot E^i_p(\mathbf{p}^0, U_i^A) > R(\mathbf{p}^0) \tag{A.8}$$

As $\sum_i E^i_{\mathbf{p}}(\mathbf{p}^0, U_i^A)$ is a feasible production, (A.8) contradicts the definition of a revenue function. □

### *Stolper-Samuelson and Rybczynski*

Consider the technology defined by the revenue function:

$$R(\mathbf{p}, \mathbf{z}) \tag{A.9}$$

where $\mathbf{p}$ is the vector of goods prices, and $\mathbf{z}$ is the vector of factor quantities. The optimal production of the economy, $\mathbf{y}$, is given by:

$$\mathbf{R_p}(\mathbf{p}, \mathbf{z}) \tag{A.10}$$

The shadow prices of the non-tradeables (factors) are given by:

$$\mathbf{R_z}(\mathbf{p}, \mathbf{z}) \tag{A.11}$$

In (A.10) and (A.11) the printing of $R$ bold reminds us that this is now a vector of partial derivatives with respect to $\mathbf{p}$ or $\mathbf{z}$. The changes in production levels caused by marginal changes in factor supplies (Rybczynski effects) are given by elements of the matrix:

$$\begin{bmatrix} R_{p_1 z_1} & R_{p_1 z_2} & \cdot & R_{p_1 z_n} \\ R_{p_2 z_1} & R_{p_2 z_2} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ R_{p_n z_1} & \cdot & \cdot & R_{p_n z_n} \end{bmatrix} \tag{A.12}$$

The changes in factor shadow prices caused by marginal changes in goods prices (Stolper-Samuelson effects) are given by elements of the matrix:

$$\begin{bmatrix} R_{z_1 p_1} & R_{z_2 p_1} & \cdot & R_{z_n p_1} \\ R_{z_1 p_2} & R_{z_2 p_2} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ R_{z_1 p_n} & \cdot & \cdot & R_{z_n p_n} \end{bmatrix} \tag{A.13}$$

With strong differentiability, by Young's Theorem, these two matrices transposed are identical.

### *Factor-Price Equalization*

Generalizing the standard Heckscher-Ohiln result to the square case with more than two factors and goods is feasible as a mathematical project, yet awkward from the point of view of economic intuition. Let:

$$c^i\left(w^1, w^2, \ldots, w^n\right) \tag{A.14}$$

for $i = 1, \ldots, n$, be unit cost functions. If factor price equalization does not apply there must exist values $w^1, w^2, \ldots, w^n$ and $w'^1, w'^2, \ldots, w'^n$ different from each other such that:

$$c^i\left(w^1, w^2, \ldots, w^n\right) = c^i\left(w'^1, w'^2, \ldots, w'^n\right) \tag{A.15}$$

all $i$.

Nikaido (1972) gives a sufficient condition to exclude (A.15) in terms of bounds on the determinants of the leading principle minors of the matrix of unit factor input requirements. See Feenstra (2004, p 65–70) for more details.

The Nikaido conditions are not particularly interesting and this for two reasons. First, the high-dimension square case is, as it appears to be, quite special. It may well be that given $N$ factors many countries will produce exactly $N$ goods. Yet these need not be the same $N$ goods in each case, and the factor-price equalization result will apply only between countries which produce the same menu of goods. Secondly, the conditions lack the simple intuitive interpretation of the 2X2 case. There it indicates no factor-intensity reversals. But the general Nikaido feature in the square $NXN$ case is a spare mathematical property. Just imagine that you are involved in an argument over whether this property will obtain. What kind of points would you make? What evidence would you bring to bear? You might even have the print-out of an observed $A$ matrix, but that is only a point observation, and if that particular matrix should pass the test what does it prove?

The following theorem, which is a direct generalization of the discussion above of the simple HOS model, is perhaps of more interest. A full-rank equilibrium is one in which all goods are produced and all factors are fully employed. The definition does not require that the number of goods and factors be equal, although that may indeed be the case.

**Theorem 5.2:** *Let $E$ be a full-rank equilibrium with given product prices $\mathbf{p}$. Let the activity levels of the processes producing the outputs be $\mathbf{x}^1$, and let the total factor inputs be $\mathbf{z}^1$. Let $\mathbf{A}$ be the possibly non-square matrix of unit factor inputs per unit output. If $\mathbf{z}^2$ is a non-negative vector which solves the equation:*

$$\mathbf{z}^2 = \mathbf{A} \cdot \mathbf{x}^2 \tag{A.16}$$

*where $\mathbf{x}^2$ is non-negative, then $\mathbf{x}^2$ maximizes $\mathbf{p} \cdot \mathbf{x}$ subject to factor supplies $\mathbf{z}^2$.*

**Proof:** It must be the case that $E$ defines a linear programme:

$$\text{Max } \mathbf{p} \cdot \mathbf{x} \tag{A.17}$$

subject to:

$$\mathbf{A} \cdot \mathbf{x} \leqq \mathbf{z}^1 \tag{A.18}$$

to which a solution is $\mathbf{x}^1$. If $\mathbf{x}^1$ is not a solution to (A.17)–(A.18) then higher value can be obtained from the same factors either by altering $\mathbf{x}$, or by using a feasible linear activity not included in the rows of $\mathbf{A}$. In either instance $E$ is not an equilibrium, contrary to assumption.

The solution $E$ satisfies:

$$\mathrm{A} \cdot \mathrm{x} = \mathrm{z}^1 \tag{A.19}$$

and:

$$\mathrm{w} \cdot \mathrm{A} = \mathrm{p} \tag{A.20}$$

where $w$ is a vector of shadow prices.

Now, by definition, and equation (A.16):

$$\mathbf{z}^2 = \mathbf{A} \cdot \mathbf{x}^2 \tag{A.21}$$

and this together with (A.19) and the linear programming duality theorem implies that $\mathbf{x}^2$ maximizes $\mathbf{p} \cdot \mathbf{x}$ subject to factor supplies $\mathbf{z}^2$, as required. $\square$

# 6

# High-Dimension Models[1]

> With three or more factors of production it is certainly not necessary that the result of trade is to make the ratios of factor prices in the respective countries more closely approach unity. Some may do so, but others may diverge depending upon complicated patterns of complementarity and competitiveness.
>
> (Stolper and Samuelson 1941–2: 72)

## 6.1 The Structure of High-Dimension Trade Models

In Chapter 5 we saw how the Heckscher-Ohlin-Samuelson trade model (HOS for short) has retained its popularity and usefulness for analysing international trade. It has sometimes seemed necessary to adapt it for such application; the work of Paul Krugman and Adrian Wood provides leading examples of how that might be done. See Krugman and Lawrence (1993) and Wood (1994). The analysis of Chapter 5 indicates that some generalizations of HOS are readily available. We saw for instance that there is a general Rybczynski result, which applies to a small country facing fixed prices for tradeable goods. This theorem associates a vector of output changes, both negative and positive, with a small change in factor supplies. The same analysis indicates how factor prices will be unaffected by small changes in factor supplies, provided that the vector of factor supplies lies within a cone of diversification, just as in the standard 2X2 HOS model.

It would be an illusion to suppose that this great leap of generality comes without a cost. We obtain some results which evoke the HOS

[1] The central analysis of this chapter is based on Bliss (2003*a*). The argument is reworked and extended. The application is new.

model, yet on closer inspection they can be seen to be only faint images of the clear and robust results with which 2X2 HOS theory rewards us. Thus factor-price equalization can be demonstrated, at least for the square case when the number of factors and goods is equal. However the axiom that supports the result lacks a natural economic interpretation, unlike the factor-intensity condition of HOS theory. Similarly, it is plain that if an increased supply of IT specialists requires increases in specific quantities of the outputs of various goods, then the production of those goods in the said proportions may be said to be intensive in the use of IT specialists. Then, as has been seen, should that particular basket of goods increase in tradeable value, then so will the price of the services of IT specialists. What has been lost here is the simple direct intuition of 2X2 HOS theory. There one good is associated uniquely with one factor, as when agriculture is labour-intensive. It is not without reason that economists cling to the root 2X2 model.

Yet that model, for all its convenience, often generates doubts and dissatisfaction. Which precisely are the two goods? Which are the two factors? Is it land and labour, as with Heckscher's original analysis? Or is it capital and labour as most textbooks state? If the distinction between skilled and unskilled labour is of crucial importance in the modern world, does that require a three-factor analysis? Or can we follow Wood and assume perfect capital mobility, dropping influences that may spring from differential endowments of physical capital? If we go beyond 2X2 we face the issue of squareness versus non-squareness. Will the number of factors and goods be expanded in parallel, or will one be larger than the other? As the fine survey by Ethier (1984) makes clear, high-dimension results mostly depend upon the imposition of some particular structure. Even 2X2 HOS theory needs structure, in the form of unambiguous factor intensity. When we go into higher dimensions more structure and less generality is required. It seems that there are no models available which are both truly general and also useful. So the choice of a model involves pragmatic compromise between the competing demands of realism and utility.

This chapter makes extensive use of a toy model. What is a toy model? It is a simple low-dimension model built for a particular purpose, with no pretensions to mathematical generality. The HOS model started life as a toy model, but so much has it become the reference model of trade theorizing that it is hard to see it today as a toy model. Here the model is formally the same as that exposed in Bliss (2003*a*). In that paper the focus is on the development of real wages in Britain during the second half of the nineteenth century. The same toy model structure lends itself readily to the analysis of contemporary situations, including trade between countries differentially endowed with both human and physical capital. Wood (1994) and Barro, Mankiw, and Sala-i-Martin (1995), again BMS, are each models with three factors and two goods (Wood), or one good (BMS). The three factors are physical capital and two other inputs. These are labelled either unskilled labour and skilled labour (Wood), or labour and human capital (BMS). The labels chosen are significant. Unskilled versus skilled labour encourages a relatively short-run view of factor supplies in which the balance of the two types of labour is fixed. This is appropriate for Wood's concern with current trading equilibrium in the world economy. Human capital indicates that labour skill can be and will be accumulated, just like physical capital. And indeed the BMS paper is concerned with the long run and asymptotic convergence.

With two three-factor models in front of us, it seems that the complexities that come with three factors can be surmounted. On closer inspection it will be seen that each of these models prunes the complications that a three-factor specification implies, so that in effect only one or two of the factors play active roles. Wood assumes that physical capital is perfectly mobile, with the consequence that his two unit-cost functions include only two prices which vary by country: wage rates for skilled and unskilled labour. To all intents and purposes we have a two-factor model. In the BMS model the input of labour is constant, so that this factor adds no extra complexity. Furthermore, physical capital is again perfectly mobile, as it represents perfect collateral. Then, as was shown in Chapter 4, only one type of capital, human capital, is subject to an optimal accumulation condition. All the intricate analysis in the BMS model is concerned with one factor: human capital.

In general when the HOS model is expanded by including more factors one has to restrict its generality so as to keep it manageable. This is done in the Ricardo-Viner model. Jones (1971) provides one of the most influential expositions. The model has three factors and two goods. Two of the factors are specific to one sector (a different sector for each such factor). Jones assumes that two of the factors are permanently wedded to their two separate sectors. Many later expositions of the model take it that the long-run equilibrium is that described by the 2X2 HOS model. However, following a shock the immobile factor (often taken to be capital) is misallocated between the two sectors. Then the specific quantities of immobile capital can be taken temporarily to be entirely different factors, just as Jones assumes.

The two specifications lead to exactly the same model in the short run. For a lengthier explanation of the relation between the Ricardo-Viner and HOS models see the survey by Jones and Neary (1984). It is better here to stay with the original Jones specification. So the factors specific to certain sectors are permanently confined to those sectors. Their confinement to one sector is not the result of short-run immobility as they are not productive elsewhere. In this model factor-price equalization does not feature, not even in the long run. That is the feature which Samuelson (1971) uses in his parallel model to support Bertil Ohlin's contention that factor-price equalization will be incomplete. Also, in this type of specific factors model, Stolper-Samuelson magnification cannot be shown.

Ruffin and Jones (1977) is another small-scale (toy) non-square trade model. A useful general discussion of higher-dimension models is included. The only model considered in detail is what the authors call the Ricardo-Viner model. This is the same as the Jones (1971) specification. Closer to the model of this chapter is Ruffin (1981), as that paper addresses the three-factors and two-goods case. Its main concern is the effect of a factor supply change on the price of another factor. However the analysis confirms the point already made by Stolper and Samuelson, that patterns of complementarity and substitutability are crucial for results.

The paper which bears most directly on this chapter is Jones and Easton (1983). In fact our toy model is a special case of Jones-Easton. In a two-goods three-factor set-up, these authors impose the restriction:

$$\frac{\theta_{31}}{\theta_{32}} > \frac{\theta_{11}}{\theta_{12}} > \frac{\theta_{21}}{\theta_{22}} \tag{1}$$

where $\theta_{ij}$ is the share of factor $i$ in sector $j$. As in the standard HOS model, (1) should be satisfied at all factor prices. For the sake of a specific case, let sector 1 be a high-tech sector (such as computers) and sector 2 be a low-tech sector (such as saucepans). The three factors are 1 capital, 2 unskilled labour, and 3 skilled labour. Then (1) implies the inequality $\frac{\theta_{11}}{\theta_{12}} > \frac{\theta_{21}}{\theta_{22}}$. That is a natural factor-intensity assumption for capital and unskilled labour. Also implied by (1) is $\frac{\theta_{31}}{\theta_{21}} > \frac{\theta_{32}}{\theta_{22}}$. The high-tech sector uses skilled labour more intensively relative to unskilled labour than does the low-tech sector. We would hardly consider any other specification. And finally $\frac{\theta_{31}}{\theta_{11}} > \frac{\theta_{32}}{\theta_{12}}$, the high-tech sector uses skilled labour more intensively relative to capital than does the low-tech sector. These restrictions might be questioned. What matters is whether these restrictions or any similar can lead to definite analytical conclusions.

Jones and Easton show that the strong factor-intensity conditions implied by a condition such as (1) leave the analysis of particularly output-price changes and factor-availability changes more complicated and uncertain than in the basic HOS case. Two observations will illustrate the issues. Suppose that at all factor prices $\theta_{21} = \theta_{22}$. That is the share of unskilled labour costs is the same in both sectors. Consider a relative price change between high-tech output and low-tech, say in favour of low-tech. The standard Stolper-Samuelson argument (now applied to capital and skilled labour) goes through virtually unchanged. The only way the relative unit cost of the low-tech output can rise is if the rental of capital rises and the wage of skilled labour falls. As usual these changes will involve magnification. The factor price of unskilled labour can do what it likes as any change has no effect on the relative cost of the two goods. Now, clearly, if we relax the special assumption $\theta_{21} = \theta_{22}$, matters are far more complicated. If, for instance, the expansion of the low-tech sector and the contraction of high-tech sector following a goods-price change greatly lowers the wage rate of unskilled labour this could assist a relative decline in high-tech unit costs independently of a Stolper-Samuelson effect involving capital and skilled labour.

With the consequences of factor-supply changes (Rybczynski effects) the two-good three-factor model is again very different from the HOS case. Take the case of fixed coefficients. Outputs are such as to fully employ factors and over a broad range are independent of goods prices. With two goods and three factors, and again assuming fixed coefficients, outputs are uniquely determined by any two fully employed factors. And only a particular supply of the other factor can be fully employed. Then a change in the supply of any factor will destroy full employment of all factors. If factors are sufficiently substitutable, factor-price changes may restore full employment of all factors. But even when that is possible the situation is very different from the Rybczynski situation when factor-supply changes are fully accommodated by output changes at constant factor prices.

## 6.2 The Three-Dimensioned Factor-Price Frontier

Consider a country producing both goods, high-tech and low-tech, with unit costs equal to given output prices $p_1$ and $p_2$. Let the factor prices of specifically capital, unskilled labour, and skilled labour be $(w_1, w_2, w_3)$, and the unit-cost function for sector $i$ be $c^i\,[w_1, w_2, w_3]$. We can now show how our simple two-good three-factor model is consistent with the possibility that the North might have higher unskilled wages than the South, and also a higher rate of return on immobile capital than applied in the South. That case might be thought realistic. If so, it comes to be in the present instance when the North has more abundant skilled labour than the South, and for that reason a lower wage rate for skilled labour. We may think of skilled labour as ingenious Yankees, or abundant land. Then the idea employed here is similar to the Temin (1966) explanation of the apparent paradox that the USA in the nineteenth century had higher real wages and a higher return on capital than did Britain.

The cost-price equations implied by our assumptions are:

$$c^1\,[w_1, w_2, w_3] = p_1 \tag{2}$$
$$c^2\,[w_1, w_2, w_3] = p_2 \tag{3}$$

Equations (2) and (3) define an implicit relationship between the $w$ values which is the three-dimensional factor-price frontier. Now we differentiate (2) and (3) with respect to $w_3$, with output prices constant. We take into account that the partial derivative of a unit-cost function with respect to an input price is the factor-input unit-output coefficient. Thus we obtain:

$$a_{11}\frac{dw_1}{dw_3} + a_{21}\frac{dw_2}{dw_3} = -a_{31} \tag{4}$$
$$a_{12}\frac{dw_1}{dw_3} + a_{22}\frac{dw_2}{dw_3} = -a_{32} \tag{5}$$

Or, writing (4) and (5) in matrix form:

$$\begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}\begin{bmatrix} \frac{dw_1}{dw_3} \\ \frac{dw_2}{dw_3} \end{bmatrix} = -\begin{bmatrix} a_{31} \\ a_{32} \end{bmatrix} \tag{6}$$

Let $D$ be the determinant of the matrix $\begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}$. Then:

$$D = a_{11}a_{22} - a_{12}a_{21} \tag{7}$$

Notice that the factor shares $\theta$ are just the input-output coefficients $a$ multiplied by the factor price $w$ and divided by the output price $p$. For that reason ratios of input-output coefficients satisfy the same inequality conditions as do the $\theta$ values. For this reason, because $\frac{\theta_{11}}{\theta_{12}} > \frac{\theta_{21}}{\theta_{22}}$, as stated above, $D$ must be positive. Now solving the linear equations (6) we obtain:

$$\frac{dw_1}{dw_3} = \frac{-\begin{vmatrix} a_{31} & a_{21} \\ a_{32} & a_{22} \end{vmatrix}}{D} \tag{8}$$

Because $\frac{\theta_{31}}{\theta_{32}} > \frac{\theta_{21}}{\theta_{22}}$, as detailed above, $a_{31}a_{22} - a_{32}a_{21} > 0$. Therefore $\frac{dw_1}{dw_3} < 0$, and we conclude.

**Result 6.1:** *Under the maintained assumptions concerning relative factor intensities, a country which has a high wage for skilled labour will have a low wage for unskilled labour.*

Now solving (6) for $\frac{dw_1}{dw_3}$ we obtain:

$$\frac{dw_1}{dw_3} = \frac{-\begin{vmatrix} a_{31} & a_{21} \\ a_{32} & a_{22} \end{vmatrix}}{D} \tag{9}$$

Because $\frac{\theta_{31}}{\theta_{32}} > \frac{\theta_{21}}{\theta_{22}}$, $a_{31}a_{22} - a_{32}a_{21} > 0$. Therefore with $D > 0$, $\frac{dw_1}{dw_3} < 0$, and we conclude.

**Result 6.2:** *Under the maintained assumptions concerning relative factor intensities, a country which has a high wage for skilled labour will have a low return to capital.*

## 6.3 Assessing the Results

Results 6.1 and 6.2 are appealing in that they seem to depict a situation similar to reality. Think of the rich industrial world (the North) as well endowed with skilled labour, and the developing countries (the South) as poorly endowed with skilled labour. It is plain that the South has a low wage rate for unskilled labour, so in that regard the model does well. It is less plausible to suppose that the South has a low return to capital, as that would lead to a capital outflow in the direction South to North. Something like that does occur. However a long-run equilibrium would not allow of different returns to a mobile factor in different regions.

If capital is perfectly mobile we are back to essentially a two-factor (skilled and unskilled labour) model, as analysed by Wood. Suppose then that technology is everywhere the same, that goods are freely mobile, and that there is no specialization. Then the abundance of skilled labour in the North will have no effect on the unskilled wage rate. Only national production levels will be affected. The North will produce more of the high-tech good relative to the South. It is the Rybczynski story again.

How reasonable are the conditions (1), repeated here for convenience in terms of input-output coefficients?

$$\frac{a_{31}}{a_{32}} > \frac{a_{11}}{a_{12}} > \frac{a_{21}}{a_{22}} \tag{10}$$

Now (10) implies three separate inequalities:

$$\frac{a_{31}}{a_{21}} > \frac{a_{32}}{a_{22}} \tag{11}$$

This says that the high-tech sector uses skilled labour intensively relative to unskilled labour when compared with the low-tech sector. Below we look at a case in which $a_{32} = 0$ which guarantees the inequality. In general we would hardly consider any other specification, as intensive use of skilled labour is almost a definition of the high-tech sector.

$$\frac{a_{11}}{a_{21}} > \frac{a_{12}}{a_{22}} \tag{12}$$

This says that the high-tech sector uses capital intensively relative to unskilled labour when compared with the low-tech sector. That is a most reasonable condition if only because the high-tech sector makes moderate use of unskilled labour.

$$\frac{a_{31}}{a_{11}} > \frac{a_{32}}{a_{12}} \tag{13}$$

This says that the high-tech sector uses skilled labour intensively relative to capital. Again the condition is guaranteed if $a_{32} = 0$. And it is reasonable even if the low-tech sector makes some use of skilled labour.

In summary, the above inequalities are reasonable mainly because a strong differentiation, in favour of the high-tech sector, with regard to the intensive use of skilled labour, is almost an inescapable feature. Then the intensity of the usage of capital falls in between that for the two types of labour.

At this point we may refer to an issue that must await Chapter 8 for a thorough discussion. It has been seen how with a three-factor model a high wage rate for skilled labour can result in low factor prices for both capital and unskilled labour. Does that high factor price, for skilled labour, need to be the market cost of a standard input? Might $w_3$ stand for the additional cost imposed on producers by a poor economic environment or malfunctioning institutions? It is an intriguing idea. To make one obvious point, while deferring more extensive argument to Chapter 8, not all poor institutions, such as bad government, can be modelled as a high unit cost for a notional input. Suppose for example that property rights are insecure, and governing party thugs will help themselves to good share of any profit they can see. This is not a problem of elevated unit costs.

## 6.4 A Toy Model with Skilled and Unskilled Labour

Despite its elegant construction, the Jones-Easton model is not easy to apply because it often leads to complications and ambiguities. So a special case of Jones-Easton generates a toy which more readily provides definite conclusions, but may yet be rich enough to prove interesting. The model supposes a division of production between low-tech goods and high-tech goods, just as Wood assumes. One of the factors, skilled labour, is used exclusively in one sector. In the Jones-Easton model that is equivalent to assuming one of the parameters $\theta_{32}$ equal to zero. The model is:

$$c^1 [w_1, w_2, w_3] = p \tag{14}$$

$$c^2 [w_1, w_2] = 1 \tag{15}$$

$$y^1 \cdot \frac{\partial c^1 [w_1, w_2, w_3]}{\partial w_1} + y^2 \cdot \frac{\partial c^2 [w_1, w_2]}{\partial w_1} = K_0 \tag{16}$$

$$y^1 \cdot \frac{\partial c^1 [w_1, w_2, w_3]}{\partial w_2} + y^2 \cdot \frac{\partial c^2 [w_1, w_2]}{\partial w_2} = L_0 \tag{17}$$

$$y^1 \cdot \frac{\partial c^1 [w_1, w_2, w_3]}{\partial w_3} = S_0 \tag{18}$$

where $c^1 [w_1, w_2, w_3]$ and $c^2 [w_1, w_2]$ are respectively the unit-cost functions for the high-tech and low-tech sectors; $w_1$, $w_2$, and $w_3$ are the factor prices in terms of the low-tech good of respectively capital, unskilled labour, and skilled labour; $y^j$ is output in sector $j$ ($j = 1, 2$); and $K_0$, $L_0$, and $S_0$ are given factor supplies of respectively capital, unskilled labour, and skilled labour.

Clearly the high-tech sector is skilled-labour-intensive relative to the low-tech sector, as it alone uses skilled labour. As the argument above has indicated, it is not completely obvious which sector will be intensive in the use of unskilled labour relative to capital. Above we favoured the assumption that the low-tech sector will use unskilled labour intensively relative to capital. The following exposition assumes that the high-tech sector has the higher capital/unskilled-labour ratio. The opposite possibility has to be kept in mind and will be considered again below.

## 6.5 A Separable Version of the Model

The model (14) to (18) is most easily understood from a special case. As often happens, once the special case has been exposited, it is not difficult to see what more general cases will look like. Suppose that the production function for high-tech output is a constant returns function of the form:

$$y^1 = \phi^1(n^h) \cdot f^h\left[k^h, \ell^h\right] \tag{19}$$

where $k^h$, $\ell^h$, and $n^h$ are the inputs of respectively capital, unskilled labour, and skilled labour into the high-tech sector. The production function is separable in the sense that the choice of the optimal amount of high-tech labour to use is independent of the other two inputs and their prices.

With the production function (19) we have a unit-cost function:

$$c^1\left[w_1, w_2, w_3\right] = c^{11}\left[w_1, w_2\right] \cdot c^{12}\left(w_3\right) \tag{20}$$

where the two $c^{1i}$ [·] functions in (20) are distinct functions as is indicated by their different arguments. To get from the general to the special model (20) replaces (14) above.

**Theorem 6.1:** *In the special model factor-price equalization does not necessarily result. If there is any substitutability between unskilled labour and capital, Stolper-Samuelson magnification is a feature of the model, but its impact is moderate relative to the two-factor HOS model.*

**Proof:** Given relative product price $p$, unit-cost price equality requires:

$$c^{11}\left[w_1, w_2\right] \cdot c^{12}\left(w_3\right) = p \tag{21}$$

$$c^2\left[w_1, w_2\right] = 1 \tag{22}$$

Given the factor-intensity assumption for capital and unskilled labour, if two countries produce both products and share the same value of $c^{12}\left(w_3\right)$, factor-price equalization follows from (21) and (22). It is as if the price of the high-tech product in terms of the low-tech product were $p/c^{12}\left(w_3\right)$

in a standard HOS model. Equally if the two countries do not share the same value of $c^{12}(w_3)$, it is as if they faced different output prices, and factor-price equalization will not feature.

With Stolper-Samuelson magnification suppose a rise in $p$, and imagine that the changes which result do not include any alteration in the value of $c^{12}(w_3)$. Then the usual conclusions of the HOS model apply, and the real wage of unskilled labour in terms of either product will fall. However if there is any substitutability between the capital and unskilled-labour inputs, the high-tech sector will increase in size; the marginal product of skilled labour $w_3$ will rise; and the high-tech sector will experience a rise of its unit-cost function in terms of $w_1$ and $w_2$. The final effect is that $p$ has risen, but so has $w_3$. Then $p/w_3$ has gone up but by less than the rise in $p$. Note that $w_3$ increases only because $p/w_3$ has risen, so the rise in $w_3$ cannot be in a greater percentage than the rise in $p$. It is as if the rise in $p$ were more moderate than its true arithmetic value. Because magnification does not depend upon the size of the increase in $p$, magnification will still follow. □

## 6.6 The Model without Separability

The discussion of the separable model, and the proof of the theorem makes clear why greater complications will be encountered if separability is not assumed. Suppose one tries to push through a similar argument to the theorem using the non-separable cost-price-equality equations (14) and (15), rather than (21) and (22). So long as $w_3$ is not altered, everything is as standard HOS reasoning. With the assumption that the high-tech sector is capital-intensive, an increase in $p$ will cause $w_1$ to rise, and with some substitution the high-tech sector will expand.

The expansion of the high-tech sector increases the demand for skilled labour, and for that reason $w_3$ will rise to cut back demand to equality with the fixed supply. Now, without separability the change in $w_3$ has a differential effect on the marginal attractiveness of the other two inputs, capital and unskilled labour. Then Stolper and Samuelson's 'complicated patterns of complementarity and competitiveness' make themselves felt.

Thus suppose the unlikely case that unskilled and skilled labour are close substitutes. So long as $w_3$ is constant the model will be essentially the familiar 2X2 HOS model. A rise in $p$ will expand the high-tech sector, and $w_3$ will rise, just as in the argument above. Now the rise in $w_3$ will cause the substitution of unskilled for skilled labour and $w_3$ will fall back somewhat. Taking all effects into account the conclusion is that $w_3$ will rise by less than in the separable case. That is a sufficient weakening of the previous argument to upset the conclusion that magnification will be present although moderated.

## 6.7 The Capital-Intensity Assumption Reversed

The argument above has throughout employed the assumption that the high-tech sector has the higher capital/unskilled-labour ratio. The opposite assumption is possible if not wholly convincing. There is no need however to rehearse the entire analysis again with the assumption reversed. The method has been demonstrated and the interested reader can work out what happens with a different assumption.

That is not to say that the assumption is inconsequential for the qualitative conclusions of the model. Take the basic instance of magnification in the separable model. Now the low-tech sector has the higher capital/unskilled-labour ratio. Suppose again a rise in $p$, the relative price of the high-tech good. With some substitutability the high-tech sector will expand, but in this case the Stolper-Samuelson effect is a magnified fall in the return to capital and a rise in the unskilled wage rate. Even so the expansion of the high-tech sector will increase the demand for skilled labour. Once more this is equivalent to a moderation of the price rise, hence a moderation but not the cancellation of magnification.

## 6.8 The Separable Model Depicts North–South Trade

Our model is a toy because it lacks mathematical generality. The term is not intended to indicate that the model is not to be taken seriously. On the contrary, it will be argued that the model is useful for the analysis of trade between North and South in a globalized world. In particular this toy model performs better than either the standard HOS model, or Adrian Wood's adaptation of that model.

Here the North is a rich country well endowed with capital in both its forms. That is to say the North is relatively well endowed with both physical and human capital, where the latter is identified with skilled labour. The South, on the other hand, is poorly endowed with each kind of capital, and therefore relatively well endowed with unskilled labour. Other assumptions are standard and the same as in the HOS model. Thus technology is the same in both countries; all factors are internationally immobile; but goods are perfectly mobile, so that the relative price $p$ is the same in both countries. Finally, for the sake of the present argument, the high-tech sector has the larger ratio of capital to unskilled labour at all factor prices.

Suppose a liberalization of trade that equalizes $p$ in both countries. Previously $p$ would have taken a higher value in the South, because the local factor endowments entail a scarcity of the high-tech good, and trade restrictions inhibited imports. Similarly $p$ would previously have taken a lower value in the North, because the local factor endowments entail a relative scarcity of the low-tech good, and trade restrictions inhibited imports.

Now Theorem 6.1 above tells the whole story. First, and crucially, factor-price equalization does not apply. This is a welcome and agreeable conclusion. It is always a cause of embarrassment that a model such as the 2X2 HOS model, which starts with assumptions that are by no means absurd, should arrive at conclusions that are patently unrealistic. Chief among these unrealistic conclusions is the equalization of real wages in all trading countries. Of course no-one would take the HOS assumptions to be exactly correct. And in Chapter 5 above we have noted numerous instances in which factor-price equalization may not apply even given HOS assumptions, or their generalizations. Even so, real wages vary hugely between trading nations, and even between trading nations that are apparently diversified in their productions. A model that depicts such an outcome as only natural given varied endowments of two types of capital is appealing.

Next note that Theorem 6.1 describes what happens when $p$ changes, either up or down. What it shows is magnification just as with the Stolper-Samuelson result in the 2X2 HOS model, but moderated in its extent. To be specific, consider the North as trade expands. In the North $p$ rises, which is the same as saying that the relative price of the low-tech good falls. Table 6.1 shows what happens to each of the factor prices, all measured in terms of the low-tech good. The directions of the price changes are indicated by arrows. An arrow pointing upwards indicates an increased factor price. An arrow pointing downwards indicates a lower factor price.

**Table 6.1.** Effects on real factor prices of a rise in $p$

| Factor | Capital | Skilled labour | Unskilled labour |
|---|---|---|---|
| Factor Price change | ↑ | ↑ | ↓ |

## 6.9 Trade and Inequality

Table 6.1 is drawn up from the point of view of the North, following a rise in $p$. It can be used to tell the story of the South, where the opening up of trade will cause $p$ to fall. In that case obviously the same table applies with the direction of each arrow reversed. If we ignore the column headed Capital, the model tells essentially the same story as that related by Adrian Wood. In the North the wage rate for skilled labour rises; the wage rate for unskilled labour falls. If those were the only changes they would amount to an unambiguous increase in inequality, not unlike the changes that have been observed in northern countries (notably the USA) in the last two decades. Now consider how that account of the change in inequality caused by the opening up of trade has to be modified by taking into consideration the arrow on the left of the table, which shows the change in the return to capital. That return rises.

It is always the case that the implications of trade for income distribution are determined by two separate facts. Firstly, how are factor prices altered? Secondly, what is the pattern of factor ownership between the various households in the economy? For an empirical analysis of the effect of trade accounting for both these influences, see Bourgignon and Morrisson (1989). Where the particular present instance is concerned, trade liberalization and the North, the issue seems to be plain. While the ownership of capital may be widely dispersed, it is unequally distributed, and large holdings of physical capital are highly correlated with substantial holdings of human capital. Human capital is the same as skilled labour in the present model. Therefore it follows that the rise in the return to capital will only accentuate the increase in inequality in the North already detailed by Wood.

Turning to the South, the story is again just the reverse of what happens in the North. The relative price $p$ falls rather than rising. And the consequences can be read from Table 6.1 when all the arrows have their directions reversed. In the South, as in the North, large-scale ownership of physical capital and large-scale ownership of skilled labour are highly correlated. For this reason the changes that bring increased inequality to

**Table 6.2.** More effects on real factor prices of a rise in $p$

| Factor | Capital | Skilled labour | Unskilled labour |
|---|---|---|---|
| Factor Price change | ↓ | ↑ | ↑ |

the North are reversed in the South, where they bring reduced inequality. The picture seems to be quite similar to the conclusions of Wood's model.

On closer inspection there are differences between Wood's two-factor HOS-style model and the present three-factor separable model. First the magnification which features in any HOS model, and hence in Wood's version of the same model, is present in the separable model, but is moderated in its magnitude. More important, possibly, the separable model allows more scope for different conclusions as its assumptions are varied. For instance, we have concentrated on the assumption that the high-tech sector has the higher ratio of capital to unskilled labour. The opposite assumption cannot be dismissed as absurd. What happens if we take that case? It is equivalent to relabelling two of the factors: capital becomes unskilled labour; and unskilled labour becomes capital. Then Table 6.1 is replaced by Table 6.2 above.

Suddenly the clear picture of the change in income distribution is replaced by a fog. In the North unskilled labour is now better off. What about the rich household that owns both capital and skilled labour? It loses and it gains. It loses because the return to the physical capital that it owns goes down. But it gains because the wage rate of its skilled labour goes up. The final implications for the income of any particular household depend upon the numbers concerned, on precise quantities and exact changes in factor prices. Over the whole income distribution anything is possible. It could happen, for example, that rich rentiers, owning mainly physical capital, will lose, while middle-class professionals, owning mainly skilled labour, will gain.

As long as we stay with the root assumption that only the high-tech sector uses skilled labour, and skilled labour enters separably into production, the above is as far as varying assumptions in the model can take us. However matters become even more complicated if the assumption of separability is relaxed. Then we let loose Stolper and Samuelson's complicated patterns of complementarity and competitiveness. There is no need to dive into these murky waters. It is enough to conclude that in general the consequences of trade liberalization in a high-dimension world allow for a rich variety of possibilities.

## 6.10 Labour Migration in the Separable Model

It is claimed above that it is a definite advantage of the separable three-factor model that it does not give us factor-price equalization. That is because factor-price equalization looks distinctly unrealistic. Also unrealistic, possibly, is the Rybczynski theorem which is a feature of the HOS model. The Rybczynski theorem leads to the conclusion that migration of factors is unnecessary, because we have factor-price equalization. But should it happen it is innocuous. The outputs of the two sectors in the country receiving a factor inflow adjust to absorb the increased supply at constant factor prices, and existing residents are unharmed.

This comforting picture does not hold with the separable three-factor model. Take a migration into a small country of skilled labour. If the result were to be as the Rybczynski account, there would be no effect on factor prices; certainly no effect on $w_1$ and $w_2$. Suppose, to simplify the argument, that there are fixed factor proportions in both sectors where capital and unskilled labour are concerned. Then the cost-price equality equations take the form:

$$a_k^1 w_1 + a_u^1 w_2 + a_s^1 w_3 = p \tag{23}$$

$$a_k^2 w_1 + a_u^2 w_2 = 1 \tag{24}$$

where $a_j^{1\,\text{or}\,2}$ is the unit input of factor $j$ ($j = k$, $u$, or $s$) into the sector indicated by the superscript. The fixed factor proportions assumption means that of all the $a$ coefficients in (23) and (24) only $a_s^1$ can vary. The consequence will be that with fixed coefficients, and so long as production remains diversified, the full-employment conditions for the two-factors capital and unskilled labour determine the outputs of the two sectors, and these cannot be influenced by the supply of skilled labour. For that reason the size of the high-tech sector will not change. Therefore $a_s^1$ must rise so that the high-tech sector can absorb the increased supply of skilled labour. That can only happen because of a fall in $w_3$, the price of skilled-labour services. A rise in $a_s^1$ and a fall in $w_3$ has an ambiguous implication for the product $a_s^1 w_3$; it depends upon the elasticity of demand for skilled labour.

Take first the case when the demand for skilled labour is inelastic. Then when the supply of skilled labour increases $a_s^1 w_3$ falls. This, as inspection of equation (23) shows, is equivalent in its effects on the prices $w_1$ and $w_2$ to a rise in $p$. Then with the high-tech sector capital-intensive the

Stolper-Samuelson theorem implies that $w_1$ will rise and $w_2$ will fall. From the point of view of both kinds of labour in-migration of skilled labour is not innocuous. It entails a decline in real wages. Plainly the other case, when the demand for skilled labour is price-elastic, is in part the mirror image of the inelastic case. The effects on the prices $w_1$ and $w_2$ are equivalent to a fall in $p$. Now $w_1$ will fall and $w_2$ will rise. Migration of skilled labour into the country is bad for capital and good for unskilled labour. A nice case is when the demand function for skilled labour has unit elasticity. Then skilled labour is harmed, because $w_3$ falls, but there are no knock-on effects on the other factors.

As often happens, the fixed-coefficient case gives a good idea of what to expect in a more general model with factor substitution. With the above analysis consider what will occur if the ratio of capital to unskilled labour can vary in each sector according to just the two prices $w_1$ and $w_2$. So we retain the separability assumption, and $w_3$ does not affect the optimal choice of the capital to unskilled-labour ratio. Now an increase in $S_0$ will cause $w_3$ to fall, and $p - a_s^1 w_3$ will change according to the elasticity of demand for skilled labour. Suppose an inelastic demand for skilled labour, so that the effect on the other factor prices is equivalent to a rise in $p$. So $w_1$ rises and $w_2$ falls. The high-tech sector expands and both sectors become less capital-intensive. The substitution of unskilled labour for capital releases just enough capital to balance the increased demand for capital caused by the expansion of the high-tech sector. All this is as standard HOS analysis. The implications for income distribution are the same as with fixed coefficients. An inflow of skilled labour harms both kinds of labour and is good for capital owners.

What happens if migration takes the form of an inflow of unskilled labour? If the supply of skilled labour was not a problem we would be back to standard Rybczynski theory. There would be no effect on factor prices because the low-tech sector would expand, and the high-tech sector would contract to absorb the increased supply of unskilled labour. In the general model a similar story applies. But now the contraction of the high-tech sector causes a fall in the demand for skilled labour, the price of which falls. What that implies depends again upon the elasticity of demand for skilled labour. If, as seems likely, the demand is inelastic, then $a_s^1 w_3$ will fall. That, as we have seen, is equivalent to a rise in $p$. One could say that the Rybczynski effect induces a Stolper-Samuelson effect. A rise in $p$ raises $w_1$, the return to capital, and lowers the unskilled wage rate $w_2$. An inflow of unskilled labour harms all workers and is good for capital.

## 6.11 Concluding Remarks

Our point of departure for the present chapter is the numerous difficulties with the simple 2X2 HOS model. That model has dominated trade theory for the last sixty years. Yet its realism must be seriously questioned. It can yield unrealistic results, such as factor-price equalization, or the Rybczynski account of factor migration. A model with only two goods evidently involves a high level of aggregation, but that may serve well where trade between two countries, or types of countries, is the focus of analysis. Having only two factors of production is more problematic. It is not clear which those factors should be, and having three factors, but allowing perfect mobility for one of them, as Wood does, may not be realistic.

An unrestrained launch into high-dimension trade theory however may not bring us applicable results. Those models yield few theorems with easy insights attached. And the reader may be choked by a complex taxonomy of cases depending upon the exact specification of the patterns of complementarity and competitiveness to which Stolper and Samuelson made reference long ago. The chapter has shown how a very limited step in the direction of generality—just one extra factor, and that employed separably in one sector—gives a model with many attractive features. In particular it looks more suitable for depicting trade between strongly different regions. It does not need specialization or factor-intensity reversals to depict economic inequality. It replicates the magnification result of HOS theory but shows how that effect is moderated by a price change for the third factor. It allows a richer menu for the possibilities that arise when trade between North and South is liberalized without making that analysis too horribly complicated.

Over sixty years ago Samuelson and Stolper wondered why labour supports protection when economic theory seems to say that it is harmful. Much more recently many economists have wondered why migration of labour into industrial countries can evoke fierce popular resistance, although much analysis seems to suggest that it is beneficial in its economic effects. One answer serves for all such examples: reality is more complicated than simple models. Such was Stolper and Samuelson's analysis. They argued that views on protection are not determined solely by judgements of its efficiency. It is essential that its consequences for income distribution be taken into account. Migration in all its implications is hugely complicated. It involves disparate economic effects, and it has powerful cultural and political implications.

That said, our separable three-factor model already shows how migration may have very different consequences for factor interests than the usual Rybczynski analysis would indicate. Depending upon the precise assumptions, factors already resident in the country can lose when labour migrates in. It is no surprise that labour of the same type as the migrants always loses. The other type of labour may lose or gain, as may capital. Here the relatively simple model confirms what intuition would suggest for more complicated cases.

# 7

# Participation and Comparative Advantage

## 7.1 Countries that do not Participate

A simple account of globalization would have it that international trade in today's world is a club to which all are welcome. True there are numerous, and scandalous, restrictions on access to rich country markets, particularly where agriculture and clothing are concerned. Yet, over a large range of goods, trade is feasible even on a large scale, as is indicated by the exporting successes of the Asian Tiger countries, and now by China and India. Why then are there countries and regions that barely participate in merchandise trade, particularly when primary production is excluded?

This question is particularly pointed in two instances:

- The Arab World
- Sub-Saharan Africa (SSA)

The UNDP report on the Arab world, United Nations (2003), notes that the merchandise exports of the region which was home to 280 million people in 2002 are lower than those of Denmark. In SSA a similar poor participation in goods trade is observed. See Collier (2003) and Ng and Yeats (2000 and 2002). It goes without saying that the causes of poor export performance in each of these regions are many and various. Most of the economies concerned are in societies that to a greater or lesser extent are failing. Poor infrastructure, in SSA massive health problems from AIDS to malaria, poor-quality education, the burden of debt, and the scourge of corruption, all drag down economic performance, and with it possibilities for exports.

In recent years the World Bank has devoted considerable attention to the economic underperformance of SSA. See Plekovic and Stern

(2003). That has included careful analysis of the particular problems and conditions of the region. There has been far less comparative analysis, although Wood (2003) offers an insightful comparison of SSA with South America (SA). What is lacking is a comparative analysis of SSA and its 'neighbour'—abstracting from the existence of the Sahara desert—the Arab world. Of course there are huge differences, as there always are when such comparisons are undertaken. Yet this comparison is in many ways particularly illuminating. To cite one justification for that claim, it is often argued that poor access to world transport networks burdens many SSA countries, in particular the landlocked countries. Certainly transport costs are an important influence on the international division of labour, and that point will be examined in more detail below. If it were the decisive point however, the Arab world would not be the economic underperformer that it is, as all Arab countries have access to the sea, and in many cases that access is good.

This chapter concentrates on those points that are amenable to close-to-standard economic analysis. Plainly that defines a limited approach. However it is interesting to see that much can be achieved via a narrow treatment of the issues. One feature that suggests that this approach may not be worthless is that it appears that a mirror image of the analysis may not be too bad as a partial account of the recent success of China. For example where we look at overvalued real exchange rates as a drag on exports, China has an undervalued real exchange rate.

Returning to the comparison with South America (SA), it is the case that the economic performance of the SA region has been something of a roller-coaster. There have been dramatic ups and downs ranging from the 'economic miracle' recovery of the economy of Chile to the largest-ever debt default by Argentina in 2001. With all that noted, SA has generally done far better than SSA. Wood (2003) argues that SA is a highly relevant comparator for SSA. By that he means that it is a better comparison than the frequently used East-Asian region. The Arab world would not serve Wood's purpose here as its performance, broadly speaking, is not much better than that of SSA; Wood's point is that land and resource abundance, as well as climate and economic geography, are more similar in SSA and SA than in East Asia.

## 7.2 Sub-Saharan Africa

If we look at sub-Saharan Africa from a traditional factor-availability angle, as with the HOS model, we see a land-rich resource-rich region that is short of labour, especially skilled labour, and short of capital. See Bloom and Sachs (1998) On one view shortage of capital is never a crucial bottleneck, as mobile international capital can always make good any shortfall. It will be seen that this last view is oversimple, but the relevant discussion can wait until Chapter 10. Assume for the time being that capital is perfectly mobile internationally. With the help of foreign capital that points to primary agricultural and horticultural exports, and despite heavy protection of the agricultural sector in the rich North, Africa has had some success there. For Kenya and Tanzania a visit to any supermarket vegetable section will confirm the point. Collier (2003) argues that abundant land raises wage rates and makes Africa uncompetitive in comparison with land-scarce regions. He has to concede however that wages in the region, expressed in international value, are low. So this is not an example of a classic Dutch disease crowding out of labour-intensive production.

SSA's heavy dependence on primary-product exports causes more than one type of problem. The primary sector has not experienced increases in productivity sufficient to keep SSA incomes growing in line with population growth and the per-capita income growth experienced elsewhere. And even if the sector had been as good at income generation as manufacturing or services there is an additional problem that primary-product prices are notoriously unstable. Collier argues that African countries have a poor record when it comes to managing unstable external earnings. He writes:

> Its [Africa's] economies are indeed shock-prone, governance is poor on average, and there is a high incidence of civil war. (Collier 2003: 140)

That reads like a list of three problems, but according to Collier's view they are closely interconnected. Resource richness tends to generate poor governance because it places great wealth straight into the hands of governments; governments that do not need to justify its use to the population, as a democratically accountable government has to do. If that loot were absolutely secure in the government's coffers, the situation might be bad but the control of resource wealth would not be an issue. In fact that control can often be contested and the result is too frequently civil war. Collier and Hoeffler (1998) show that there is a correlation between civil conflict and natural resource wealth. The sad history of Sudan over the last twenty years provides a perfect illustration of the point, as does the horrible struggle going on in Iraq at the present time. The reluctance of the Sunni Arabs to accept a constitution for Iraq is explained only partly by their minority status in the country. Constitutions can be and have been designed to take care of that problem. Equally a constitution could divide national oil wealth in any way desired, notwithstanding the fact that the central region where most Sunni Arabs live has no oil. Presumably, enough Sunni Arabs doubt that an acceptable division of oil revenue can be realized and maintained for them to prefer, and impose, civil conflict.

These problems aside, why does the instability of foreign earnings give rise to difficulties in addition to those that would spring from a constant stable flow? A simple basic point is that expenditure financed from a fluctuating income stream can be optimally managed. At the most trivial level, anyone whose salary is paid monthly learns to do just that. The optimal choice of expenditure is more difficult when the income source varies randomly, but again an optimal policy can be designed. That is easier to do in the classroom, where the stochastic process generating income is usually known and unambiguous, than it may be for a real-life government. Take the sudden arrival of a resource boom. Is it a temporary blip that will be reversed after a short time, or might it represent a permanent shift in the price level? It needs an optimal Bayesian strategy to respond to that situation, and it is evident that when the reduction of an established expenditure level is particularly costly the best policy will incline to the cautious wait-and-see direction.

Collier argues that the management of fluctuating income by SSA governments is remarkably poor and very far from the optimal policy sketched above. So much is this the case that better management of expenditure is held to be as important as a movement away from primary-product dependence. In most African countries a significant shift in the composition of exports will be difficult to achieve. An improvement in the quality of expenditure planning where this is initially at a low level may be more easily attained, and that may be the best direction towards which policy improvement should be directed. Collier writes:

> Booms do not translate into sustained increases in income—they are missed opportunities—whereas crashes produce devastating long-lasting declines. (Collier 2003: 141)

Note that Africa is a huge and greatly varied continent, for which reason broad-bush generalizations can miss the mark where individual countries are concerned. Thus Rwanda is not land-abundant, and Kenya which used to be is no longer so well described by that term. Further, there is trade which the simple HOS model is ill designed to describe. Some SSA countries are well endowed with minerals, including diamonds, gold, and oil. These sectors generate exports but often at a cost to alternative exports, notably manufacturing.

Wood and Mayer (2001) perform a cross-section regression analysis of the SSA countries in which they examine how far the ratio of the exports of manufactures to the exports of primary products can be explained by domestic ratios of factor supplies. The paper appeals to what it admits to be informal HOS theory. This can be summarized by the statement that while equal technology and factor-price equalization may not feature in reality, still countries will tend to export goods that use intensely factors with which they are relatively abundantly endowed.

As this is informal theorizing intended only to motivate regression analysis that is free to accept or reject the postulated relationship it could be held that it is out of order to ask what exactly is the theoretical model. Even so, that line of enquiry throws up some interesting points, so it is reviewed briefly here. As is well known, the association between factor endowments and trade patterns provided by the HOS model depends upon an assumption concerning national patterns of final demand. These should be the same in all countries and not subject to awkward income or distributional effects. If a country highly endowed with capital has its demand strongly biased towards the capital-intensive product, it will produce that product at a high level, but may not export it. Similarly, if the owners of the scarce factor have a demand biased towards the good produced with intense use of the other factor, then the relatively high production of that good may not translate to high exports.

All this concerns the basic 2X2 HOS model. However that is not the model that Wood and Mayer have in mind. They model the ratio of the exports of manufactures to the exports of primary products. That is two goods, yet there must be a third good at least, as the country will import something in return for its exports. To keep things simple we could make that third good one that cannot be produced domestically. So only two final outputs are needed. On the other hand this is undoubtedly a three-factor model. The factors are land, which includes mineral resources, plus labour skilled and unskilled. Using the sensible simplification according to which scale as such makes no difference, factor abundance is reduced to two ratios: land per head, and skilled labour per head, the heads here being unskilled labour. As Chapter 6 has taught us, two-good three-factor models are potentially of formidable complexity and ambiguity. This is on account of Stolper and Samuelson's 'complicated patterns of complementarity and competitiveness'.

There is no need to go over these issues again in detail. To see how problems can arise consider the following case. At the start the country has no skilled labour. Its land and unskilled labour are fully employed, and the pattern of production is decided by the relative use of these two factors in primary and manufacturing production. Skilled labour is only employable in the manufacturing sector. Now some skilled labour arrives in the country, perhaps as returning educated migrants. The factor price of skilled labour drops from infinity to a finite value, and the manufacturing sector takes it in and expands its output. With the simple HOS story the expanding manufacturing sector would take land and unskilled labour from the primary sector. But suppose that skilled labour is a good substitute for land. Then it is possible that to maintain full employment of all factors *the manufacturing sector will have to contract*. It needs to release the unskilled labour that the primary sector requires for it to take up the land that the manufacturing sector has thrown onto the market because of factor substitution.

The reader may feel impatient at this point. Wood and Mayer made simple direct assumptions and now a theorist is showing that other assumptions give different results. So what. The regression works and that vindicates the assumptions that underlie it. Unfortunately it is not that straightforward. If a solid unquestionable theory says that the data-generating process for three variables $x$, $y$, and $z$ is:

$$x = \alpha y + \beta z + \epsilon \tag{1}$$

then regression analysis will extract good estimates of $\alpha$ and $\beta$ even if the sampled values of $y$ and $z$ are non-random. If the true data-generating process is different; if for instance the true relationship is non-linear, then the estimated values of $\alpha$ and $\beta$ will be unreliable and will be sensitive to the sampled values of $y$ and $z$.

Whatever the importance of the above technical points, the story that Wood and Mayer tell concerning the SSA region is similar to Collier's account and is plausible and compelling. The failure of the SSA region to generate a good volume of manufacturing exports can be put down, to a considerable extent, to a lack of skilled labour. Here it is plain that skilled labour does not mean rocket scientists. It means workers with the level of literacy, numeracy, flexibility, and work-discipline that modern production of high-quality manufactures demands. Thus far the argument has taken it for granted that Africa's choice is between primary-product exports and manfactured exports. Could some SSA countries leap straight into service exporting? The good English-language skills of some SSA countries would favour that development. And the success of the SOFT Tribe company in Ghana stands as a single but striking example. Watch this space, as the expression has it.

## 7.3 Transport Costs

An influence on international trade flows that HOS-style models deliberately ignore is transport costs and closeness to markets. Yet such effects are sometimes of great importance. The so-called gravity models of trade place great emphasis on closeness to markets, usually defined as crude 'as a bird flies' distance. For a more sophisticated approach on similar lines see Redding and Venables (2002). Once one starts to think seriously about what precisely constitutes transport costs, the points at issue are seen to be several and complex. Take the case of large landlocked countries, where SSA provides several examples. A standard argument says that being landlocked inhibits international trade simply because the transport of goods by road is far more costly than is sea transport.

As the unit cost of shipping goods by sea is extremely low in today's world, the protection provided to European or US car manufacturers by the fact that Japan and Korea are a long way away from their home markets is rather insignificant. That is just one case. Per ton, cars are high-value durable products, and it matters little to the dealers if it takes two months or more from final assembly to remote delivery. By way of contrast consider the favoured location for the production of fashion clothing for the European market. If only cost and quality are taken into account, China is the best place to locate production. Yet recently wholesalers have redirected some of their orders to Eastern Europe. The reason is that fashion clothing is a 'must have the latest now' product, and a retailer can suffer badly if an up-to-the-minute style is not stocked. Shipping clothing from China to European ports takes twenty-one days, and that is too long for such a fiercely competitive sector (report in the *Financial Times* 30 Sept. 2005). On geography and economic development see Gallup, Mellinger and Sachs (1998).

## 7.4 The Arab Middle East

When viewed from a resource-availability perspective the Arab Middle East shows similarities and contrasts in comparison with the SSA countries. Any apparent land richness is illusory, as much of the land is of poor quality, where it is not simply desert. Water is a dominating resource issue and one of growing significance as population growth, urbanization, and climate change make water ever more scarce. Some countries are resource-rich, especially with natural gas and oil. However these endowments are unevenly distributed. Where they are abundant, as in Saudi Arabia and the Gulf States, manufacturing is almost non-existent. The region as a whole is labour-abundant, but the skill level attained by that labour is poor. Despite high levels of participation, by young males at least, the quality of education is low. There is too much rote learning, and independent thought and speculation are not encouraged. Scientific education is particularly poor. The 2003 UNDP report, United Nations (2003), already cited, provides a detailed review that supports this negative assessment. See Chapter 4 of the report, *Building human capabilities: education*. Numerous newpaper reports on youth unemployment, and sometimes unemployability, in Arab countries confirm the observation, made also in the UNDP report, of a mis-match between employers' needs and what the education system provides. If the region is short of capital this is mainly because other features, such as labour quality and insecurity of property rights, depress a rate of return that would otherwise attract international capital.

The region provides good illustrations of a problem that is encountered also in SSA countries—*rent-looting*. This will be discussed more fully in Chapter 8. Put simply, rent-looting is the disincentive to wealth creation and risk-taking that accompanies insecure property rights and the absence of due process. To explain the idea fancifully, imagine that Bill Gates had been Egyptian, and pretend that Egypt had all the labour skills and infrastructure that made *Microsoft* a feasible project in the US. We are supposing that one individual heads a private organization in Egypt that is, or could become fabulously wealthy, commanding a big slice of national wealth. So far imagination has been given free reign, but it would be asking too much to suppose that the autocratic cash-strapped Egyptian government, let alone corrupt parties that it allows to operate, or cannot stop from operating, would permit this organization to flourish and accumulate massive wealth without their dipping their hands, or their buckets, or their heavy-lifting gear, into the cash pool.

It is true that the seizures of assets that took place under Nasser are no longer prevalent and it is worth noting that recently, but only recently, Egypt has been able to sustain a seriously large enterprise. *Orascom Telecom Holdings* is worth well over one billion dollars, and flourishes both within Egypt and throughout the region. Its founder Neguib Sowiris has been able to overcome the twin perils of state interference and its companion rent-looting. Sowiris is a Christian, with secularist views entirely in accord with those of the Egyptian government, and this may explain in part how he has been able to avoid trouble. Yet the wisest conclusion at the time of writing may be that the world changes, and that fleet-footed sectors, such as telecommunications can sometimes outwit the bureaucrats and succeed. It may be relevant here that the imagined Egyptian Bill Gates would have been a profound innovator, whereas Sowaris is a copier, albeit a talented copier.

Egypt's problem is not that it cannot grow a *Microsoft*. The problem is the lack of any flourishing innovative manufacturing sector, and even its agriculture has not grown in line with its rapidly expanding population. Rent-looting or the fear of rent-looting is part of the explanation. An even greater role is played by the stifling bureaucracy that Nasser put in place with his programme of Arab socialism. Centralized bureaucracy and rent-looting are close cousins. This is because bureaucratic power allows control of wealth, which is nearly as good as ownership. Furthermore, through bribery and corruption bureaucracy translates to rent-looting, allowing the corrupt agent to walk away with the wealth that enterprise has created.

The same picture that is seen in Egypt is replicated in Syria. In both countries there has been much talk of reform but little action. It is one of the ironies of contemporary history that the centralized planned socialist economy that fell apart in the Soviet Union sees its image maintained in some Arab countries. This is in contrast to most SSA experience, where the socialist model has never enjoyed much support. Corruption, however, is rife in SSA countries. And rent-looting affects not only manufacturing but also primary production, especially high-value mineral extraction. The corrupt diversion of oil revenue in Nigeria has resulted in great poverty in one of SSA's richest nations. Recently Zimbabwe under Robert Mugabe's rule has experienced state-organized rent-looting, not unlike that seen in many socialist countries at the start of their socialist experiments, but here implemented in part via the legal process by what is supposed to be a legitimate democratic government.

Bad government is considered in greater depth in Chapter 8.

## 7.5 Semi-Tradeable Goods and Participation

We have seen above the division between tradeable goods and non-tradeable goods. That style of modelling fits most easily with neoclassical general equilibrium theory, where that theory is translated to a many-country case. There is one set of world markets for tradeable goods, and in each country markets for that country's non-tradeable goods. In each case those markets are perfect neoclassical markets in which goods of unambiguous type and quality can be bought and sold at prices fixed for the individual seller. This is an extension of the HOS model. It is plainly a gross simplification of reality, yet one that in the right context can get to the heart of matters and serve understanding well.

While the neoclassical markets model has done good service, there are many cases in which its deficiencies are all too clear. Over the last thirty years much analysis in a different style has been developed. This analysis usually goes under the name of the Economics of Imperfect Information. In many real-life markets the nature and quality of the goods on offer is far from being clear and definite. The most famous treatment of this type is Akerlof (1970). The problem of the uncertain quality of goods has not figured much in international trade theory, but is of growing relevance. The international sourcing of goods has undergone great change and it continues to change. China and India are now volume exporters of many products to world markets, following the example of the Asian Tigers (Korea, Singapore, and Taiwan). The success of these countries poses a critical question for underperforming countries or regions: why could you not do the same?

That question is particularly pointed for the Arab world and for sub-Saharan Africa. We have seen above how the availability of factors and transport costs are part of the answer, but another issue is what might be called market access. Sale into rich-country markets requires a reputation for high quality and reliability, and this represents an overhead cost. Who, offered a notably cheap box of nails manufactured in Uganda, would want to run the risk of purchasing it? Yet similarly priced boxes of nails made in China fill the shelves of rich-country hardware stores. This issue is similar to the evaluation of the so-called Armington Assumption; see Armington (1969). Put simply, the question is how closely a country's tradeable imports are close substitutes for the same country's tradeable exports.

Insight into what is going on here can be obtained from an old and neglected paper: Kaldor (1960). In his discussion of the role played by advertising Kaldor looks at how the consumer is assured of the quality of the products she buys. In traditional retailing the shopkeeper guarantees quality. If the consumer buys flour that is found to be adulterated she goes next time to another grocer. Then, as early as the nineteenth century, branded products start to appear, and advertising communicates directly to the consumer concerning their quality. Now if the consumer buys a tin of *Horlicks* that she finds unsatisfactory she avoids that product, but the grocer who sold it to her is irrelevant. Finally, with outstanding foresight, for the original version of the paper dates from the 1940s, Kaldor suggests that volume retailers might regain the task of authenticating quality. Something much like that has come to pass, for today supermarket (and other retailer) own-brands command a huge slice of the relevant markets. Typically these products are cheap and they are not directly advertised. One might say that the UK supermarket chain *Tesco* does not sell ice-cream; it sells a one-stop shopping experience, and that product is widely advertised. If *Tesco*'s own-brand ice-cream is not good that detracts slightly from the value of shopping at a *Tesco* supermarket. In that sense *Tesco* has become an authenticator of quality much as Kaldor predicted.

The situation just described is very different from the neoclassical perfect-market story. And a crucial aspect of the difference consists in the fact that a reputation for good quality demands the outlay of resources and it cannot be purchased economically for a low level of sales. The idea that marketing is an active costly activity is taken for granted in business schools. Here we need to consider how it should be taken into account in explaining differences in trade performance between countries and regions.

The jump to market access can take more than one form. Sometimes it is producer-led. An Indian manufacturer of engineering tools displays his products at a trade fair, and visitors to that fair are convinced of the quality of the products, and orders follow. Entry in this form, or using other methods of marketing, is expensive, and demands high resources and a commitment to risk-taking by the producer or his financiers. Often rich customers are in a better position than producers to bear the costs of quality assessment and technology transfer. These buyers seek out cheap sources of quality products and they might be said to be doing the marketing work for the producers. Discussions of current international competition often suppose that goods automatically get produced where they can be produced most cheaply after taking transport costs, tariffs, and other add-on costs into account. This view calls to mind an opinion attributed to Ralph Waldo Emerson:

> If a man . . . make a better mousetrap than his neighbour, tho' he build his house in the woods, the world will make a beaten path to his door.

As it is doubtful whether this was ever true of mousetraps, so it must be questioned whether non-participation in international trade may not be attributed in part to a failure of the countries concerned to market their potential as sources of manufactured products. There is a free-rider feature to this situation because success by one sector promotes opportunities for other sectors. For that reason export-promotion exercises are usually undertaken by public bodies. Both Arab and SSA countries have done far too little in this regard, too often relying on the self-justifying belief that there is nothing to promote. Even so, without going to the extreme view of Emerson, it is hard to believe that the main problem for these regions is that the world has failed to beat paths to their doorways. Far tougher and more structural problems largely explain their failure to compete in world export markets. Some of these problems consist of economic environments unfavourable for competitive production. Mention of these has been made above. The concept of the economic environment and its formal analysis is examined extensively in Chapter 8.

In simple neoclassical theory there is no discontinuity at the point of zero participation; no large difference between a low level of exports, and the zero level. That this is a dubious feature is not surprising. In connection with the computable general-equilibrium approach to modelling (CGE models), in particular for the case of NAFTA, it is already known that the modelling of export categories not active before trade liberalization is peculiarly difficult. See Kehoe (2005). But note that this problem does not arise because the CGE models include a structural discontinuity at zero exports. Rather the difficulty is that these models make heavy use of the data for the initial, preliberalization state in order to estimate what the postliberalization situation will look like. With no exports present initially, crucial data is missing.

These remarks provide this author with the opportunity to explain the neglect of CGE models in his book, despite the fact that trade applications of the technique often employ a specific-factors specification, similar to that seen in Chapter 6. That is done because a specific-factors approach works better in the CGE context. Without that specification, model responses to policy changes can be excessive in magnitude. It is obvious that a CGE model cannot simultaneously estimate the structure of the world, and also model parameters. So no proof has been provided that the specific-factors story is the truth. CGE models can be used to predict, or they can be used to simulate what is known to have happened. In either case they remain problematic. One immovable problem is that critical model parameters, such as demand elasticities, have to be fed into the exercise. Reliable estimates of these numbers are frequently unavailable, and sometimes guesstimates have been used to fill the gaps. For whatever reason, CGE modelling is out of fashion, relative that is to its popularity in the 1970s. But note that any numerical estimates of the consequences of a policy change must be based on something like a CGE approach. On CGE modelling, see Abrego and Whalley (2005) and Devarajan and Robinson (2006).

## 7.6 Comparative Advantage

It is often claimed by economists in textbooks and elsewhere that every country must have comparative advantage in something. This is an extension of Ricardo's original argument. What precisely does this claim amount to? Comparative advantage is about low relative costs for tradeable goods, sometimes interpreted to mean a low relative autarky price for a tradeable good. In the light of this proposition what is one to make of countries that barely participate in international trade? If a small country is open to trade its tradeable goods prices are already aligned to world prices. If one thinks about a hypothetical autarky situation it is possible that the country may produce nothing of international-marketable quality. Once it opens up to trade the strong demand for imported products should drive the development of marketable exports. Sadly with a dysfunctional economy that export may turn out to be the local relatively skilled labour that will migrate abroad. This can lead to the dependent economy, one for which remittances and institutional transfers sustain an economy that would otherwise be scarcely viable. The importance of migrant labour and remittances is well illustrated by the Arab world. Many observers of that region note that the oil wealth is distributed most unequally, ranging all the way from superabundant countries, such as Saudi Arabia, to countries without any oil, including Jordan. While true in itself this observation fails to note that the entire region has shared to some extent in the oil wealth via the mechanism of migrant labour. Where teaching and administration are concerned the Arab language has favoured Egyptians and Palestinians in particular over non-Arab expatriots. This external source of funds has facilitated the non-participation in merchandise trade of the Arab region described above.

The point about marketable quality is similar to one made by Macartan Humphreys in his Oxford M. Phil. thesis, see Humphreys (2000). He asks whether labour-intensive activities are helped when a labour-abundant country opens up to trade. He argues that the labour-intensive product may be displaced by a more attractive capital-intensive product. The plastic bucket displaces the local wooden bucket. A few traditional buckets may be sold to tourists, but this in no way compensates for the fall in demand caused by the inflow of plastic buckets. The home country cannot produce its own plastic buckets as this requires capital and large-scale production.

## 7.7 Slaying the Participation Dragon

The analysis laid out above contains many suggestive ideas that may help to explain low participation in trade. Goods may be standardized and there may be no difficulties in marketing the product on account of its national origin. This is the case with many primary products. Even when these vary, as with particular oils and their sulphur content, quality control is not a problem and may in any case be handled by FDI arrangements. That explains why the export of primary products, particularly oil, is a relatively easy way to participate in international trade.

Transport costs are a major problem for trading even standardized products. In such cases getting the product to a seaport may be the main mechanism by means of which a good is made completely tradeable. These costs are at their highest for *landlocked countries* and/or countries with *poor transport infrastructure*. Those problems apply with great force in many African countries. That South Africa is the best trade performer among the sub-Saharan African countries is unsurprising and is explained by several factors. One of these is relatively good transport links. However several poorly performing West-African countries have good ports but often dreadful road/rail systems. On some implications of space and transport see Krugman, Venables and Fujita (1999). Wood and Jordan (2000) examine the contrast in manufacturing-export performance between landlocked Uganda and landlocked Zimbabwe.

The simple discussion above makes it appear that if one firm can make a profit from exporting once it pays the overhead cost of market entry, then it will go ahead. And if a domestic firm cannot raise the capital, a multinational can do the job. That is too simple for at least two reasons. The model has certainty while in reality large risks attach to entering a market. And the increasing returns implied by a fixed overhead cost mean that the 'toe in the water' approach to entry may stand no chance of success. The relatively small dispersed populations of many African countries make a large-scale jump into exporting particularly unattractive.

Costs of entry are not necessarily constants. They may depend upon history and upon the participation of other firms. Once people get the idea that the Chinese can make quality machine parts, it is far easier for other Chinese producers to enter the same or similar markets. Then we may encounter waiting games, as analysed by Bliss and Nalebuff (1984), when firms free-ride to avoid the highest 'first-in' cost of entry. In sharp contrast to Africa, China has set the manufacturing bandwagon rolling, so that more types of products are produced, partly on the back of previous successes.

## 7.8 Concluding Remarks

A vital aspect of the globalized trading system is the export of non-primary products from the poor countries of the 'South' to the rich industrialized 'North'. How did this come about? Three influences are important:

1. A reduction in protective trade barriers in the North.
2. The removal of gross anti-export distortions in the South, such as controls and tariffs on intermediate inputs.
3. Technical changes in the North that facilitate outsourcing. For example motor-vehicle assembly is more disaggregated, and techniques have been developed that make it possible to have Indonesian workers making jeans to a precise Kalvin Klein specification for sale in the US.

Arguably the third item on the above list is of greatest importance. It means that the problem of surmounting the initial barriers against market access are dealt with by rich-country buyers rather than poor-country sellers. We go to a department store and buy a frying pan and we rely on the store and its reputation to guarantee the quality of the product. That the pan was made in China does not concern us much, just as Kaldor (1960) noted.

Simple economic analysis by itself will never resolve Africa's trading problems. At best it offers partial insights. The economic geography of sub-Saharan Africa is particularly unfriendly to external trade, with huge sparsely populated territories, poor transport networks, and landlocked countries. That said, the Arab world including North Africa, mentioned above, does not suffer from those particular problems to the same extent, yet equally does not participate much in export trade. South Africa is an interesting case in point. Against the background of sub-Saharan Africa, South Africa appears as a success. Viewed more broadly its performance is less impressive.

It is unlikely that simply copying success will be the route that is needed. Even so, successful examples can be suggestive. The economic miracle in China started as a coastal phenomenon, and although that is changing, it remains true that the greater part of the industrialization is a seaboard city activity. How much that might be replicated in Africa is questionable, although South Africa has much of what is required.

# 8

# Institutions, Failing States, and Corruption

## 8.1 The Economic Environment

The following story is related by the great Islamic historian Ibn Khaldun:

> The philosopher Ibn Baja happened to recite at an Andalusian court a poem which made its ruler so enthusiastic that he vowed that Ibn Baja should go home walking on gold. Ibn Baja found himself in a predicament: to refuse or accept the gift was equally dangerous. To refuse would offend and anger, while to accept would make Ibn Baja so rich that his life might henceforth be at risk. Ibn Baja resolved the dilemma by begging for two pieces of gold which he proceeded to insert one in each shoe, and so walked home literally on gold, thus making the ruler's vow come true, and yet preserving himself from the perils of great wealth.
>
> (quoted by Kedouri 1992: 13–14)

In its way this tale depicts the concept of an economic environment. Ibn Baja's decision viewed as a maximization problem is trivial. Assuming that more gold in its own right is better than less, and there is nothing in the story to indicate the opposite, he should accept the ruler's offer. His difficulties arise from perceived reactions to either decision: the ruler's anger or his murder by thieves. To view it from a different angle, the story illustrates how autocratic rule and insecure property rights discourage the accumulation of wealth.

In Chapter 7 we re-examined the concept of comparative advantage. That analysis introduced effects that might well go under the title economic environment. The argument stayed as close as possible to the classical trade model. So participation in international trade involves a technical transformation much like any productive activity. Only the non-convexity of such transformation processes separated the analysis from standard competitive-equilibrium theory. That small difference forced the analysis to look at models that resemble monopolistic competition theory. The same change in turn broadens the scope of the political economy of production decisions. Now production requires a large-scale commitment of resources, and with that a necessary quantum of risk that cannot be moderated by limiting the scale of the gamble.

These ideas help a little with the understanding of the situation of the countries that fail largely to participate in world trade, as well as other failing dysfunctional economies. These economies are to be found numerously in sub-Saharan Africa and the Arab world. These regions include many failing states. The term immediately suggests social and governmental failure, corruption, and incompetence. To take the argument of Chapter 7 further it is necessary to elucidate formally what it means to the individual producer to operate in any economic environment. Then the failing state will be a special case, albeit the one of particular interest here. This chapter considers the situation of the individual producer in more detail. The need is to represent the problems of economic activity in difficult circumstances. For this purpose received models of competitive economic equilibrium are particularly inappropriate.

In the neoclassical competitive theory the producer's external environment is entirely defined by a vector of prices. Then if there are problems they must spring from incorrect prices. The approach to development economics that locates the source of many problems in wrong prices is a fruitful one. See in that connection Little (1982) and Little and Mirrlees (1972). For price distortions in poor-country labour markets, see Sen (1966). It is no surprise that comparing actual and shadow prices is a fruitful technique, because frequently in developing economies prices are wrong. However over the last twenty years that type of problem has declined in importance, even if it has by no means disappeared. The explanation is that many of the most gross distortions of prices in the past were the result of really bad government policy, especially highly distortionary taxation, and ill-chosen and excessive tariffs. The worldwide movement towards trade liberalization has produced much improvement in that regard, and this change demands a redirection of emphasis away from a simple wrong-prices account of contemporary development problems. A leading question today is why the shift towards more liberal economic policies has not been rewarded with more success, at least not in those countries that have continued to stagnate.

An exclusively price-based treatment of the producer's economic environment excludes features of that environment that cannot be reduced to prices. In the trade general-equilibrium models, including the HOS model and its extensions, all that matters to the firm is the prices that prevail in the markets in which it operates. Its production technology is given independently of the decisions of other agents in the economy, including large actors such as the government, except where the government affects prices. By making everything depend upon prices the trade general-equilibrium models exclude many important considerations.

## 8.2 Infrastructure and Government

One type of influence excluded by a pure price treatment is infrastructure. Infrastructure provides important instances of public goods. If an economy has a dreadful road system one might just be able to capture that feature by putting in a high implicit price for transport. Yet it would not be feasible to express in terms of a price the implications of a public electricity supply that cuts power on average four hours a day, and at unpredictable times.

A failing and incompetent electricity-supply system indicates a country that is not well governed. Government matters and in many ways. The literature on convergence has placed emphasis on the size of government as a proportion of national product, on the protection of property rights, and on democracy. See Barro (1991). Good performance on these indicators is argued to promote economic growth. Analysis in this style is open to questions discussed already in Chapter 4. In brief, do the indicators measure accurately what they purport to measure; and even if they measure well, how far are these features of a successful well-developed economy, rather than causes of its success?

## 8.3 The Mystery of Capital

A powerful case for the inhibiting effects on economic development of oppressive bureaucracy and poorly defined property rights is made by de Soto (2000) in his book *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*. Indeed this author shows that the two features are intimately connected. When it requires huge inputs in terms of time and expense to establish solid legal rights to property and to conduct a business, then these rights will remain tenuous and unclaimed. Many consequences follow. People may not choose high-quality and expensive residences, and to improve existing homes or workshops, when their legal rights to live in or to use these structures are non-existent. Why paint the front door when a knock on that door may announce the driver of a demolition vehicle, who shows an order to demolish the house? Shanty towns, de Soto argues, do not demonstrate simply the incurable poverty and lack of initiative of their residents. Shanty-town living is in part a rational response to insecure property rights.

In *The Mystery of Capital* de Soto provides numerous spectacular estimates of the costs that must be paid to obtain full legal recognition of ownership rights, or rights to operate a business. For Peru, eleven months and 207 admininstrative steps in many government departments, to obtain legal authorization to build a house. In Egypt the same process takes between six and ten years, and involves 77 procedures in 31 different agencies. Similar claims are made for other countries, and also for the costs of registering a business. These estimates could be too high to a serious degree without undermining the conclusion that the costs of 'making it legal' are prohibitive.

For informally owned assets de Soto coins the term 'dead capital'. His estimates of its prevalence in poor countries are dramatic. In Peru 81 per cent of rural property and 50 per cent of urban property is informally owned. In Egypt the equivalent shares are 83 per cent of rural and 92 per cent of urban property. In Haiti a staggering 97 per cent of rural property is claimed to be owned outside the legal system. The point that inaccessible property rights discourage the improvement of dwellings has been noted above. It also encourages unsafe developments. In Egypt there has been more than one case of the illegal 'improvement' of government-owned apartment buildings by adding additional floors that are then rented out, or allocated to relatives. This activity would be less attractive if official permission to build on vacant plots were easier to obtain. A tragic consequence of illegal 'height enhancement' to residential buildings is seen when the entire structure collapses, with heavy loss of life. This has happened on more than one occasion.

One can add that the precise insecurity that attaches to irregular residences is a function of the politics of the nation. This author has visited an acquaintance in Delhi, giving the taxi driver an address including the words 'unauthorized colony', one of those terms that Hindi speakers do not bother to translate from English. The more prosperous of these 'colonies' are smart by the standards of slums, and have been provided with drainage and electricity by the city authorities. That seizure or demolition are out of the question here owes everything to the fact that

India is a democracy. The unauthorized residents have votes. It is true that India has seen forcible slum clearances, but it is no surprise that these have been exceptional and have targeted small, isolated, and vulnerable groups.

A leading point developed by de Soto is that insecurity of property rights leads to shortage of capital. What is not owned solidly and legally, cannot be used as collateral for a loan. In rich countries people frequently finance a small-business project by taking a mortgage on their home. While this is highly risky, small-business investors are typically risk-loving and possibly over-optimistic concerning their chances of success. And notwithstanding the high costs to individuals of business failure, the proportion of these inititatives that succeed make a large and vital contribution to economic growth. The gamble that is a new business initiative is unavailable to the poor who lack any resources against which they can borrow, and this represents an important non-convexity, and a potential poverty trap, of the type examined in Chapter 3. The case laid out by de Soto includes a radical reform and development of property rights, in an attempt to cut straight through the block that weak property rights put in the way of economic growth. It is easy to dismiss this programme as wildly ambitious and absurdly infeasible. Yet de Soto's book makes a compelling case that successful development will not happen unless this reform can be achieved, at least eventually.

It is helpful to unpick de Soto's argument into two distinct components. First he offers eloquent testimony to the crushing burden of bureaucracy in the countries he examines. Secondly, de Soto has his own particular idea, that mushy property rights inhibit the growth of effective capital markets. Even if one questioned the importance of the second point, the first would retain huge importance. For that reason, when de Soto is depicted, in Latin America in particular, as reactionary, it should give pause for thought. If it is reactionary to expose the massive costs of strangulating bureaucracy, then we are in trouble.

Easterly (2005) provides a different account of the stifling effects of bureacracy. His bureaucrats are in the aid agencies of the rich West, and in these countries' governments. The grotesque inefficiency of poor-country donation leads to the outome that 300,000 people worldwide are raised from poverty, that is just above $365 per annum, by an expenditure of over $3000 per capita. Reforms that would improve this are not difficult to find, and Easterly suggests a few. The sad truth is that donor governments do not care much about poverty in distant countries, and are content to go through the motions.

Richardson (2004) argues that insecurity of property rights is a dangerously infectious problem. He claims that the land-reform programme for Zimbabwe, instituted by Robert Mugabe from the year 2000, which amounted to land confiscation, had much larger effects on the economy than might be expected, given that the farming sector was only 18 per cent of the economy. His point is that investors, observing that the ownership of land had become insecure, will infer that all property rights have become shaky. The result will be depressed investment and capital flight.

These compelling and depressing stories raise a pointed question. Is it that the authorities fail to recognize how disastrous are the institutions and policies that they allow to be, and to stay, in place? Or are there reasons why these social diseases (for that is the appropriate term) exist and endure? No doubt the correct answer is that it is a bit of each. Hernando de Soto notes that many popular beliefs amount in essence to the denial of the problems that he is at pains to advertise. For example, the idea that those people who live in the slums or *favelas* are hopelessly unenterprising and irrational, disguises the disincentives to change that these people face, precisely on account of the insecure property rights detailed above. Those disincentives are the same as the disincentives facing the imagined resident of Harlem depicted in Chapter 1. Yet even clear-sighted administrators may have strong, if sinister, motives to allow a catastrophic legal-administrative system to persist. There are various reasons why this may be the case.

First, bureaucrats are the last people to come out against bureaucracy. Secondly, in association with the appalling costs that oppressive bureaucracy imposes on its victims, there are huge benefits to the bureaucrats themselves. On the one hand, there is safety in numbers. There is nothing that bureaucrats fear so much as scandal and exposure. Permission has been given for a residence to be constructed, and now some kind of embarrassing side-effect of this project is exposed. Who is to blame? If this is Egypt, and 77 procedures in 31 different agencies were involved, it will never be possible to isolate, to blame, and to sack any particular bureaucrat. The complexity of the bureaucratic spider's web provides shelter for everyone involved.

But there is also the question of graft and corruption. De Soto counts the cost of surmounting the bureaucratic mountains that he describes so eloquently in terms of hours spent ploughing through formal procedures required to obtain the piece of paper that gives the desired formal legal entitlement. In reality these procedures often proceed more rapidly when they are oiled by bribes. That point does not undermine the fundamental argument. It matters not whether property ownership is unavailable on account of its huge cost in terms of time, or whether it is unavailable on account of its huge cost in terms of dollar bills in brown envelopes. This is from the point of view of the subjects of bureaucratic procedures. But from the point of view of the bureaucrat, bribes are a crucial plus. They will frequently comprise a source of income of equal, if not greater, importance than the official above-board salary. If 90 per cent of possible applicants for permits do not bother to apply, because of the prohibitive cost of bribes, that leaves a lucrative 10 per cent who will pay up and provide the bureaucrat with a more comfortable living.

## 8.4 Institutions, Religion, and Government Again

Thinking broadly, the range of things that may influence economic development is wide. North (1990 and 2005) argues for the critical role that institutions play in fostering, or inhibiting economic development. His treatment of institutions is fairly broad-brush. He lays his emphasis on security of property, as does de Soto, and on the absence of government interference in economic life. These influences, like the influence of culture or religion, have to be translated into specific features of the producer's environment. North compares institutions to *rules of the game*. While that comparison is suggestive it needs to be expressed as an exact formulation of the agent's decision problem.

Central to North's case is the assertion that good institutions lower transaction costs. Transaction costs can mean many things, and come from many sources. Access to the sea, and cheap sea transport, can be critical: witness the success of Venice, the Netherlands, and Britain, once it had secured sea routes widely, and eventually worldwide. Of greater importance in the early stages of industrialization is the absence of taxes on trade. The Middle East was riddled with these. It is not by chance that the Arabs gave us the word *tariff*. Early modern England was also afflicted with taxes on trade, as Adam Smith indicated. Furthermore, insecurity of exchange is a tax on exchange. If one in ten mail coaches are looted by highwaymen, then exchange using mail coaches attracts a 10 per cent tax on average.The big aching question of explanatory economic history: *Why not China*?, is answered in part in terms of transaction costs, and certainly not by scale, where China enjoyed a large comparative advantage. On China's failure to industrialize, see Landes (1998: ch. 21). In the case of China, Landes's dismissive attitude towards societies that did not modernize may be somewhat misplaced. The Chinese established a system of canals, to take rice from the south to the north, with locks and gates, over one thousand years before Europe discovered the same technology.

A parallel point applies to Robert Barro. When in Barro (1997) he examines the influence of religion on economic growth, via his growth regressions, he does not tell us *how* religion affects economic growth, but only *how much* it affects economic growth. The same is true of the analysis in Barro and McCleary (2003) in which religion is at the centre of the analysis, though that paper fills in a great deal more detail. The conclusion is that belief in God and an afterlife are positive influences for economic growth, but that attendance at church services has the opposite effect. Such curious conclusions can fuel uninhibited speculation. Could it be that church attendance consumes time that would be better devoted to book-keeping? Or is this another smoke-and-mirrors effect from the curious stage of the cross-section analysis, with its correlated variables and missing variables?

A loosely specified approach to religion and economic performance is in contrast to the famous essay by Max Weber (1930). Weber argued that Protestant Christianity played a crucial role in the generation of capitalism. Specifically, Calvinism produced a powerful individualistic ethic in which economic success is taken as an indication of election to salvation, and hard work is given the highest approval. Capitalism is the perfect vehicle for the laborious accumulation of wealth not directed to its hedonistic enjoyment. Many people who do not share that particular religious position, find it puzzling that one should get out of bed early to demonstrate that one is among God's elect, while knowing for certain that one cannot affect his choice. Such paradoxes give to religious belief its savour.

Weber's is a *how* argument. He claims to reveal the mechanism by means of which religion promotes capitalism, and with it economic growth. Against Weber, Tawney (1938) argues that Catholic Europe was already experiencing a capitalist revolution before the Reformation. Also Protestant sects aside from Calvinism promoted an ethic of hard work and sobriety, in particular Methodism. Tawney seems to want to have it both ways with regard to timing. He questions the priority of Calvinism with regard to capitalism, then emphasizes Methodism, surely a result of the frantic urbanization caused by the industrial revolution; hence never a cause of Capitalism.

Landes (1998: ch. 12) notes the growth of the opinion that Weber had it seriously wrong. Yet he writes:

> Indeed it is fair to say that most historians today would look upon the Weber thesis as implausible and unacceptable: it had its moment and it is gone.
>
> I do not agree. Not on the empirical level, where the records show that Protestant merchants and manufacturers played a leading role in trade banking and industry. (*ibid*. p. 177)

Probably there will never be a final resolution of the questions raised by Weber's musings. The critical question, as always, is the how part of the argument. Much economic progress has been contributed by minorities, and often a minority is defined by its religious beliefs. There can be an advantage from exclusion. Thus English Non-conformists during the industrial revolution were excluded from the ancient universities, with their concentration on a sterile classicism. That could well have been advantageous, as it enabled them to pioneer advances in applied science outside the most prestigious universities. But in that case, why did Catholics not show a response similar to that of the Non-conformists?

If exclusion is not in itself the explanation, could religion itself do the job? Here the difficulty is to make firm connections between religious belief and action, as connections that run from religion to action. Take the example of Isaac Newton. He changed the way that mankind viewed the physical world, and he was strongly religious in an unorthodox style. Newton was a Unitarian: he rejected the idea of the Trinity; the elusive tripartite specification of the deity that is found in standard Christianity. Can we imagine that Newton's Unitarianism might explain his extraordinary creativity? Surely not. More plausibly, his fantastic intelligence explains everything. Like many exceedingly bright individuals, Newton could think for himself. And it does not need amazing intelligence to conclude that the doctrine of the Trinity is riddled with difficulties. As to Physics, Newton built on what had been done before; notably on Galileo's principle of inertia, which became one of Newton's fundamental laws. Galileo, let it be said, was a Catholic, and a loyal son of the Church.

The subtitle of de Soto's book starts with the word *Why*, but this is not a question. The author proposes a clear and confident thesis, already discussed above. By contrast, Landes (1998) does propose a question in his subtitle: Why are some rich and some poor? In what amounts to an economic history of the world, no simple underlined answer to this last question emerges. Even so, it is not difficult to discern the broad shape of Landes's opinion on the issue. To put it simply is to throw away much that is both subtle and important. Running that risk, one can say that Landes's theory is modified Smith, where the term Smith refers to the author of *The Wealth of Nations*. Adam Smith thought that all could prosper, given only good non-intrusive government. He was aware that culture could make a difference, could indeed inhibit the accumulation of wealth. His enlightenment optimism, however, caused him to believe that cultural barriers could only form a temporary block to progress. Landes is less optimistic. His discussion of the Arab Middle East, although historical, will leave the reader with little expectation that things are likely to improve, not in the near future in any case.

## 8.5 A Three-way Classification

Any answer to the how type of question demands micro analysis. Therefore micro theory is an appropriate starting point. In Chapter 5 we could work with the aggregate economy-wide revenue function. This is the maximized sum of the revenue functions of individual competitive producers. If individual producers each maximize revenue subject to their separate production possibility sets, then the sum of their revenues is maximized subject to the economy's production possibility set, which is the sum of the individual production sets.

When we come to consider the economic environment it is no longer useful to work at a highly aggregated level. The economic environment is like the weather. While it may arise from grand global forces, its impact on the individual depends upon time, place, and individual circumstance. If it rains on a London resident it is only of secondary interest to that person if it is raining in Glasgow too. And that London rain will have different consequences if it falls from 1 a.m. to 6 a.m. in comparison with the consequence of it falling from 11 a.m. to 4 p.m. while the subject would like to watch a cricket match.

So we start with the economic environment of an individual producer. Let that producer be small at this stage of the argument. Small has the standard meaning here: the producer cannot influence the prices at which he trades, or he behaves as if he cannot influence prices. Now a three-way classification of influences on the producer is as follows.

1. Prices, as in the standard competitive model.
2. Non-price effects that can be expressed as a fixed vector of levels, parametric to the producer, just like competitive prices.
3. Non-price effects that represent reactions to the producer's actions.

## 8.6 Prices and Economic-Environment Variables

To start with prices in the standard competitive model, this gives us the revenue-function specification of Chapter 5. The effect of prices is entirely captured in the revenue function.

$$R(\mathbf{p}) \tag{1}$$

We have seen in the same chapter how this function can be modified by taking into account explicitly the effects of factor supplies. In that case the revenue function has to be the aggregate function for a whole economy, as individual competitive producers do not have fixed-factor supplies. We can interpret (1) as the revenue function for a single competitive producer. If we modify it to become:

$$R(\mathbf{p}, \mathbf{v}) \tag{2}$$

where $\mathbf{v}$ is a vector of economic environmental variables, we have maximized revenue, modified by these environmental variables. It will be seen that $\mathbf{v}$ in equation (2) is similar to the vector of factor supplies $\mathbf{z}$ that moderated the revenue function in Chapter 5. That comparison points up more than a similarity. In certain respects it is formally an exact equivalence, but there are differences also. From the revenue function (2) comes a vector of shadow prices of environmental variables:

$$R_{\mathbf{v}}(\mathbf{p}, \mathbf{v}) \tag{3}$$

These shadow prices will apply to an individual producer, and that already reveals a difference from the earlier theory. There are markets for mobile factors, and if these function efficiently, factor shadow prices will equal market prices, and will be equal for all producers. For the environmental variables it is different. They are equal for all producers and their marginal consequences will vary greatly between different producers. Their cost is like the benefit of a public good. It is obtained by summing the shadow prices of environmental variables across all producers

$$\sum_i R^i_{\mathbf{v}}(\mathbf{p}, \mathbf{v}) \tag{4}$$

where the subscript $\mathbf{v}$ denotes partial differentiation. If the revenue function is that of the whole economy, that is the sum of all competitive producers' functions, then equations (2) and (3) apply, and the similarity to the theory of factor-augmented revenue functions is complete.

Given that equivalence it is no surprise to find effects that look very like the Rybczynski effects of Chapter 5 appearing again in this new context. Should there be an infinitesimal alteration in the environmental vector $\mathbf{v}$ the consequent change in outputs is given by:

$$\left[R_{\mathbf{pv}}\left(\mathbf{p},\mathbf{v}\right)\right] \tag{5}$$

where the subscripts to $R$ denote partial differentiation, and (5) is a matrix whose typical element is a second-order partial derivative:

$$\frac{\partial^2 R}{\partial p_i \partial v_j} \tag{6}$$

This is equal to the rate of change of the $i$th output with respect to the $j$th environmental variable. This analysis is abstract and it is hard to picture the kind of concrete effects that it depicts. A huge range of possibilities is embraced. Consider just one example. Let one element of $\mathbf{v}$ be rainfall, and let rainfall decline, perhaps due to climate change. Then the model says that the entire vector of competitive outputs will change, some elements will decline, some will rise. The output of salad crops is likely to fall, the output of drought-tolerant crops will rise. A column of the matrix (5) indicates the magnitude of these effects for all outputs.

The symmetry condition of Chapter 5 now says that:

$$\frac{\partial^2 R}{\partial p_i \partial v_j} = \frac{\partial^2 R}{\partial v_j \partial p_i} \tag{7}$$

The left-hand cross-partial derivative shows the effect on the output of the $i$th good of an infinitesimal increase in the $j$th environmental variable. Then equation (7) says that this number is equal to the effect on the shadow price of the $j$th environmental variable of an infinitesimal increase in the $i$th price. The reason for this result is exactly as that examined in Chapter 5: the effect of an output price change on the shadow price of an environmental variable is just the alteration in the output concerned caused by that environmental variable change. In Chapter 5 this comparison was for the entire competitive economy. Here the revenue function is for one firm. So long as the firms are price-takers the equivalent measures can be obtained by summing over all firms.

If these comparisons between the competitive trade theory of Chapter 5, and the economic environmental theory of the present chapter, appear to be simple and reassuring, a word of caution is required. Imagine a small change in a component of the vector **v**. Make that a reduction in rainfall, so that we have a definite example in mind. This is a qualitatively different change from an alteration in the supply of one factor that we looked at in Chapter 5. There we were dealing with linear activities that separately and independently employ the factors that are available. So an increase in the labour supply, for instance, requires a change in activity levels that will employ that additional labour and leave demand for all other factors unchanged.

With a change in rainfall we are in different territory. This is not like a factor that must be allocated between different sectors. It is a measure of a commonly experienced climate. So all sectors are affected by this change, each in its own manner, and to a greater or lesser extent. It may well be the case that reallocations of separately allocated factors will follow from a reduction in rainfall. Agricultural activities may decline and the factors that these sectors release will be employed in, and will expand the size of, other sectors, in standard Rybczynski style.

We are more interested in the present chapter in environmental variables that are more directly the product of local government actions and social institutions than is rainfall. So imagine as a pertinent example de Soto's leading variable, bureaucracy. The vector **v** may incorporate various features of the administrative regime to which producers are subject. That would include rules and regulations, as well as time that must be expended in dealing with officials, obtaining their permissions, and even the cost of paying them bribes. Most rules and regulations are not well represented by continuous variables, but set that point aside for the moment. Then $R$ is revenue after bureaucracy has had its depressing effect. Notice that bureaucracy and corruption can resemble each other closely when their consequences for profitability and incentives are under consideration. Consider two cases. In the first the entrepreneur has to expend \$1000 worth of his time filling out forms and waiting for them to be checked and rechecked at the office of an official, in order to comply with a cumbersome regulation that is genuinely the law. In the second case, the same entrepreneur has to pay a bribe of \$1000 to get a licence for his premises to be used for production. To our entrepreneur these two cases amount to the same thing: \$1000 is \$1000.

Once again changes in bureaucracy induce alterations in outputs, the Rybczynski-style effects in that connection. An increase in an element of

**v** that represents the harshness of bureaucratic dominance triggers a full vector of Rybczynski-like output changes. If being small, or protected by the elusiveness of a particular activity, lightens the burden of bureaucracy, such activities will flourish. They are like Ibn Baja in the tale that opened this chapter; protected by obscurity from the perils of a hostile environment. This last point is of leading importance. That bureaucracy or its cousin corruption presses down on enterprise like a heavy sack on the back is obvious. It is more important to take into account the uneven and selective manner in which bureaucracy imposes its weight upon different and various activities. An evident and important conclusion is that heavy bureaucracy tends to encourage informal and unauthorized economic activity at the expense of formal and legally conformist activity. One can conclude that excessive bureaucracy encourages the 'black' economy at the expense of its opposite, which could be called the 'white' economy; although strangely that last term is not used.

In rich advanced countries it is argued frequently that bureaucracy is especially onerous for small enterprises. This is a pure scale effect. Despite some remissions for small businesses, form-filling is to a great extent a fixed overhead cost. It is less work to prepare tax returns, or social security deductions, for four employees than to do the same task for 400 employees. But it is nothing like 99 per cent less costly. For that reason the growth of bureaucratic regulation in many industrial countries in recent decades, in the form of more legislation and more reporting requirements, may have discouraged small enterprise, and particularly the start-up of new small businesses. See in this connection Ketter (2004).

With poor developing countries the picture may be significantly different. There the reach of bureaucratic interference may be far less complete than in a rich country. Control of the economy less typically takes the form of legislation that is largely implemented by circulation of the consequent statutes. This, plus a general understanding that laws should be obeyed, produces widespread compliance supported by only light bureaucratic oversight. This difference should not be exaggerated. In many rich countries including the Southern European nations, respect for the law, and for government in general, is far lower than is the case in the European North. Even so, wealth makes elaborate bureaucracy feasible, as it never is in poor countries. On how private agents may organize alternatives to government, see Dixit (2004), Gambetta (1993), and Grief (1993).

In poor countries government is often both weak and interfering at the same time. And weakness frequently promotes interference, because weakness generates insecurity, and insecurity leads the government to try desperately to control everything that it can. Weakness plus a strong desire to control makes the government control what it can, and that is usually the large organization. So where economic enterprise is concerned, bureaucracy is often more burdensome for the large enterprise, simply because that is where limited bureaucratic effort will show its highest return.

Our argument has slid gradually, perhaps almost unnoticed, from the simple formalism of the small firm and its given economic environment, to the interaction between large actors. Large here refers to particular local context in which these actors play out their roles. So it is that the bureaucrats of the above argument are like hunters, seeking out the firms that will give them the best hunting for a given effort. And the firms react, not to a fixed environmental vector, that confronts them like laws in a statute book, but to the reactive maximizing behaviour of their bureaucratic adversaries.

## 8.7 Rent-Seeking and Rent-Looting

In one of the most seminal papers ever on the boundary between the economic and the political, Kreuger (1974) gave us the concept of rent-seeking. See also Bhagwati (1982), Bhagwati, Brecher, and Srinivasan (1984), and Tullock (1967). Kreuger's idea is that lobbying or pressurizing to obtain particular legislative or administrative results is an economic activity, like any other, and subject to the same rules and principles. Suppose, by way of illustration, that a firm can employ five extra workers, to keep its machines oiled and well cared for. Were it to do that, its output would increase by 10 per cent. Alternatively the same five employees can lobby legislators to obtain a tariff that will increase its sale price by 10 per cent. Between those two possibilities the firm will be indifferent. Both increase revenue by 10 per cent for the same cost; they are equally profitable. The first strategy is efficiency improvement, economic growth if you like. The second is rent-seeking that allows the firm to benefit from increased economic distortion. In the terminology of Bhagwati, Brecher, and Srinivasan (1984), it is directly unproductive. In the example the two allocations of labour are equally profitable, and they are not strictly competitive, as both can be implemented simultaneously.

Imagine instead that the technology of rent-seeking uses many factors, as required, and that it is described by a constant-returns-to-scale production function. Now rent-seeking will be expanded until a greater extension is not economically sensible. When that state has been reached, rent-seeking will only add revenue equal to its cost. It will profit the firm nothing, but it will cost society the distortion caused by additional protection, or regulation.

All this is established theory, yet notice that the rent-seeking model applies to an economy with a strong legal structure. Laws are applied and are obeyed; they have to be manipulated to a firm's or individual's advantage. Lobbying in the US, and in the EU, fits this description well. That is not to say that lobbying and political manipulation are worthless where the ruler is autocratic, and laws are freely ignored. It is to say, however, that the game is different under an autocratic regime. Dictators can be bribed, with money, or with non-monetary pay-offs, and if their actions create rents, it does not matter whether these are the result of a legal change, or the result of an autocratic dictat.

What does distinguish law-bound societies from their opposite, is the subject of this section. Laws affect the economic environment in a manner to which agents react as passive players. In that regard they are not unlike prices. Real insecurity of property, with an unaccountable government, however, creates a different situation. Now *Rent-Looting* becomes feasible. The government can invent the rules as it goes, and any concentration of wealth can be a target for confiscation, specially if the wealth is held by a vulnerable group or person, without political power to repay bad treatment. See in this connection Schliefer and Vishny (1998).

Some of the most striking examples of the intimate connection between autocracy and the insecurity of property are provided by Islamic societies in history. Ibn Baja's story has already been mentioned. Landes (1998: 394–5) picks out Moghul India and the Ottoman Empire.

> Both these entities were aristocratic (despotic) empires in the classical mold: societies divided between a small elite and a large mass of fleeceable subjects . . .
>
> Below the divider, people had no rights, no security; only duties and submission. Resistance was next to impossible. The only escape from abuse was to fly or hide—the invisibility of nobodyhood. As one of the caliphs in Baghdad is said to have said: 'The best life has he who has an ample house, a beautiful wife, and sufficient means, who does not know us and whom we do not know'.

The reader may think that in modern times pure rent-looting is somewhat unusual. One thinks of Jews in Nazi Germany; East African Asians under Idi Amin; or white farmers in Zimbabwe. In all these instances the shift to hostility towards the group that suffered, happened relatively rapidly, and probably was not well anticipated. Far more important may be the wealth accumulation that never happened for fear of the rent-looting that it might have triggered. A possible example may be the Christian Arabs, who have migrated in large numbers from the Lebanon, to the USA and Australia, among other places. This group has felt increasingly insecure in a country in which the original Christian predominance has steadily lost out to increasing Muslim numbers and assertiveness. While motives for migration are never simple or all of a kind, worries about security of their property may be one factor involved.

Certainly the autocratic nature of contemporary Arab countries is a serious inhibition to economic growth. While mass confiscation of property, as happened under Nasser in Egypt, is not seen today, it is the case that the free utilization of private wealth is restricted. One recurrent problem, in Egypt and elsewhere, is that anyone who offends a powerful individual can find himself faced by endless troubles, manifested via the bureaucratic apparatus and in other ways.

## 8.8 Corruption

Another factor that has attracted attention is corruption. For surveys, see Aidt (2003), Bardhan (1997), Jain (2001), and Rose-Ackerman (1978). Corruption is a feature of what one might call the social infrastructure of an economy. Once again we need a how description as well a how much estimate. In this case it might appear that the how part is trivial, but that is not the case. Just to say that corruption has real allocational effects is not to say nothing. That is to deny the opinion that corruption only redirects rents without altering what economic agents do, and that point does need to be underlined. See Bliss and di Tella (1997) and Schliefer and Vishny (1998).

Many people will attribute corruption to culture, arguing the Northern European cultures are uncorrupt relative to Mediterrean cultures, or to addresses further afield, on account of the Protestant ethic, or something similar. Obviously there is much truth in that way of viewing the problem, but two contrary points deserve emphasis. First corruption is not one-dimensional. For example, US politics is notoriously corrupt, as the pork-barrel politics of Washington indicate. By contrast, try offering a US public officer a bribe. The US administration of government is among the least corrupt in the world.

Another important point is that corruption is always endogenous in the sense that corruption may breed corruption. Consider the following fanciful story. An Oxford professor, taken here, if my readers will excuse me, to be the essence of propriety and other-worldly inclination, is approached by a wealthy graduate student, who asks the professor to write his thesis for him, and offers in return a large monetary payment. Obviously the professor will decline the corrupt offer. Now, without changing the basic situation in any way, and with no alteration to the professor's moral position, change the environment within which our characters are operating as follows.

First suppose that writing other's theses for money is the general practice in the university at the time. So refusing to help the student will not force him to write his own thesis. It will simply redirect him to another academic, one less suited for the task, and the student will do less well than his fellows, and might even be said to have suffered an injustice. Secondly, suppose further that Oxford academics are exeedingly ill-paid; not as they are in fact underpaid, but so dreadfully that our imagined professor is working as a shelf-stacker at night, just to put food on his family's table. Now to refuse the bribe will be an injustice, not just to the student, as we have seen, but also more gravely to the professor's family. He accepts the student's offer.

What this parable-like story has done is to show that corruption is not just a product of the moral weakness of the individuals who engage in it. To be uncorrupt in a corrupt society may demand an almost saint-like commitment to propriety. As the guilty often say in their defence: everyone was doing it.

It is often proposed that the elimination of corruption should be a leading objective of the reform of poor countries. While it is hard to disagree, the manner in which the proposition is stated embodies a misconception, and a common misconception as well. Corruption is rarely a disease; it is almost always a symptom. Obviously, to stay with the medical analogy, it is sometimes appropriate, even essential, to treat a symptom, as with high fever. So anticorruption drives can contribute something.

To see the limitations of anticorruption drives, however, the disease itself must be identified, and the disease is weak government. Rampant corruption is nearly always a sign of a weak, shakey government. Examples are legion. Imperial Rome after Augustus degenerated into widespread corruption, largely because power ebbed away from the Emperor. The Abbasid Caliphate was infested with corruption. How can a government with apparently unlimited powers, including the absolute control of life and death, be called weak? The answer is illuminating. The Caliphs were weak on account of lack of information, not for want of legal instruments. Isolated and remote, they depended upon their officials to keep them informed, and these officials had neither the means nor the motives to make that information accurate. The court was a snake-pit of intrigue and competition, among officials who could be dismissed and ruined at any time, and who consequently played a short-term game. They accumulated bribe money as fast as possible, and avoided risky actions, which likely included giving the Caliph honest, accurate advice.

By way of contrast, consider the vigorous anticorruption drives that cleaned up New York's police under LaGuardia's mayorship from 1934, or that which purged Hong Kong of much corruption in 1974. Both of these were the result of strong government, with the power to investigate, punish, and dismiss.

A final example is particularly illuminating, because it allows us to observe a transition from rampant corruption to its elimination. England under the Hanoverian Kings George was among the most corrupt societies ever. Positions, from ecclesiatical livings, to parliamentary seats, to military commissions, to votes, were for sale. Colonialism, especially in India, was hard to distinguish from licensed looting. And numerous institutions that inhibited economic progress, such as tolls and taxes and restraints on trade, generated rents and locked in interests to preserve them. By allowing all this, the Hanoverian rulers bought support across a wide range of influential people. This fitted with the fact that Hanoverian rule was weak. It was established by foreign invasion, albeit an invasion widely welcomed. Its continued legitimacy derived from the fear that any alternative might be far worse; always a dubious justification for rule. Finally, spreading industrialization was shifting the weight of economic importance, away from the landed interest, that exercised power in Parliament, towards the new industrialists, who prior to the 1832 Reform Act were hardly represented.

Contrast the above account with Victorian England in the second half of the nineteenth century. Although Victoria was a Hanoverian, she was far better integrated into British culture than her predecessors. Victoria became popular, especially after her marriage to Albert, who despite his 'foreignness' was skilful at courting public support. Reforms hugely enhanced the legitimacy of Parliament; the Civil Service was reformed; and the Government increased its power. The extraordinary result is that Victorian England was one of the least corrupt societies that the world has ever seen. England is one of the cases discussed in Acemoglu and Robinson (2005).

The reason why a strong government will not tolerate corruption is not hard to comprehend. Corrupt agents siphon off revenues that might otherwise be captured by government. Why should government tolerate this, when it does not depend on these robbers for support? If it is conceded that corruption is a symptom, and the disease is weak government, what are the implications for policy? Is not making government stronger an even tougher policy objective than the elimination of corruption? Not necessarily. Generally speaking, undemocratic governments lack legitimacy and are weak; even when they cannot be removed. As poor country governments become more democratic, as is the trend, they may find the problem of corruption easier to address. It need not be added that democracy does not guarantee strong government. It depends a great deal on structure, and particularly upon how divided is the society concerned, and how its constitution deals with those divisions. As has been noted, US governments are weak. The vastness of the nation and the division of powers ensure that result.

## 8.9 A Model of Endogenous Corruption

The next model shows how corruption may be endogenous for purely economic, resource-allocation, reasons. It owes much to dal Bo′ and dal Bo′ (2004), although the present model is simpler than the dal Bo′ construct, but achieves a richer menu of possibilities, on account of its design. The model is similar in conception to Acemoglu (1995). It dispenses with some irrelevant features of that model, and at the basic level is more general.

We consider an economy with one unit of labour. This labour may be applied to production, or it may engage in corrupt rent-looting, called here *product diversion*. This product diversion takes the form of applying labour to creaming off a share of the honestly produced output. The one unit of labour is always employed. Of that total $l$ units work at producing output; while $m$ units apply themselves to diversion.

The return to producers is:

$$f(l)\,[1 - h(m)] \tag{8}$$

where $f(\cdot)$ and $h(\cdot)$ are increasing concave functions. As $h(\cdot)$ is the share of product diverted by the looter, it necessarily takes values on the range $[0, 1]$. The gross output is $f(l)$, while $f(l)\,[1 - h(m)]$ is the output that remains after a share $h(m)$ has been diverted. The return to corrupt agents is:

$$h(m)\, f(l) \tag{9}$$

The Nash equilibrium must satisfy the condition that the marginal productivity of labour is the same whether it works to produce or to divert:

$$f_1(l)\,[1 - h(m)] = h_1(m)\, f(l) \tag{10}$$

where subscipts denote differentiation. Or, rearranging and taking into account $l + m = 1$:

$$\frac{f_1(1 - m)}{f(1 - m)} = \frac{h_1(m)}{1 - h(m)} \tag{11}$$

The question of whether this model might have multiple equilibrium solutions turns on whether equation (11) may have many solutions in $m$. Note that, because $f_1(\cdot)$ is concave, as $m$ increases, $f_1(1 - m)$ becomes larger; while $f(1 - m)$ falls in value. As a consequence, the left-hand side is monotonically increasing with $m$.

As $m$ increases, both the numerator and the denominator of the right-hand side of (11) fall in value. That might suggest that the response of the right-hand side of (11) to a rise in $m$ will be ambiguous. That this is indeed the case is confirmed by differentiating the right-hand side with respect to $m$ to obtain:

$$\frac{h_{11}(m)\,[1 - h(m)] + h_1(m)^2}{[1 - h(m)]^2} \tag{12}$$

The sign of (12) is the same as the sign of:

$$\frac{h_{11}(m)\, m}{h_1(m)} \frac{1 - h(m)}{m} + h_1(m) \tag{13}$$

Or,

$$-\eta \frac{1 - h(m)}{m} + h_1(m) \tag{14}$$

where $\eta$ is the elasticity of the marginal product of labour in diverting product, regarded as a positive value. As $\eta$ may in principle take any positive value, and may vary with $m$ in any reasonable manner, the ambiguity of the sign of the response of the right-hand side of (11) is confirmed.

For multiple solutions to (11) it is not even required that the right-hand side should sometimes decrease with $m$. It is enough that the slope of the right-hand side should vary between above and below the slope of the left-hand side, so as to permit multiple intersections of the two curves. Any of these intersections will be an equilibrium, with the return to labour equal in the two activities. However some of these equilibria will not be stable, because a small fall in $m$, meaning more labour employed in production, may lead to the return to productive employment rising above the return to diversion. As is always the case with such models, stable and unstable equilibria alternate as $m$ varies. What matters is that the possibility of multiple, hence different, stable equilibria is firmly established.

A specific case of the model with a continuum of equilibria is the following:

$$f(l) = 1 + l \tag{15}$$

$$h(m) = 0.5m \tag{16}$$

With this specification, (10) becomes:

$$[1 - 0.5m] = 0.5\,[2 - m] \tag{17}$$

which is satisfied for all values of $m$ between 0 and 1. Any level of corruption is an equilibrium. The production function (15) has positive output for zero labour input. If this feature is seen as undesirable, the function can be modified for small values of $l$ to make it pass through the origin. Or the positive intercept at $l = 0$ might be interpreted as resource abundance, such as revenue from a foreign multinational exploiting a national oil endowment.

Take two stable equilibrium solutions for this model. For the same specification, both low and high levels of diversion are equilibria. What is the intuition of this result? In a poor economy riddled with corruption, the gross marginal product of labour in producing output is high, but the net marginal product is low, because there are many looters diverting the output. The marginal product of labour in corruption is also low, because there are many corrupt agents, and little output to be taxed. In the richer-economy solution everything is just the opposite. The gross marginal product of labour in producing output is low because there is high employment, but the net marginal product is high, because there are few corrupt agents to divert the output. The marginal product of labour in diversion is also high, because there are few corrupt agents.

## 8.10 Religion and Failure of Capitalism

An environmental vector **v** embraces so many possibilities that it seems that only a paucity of imagination could limit what it might include.

From the practical point of view, however, the elements of **v** should be measurable, otherwise we are back to the problem with religion already discussed above. While formal religious adherence is readily measurable, deep religous engagement is less accessible. The example of Italy, supposedly an overwhelmingly Catholic country, illustrates this point. And even where true religiosity is measurable it has no place in the environmental vector. It is absurd to suppose that the belief that Jesus was born of a virgin mother, common to Christians and Muslims, might directly influence the revenue function. The hard work consists in identifying aspects of the producer's economic environment that are measurable and impinge directly on profitability.

## 8.11 Reactions to the Producer's Actions

The final item in our three-way classification is non-price effects that constitute reactions to the producer's own actions. Until now the producer has been represented as an atomistic agent in a given environment. That environment is fully defined by two vectors, the price vector **p** and the non-price vector **v**. These influence how the producer behaves but are not affected by what the producer does. One might say that the environment so described is passive, at least as far as the indvidual agent is concerned. In some cases the economic environment does not seem to be passive at all. It reacts to what the producer does and the position becomes the same as that analysed using game theory.

At this point it is germane to show how sometimes non-price effects that constitute reactions to the producer's own actions can be taken into account by means of the fixed vector **v**. Once that has been done we can then go on to see why this same approach is not a good one. Take an absurdly simple specification of corruption, or equivalently insecure property rights. Suppose that corrupt agents rent-loot any profit in excess of a fixed value $P$, and assume that all agents are aware of this. Without this corrupt intervention the revenue function of a producer would be a simple standard:

$$R\left[\mathbf{p}\right] \tag{18}$$

Then with corruption the revenue function becomes:

$$R\left[\mathbf{p}, P\right] = \text{Min}\left\{R\left[\mathbf{p}\right], P\right\} \tag{19}$$

Here the reaction to the producer's choice is represented simply by a single parameter. It is not even a vector in this case, just the scalar $P$. Now if the producer is maximizing profit, then (19) represents maximum profit, just as does (2), repeated here for convenience as (20).

$$R\,(\mathbf{p},\mathbf{v}) \tag{20}$$

Thus (19) is a special case of (20) with $P$ replacing the more general vector $\mathbf{v}$. The example illustrates a more general possibility that the effect on maximized profit of the interactions between the agent and his economic environment might be reduced to a fixed vector of parameters. In that case the revenue function is a *reduced form*, one might call it the value of a game. And surely an indirect function such as a revenue function is always a reduced form. It gives a maximized value without displaying the process that achieves that maximization. In the context of producer decisions and reactions from others, this is not a good road to go down. We want to see how the game is played and how the result is decided, and an expression of the value of the game fails to illuminate those features.

As an illustration of what must be modelled, take the example of rent-looting already mentioned above. Treated as a game played between a rapacious government and an individual economic agent; what is the structure and what are the moves available to the players? The producer plays a risk-avoiding defensive strategy. The rewards of ambitious risky strategies are severely diminished. Rent-looting chops the top off high returns, with the consequence that agents will appear to be more risk-averse than their underlying preferences may truly be. More generally, like the caliph's subjects, or the peoples of the Soviet Union, the rule is keep your head down; do not attract the attention of the authorities; do nothing to stir up trouble. In the Soviet case there was little scope for capitalist innovation, but the same principles handicapped the administrators of state-owned enterprises. Their incentives to improve performance were blunted. To criticize the centralized planning system was to invite trouble, and in return for no reward. It is no surprise that the abandonment of communism was chiefly a top-down process. Even for the highest leaders, reform implied large risks. They could see, however, that the alternative was becoming hopeless, and successful reform promised gains.

# 9

# The Real Exchange Rate

## 9.1 Non-Tradeables and the Real Exchange Rate

The term *real exchange rate* is used so liberally in the economics literature that one might think that it has a unique unambiguous meaning. In fact it is used both to indicate relative price levels for traded goods in different countries, and also as a theoretical variable that measures the relative price for one country of its non-traded goods in comparison with traded goods.

For the time being assume that goods are divided strictly and completely between tradeable goods that move freely in international trade, and non-tradeable goods that cannot be traded at all. This assumption is used by Dornbusch (1980). The assumption entails that the quality of traded goods, if it varies, can be readily and costlessly assessed. The non-traded goods may include immobile factors, as in the HOS model. To keep the argument simple at the outset, assume one non-traded good.

A mixed price-quantity revenue function can be written:

$$R\,[\mathbf{p}, x] \tag{1}$$

where $\mathbf{p}$ are traded goods prices, and $x$ is the net output of the non-traded good. The revenue function is for a small country, so $\mathbf{p}$ is fixed. If the country can affect the prices at which it trades the analysis remains valid so long as individual producers are price-takers. It may however make the argument easier to understand if the reader concentrates on the small-country case.

Now:

$$-\frac{dR\,[\mathbf{p}, x]}{dx} \tag{2}$$

gives the shadow price of the non-traded output. Let the price of non-traded output in international value be $q$. What happens if $q$ is not equal to (2) when $x$ takes its optimal value? To be specific, let that $q$ be far higher than (2). Then we can write the revenue function in normal form as:

$$R\,[\mathbf{p}, q] \tag{3}$$

Full equilibrium obtains when $q$ is flexible. Then the two variables $q$ and $U$ are determined by:

$$\mathbf{p}\left\{R_{\mathbf{p}}\,[\mathbf{p}, q] - E_{\mathbf{p}}\,[\mathbf{p}, q, U]\right\} = 0 \tag{4}$$

which is balance-of-trade equality. Also:

$$R_q\,[\mathbf{p}, q] - E_q\,[\mathbf{p}, q, U] = 0 \tag{5}$$

is local market clearing for the non-traded good.

With $q$ a predetermined fixed price the local market for the non-traded good need not clear.

$$E_q\,[\mathbf{p}, q, U] \neq R_q\,[\mathbf{p}, q] \tag{6}$$

Then $U$ is determined by (4) alone and either buyers or sellers are rationed in the non-traded goods market. This is the situation as depicted in the *Dutch Disease model*, and in some *IMF adjustment programmes*. This last case requires an inflexible price. Then disequilibrium can be the result of nominal exchange rate misvaluation.

**Theorem 9.1:** *In a one-consumer economy misalignment of the real exchange rate q cannot increase utility.*

**Proof:** When equations (4) and (5) are satisfied the real exchange rate is not misaligned and there is a standard general equilibrium. Then the allocation is Pareto optimal, and no other feasible allocation can give the single consumer greater utility. □

Normally an exchange-rate misalignment will lower utility. If exchange-rate misalignment means an allocation that cannot be supported by any price system, as in (4) and (5), then this allocation must be inefficient. That follows from the *fundamental theorem of welfare economics* (every efficient allocation can be supported by a price system).

An adjustment of the real exchange rate can be brought about by means of **either** a nominal exchange rate revaluation **or** by alteration of prices of non-tradeables in the domestic currency. If the second takes place smoothly and quickly, the second is redundant.

## 9.2 Changes in the Terms of Trade

For developing countries in general, and for SSA countries in particular, changes in their terms of trade are of great importance. These define such crucial events as the Kenya coffee boom 1976–9. Many of the largest issues raised by terms-of-trade changes arise when the change is, or is likely to be, temporary. The static analysis of the present chapter is ill-suited to throw much light on that case. For Africa see in this connection Bevan, Collier, and Gunning (1992) and Collier (2003).

To see the effect of a terms-of-trade change on the equilibrium real exchange rate let $\mathbf{p}$ be a function of a parameter $\eta$. Then a change in $\eta$ can represent any change in the external terms of trade as required. Now differentiate equations (4) and (5) totally with respect to $\eta$ to obtain:

$$\frac{d\mathbf{p}}{d\eta}\left\{R_p\left[\mathbf{p},q\right]-E_p\left[\mathbf{p},q,U\right]\right\}+\mathbf{p}\left\|R_{p\mathbf{p}}\left[\mathbf{p},q\right]-E_{p\mathbf{p}}\left[\mathbf{p},q,U\right]\right\|\frac{d\mathbf{p}}{d\eta}$$
$$+\mathbf{p}\left\{R_{pq}\left[\mathbf{p},q\right]-E_{pq}\left[\mathbf{p},q,U\right]\right\}\frac{dq}{d\eta}-E_{pU}\left[\mathbf{p},q,U\right]\frac{dU}{d\eta}=0 \tag{7}$$

$$\left[R_{q\mathbf{p}}\left[\mathbf{p},q\right]-E_{q\mathbf{p}}\left[\mathbf{p},q,U\right]\right]\frac{d\mathbf{p}}{d\eta}+\left\{R_{qq}\left[\mathbf{p},q\right]-E_{qq}\left[\mathbf{p},q,U\right]\right\}\frac{dq}{d\eta}$$
$$-E_{qU}\left[\mathbf{p},q,U\right]\frac{dU}{d\eta}=0 \tag{8}$$

In the first line of (7) the term $\left\|R_{p\mathbf{p}}\left[\mathbf{p},q\right]-E_{p\mathbf{p}}\left[\mathbf{p},q,U\right]\right\|$ is the matrix of net supply changes induced by small alterations in tradeable prices $\mathbf{p}$. The first line of equation (7) shows the effect on the balance of trade of a small change in $\eta$; that is a small alteration of the external terms of trade of the kind under consideration. For convenience this term is denoted $BOT$. The first term in (8) shows the effect of the change in the terms of trade on the net supply of the non-tradeable good. This is denoted $SNT$, short for 'supply of non-tradeable'. Rewriting (7) and (8) in matrix form gives:

$$\begin{bmatrix}\mathbf{p}\left\{R_{pq}\left[\mathbf{p},q\right]-E_{pq}\left[\mathbf{p},q,U\right]\right\} & -E_{pU}\left[\mathbf{p},q,U\right]\\ R_{qq}\left[\mathbf{p},q\right]-E_{qq}\left[\mathbf{p},q,U\right] & -E_{qU}\left[\mathbf{p},q,U\right]\end{bmatrix}\begin{bmatrix}\frac{dq}{d\eta}\\ \frac{dU}{d\eta}\end{bmatrix}=\begin{bmatrix}-BOT\\ -SNT\end{bmatrix} \tag{9}$$

For the matrix in the first line of (9) the following assumptions are natural and reasonable. The top left-hand term is the effect on the balance of trade of an appreciation in the real exchange rate and will be assumed negative. The top right-hand term shows the negative of the value of the tradeable goods needed for a small increase in utility. This will be negative if tradeable goods are normal goods in aggregate. The bottom left-hand term is the effect on net supply of the non-tradeable good of an appreciation in the real exchange rate, and is assumed positive. Finally, the bottom right-hand term is the negative of the change in the consumption of the non-tradeable good required for a small increase in utility. This is positive when the non-tradeable good is a normal good. All these assumptions imply that the determinant of the matrix, denoted $D$, is positive.

What is the relation between the signs of respectively $BOT$ and $SNP$? The natural assumption is that they have opposite signs. To illustrate this point take the case of a coffee boom for a coffee-exporting country. Plainly $BOT$ is positive in that case. On the supply side the higher coffee price pulls production away from other outputs, including production of the non-tradeable good. On the demand side the higher coffee price shifts demand towards other goods, including again the non-tradeable good. Then a positive $BOT$ will usually be accompanied by a negative $SNP$.

Solving for the effects of small changes in $\eta$ using Cranmer's rule gives:

$$\frac{dq}{d\eta} = \begin{vmatrix} -BOP & - \\ -SNT & - \end{vmatrix} \Big/ D \tag{10}$$

$$\frac{dU}{d\eta} = -\begin{vmatrix} - & -BOP \\ + & -SNT \end{vmatrix} \Big/ D \tag{11}$$

To be specific, suppose $BOT < 0$, which corresponds to an unfavourable change in the terms of trade, and suppose $SNT > 0$ as argued above. Then (10) and (11) imply that the equilibrium real exchange rate will depreciate and utility will decrease. This is an entirely static model and it should be interpreted with care as the following argument indicates. Assume $BOT > 0$, an improvement in the terms of trade that can be thought of as a resource boom. Now the model says that the real exchange rate should appreciate just as it did in the Netherlands in the original Dutch disease boom. That episode was characterized by inflexible prices and immobile factors. These entailed costs and problems not demonstrated by the present model. Even so, the avoidance of any real exchange-rate appreciation is unlikely to be an optimal policy even in a more comprehensive model. Similarly some increase in utility is natural and sensible given a favourable change in the terms of trade. That said, a static model overestimates the optimal increase in utility.

## 9.3 Multiple Real Exchange Rates

One final point deserves mention. The formal analysis above has availed itself of the convenience that flows from assuming that there is only one non-tradeable good. Inspection of equations (4) and (5) will convince the reader that it is easy to generalize the argument to any number of non-tradeable goods. When that is done equation (5), which shows local market clearing for the unique non-tradeable good, is replaced by as many local-market-clearing equations as there are separate non-tradeable goods. These equations, together with equation (4) which is unaffected, then determine a vector $\mathbf{q}$ of real exchange rates, one for each non-tradeable good. The result is a fascinating possibility, namely that a country might have some of its real exchange rates misaligned, but not all. Or its various real exchange rates might all be misaligned but to different extents, or even in different directions.

These thoughts might be explored fruitfully via an examination of the condition of the UK economy in the decade from 1995. Many people argued that towards the end of that decade the UK housing market had become seriously overvalued. How far that is true is controversial and the detailed points that need to be examined are outside the scope of this book. For the sake of discussion take it as given that the UK's real exchange rate for housing was overvalued in the early years of the twenty-first century. The general equilibrium analysis of multiple real exchange rates shows why it is fallacious to argue that housing cannot be overvalued when it is priced in a free market of willing buyers and sellers. When a real exchange rate is misaligned there must be some market that is not clearing, but it does not have to be the market with its real exchange rate out of line.

For the UK in the period under consideration the balance of trade appears to be in severe disequilibrium. If that is indeed the case, it might be the result not simply of an overvalued nominal exchange rate between the pound sterling and other currencies, but also an indirect consequence of an inflated housing market. Does a similar argument apply to SSA countries or the Arab world? A case in point might be Egypt, which like the UK has a serious balance-of-trade deficit. In the past, and even today, the price of bread has been maintained at a low level by government policy. Freshly baked bread from a Cairo bakery is effectively a non-traded good, so this is a case of an undervalued real exchange rate. As the law requires bakers to provide bread at the fixed price that market is forced to clear. However it is not possible to force all markets to clear when prices are wrong. Another example from the Arab world is provided by the current situation in Jordan. Following the breakdown of order and security in Iraq, close to one million Iraqis have crossed into Jordan. These are not random Iraqis; they come from the wealthy, and they bring with them huge amounts of money. As a result, property prices in Jordan have rocketed up, far beyond the level appropriate for a medium-low-income country. One could say that the housing real exchange rate in Jordan is overvalued beyond that of the Jordanian dinar itself.

It will nearly always be the case in practice, at least in the short term, that real exchange-rate revaluations will be effected by an alteration in one variable: the nominal exchange rate. For that reason the disequilibrium situations created by multiple real exchange-rate misalignments will not be cured by a nominal revaluation, and may in some cases be made worse. These considerations may be of secondary importance when the nominal exchange rate is badly misaligned. It is argued that such is currently the case for China, as evidenced by that country's rapid accumulation of foreign exchange reserves. Is that the correct measure of disequilibrium? In a world of fluid capital mobility current accounts need not balance, so if all the Chinese surplus were going straight into capital purchases, would we see equilibrium? It needs a multi-sector CGE model to estimate the nominal revaluation that would be needed to reduce China's surplus by a specified amount. See Willenbockel (2006).

## 9.4 General Equilibrium Consequences of a Housing Boom

We can use the above analysis to see what happens when the price of housing is 'wrong' (i.e. not equal to its all-markets-clearing value). The model has two non-tradeable goods. One is housing with price $h$. The other represents all non-tradeable goods apart from housing, and has price $q$. The model is:

$$B = \mathbf{p}\left\{R_{\mathbf{p}}\left[\mathbf{p}, h, q\right] - E_{\mathbf{p}}\left[\mathbf{p}, h, q, U\right]\right\} \tag{12}$$

which defines the balance of trade, not required to be zero. Also:

$$R_h\left[\mathbf{p}, h, q\right] - E_h\left[\mathbf{p}, h, q, U\right] = 0 \tag{13}$$

$$R_q\left[\mathbf{p}, h, q\right] - E_q\left[\mathbf{p}, h, q, U\right] = 0 \tag{14}$$

show market clearing for the two non-tradeables markets. Now start with $B = 0$. Equations (12)–(14) then determine $h$, $q$, and $U$. We call these values the full-equilibrium values. Starting from those values, let $U$ increase and allow $B$ together with $h$ and $q$ to take the values required to satisfy equations (12)–(14).

## 9.5 Multiple Real Exchange Rates: An Empirical Study

Data on multiple real exchange rates is extremely scarce. In particular data on house prices converted to the dollar does not exist. There is data on house prices but only for a small number of industrial countries, and then only for the growth rates. To obtain a wide coverage I have had to turn to two exotic data sources. Table 9.1 shows first the BigMac index published by the *Economist* magazine for the 63 countries covered. This supposedly gives the price of a standardized product in dollars for each country. The idea is that a high BigMac cost will indicate that the country concerned

**Table 9.1.** Prices of Two Non-Tradeables

| Country | BigMac | Lodging |
|---|---|---|
| Argentina | 1.64 | 170 |
| Australia | 2.5 | 129 |
| Brazil | 2.32 | 148 |
| Britain | 3.44 | 274 |
| Bulgaria | 1.88 | 150 |
| Canada | 2.63 | 162 |
| Chile | 2.53 | 130 |
| China | 1.27 | 145 |
| Columbia | 2.79 | 122 |
| Costa Rica | 2.38 | 104 |
| Czech R. | 2.3 | 209 |
| Denmark | 4.58 | 156 |
| Dom. Rep. | 2.17 | 108 |
| Egypt | 1.55 | 145 |
| Estonia | 2.32 | 115 |
| Euro Area | 3.58 | 204 |
| Fiji | 2.5 | 125 |
| Georgia | 2 | 170 |
| Guatemala | 2.2 | 132 |
| Honduras | 1.95 | 122 |
| Hong Kong | 1.54 | 232 |
| Hungary | 2.6 | 146 |
| Iceland | 6.67 | 207 |
| Indonesia | 1.53 | 121 |
| Jamaica | 2.7 | 130 |
| Japan | 2.34 | 189 |
| Jordan | 3.65 | 117 |
| Latvia | 3.92 | 134 |
| Lebanon | 2.85 | 126 |
| Lithuania | 2.31 | 126 |

**Table 9.1.** (*Continued*)

| Country | BigMac | Lodging |
|---|---|---|
| Macau | 1.4 | 135 |
| Macedonia | 1.9 | 143 |
| Malaysia | 1.38 | 76 |
| Mexico | 2.58 | 181 |
| Moldova | 1.84 | 140 |
| Morocco | 2.73 | 150 |
| N. Zealand | 3.17 | 166 |
| Nicaragua | 2.13 | 115 |
| Norway | 6.06 | 155 |
| Pakistan | 2.18 | 178 |
| Paraguay | 1.44 | 116 |
| Peru | 2.76 | 135 |
| Philippines | 1.47 | 127 |
| Poland | 1.96 | 176 |
| Russia | 1.48 | 256 |
| Saudi A. | 2.4 | 153 |
| Serbia | 2.08 | 251 |
| Singapore | 2.17 | 135 |
| Slovakia | 2.09 | 175 |
| Slovenia | 2.58 | 121 |
| S. Africa | 2.49 | 153 |
| Sri Lanka | 1.75 | 93 |
| S. Korea | 2.49 | 204 |
| Sweden | 4.17 | 244 |
| Switzerland | 5.05 | 184 |
| Taiwan | 2.41 | 181 |
| Thailand | 1.45 | 125 |
| Turkey | 2.92 | 153 |
| UAE | 2.45 | 121 |
| Ukraine | 1.43 | 185 |
| Uruguay | 1.82 | 92 |
| USA | 3.06 | 134 |
| Venezuela | 2.13 | 157 |

has an overvalued nominal exchange rate in purchasing power parity terms. Unfortunately this simplistic and inaccurate view has degraded the data-collection process. Thus only one observation is recorded for the Eurozone, presumably because there is only one nominal exchange rate to be over- or undervalued. It would seem that the *Economist* believes that BigMacs must cost exactly the same in all Eurozone countries, though surely that cannot be the case.

The other data set is labelled 'Lodging'. This gives the maximum that anyone on US Government business can claim for one night's lodging. The data is assembled by the US State Department and published, and updated regularly, on the Internet. These, frankly, are hotel costs, not house prices. The two measures, hotel charges and house prices, are not closely connected and may diverge considerably. That said, the lodging variable is the price of a non-tradeable service. Were it more or less correct to assume that there is only non-tradeable good, then our two variables, BigMac and Lodging, would be highly correlated.

In fact the correlation between BigMac and Lodging is 0.24 and it just fails to be significantly different from zero at the 5% level. This is an enlightening observation because it contradicts *The Economist*'s view that only misalignment of the nominal exchange rate can explain a departure from purchasing power pority (PPP).

## 9.6 The Political Economy of the Real Exchange Rate

Theorem 9.1 is based on a typical one-consumer argument. With many consumers the level of the real exchange rate involves conflicts of interest, just as with free trade. Real exchange-rate overvaluation is a common phenomenon, particularly in developing countries. Often one or all of three reasons will help to account for this:

1. To control inflation countries peg their nominal exchange rates to a hard currency. This does not immediately moderate inflation, but if the peg holds the domestic price level will eventually stabilize at a high relative level. Then an overvalued real exchange rate is a consequence of inflation control.
2. An overvalued real exchange rate favours some members of the economy even while it harms others. In particular cheap imports are in the interest of many of the westernized urban middle classes, the very people who tend to enjoy excessive influence in the imperfect political systems that are found everywhere.
3. In poor dysfunctional economies the rich often hold their wealth in foreign currency. When they cannot or do not do that, they will not wish to see a nominal devaluation reduce the international purchasing power of their wealth, even if it leads to an improved flow equilibrium. Then the point made in (2) above applies. Those with an interest in maintaining an overvalued exchange rate may enjoy a political influence far in excess of their numbers.

Examples of clearly overvalued real exchange rates are not hard to find. They would include the pound sterling during the UK's unfortunate participation in the European Exchange Rate Mechanism that ended in

1992. Another example is the position of Thailand prior to its exchange rate crisis in 1997. Hinkle and Montiel (1999) in their large-scale study of exchange-rate misalignment identify further examples in SSA. These are the Cote D'Ivoire and the whole of the CAS zone in the period 1980–92.

The reader will notice that all the examples just exhibited seem to belong to the first reason for exchange-rate misalignment in the list given above. They are all cases of hard-currency pegging, and they seem to illustrate how such a policy can come badly unstuck when internal or external conditions alter. The other reasons listed, which can be summarized as the self-interest of powerful actors seem to lack a pure ideal example. While that may be true, the biased self-interest effect will be found, on close examination, to be present in cases that seem at first to be purely about hard-currency pegging. Consider the UK in the Exchange Rate Mechanism. Surely at the outset that was a policy initiative entirely designed to buy low-inflation credibility for the UK. And credible low inflation could be said to be in the general interest of all UK citizens. Perhaps so, but following German unification the policy came to have effects that entailed strong conflicts of interest between different groups. Put simply, an overvalued exchange rate is rather nice for those who remain employed, especially if they enjoy foreign travel. Equally it is good for the politicians in power whose credibility is at stake. On the other hand it is bad for the unemployed hoping to find employment. Thus the maintenance of an overvalued exchange rate, originally the result of anti-inflation policy, came to have a decidedly political-economic aspect.

## 9.7 The Resource Curse

An influential study, Sachs and Warner (1997*a*), examines the relation between resource richness and economic growth. The authors use a wide cross-section of countries for the period 1970–89 and claim that resource richness is negatively correlated with economic growth. In this connection see also Auty (1990) and (2001), Gelb (1988) and Hausmann and Rigobon (2002). Sachs and Warner assert that the negative association between resource abundance and economic growth survives the inclusion of additional standard variables in the growth regression, in particular measures of trade openness and of the quality of bureaucratic administration. Several oil-rich Arab states are not included in the sample for lack of data, but their growth performance is even worse than included resource-rich nations. This is an ambitious study, well designed to provoke debate and speculation. It is not without problems. Some coefficients are barely significant. Also the trade-openness and bureaucratic-quality variables measure what is required fairly imperfectly. For instance one oil-exporting state with no manufacturing to protect may well be 100 per cent open, where another resource-rich nation uses the revenue that its resource exports generate to subsidize local manufacturing, and is counted as less open.

These little points hardly matter as the grand question remains: why should growth be lower in resource-rich states? One known consequence of resource richness is the so-called *Dutch disease*. A resource discovery or boom increases the demand for domestic non-tradeable goods, either because the resource sector requires these goods, or from the income effect from local increased prosperity. Non-traded goods are then drawn out of the non-resource traded-goods sector. This constitutes a disease only because structural adjustment is painful, particularly for factors that must move, or whose price falls. Notice that this account includes no obvious association with economic growth beyond a temporary decline that may follow from factor unemployment.

Sachs and Warner suggest that an endogenous-growth explanation may lie behind their findings. If resource sectors generate less endogenous growth than do other sectors, then resource abundance may depress growth simply by shrinking the alternative good-for-growth sectors. That cannot be the complete story for the SSA countries, as an SSA dummy variable is typically significantly negative in cross-section regressions, while these countries are resource-abundant, or not, to a greatly variable extent.

## 9.8 Concluding Remarks

Real exchange-rate theory was developed in the 1970s, and the fundamental ideas require no amendment today. This chapter adds a new angle. It shows that the assumption that there is only one non-tradeable output is a useful simplification, but one that hides the potentially important point that there can be multiple real exchange rates, and that their misalignments may be in opposite directions. In Chapter 13 we will use an endogenous growth model to investigate the relationship between steady-state terms of trade and the rate of growth. The resulting analysis is complex, but in the most compelling case for a small developing country, improved terms of trade are associated with a lower growth rate.

# 10

# Mobile Factors and Urbanization

## 10.1 The Competitive Trade Model Treats Factor Mobility

In Chapter 5 we saw how the Heckscher-Ohlin-Samuelson trade model (HOS for short) can be used to support the counter-intuitive conclusion that factor mobility is largely without consequence, at least where factor prices are concerned. With two countries within the same cone of diversification, they will share common factor prices and common techniques of production. Suppose in that case that there is a small change in the factor supplies of one or both these countries, caused possibly by the migration of one of the factors from one country to the other. Then there will be a similarly small alteration in activity levels, to accommodate the new factor supplies, and no change in factor prices.

Here is a good point at which to bring together the various reasons why the picture just sketched may be misleading. Even within the narrow confines of the 2X2 HOS model, the assumption that our two countries will be found to occupy the same cone of diversification is seriously restrictive. There may be multiple cones of diversification, as when a factor-intensity reversal creates that result. In any case our two countries may not both have their production diversified between the two sectors. When that happens factor prices will be different, a mobile factor will have an economic motive to migrate, and migration will affect factor prices.

When we move beyond two factors and two goods, a fundamental change in the landscape is observed. It is the case that there are still cones of diversification, and within those cones are pairs of countries that differ in their factor endowments but not in their factor prices. These differences in endowments are accommodated by alterations in activity levels, just as with the Rybczynski effect in the simple HOS model. The radical difference that comes with higher dimensions springs from what is in essence a topological fact. In a two-dimensional space, as is represented by a standard two-axis graph on a flat page, the only cone that has no interior points is a straight line through the origin. Any two countries that occupy such a cone can differ only by their respective scales. In a constant-returns world, scale is fairly irrelevant, so the two countries do not differ in any interesting manner.

Now think about a three-dimensional space. Such a space is defined by the positive orthant of an infinitely extended space with its origin at a corner of a room, and the indefinitely extended two walls and the floor that meet there, forming its boundaries. Imagine a plane structure with straight sides meeting at a point. This is made of flat two-dimensional board, infinitely extended like the space in which it is positioned. When the pointed end of this structure is placed into the room corner, the board defines a two-dimensional sub-space of the room. This structure is a convex cone because it includes any convex combinations of its points. Also any linear projection of a point from the origin is contained within it.

The axes of our space measure factor endowments. For any pair of countries whose factor endowments place them precisely on the board the standard Rybczynski analysis applies. Take a separable three-factor two-good case, like the one examined at length in Chapter 6. The third factor is only used in Sector 1. The unit input of factor $j$ into sector $k$ is $a_j^k$. The output level of sector $k$ is $y^k$. Then with the total supply of factor $j$ equal to $z_j$, full employment of the three factors requires:

$$a_1^1 y^1 + a_1^2 y^2 = z_1 \tag{1}$$

$$a_2^1 y^1 + a_2^2 y^2 = z_2 \tag{2}$$

$$a_3^1 y^1 = z_3 \tag{3}$$

What happens if migration takes the form of an inflow of unskilled labour? If the supply of skilled labour were not a problem we would be back to standard Rybczynski theory. There would be no effect on factor prices because the low-tech sector would expand, and the high-tech sector would contract to absorb the increased supply of unskilled labour. In the general model a similar story applies. But now the contraction of the high-tech sector causes a fall in the demand for skilled labour, the price of which falls. What that implies depends again upon the elasticity of demand for skilled labour. If, as seems likely, the demand is inelastic, then $a_s^1 w_3$ will fall. That, as we have seen, is equivalent to a rise in $p$. One could say that the Rybczynski effect induces a Stolper-Samuelson effect. A rise in $p$ raises $w_1$, the return to capital, and lowers the unskilled wage rate $w_2$. An inflow of unskilled labour harms all workers and is good for capital.

The above discussion confirms a finding of Chapter 5. This says that close to any country that is fully diversified with regard to production will be found possible countries, which may or not exist in reality, that share the same factor prices but have different production levels. In our higher-dimension discussion that statement is consistent with another claim. *Close to any country that is fully diversified with regard to production nearly all possible countries will not share the same factor prices*. Here nearly all means all, excluding a measure-zero subset. If we relate this conclusion to our discussion of migration we can say that in nearly all neighbouring countries factors will have an economic incentive to migrate. This conclusion is coming from an analysis in which technology is the same in all countries and there are no increasing returns to scale, though both these last features are powerful motors of factor migration.

## 10.2 Perfect Factor Mobility

If we lived in a world of perfect factor mobility, meaning that all factors could migrate between countries costlessly in search of higher returns, then the only possible equilibrium would be one in which the returns to all factors present in positive quantities would be equal in all countries. The qualification 'present in positive quantities' is needed to allow for the possibility that the return to a particular factor might be low in a certain country, when all that factor will migrate abroad. It may well be the case that the salary that a particle physicist can earn in Chad is low, and there are no particle physicists in Chad. This case requires that the factor concerned be inessential, as is presumably the situation with particle physicists.

The existence of a world equal-returns equilibrium requires only simple convexity and differentiability conditions. Let there be $n$ countries indexed by $i$. The national product of the $i$th country is:

$$f^i\left[x_1^i, x_2^i, .., x_n^i\right] \tag{4}$$

where $x_j^i$ is the quantity of the $j$th factor present in country $i$, and all the $f$ functions are concave. Let world supplies of the various factors be:

$$X_1, X_2, .., X_n \tag{5}$$

**Theorem 10.1:** *An equal-returns allocation of world factors exists in which the marginal products of all factors present in positive quantities are equal for all countries, and this allocation maximizes the sum of national products across all countries.*

**Proof:** Consider the maximization of:

$$\sum_i f^i \left[x_1^i, x_2^i, .., x_n^i\right] \tag{6}$$

subject to:

$$\sum_i x_j^i \leq X_j \tag{7}$$

all $j$, with the $f$ functions continuous and concave. As (6) is continuous on a compact support this problem must have a solution. That solution must maximize the Lagrangean:

$$\sum_i f^i \left[x_1^i, x_2^i, .., x_n^i\right] + \sum_j \lambda^j \left(X_j - \sum_i x_j^i\right) \tag{8}$$

This requires that:

$$\left[\frac{\partial f^i \left[x_1^i, x_2^i, .., x_n^i\right]}{\partial x_j^i} - \lambda^j\right] x_j^i = 0 \tag{9}$$

for all $i$ and $j$. Equation (9) is a complementary-slackness condition. To satisfy the equation either $x_j^i = 0$, or:

$$\frac{\partial f^i \left[x_1^i, x_2^i, .., x_n^i\right]}{\partial x_j^i} = \lambda^j \tag{10}$$

when (10) says that for all $x_j^i > 0$ marginal products are equal in all countries, as required.

Finally note that when the $f$ functions are each concave, then (6) is a concave function of all the $\left(x_1^i, x_2^i, .., x_n^i\right)$ values taken together. Now when (9) is satisfied there must be a global maximum of (6). If not a maximum is at $\left(\tilde{x}_1^i, \tilde{x}_2^i, .., \tilde{x}_n^i\right)$ for the various values of $i$. Then $\left(x_1^i, x_2^i, .., x_n^i\right)$, again for the various values of $i$, is a local maximum. The value of (6) can be evaluated along the line:

$$\lambda \left(x_1^i, x_2^i, .., x_n^i\right) + (1 - \lambda) \left(\tilde{x}_1^i, \tilde{x}_2^i, .., \tilde{x}_n^i\right) \tag{11}$$

for all $i$, when (6) becomes a function of $\lambda$, denoted $F(\lambda)$. Starting at $\lambda = 1$ the function $F$ must first decline and then increase. This contradicts the concavity of (6). □

In a concave world with perfect factor mobility Theorem 10.1 shows that an equilibrium exists that equalizes returns for all factors in all countries, and that any such equilibrium will maximize world production. The equilibrium need not be unique. It might happen that two of the countries have identical $f$ functions with constant returns to scale. Then the particular allocation of factors between these countries is indeterminate, but also unimportant for world production. This is a case of concavity that is not strict concavity. True non-concavity allows for completely different possibilities. Now equalization of factor returns need not imply global production maximization. Imagine that a certain country has huge potential once its human population has reached a critical level, but below that level is inefficient and unproductive. There might then be two possible outcomes for this country. In one it is unpopulated, with the wage rate that it could provide below the world level. This is the corner solution permitted by equation (9) above. In another case it is well populated and pays the world wage rate.

In general complete factor-price equalization requires that all factors be completely mobile, but there is one important exception to that requirement. We have seen this already in Chapter 5 above. When all countries share the same $f$ function with constant returns to scale, only ratios of factor inputs matter, and one factor can stay where it finds itself, while the other factors move to establish the required ratios. That is the type of case that people have in mind when they argue that increased mobility of capital in the modern world may weaken the incentive for labour to migrate. This point is in addition to the conclusion that sometimes at least, trade may bring factor prices closer together, and once more moderate the incentives for labour migration.

## 10.3 Barriers to Capital Mobility

The world depicted in the previous section, in which all factors are perfectly and costlessly mobile, is a fanciful creation. It is evident that we live in a world in which the barriers to factor mobility are enormous. That is particularly clear where labour is concerned. Even for labour however, two questions must be considered. First, what precisely are the crucial barriers to international labour mobility? Are they mainly legal-administrative restrictions, or are other constraints on mobility equally important? Secondly, why does capital, that should be more readily mobile than labour, not make up for deficiencies in labour mobility, by going to the labour rather than waiting for labour to come to it? For the latter process to be fully effective we must have constant returns to scale, or something close to constant returns to scale. But we do not need that everywhere in all sectors. It is enough that there be substantial sectors open to freely mobile capital that can offer employment to the national labour supply at a high wage rate.

Aside from constant returns to scale, another condition that must be satisfied if labour mobility is to be unnecessary is that technology should be as freely mobile internationally as is capital. We have seen already in Chapter 8 that there are numerous and various reasons why technology may not move easily from one country or region to another. Technology transfer is not a matter of sending blueprints and technical descriptions through the mail. It involves infrastructure, the cultures and customs of labour forces, and many other influences that are far outside what is normally understood by technical knowledge. Even climate is involved. In the tropics air-conditioning may be necessary for the production of a pharmaceutical product, where in Europe air-conditioning in a similar factory would be a luxury.

In the realization that its potential role may be of the greatest importance, we now consider capital mobility, and in particular barriers to capital mobility. It is a truism that investment in all countries of the world must be equally attractive to international capital. That statement must be qualified by the same complementary-slackness condition as was exposed in equation (9) above. That is to say that a country may offer a lower attractiveness to international capital than the going rate, if it attracts no international private capital. There are countries in the world that receive no foreign private investment from standard sources. Sometimes in such cases there is an inflow of private capital from expatriates who wish to retain an interest in their birth country. They may be motivated by a wish to return eventually, or they may wish to take advantage of their insider knowledge and networking options, to gain a return that is not available to world investors in general.

In labour economics there is a well-established principle known as *equality of net advantages*. This means that in a free labour market different occupations that demand equivalent skills need not pay the same, or even closely similar, wage rates. What matters is the entire employment package considered in all its aspects. So night work pays more than identical day work. Hard dirty work pays more than clean desk work. Insecure temporary employment pays more than secure employment. Jobs with good pension packages pay lower current wages than do jobs without such packages. And so on. All these comparisons are valid only with an 'other things being equal' proviso. Clearly there are easy attractive jobs that pay high wages, because they demand unusual skills. Also some labour markets perform a sorting function. On average actors earn very little, yet a small elite in the profession enjoys great prosperity. The trouble is that actors do not know in advance what their success will be. Worse still, the profession may well attract the congenitally overoptimistic and risk-loving. For these at least the net advantage of the acting profession should be the same as their alternatives.

The same equality of net advantages principle applies as well to international capital as it does to local labour markets. When we say that investment in all countries of the world must be equally attractive to international capital, we should add that equal attractiveness should be taken to mean equally attractive taking into account all the net advantages or disadvantages that may apply. The implication of that simple equality is that features that importantly affect the net attractiveness of investing in a country, similarly affect the cost of capital to local producers. Imagine as an illustration, that foreign investors are worried by the threat that their capital may be expropriated. Perhaps local politicians have been delivering speeches hostile to foreign capital and threatening its confiscation. In that situation investors will look for an unusually high return to compensate for the risk that there may be no return at all. And if international capital demands a high return, then local investment will pay a high price for its funds.

The net advantages of investment must be distinguished from differences in national technology and the quality of local factor inputs. This is a distinction for the economic theorist, and it could easily be lost on an investor. Suppose for example that the quality of the labour force in a particular country is poor from a factory owner's perspective. Local labour is ill-disciplined. It engages in part-time agriculture, with the consequence that workers frequently fail to attend for work when the fields demand attention. Other workers, or sometimes the same workers, are not infrequently drunk, with the expected reduction in their productivity. These features will certainly discourage investment, and the resulting scarcity of capital will keep up its rate of return, notwithstanding the poor quality of local labour. The crucial point here is that this is not a problem of a high cost of capital, and local investors need pay no more for capital than do other investors anywhere. It is just that given the standard cost of capital, less will be demanded where local production conditions are unfavourable.

The world market for international capital is only one instance of a capital market, and it is subject to some of the principles that apply to any capital market. The modern view is that imperfect and asymmetric information are crucial features of capital markets, and that these cannot be understood without taking the nature of information and its distribution into account. On asymmetric information in a standard market, see Akerlof (1970). The seminal paper is Stiglitz and Weiss (1981), which shows how imperfect information leads to credit rationing. See also Bell (1988) for the implications of this and other problems for rural credit markets. In general, investors prefer to place their capital where the quality of their information is good, and they are reluctant to place it where their information is poor, and the risk consequently is perceived to be large. Local information is nearly always of better quality than information concerning the remote. That leads to the conclusion that capital markets are always segmented, with some segments provided with capital on more generous terms than others. This observation is part of the explanation of why countries that save a lot also invest at a high level. There is no reason why that should happen in a world of fully integrated capital markets.

The distinction explained above between the cost of capital and technology is of particular interest to the economic theorist because it turns on the question of which assumption of the standard HOS model is violated. Complex patterns of influence on the net advantages of investment show themselves at the level of the simple production model as imperfect capital mobility. Variations in technology, or production functions, between countries imply just that for the HOS model. The assumption that the same production functions apply to all countries is discarded. The standard HOS model assumes that there is no capital mobility at all. This assumption is often contrasted with an alternative assumption under which capital is perfectly mobile internationally. A net advantage approach takes neither of these two paths. It says that capital is internationally mobile, but that mobility is far from perfect. When we turn to technology the HOS model says that this is the same in all countries, and that presumption is equally questionable. The economic historians who write about technology transfer and differential rates of factor-productivity growth between nations, would have nothing to discuss if technology were the same in all countries. See in this connection David (1975).

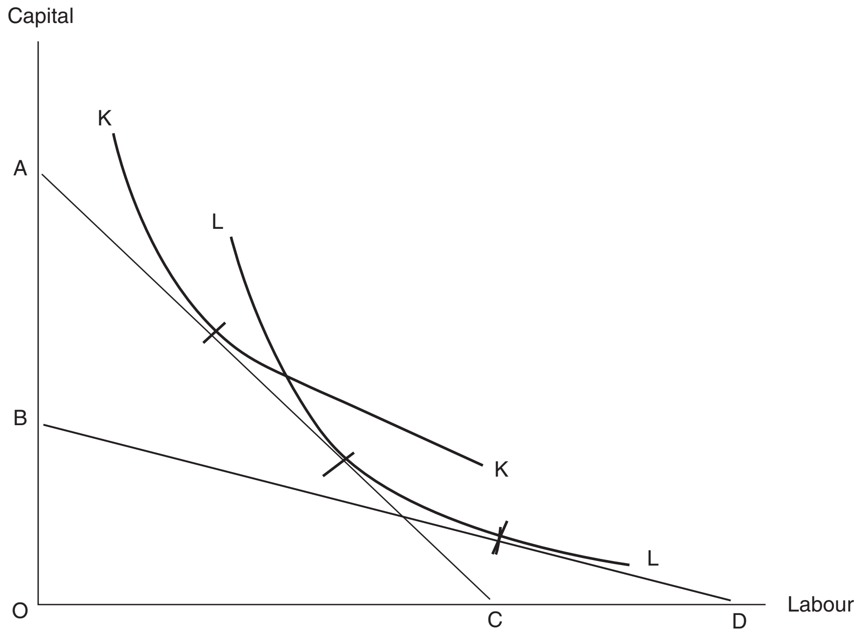



**Figure 10.1.** Equilibrium with Capital Rationing

What precisely imperfect capital mobility implies for an HOS-style model depends upon how exactly the possibilities for capital mobility are specified. Take two contrasting cases. In the first, Country A can attract foreign capital. However foreign investors will never allow their ownership of the national capital of that country to exceed 50 per cent. If in comparison with the world norm the country is capital-scarce, then it will attract capital, as far as doubling the stock if that does not bring it beyond the world norm. Given the capital stock thus determined, the solution of the model is exactly the same as for any HOS model. The capital stock is fixed and the division of its ownership between home and foreign is without relevance. In the second case foreign capital is available without limit, but foreign investors require a higher return than they demand for home investments. This fixes the rate of return to capital.

Figure 10.1 shows the effect on the HOS solution of a fixed return to capital higher than that which applies to the diversified equilibrium. The unit-value isoquants for respectively the capital-intensive sector and the labour-intensive sector are KK and LL. The line of unit-value factor quantities in the diversified equilibrium is AC, and a unit-value quantity of capital is OA. The diversified equilibrium is the outside world in this argument. It refuses to give our country enough capital to bring the cost of capital down to the level represented by the distance OA. Instead the home price of capital is measured by OB. This is a smaller quantity of capital: a higher price. Now the only tangent from B to the envelope of the two isoquants is the line BD. This is a non-diversified outcome. Only the labour-intensive sector produces. Capital shortage leads to a lower wage rate and possibly, if the labour-intensive sector is interpreted as primary production, to an undue dependence on primary exports.

The figure gives an answer, albeit its own particular answer, to the question of what effect capital mobility has on international trade. For that simple 2X2 HOS model, cheaper capital does not much affect trade. This is because it is only when the cost of capital has fallen to the advanced country level that the country can diversify into capital-intensive production. Low wages do not achieve anything as long as capital is costly. This is an absurdly simple model, and with only two sectors it is hard work to relate it to the real world. It may not be correct to think of the labour-intensive sector as primary products. It might be simple manufactures as opposed to complex manufactures. Then the model will say that countries with poor, that is costly, capital markets will have to export the simplest manufactures if they can export manufactures at all. There may be a clue there to the success of China as an exporting nation. Capital for China's manufacturing comes from domestic saving, and given the problems of imperfect information noted above, that means that capital for China's manufacturing is cheap. That makes the breadth and sophistication of Chinese manufactured exports somewhat easier to understand.

## 10.4 Barriers to Labour Mobility

Judging by press commentary and popular discussion in the West one would think that the most important aspect of labour migration in today's world is migration from poor countries into Europe and the USA. The same sources might leave the impression that formal barriers to the movement of labour, immigration controls and green-card rules, as well as drives against illegal immigration, are what matter most. None of this is correct. The largest movements of populations in the world today are inside the Third World, the poor developing countries. They are between those countries on a huge scale, and within those countries on an even larger scale. A leading importance attaches to migration from rural to urban areas. This massive redeployment of humanity has increased considerably in the last two decades, but has been happening for far longer. See Williamson (1988) for more detail.

While politicians frequently complain that labour migration is far too easy, the most striking fact is how difficult it is for most people in most places and at most times to migrate. The typical human being will stay where he or she is born unless conditions in the place of birth become so stressful that migration commends itself. There is always a minority of footloose people who must see the world and who migrate mainly to get away from their birthplace, but they are the exception. As many writers have pointed out, the difficulty that attaches to migration varies greatly across individuals. Migration is easier for the young than for the old; easier for the educated than for the uneducated; easier for males than for females; easier for those who are weakly integrated into their existing society than for those deeply integrated, with many ties and local advantages that must be abandoned with migration.

At home a man is, as the phrase has it, at home. He speaks his birth language, which is a local language. He knows the local culture intimately, and can operate within it easily and accurately. If the local culture is oppressive and discriminates against him, that may bring thoughts of migration to his mind. This is a leading motive for migration, but escape in that form may come at a high price. It may mean leaving family and friends; it may imply going to a place where qualifications are not recognized; where language is a huge problem; and where the hostility towards a minority that is left behind is replaced by a new hostility towards migrants and outsiders.

On occasion migration involves truly horrific risks. An illustration of this point is provided by the migration of 'illegals' from SSA countries into Europe. This requires that two huge hostile spaces be traversed: the Sahara Desert, and a sea; either the Straits of Gibraltar, or the perilous crossing from Mauritania to the Canary Islands. Each year thousands of Africans lose their lives on these journeys, drowned when makeshift boats overturn or sink. Their bodies washed onto the beaches of Europe or Africa, provide eloquent testimony to the barriers to labour mobility, and also to the terrible drives behind desperate men whose native lands offer them no hope whatsoever.

It is no surprise then that migration takes place most readily when it involves movement into an expatriate community of people who share culture, language, and religion. If this community as well as the migrants, already has a good grounding in the language of the destination country, then that makes long-distance relocation easier. There are numerous examples of migrations that form expatriate communities. Sometimes these act as a route to fairly complete absorption into the local culture. This happened with the emigration of Jews from Poland and Russia to the United States, at least in the second or third generations. Sometimes the migrant community forms an isolate. It keeps its own culture and language to such an extent that it fails to integrate. Complete cultural isolation is rare, especially where locally born individuals are concerned, but partial isolation can happen. Some Muslim minorities in Europe fit this description, with current well-advertised consequences.

No mention has been made so far of formal legal barriers to migration. Of course these are of great importance. That said, they are of limited effectiveness, especially and obviously when internal migration is concerned. It is estimated that there are eleven million illegal migrants in the United States, mainly from Mexico and Latin-American nations further south. Elthier (1986) models illegal immigration. These migrants have in most cases crossed the extended and porous Mexico-US border. This border is policed ineffectively by the US immigration service, and protected more effectively for much of its length by the hostile and waterless environment of the southern US deserts. Latino migrants to the US provide an example of migration into expatriate territory. In California and in Texas there are areas where English is a minority language. The rapidity with which integration occurs is in part a function of policy, and in this regard US government policy has been foolish to the extent that it has permitted school teaching to be conducted exclusively in Spanish. The importance of language is underlined by the fact that many migrants into the EU make for Britain, often crossing many EU countries to arrive there. The attractions of Britain include relatively easy labour markets. Above all, the English language, which is often the second language of migrants, is what recommends this cold and rainy island to the migrant.

Another misleading picture concerning migration has it that it is the movement that is costly. Once that cost has been paid, this view has it, entry into the local labour market is the reward. This is seldom the case. Rarely do labour markets, in particular those accessible by migrants, take the Walrasian form. Here Walrasian means that all would-be sellers of labour constitute a homogeneous supply, with the wage rate equating supply and demand. Most labour markets are managed, in the sense that access is rationed by legal regulations, or by various barriers that range from trade-union control of entry, to the criminal exclusion of the unacceptable. Not infrequently, entry to a labour market demands the payment of bribes, or more restrictively still, membership of a clan, or a network based upon familial relationships, sometimes surprisingly remote. The current political debate has created a character who goes by the name of 'the economic migrant'. He plays opposite his antithesis, the asylum-seeker. Simple dichotomies are always appealing to the mind: generosity versus self-advertisement; freedom versus licence; carrot versus stick; trees versus wood. In the same spirit, is migration driven by the pursuit of economic gain, or is it a flight response to political and racial oppression? No doubt there are examples that represent either case in its purest form. Yet these are unusual.

Take an instance for which flight seems at first to be the whole story. There were massive migrations out of Europe to the New World following the potato famine of the mid-nineteenth century. Famously these migrants came from Ireland, but Scotland and mainland Europe were similarly affected. When a subsistence monoculture collapses, people have to up and go. Where do they go? For really poor people transport costs are a serious consideration. In Ireland a family would often club together to assemble one boat fare to send one son to America. That son would then send money home to pay for more fares. Often, however, the elderly would never be moved. It was more economic to support them at home with remittances. This entire picture depends upon the fact that wages in the US were high, so an economic motivation plays an essential role.

The above account is typical because motives are nearly always mixed. If someone flees from A to B, then B must be perceived as being better than A. It is conceivable that B is only preferable because it features no persecution, but that is improbable. Usually there is some choice concerning B and from the menu of possible destinations the choice will take into account economic benefit. The exception to that rule is seen with what one might call 'cross-border flight'. Africa provides examples of the case when political conflict leads to people fleeing across a border to the nearest safe haven. The movement of refugees from Sierra Leone to Guinea provides one of several examples. War generates the clearest examples of pure flight migration, simply because war is the cruellest stick.

## 10.5 The City and the Country Trade

The classical model of international trade that was laid out in Chapter 5, and extended in Chapter 6, was about trade between different countries, each one conceived as a perfectly integrated economic space: that is a single economy with only one set of prices, and with perfect mobility of all factors within the country concerned. For any country so conceived, issues of internal migration and of urbanization cannot arise, and can certainly pose no problems. It is evident that this is a hopelessly inaccurate description of any, and certainly of most, real-life nations. In particular, the rural-urban divide in many developing countries creates a dual economy situation in which, for much time to come at least, we have essentially two economies that have not integrated fully, but which do coexist, and do trade with each other.

In this case the classical trade model can usefully be applied to analyse the relations between city and country. We can think of these two regions as two 'countries', and model their interaction on precisely the same lines as the international-trade model. Now the assumption of the same technological possibilities in each region is by no means innocuous, but is surely no worse than the same axiom for different countries in the world economy. Similarly, the specification according to which factor supplies are given and fixed in the two regions, cannot be strictly correct. Even so, it will do well as a short-run description, and it will elucidate the pressures towards inter-regional migration of the factors, that will make themselves evident over time.

When we come to fill in the detail of this proposed model of inter-regional trade, we confront once again the same question of specification that we met above in Chapter 6. Assuming that we can get by with just two products, how many factors do we need? As a first attempt, we can try to specify the model with just two factors, capital and labour. Then the urban sector is relatively well endowed with capital, the rural sector with labour. Given how imperfect capital markets function, and the initial imbalance of labour distribution standardly assumed in migration models, this is not unreasonable. With intersectoral trade the urban sector tends towards specialization in capital-intensive formal economic activity, including manufacturing. The rural sector tends towards specialization in labour-intensive informal economic activity, including traditional manufacturing and, of course, agriculture.

There are two possibilities now, according to whether the rural sector specializes or diversifies. If its economic activity is diversified between formal and informal economic activities, even if the formal is only a small part of the total, then we are back to a classic factor-price-equality outcome. Real wages in the rural sector will be just as high as in the modern-urban sector. This picture looks quite like the position in a developed economy with a substantial agricultural sector, such as New Zealand.

There perhaps sheep farmers earn as much as urban factory workers, once net advantages are taken into account. Then migration from the country to the city would cause sheep farming to shrink in scale, and urban factory activity to increase in size, with no change in factor prices, in standard Rybczynski style.

For a country such as New Zealand the size of the labour force in the rural sector being relatively limited, and that does not well describe the position in many poor countries, from China to Latin America. Otherwise we will find that the rural sector will specialize in labour-intensive informal economic activity. There will not be equalization of factor prices, and the wage rate in the rural sector will be lower than the wage rate in the urban sector. Were there to be costless labour migration, the stock of labour in the rural sector would shrink rapidly, until the rural sector no longer specialized in labour-intensive informal economic activity. As labour migration is not rapid and costless, there will be a long transition during which the model of Todaro (1969), or something like it, will account for the rate at which labour will migrate from the rural to the urban sector.

An alternative interpretation of Figure 10.1 illustrates the position when the city and country trade. Ignoring the line BD, the figure shows a standard HOS-style equilibrium. Factor prices are shown by the line AC. These prices apply in the city, where production is diversified. The country specializes, and its factor prices are given by BD.

The above argument, using a standard HOS model, with two factors, capital and labour, may leave some feeling of discomfort. Informal economic activity is represented simply as employment that uses labour intensively relative to capital; this in comparison with the economic activity of the urban sector. The rural sector, however, is inevitably associated with agriculture. And agriculture, surely, is an activity that makes intensive use of land. So we are back to models that were treated in detail in Chapter 6: non-square economies.

Remember that even a model as simple as one with three factors and two goods has the potential to be horribly complicated, with those patterns of complementarity and substitutability causing their characteristic trouble. The route that proved more fruitful in the earlier discussion in Chapter 6 is a specific factor specification. That seems to be natural again in the context of trade between city and country. Suppose that the city is relatively well endowed with skilled labour, a wholly reasonable assumption. The other two factors are unskilled labour and capital. We adopt the assumption that urban formal activity is capital-intensive relative to unskilled labour. Finally, skilled labour is only employed in the high-tech, and mainly urban, sector.

Now the results are the same as those already laid out in detail in Chapter 6. When free trade between city and country operates, the city has a higher wage for unskilled labour, and also a higher return to capital. All this with fixed and immobile factor endowments. From the findings concerning factor prices the pressures towards migration of factors are evident. In so far as labour can migrate, it will want to move from country to city. The same preferred direction of movement will apply to capital, and this will be the case however much the rural sector may be undersupplied with capital relative to its perceived needs, in particular for the alleviation of rural poverty.

A problem with the model just presented is possibly one that applies to all applications of the HOS model, or its extensions; but it is visible and evident where trade between city and country is under examination. The rural sector has its comparative advantage in an activity that employs unskilled labour intensively, and no skilled labour at all. That very same activity is to be found in the city. There is no problem, of course, in supposing that the city embraces low-skill labour-intensive activities. Yet to have these technologically identical to a similar activity in the rural sector, which has to be agriculture, is far-fetched. No doubt there is city-style agriculture: growing tomatoes or salad crops under glass might be an example, as would urban milk production. But then, while these are certainly instances of agriculture on a broad definition, they are not identical with the land-intensive agronomy of truly rural areas. To really face these issues we would have to increase the number of goods distinguished in the model to at least three. We step aside from the huge complications that would be involved, while admitting that, as a consequence, our models are at best a rough approximation to reality.

## 10.6 The Political Economy of Factor Migration

Think of the rich industrial world (the North) as well endowed with skilled labour, and the other region (the South), as poorly endowed with skilled labour. From the discussion above we know that the North may be the urban sector of one country, and the South may be the rural sector of the same country. It is plain that the South has a low wage rate for unskilled labour, so in that regard the model does well. It is less completely compelling to suppose that the South has a low return to capital, as that would lead to a capital outflow in the direction South to North. Something like that does occur. However a long-run equilibrium would not allow of different returns to a mobile factor in different regions.

If capital is perfectly mobile we are back to essentially a two-factor (skilled and unskilled labour) model, as analysed by Krugman and Wood. Suppose then that technology is everywhere the same, that goods are freely mobile, and that there is no specialization. Then the abundance of skilled labour in the North will have no effect on the unskilled wage rate. Only national production levels will be affected. The North will produce more of the high-tech good relative to the South. It is the Rybczynski story again.

At this point we may bring in issues already treated in Chapter 8. It has been seen how with a three-factor model a high wage rate for skilled labour can result in low factor prices for both capital and unskilled labour. Does that high factor price, for skilled labour, need to be the market cost of a standard input? Might $w_3$ stand for the additional cost imposed on producers by a poor economic environment or malfunctioning institutions? It is an intriguing idea. To make one obvious point, referring back to the argument of Chapter 8, not all poor institutions, such as bad government, can be modelled as a high unit cost for a notional input. Suppose for example that property rights are insecure, and governing party thugs will help themselves to good share of any profit they can see. This rent-looting is not a problem of elevated unit costs.

Stay with the straightforward interpretation of the model, with all factors being factors in the normal sense. Then the three-factor structure allows for production diversification without the equalization of factor prices. Specifically the North will have a low return to skilled labour in comparison with the South. The return to the other two factors will in each case be relatively higher in the North. The latter property looks quite realistic, but is it sensible to take seriously a model that predicts lower real wages for skilled labour in the North? That conclusion is not realistic, but it arguably is to be explained by the fact that technology, in the sense of technological possibilities, is not the same in North and South. There was a time in the past when doctors in poor countries mainly treated rich patients, and were quite likely to enjoy a higher standard of living than their Northern fellow professionals. Today the economic efficiency of different regions has diverged to such an extent that medical staff migrate in large numbers from South to North.

What are the welfare consequences of the migration of labour from North to South? The separable three-factor model already shows how migration may have very different consequences for factor interests than the usual Rybczynski analysis would indicate. Depending upon the precise assumptions, factors already resident in the country can lose when labour migrates in. It is no surprise that labour of the same type as the migrants always loses. The other type of labour may lose or gain, as may capital. Here the relatively simple model confirms what intuition would suggest for more complicated cases. Few would doubt that the considerable flow of violinists from the ex-Soviet Union to the West has made life tougher for western violinists. That particular group represents a negligible political weight. Where mass migration is concerned the political effects can be huge. In France they toppled a left-leaning government. Of course while economic arguments of the 'stealing our jobs' variety figured in the debates of the time, they were diluted considerably by claims to do with national identity, not to speak of chauvinism and racialism.

An assessment of the effects of migration from a purely national point of view is not one that should commend itself to the economist. Even when there are losers in the home country, the migrants, assuming only that their decision to migrate is economically motivated and well informed, will gain. And that gain reflects an increase in world production, when workers move from a lower marginal product to a higher. To allow policy to be driven by potential losers, without considering methods of accommodating change and compensating those losers, is the essence of short-sighted populist politics.

## 10.7 Internal Migration and Urbanization

The world has recently passed an important milestone. Today a majority of its population is urbanized; it lives in towns or cities. This fundamental and drastic change has often been noted, but its full implications have yet to be worked out and absorbed. Probably the strongest implications are political. Urban dwellers exert political influence differently, and often more strongly, than do their rural cousins. Much of this is accounted for by the fact that it is easier to cause a lot of trouble in a city than in the country, and for given trouble governments are more embarrassed by urban unrest than by rural unrest.

The consequences of mass urbanization have received attention, notably from the World Bank. See Henderson (2002), World Bank (2002), and Williamson (1988). The focus has been more on the environmental costs of rapid urbanization, and on urban poverty, than on the external trade implications of urbanization. It is obvious that developments in external trade can drive rapid growth of cities. Examples of urbanization powered by trade include Manchester in the Industrial Revolution. The new technology of cotton weaving, plus the availability of water power, and economies of scale, threw up a huge industrial urban centre. As economic historians have noted, Manchester quickly came to depend on the export of cotton overseas, and given that, the city grew further. Chicago originally became huge as a meat-packing town. Its stockyards swallowed thousands of cattle transported from the plains. The resulting meat could only be sold on by its long-distance transport, as canned or frozen beef. The trans-Atlantic export of frozen meat became feasible because of technological developments in freezing and shipping. It remained costly, so only the best cuts were exported. That left cheaper beef for the local population to consume.

While trade can drive urbanization, in other cases cities grow for their own reasons, often as the seat of a central government. A large city quickly ceases to be self-sufficient, particularly with regard to food. Then a problem arises, one that was hardly ever solved satisfactorily prior to the arrival of modern industrial capitalism: how to feed the thousands of hungry mouths of the city-dwellers? To get the food in, the city must offer something in return. As the crudest solution the city uses brute force to take what it needs: Ancient Rome tried this, with its plantations in Sicily, and later Egypt, worked by slaves. The generation of a substantial surplus from slave farming is feasible, given fertile land, but can be dangerously unstable. Rome saw its slave revolts, just as Baghdad did later. In comparison with these examples, the history of Ancient China is remarkable. Northern China was fed successfully for centuries with rice grown in the south, and transported to the north using sophisticated canal transport.

History fails to tell which and when were the world's first massive cities, because the numbers at issue can only be guessed. Rome under Augustus had one million inhabitants. Baghdad under Haroun Al-Rashid was massive, in the range one to two million. London in 1800 reached one million, after which it remained Europe's largest city until it was overtaken by Istanbul in the twentieth century. The sizes of the possibly giant cities of Ancient China remain unrecorded. The population of ancient India's Vijayanagar is unknown, but it is said to have been enclosed by twenty-four-mile walls, which surely indicates a huge number of inhabitants.

Often the city offers government in return for the resources that it sucks in. Yet there is little demand for government as such. The periphery can do that for itself, thank you, or imagines that it can. Military protection is less readily decentralized, and it was military protection that the Roman Empire offered its subjects during its successful centuries. Yet as the Empire expanded, and it expanded inevitably, because the frontier always needed to be pushed out further in the attempt to stabilize the Empire, even military defence was better organized locally than centrally. Hence the *foederati*, border tribes who were given membership of the Empire in return for organizing their own defence, an arrangement replete with the possibility of instability.

Where Rome had problems, Baghdad faced similar difficulties. Like Rome it had its barbarians, in this case Central Asian nomads. But defence against these raiders by pushing the defensive frontier far out was not feasible. Direct defence of the city itself demanded a large body of horse-mounted troops, a hugely expensive outlay for a city not able to feed itself easily. And while membership of the Roman Empire was something that most barbarians desired strongly, those outside the Abbasid Caliph's rule had no comparable reason to close a peaceable deal to get inside.

The Roman Empire lasted many hundred years, even in its western manifestation. The Abbasid Caliphate in Baghdad lasted 300 years, with fading vitality for much of that period. Like the Roman emperors before them, the Abbasids became dependent upon foreign soldiers for their protection. It was these Seljuk Turks who eventually displaced them. Baghdad persisted as one centre in an increasingly fragmented Islamic world. Its final end came in 1258, when it was sacked by Mongol invaders.

Both examples, Rome and Baghdad, remind us that in the past cities often had limited lives. It is difficult to imagine that modern megacities, such as Cairo, Lagos, or Mumbai, could decay to a fraction of their former sizes in a similar manner. Two reasons support this view. First, when cities failed in the past, the countryside was able to absorb much of the population, and the population growth, that would previously have peopled the city. Secondly, the city now has something to sell to the outside: industrial products. In the past the city sold services of dubious and unstable value, such as a bit of peace in return for the payment of taxes. Manufactures or sophisticated services are of more solid value, and they may save cities in some form from catastrophic decline. The city seems secure as an institution, unless some barely imagined horror in the future, disease or violent conflict, should undercut its viability.

## 10.8 Concluding Remarks

Migration terrifies many people. Foreigners entering our country will bring no good. They may bring disease, alien culture or religion, and they are likely to exploit the social-security system of the home country. If they find employment, they will push down the wages of the local labour with which they compete. In addition, huge cities, whether populated by external or internal migrants, present a frightening aspect. The external diseconomies of mass humanity are concentrated to a toxic extent in the great city. Traffic jams, shanty-town slums, urban pollution, crime, and prostitution; all these make up the typical mental picture of the modern Third-World megacity.

Not one of the stereotypical ideas listed above is entirely false, but all are one-sided. Migrants, whether internal or external, bring important benefits as well as problems. They fill gaps in the labour market; they are typically young and relatively well qualified; they are alien, but that can imply good things, witness the 'ethnic' restaurants of Britain and US cities. While possibly depressing some local wages, they can increase other wage rates, and the return to capital. Migration can increase world production and efficiency. In addition, and crucially, migrants enhance their own welfare. Occasionally migrants may choose to move, with that decision based on misinformation, and find themselves worse off than if they had stayed put; but that is exceptional. World Bank investigations show that migrants usually improve their situations. In the city, wages are higher, of course. But could it not be the case that the higher nominal wages fail to compensate for the higher costs and disutilities of city life? No: that picture cannot be reconciled with life expectancy higher in the city, infant mortality lower, educational attainments better. Purely forced migrants, those fleeing from war or persecution, are a different case, whether they cross an international border or not. Their lives are not infrequently ruined; they lose land and livings; they end up surviving in camps on humanitarian handouts with little prospect of gainful employment. Harmful effects of migration more often apply to the donor country more than to the destination country. The stripping of poor countries of their qualified medical personnel, to feed the hungry health systems of rich countries, has attracted much adverse comment.

International mobility of capital is in many ways parallel to the mobility of labour. In principle it is efficient and welfare-enhancing. Often the problem is that there is not enough of it, not too much. Capital mobility depends greatly on good information and confidence, and these are uncertain and delicate features. A large problem in recent years is that capital mobility has been of the wrong kind, too short term, and dangerously unstable. Sometimes, though not always, this poor structure of borrowing has reflected imprudent speculation by borrowing countries. For more on this, see Chapter 14.

# 11

# International Trade Rules, Politics, and the Environment

## 11.1 Why We Need Trade Laws

The question that forms the first heading of this chapter admits of a simple, and short, answer: because actors are large. To confirm that this is the correct answer, one only needs to go back to the Arrow-Debreu general equilibrium model. In that model agents simply trade in order to achieve the best result possible from their private point of view, given the prices that they face. The same is true of the HOS model, and its derivatives, all of which are special cases of the Arrow-Debreu model. It makes no difference to that model if some of its agents are large, provided that these large agents maximize at given prices.

Once agents are large, however, it becomes artificial to suppose that they will fail to understand the fact that the prices at which they trade are not independent of their choices and actions. Given that realization, the decentralized market system no longer functions efficiently, and then the case for rules to improve market outcomes becomes compelling. The primary argument for rules to govern how trade is conducted, however, arises from the manner in which unregulated governments will behave if their choices are not restricted. Governments are always and everywhere large agents, at least in the sub-economies that they govern, and they never fail to recognize this fact.

Everything comes down to governments, because they are in principle sovereign over the private agents of their economies. Suppose, for instance, that a domestic producer is large in the world economy: it is a *Microsoft* or a *Boeing*. If the home government gives this company free rein to do as it pleases, allowing it to exploit its market without restraint, that is a policy decision. In most cases the actions that trade rules seek to regulate consist of more active and positive interventions by governments. In particular, governments undermine free trade; they erect tariffs, they impose quotas, they subsidize exports, they favour domestic contractors over the foreign.

If governments want to do all these things, why should they enter into, sometimes actively promote, agreements; let us call them trade laws, to stay with one term; that inhibit precisely these actions? Here is the answer:

**A General Principle:** *Whenever large agents interact, it frequently pays them to establish a cooperative equilibrium. This is because its opposite, the Nash equilibrium that results from unrestricted self-seeking behaviour, is often inefficient.*

The above statement is carefully qualified, and does not presume to call itself a theorem. This is because one can invent games for which the Nash equilibrium is efficient. For example, the players control respectively $x$ and $y$, each chosen on the closed interval [0,1]. The pay-off to each player is:

$$x \cdot y \tag{1}$$

and the unique Nash equilibrium is (1,1). This game represents a situation in which ideal cooperation is self-policing. The players gain individually from any extra contribution they make to the project, the value of which is given by (1). The crucial feature here is that by maximizing without regard to the interest of his partner, the individual in fact serves the interest of that partner. That feature makes this example exceptional.

In the last example the variables controlled by the two players are complementary in the following sense.

**Definition:** *The actions of the players in a non-cooperative game are* complementary *if an increase in the level of the variable controlled by one player, measured in the direction that increases that player's pay-off, weakly increases (increases or does not decrease) the pay-offs to all other players.*

Where international trade policy is concerned, the interests of the participating parties are usually not so beautifully aligned as in the above example, or as in the fully complementary case. Governments intervene to influence trade for numerous reasons. These include:

1. Where terms of trade are variable, tariffs can move them in favour of the home country. This is the classic optimal-tariff model. The improvement in the terms of trade of the home country worsens

the terms of trade of other countries. The control variables are not complementary.

2. Although tariff protection is inefficient in a one-consumer economy, where consumers are many and diverse protection may favour one group, and the government may give in to the temptation to pander to that group. By doing that it increases its own pay-off, but worsens the condition of other countries. Again, the control variables, tariff levels in this case, are not complementary.
3. When the government, for whatever reason, cannot organize the efficient collection of taxes without causing excessive distortions, or political difficulties, it may resort to tariffs as the only feasible means of acquiring revenue. This case applies to some SSA countries, for which customs duties are almost the only practical way of generating revenue. This is another case of self-serving choices harming others. That is not complementarity.
4. Where the home government fails to organize its macroeconomic policy effectively, or fails to implement the domestic adjustment interventions that are needed to produce full employment, it may have recourse to 'beggar-my-neighbour' protection, with the aim of boosting home employment. This case describes well the situation during the Depression years of the 1930s, when many countries protected heavily with the aim of exporting unemployment. Plainly beggar-my-neighbour policy does not involve the use of complementary variables.

In each of the above cases the benefit that the home country derives, or perceives itself to derive, from its breach of good free-trade behaviour, has unfavourable implications for the country's trading partners. They benefit less from the exploitation of their own comparative advantage; the terms of trade that move in favour of the home country, move against the trading partners; the unemployment that the home country reduces, at least temporarily, is reduced at the cost of higher unemployment in the other countries. On the politics of free trade agreements, see Grossman and Helpman (1995).

## 11.2 Trade Rules: History and Reform

This book is not the right place for a thorough and wide-ranging account of the history of trade regulation, and related developments, since the

Second World War. Existing references that do that job most adequately include Baghwati (2002). What concerns us here are the grand structural features of modern trade laws: what they did; how they did it; and how subsequent developments have both reinforced, and also undermined, those developments. Our current trade rules owe their origin to the Bretton-Woods Conference. That meeting, with war still vigorous, but with its end in sight, aimed to create a legal and institutional architecture to meet the needs of a war-shattered world, at last at peace.

The four pillars of the post-war institutional 'make-over', to employ a contemporary term, are the United Nations (the UN), the International Monetary Fund (the IMF), the World Bank, and the General Agreement on Tariffs and Trade (the GATT). Of these four, none could be well described as completely forward-looking. Policy makers are like generals; they are always fighting the last war. The world in 1944 was still obsessed with the horrors of the 1930s: mass unemployment; the failure of the League of Nations to confront the rise of Fascism; the protectionism and the collapse of world trade; the competitive devaluations; all these were seen as leading directly to the War, and as experiences never to be repeated.

The UN was intended to be a League of Nations that worked, but it failed to foresee the Cold War. The IMF was intended to establish a more-or-less-permanent regime of mainly fixed exchange rates, but was not able to cope with US profligacy during the Vietnam War. The World Bank was initially concerned with post-war reconstruction, mainly in Europe. It graduated to become a development agency, chiefly for the Third World, and its problems can be attributed to a great extent to the fact that this last theatre, with its huge structural problems and rampant corruption, is far harder to sort out than even a war-bruised Europe.

The GATT had as its intention the establishment of good behaviour with regard to trade. That meant no tariff escalation, which the 1930s had seen. While it permitted protection in response to shocks, without defining what counts as a shock, it imposed non-discriminatory interventions, preferably in the form of tariffs. Its designers tried to mould it to make it acceptable, especially to the US. So agriculture was explicitly excluded, a decision with huge long-term adverse consequences, because it was thought that the US Congress would never ratify a treaty that included agriculture. But then the US never ratified the treaty in any case, and the GATT survived for four decades, albeit seriously weakened, as an unconsummated legal process, supported by a minimal administration. A short helpful summary of the GATT rules is provided by Feenstra (2004: 176–8).

It is unfair perhaps to go over the various things that the GATT architects did not foresee. This is not just because hindsight always has the advantage over a contemporary view. One also needs to take into account the shape of the world in 1945; and, of equal significance, the relatively underdeveloped state of the theory of international trade at that time. On the first point, recall that the post-war world was still to a huge extent a colonial world. It was parcelled up into the colonial territories of various European countries. Typically these colonial subsets of the world of nations enjoyed preferential trade arrangements with the colonial power. A danger felt especially by the US negotiators is that these imperial preference systems would expand and replicate in the post-war environment. For these largely self-serving reasons, the GATT was born equipped with its most brilliant and successful feature: the *most-favoured-nation principle* (the MFN rule). This decrees that any cuts in tariffs offered to any member nation must be made equally available to every member country. The MFN rule was never pure and absolute. For example, temporary protection in response to claimed dumping by another country could legally impose non-MFN tariffs against just that country's exports. Yet all the major tariff-cutting exercises, such as the Kennedy Round of the 1960s, that brought down the tariff levels of the industrial countries, were MFN cuts, and were far more effective on account of that feature. On the various ways in which the GATT system has been eroded over its long life, see below.

In 1945 most of what we now recognize as the theory of international trade did not exist. That means in particular that the so-called theory of the second-best, the computation of optimal policies for distorted economies, had yet to be born. It followed quickly with Meade's great volume on trade and welfare, Meade (1955). An instance of a second-best question is the customs-union problem. Those who drafted the GATT agreement had to take a position with regard to the formation of a customs unions among a group of nations. This is second-best because it is not the GATT ideal of global free trade. Yet it seems to be a partial movement in the right direction, because trade between the nations involved is evidently freed.

The authors of the GATT were lawyers more than economists: lawyers motivated by the simple idea that liberal trade is a good thing. The legal pen can be recognized most easily in the treaty in the form of the tort-law concept of damages imposed on other parties by irregular trade practices, and the associated definition of legal forms of redress. The lawyers' intuitions served them fairly well with regard to the customs union issue. They recognized that customs unions could promote gainful trade, but they also saw that some of those excluded from the union might reasonably feel that they had been harmed. In consequence the GATT decreed customs unions to be legal provided that:

> ...the purpose of a customs union or of a free-trade area should be to facilitate trade between the constituent territories and not to raise barriers to the trade of other contracting territories with such territories. (Article XXIV)

It needs some precise economic analysis to show the conditions under which a customs union can meet simultaneously the two requirements: gain to members, no harm to outsiders. And in 1945 that analysis did not exist. The required theory followed quickly, starting with Viner's great paper, Viner (1950), and followed by numerous works that adapted and extended that first sharp dissection of the basic issues. See, *inter alia*, Lipsey (1957) and Vanek (1965). The flood of writing that built on Viner added a great quantity of detail and subtlety; yet the first shining insight remains the crucial indispensible point. A customs union can be *trade-creating* and/or *trade-diverting*. Which of these it is, and to what extent, depends upon its structure and upon its design. Trade creation means an expansion of comparative-advantage-based exchange within the union. Trade diversion means the re-routing of imports from outside the union to higher-cost sources inside.

The European Common Market (ECM), as it was titled at that time, provides a good illustration of the distinction. Research studies (see, *inter alia*, Balassa (1967)) show that the ECM was trade-creating with regard to manufactures, but trade-diverting with regard to agriculture. Neither conclusion is surprising. Industrial protection was strong within Europe at that time, with the consequence that cutting it to zero among the ECM members had a large trade-expanding effect. Agriculture in contrast involved products for which comparative advantage varies significantly within Europe, but for which the lack of comparative advantage in comparison with the outside world is far more marked.

The ECM violated the MFN principle by cutting tariffs selectively. For agriculture it instituted permanent dumping of production on world markets. Dumping is always a controversial issue and difficult to decide in some cases. It means selling at a lower price in foreign markets than in the home market. Whether and when private producers would do that is not easy to determine, but the EU's agricultural dumping is state-financed and unambiguous. Looking back one is tempted to ask: how did Europe get away with it? As is frequently the case, the answer is politics; specifically Cold-War politics. The US, the only power that could conceivably have resisted the formation of the ECM, in fact favoured it. A united Europe was seen to be a more effective bulkward against Soviet Communism, and that was judged to be a more significant point than were distortions to international trade.

Sadly the formation of what was eventually to become the European Union (EU) was part, but only part, of the story of the erosion of the GATT ideal. International trade regulation for four decades became a field on which the world's great powers played out a game in which they pretended to promote the free-trade ideal, while in fact pursuing their own self-interests as far as they could.

The undermining of the GATT has included several developments. Among these are uninhibited abuse of the right of any country to retaliate against dumping, so that the dumping label is attached to imports that happen to be upsetting to local interests. The EU is a particularly flagrant offender in this regard, but it is not alone in its resort to this malpractice. At the time the present volume goes to press heavy tariffs have been proposed by the EU Commission on imports of selected footwear from China and Korea, this protection being justified by 'overwhelming evidence' that state subsidies are giving an unfair advantage to these exporters. A drift towards managed trade, in the form of so-called voluntary export restraints, which are in fact imposed by the importing country; abuses of safety and health regulations to in fact protect home producers; and the use of import quotas, that according to the GATT rules should be a temporary emergency intervention, as a more-or-less permanent feature.

Nothing undermines the MFN principle as insidiously as do bilateral trade agreements, and these have proved to be increasingly popular with the 'bully-boy countries', meaning by that term the large powerful players on the international trade scene, notably the EU and the US. The EU has signed numerous bilateral trade agreements with neighbouring countries, that recall the unequal treaties between Britain and China in the nineteenth century. The US completed a free-trade agreement with Australia that explicitly excluded agriculture. The NAFTA agreement does not treat Canada and Mexico symmetrically. Canadian agricultural products enjoy easier access to the US than does equivalent Mexican produce. And so on.

In contrast to the bleak picture sketched above, the establishment of the World Trade Organization (the WTO) is a welcome and encouraging development. This is an extension of the GATT system. This time, however, it has been ratified by the US Congress, and it is provided with a far better administration, and improved paths to legal redress. The gross inequalities between large and small countries that characterized the GATT system are less marked with the WTO, and the proportion of complaints coming from developing countries has greatly increased. These last complaints now amount to one-third of the total. Some of the complaining developing countries, Brazil for instance, represent large economies. Even small countries can make use of the WTO, although they are often handicapped by high legal costs and a lack of local expertise for a case which will see US-government lawyers on the other side. One might say that this legal imbalance is not different in kind from the way in which an average American is disadvantaged when it comes to suing WalMart. No system is ideal, but some recourse is better than none. In trade-reform negotiations the developing countries have attempted to coordinate their negotiating positions, and this has increased the weight of their influence. A problem here is that the developing countries often have divergent interests. Thus Brazil, as a large sugar exporter, favours cuts in industrial-country sugar protection. Countries with bilateral agreements for their sugar exports, however, prefer limited exports at artificially high prices.

When the principle of free trade is widely violated, and often in ways that harm larger national interests than the interests that they serve, a question that demands an answer is: why cannot the world do better? There are endlessly complicated political-economic answers to this question, that mainly take the form of noting that a nation is not a single maximizing entity, but is rather a loose coalition of competing interests that find an equilibrium that may well not be efficient for their interests. Farmers, for example, are protected because of their well-organized political weight. Consumers who pay the bill at the supermarket checkout, are concerned with other things, and poorly informed as to how much extra they are paying. In the next section we show how protection can arise even when each nation is a rational unified maximizer.

## 11.3 Nash Equilibrium in an Optimal-Tariff Game

The model examined below provides an example of the harmful consequences of uncoordinated individually maximizing choices, when the choices at issue are not complementary. It is a special case of a standard optimal-tariff model. To keep the illustration simple, the countries involved are confined to a binary choice between free trade, meaning tariffs equal to zero, and protection, meaning a tariff level equal to 1. If tariff levels are chosen freely, the model has no uncoordinated Nash equilibrium. Tariffs escalate upwards without limit, until trade is completely extinguished, with disastrous effects on economic welfare. This is a feature of the particular, and simple, specification of the model; chosen to allow the easy and uncomplicated development of the exposition. It is a pure exchange model; there is no production to take into account; and preferences take the most simple linear-logarithmic form.

Normally tariffs will reach ceilings, and the harm they cause will be limited, if possibly still severe. None of this matters for the fundamental point that needs to be made here: with uncoordinated choices of policy, protection can be rational from the point of view of the individual countries that choose their policies. Only when this conclusion is understood can the need for collectively negotiated trade rules be appreciated.

There are two countries, A and B. Each starts with one unit of a commodity, denoted by $X$ for A's commodity, and $Y$ for B's commodity. The utility function is the same in each country:

$$U = xy \tag{1}$$

where $x$ and $y$ are the quantities of respectively $X$ and $Y$ consumed in the country concerned. The Lagrangean for utility maximization is:

$$xy - \lambda\left(I - p_x x - p_y y\right) \tag{2}$$

where $I$ is income, and $p_x$ and $p_y$ are the prices of respectively commodity $X$ and $Y$. The maximization of (2) requires:

$$y - \lambda p_x = 0 \tag{3}$$

and:

$$x - \lambda p_y = 0 \tag{4}$$

From (3) and (4):

$$\frac{y}{x} = \frac{p_x}{p_y} \tag{5}$$

The world price of $Y$ in terms of $X$ is $p$. Country A imposes a tariff $t_A$ ($t_A = 0$ or 1) on imports of $Y$. The motivation of this decision is the optimal tariff. A tariff can improve the home country's terms of trade. Will that benefit exceed the cost: reduced participation in trade?

The demands satisfy:

$$\frac{x_A}{(p + t_A)\, y_A} = 1 \tag{6}$$

where $x_A$ and $y_B$ are the demands of country A. Similarly a subscript 'B' denotes country B.

Substituting (6) into the budget constraint gives:

$$1 + t_A y_A = x_A + (p + t_A)\, y_A \tag{7}$$

where income is 1 plus the tariff revenue $t_A y_A$. The demands for $X$ and $Y$ are solved from (6) and (7) to give:

$$x_A = \frac{p + t_A}{2p + t_A} \tag{8}$$

$$y_A = \frac{1}{2p + t_A} \tag{9}$$

Country B imposes a tariff $t_B$ on its imports of $X$. Now income is $p$ plus the tariff revenue $t_B x^B$. Symmetrical calculations give its demands:

$$\frac{1 + t_B}{p}\frac{x_B}{y_B} = 1 \tag{10}$$

Substituting (10) into the budget constraint gives:

$$p + t_B x_B = (1 + t_B)\, x_B + p y_B \tag{11}$$

Now (10) and (11) together give:

$$x_B = \frac{p}{2 + t_B} \tag{12}$$

and:

$$y_B = \frac{1 + t_B}{2 + t_B} \tag{13}$$

Market clearing in either market implies clearing in the other market. So world-trade equilibrium requires:

$$\frac{p + t_A}{2p + t_A} + \frac{p}{2 + t_B} = 1 \tag{14}$$

Rearranging (14) gives:

$$(p + t_A)\,(2 + t_B) + p\,(2p + t_A) = (2p + t_A)\,(2 + t_B) \tag{15}$$

or:

$$2p + pt_B - 2p^2 - pt_A = 0 \tag{16}$$

The solution for $p$ is:

$$p = \frac{2 - t_A + t_B}{2} \tag{17}$$

Now the indirect utility functions of the two countries, $V_A$ and $V_B$, can be expressed in terms of the two tariff levels:

$$V_A = \frac{1}{2}\frac{2 + t_A + t_B}{(2 + t_B)^2} \tag{18}$$

and:

$$V_B = \frac{1}{2}\frac{(2 - t_A + t_B)(1 + t_B)}{(2 + t_B)^2} \tag{19}$$

Notice that the expressions (18) and (19) are notably asymmetric in the two tariff levels $t_A$ and $t_B$. What explains this? The answer is that the model, while symmetric between the players, with regard to their endowments and their preferences, is as formulated, asymmetric between the two sides on account of the choice of the numeraire. The price of A's good is always 1, while B's good has a variable price $p$.

Given that their asymmetry is now clarified, everything we need to know is contained in (18) and (19). Recall that we are concerned only with the country's choices between tariff levels 0 and 1. From (18) and (19) we can compute the pay-off matrix to the two players.

In Table 11.1 each country chooses to play zero or 1. The pay-offs to A are shown by the lower columns; the left-hand if B plays 0; the right-hand if B plays 1. The pay-offs to B are shown in the higher columns; the left column if A plays 0; the right column if A plays 1.

**Table 11.1.** Pay-offs in the Optimal-Tariff Game

| | | | Country B | | |
|---|---|---|---|---|---|
| | | | 0 | | 1 |
| | | | $\frac{1}{4}$ | | $\frac{1}{2}$ |
| Country A | 0 | $\frac{1}{4}$ | | $\frac{1}{6}$ | |
| | | | $\frac{1}{8}$ | | $\frac{2}{9}$ |
| | 1 | $\frac{3}{8}$ | | $\frac{2}{9}$ | |

Inspection of the table shows that the choice of 1 is a dominant strategy for both players. So (1,1) is the unique Nash equilibrium. The pay-offs to the players are $(\frac{2}{9}, \frac{2}{9})$, less than the pay-offs $(\frac{1}{4}, \frac{1}{4})$ that the players would receive if they coordinated on the strategies $(0, 0)$.

In this stylized example the strategy (0, 0) represents free trade, and the strategy (1, 1) is protection. The structure revealed is that of the familiar prisoners' dilemma game. Protection is the individually rational choice, although its consequence is a collective loss, because free trade dominates when the benefits to the players are computed. For this reason a direct and evident implication of this analysis is that there are situations in which unilateral free trade is not the best strategy, while negotiated free trade is universally beneficial. The reason in this case is the optimal-tariff effect, that protection improves the home country's terms of trade. Yet the lessons are perfectly general. Whatever the individual motive for protection, it causes negative external effects, and coordinated control of policy choices is of general benefit.

Despite its familiar structure, the example can illuminate more than the basic prisoners' dilemma insight. In the game as specified the players just choose between 0 and 1. Even when they move to a negotiated outcome, their choice is simply between coordination on (1,1), or uncoordinated protection. Now imagine for a moment that one of the players is a skilful trickster who can talk his negotiating partner into adopting free trade, while reserving for himself the right to protect. Let that artful negotiator represent Country A. Then the table shows that A gets $\frac{3}{8}$, better than the $\frac{1}{4}$ that all-round free trade provides.

Viewed in the context of our absurdly simple model, the idea that one side might steal the benefits of free trade without conceding anything to the other side, is just ridiculous. Real-life trade negotiations, however, are hugely complicated. They cover a wide range of goods and services, and endless details concerning tariffs, quotas, subsidies, product specifications, safety and health, and much more. Negotiations to regulate such a complex territory resemble a game of chess, in which the parties endeavour to outwit their opponents, employing any diversion and entrapment that serves their purposes.

This game is a dangerous pastime. It risks losing the gains from general trade liberalization for the far less important benefits that protection offers. That leaves unanswered the question of how far the current failure of the rich and powerful countries to achieve a new Doha round trade-liberalization agreement is due to entanglement in overcomplex negotiations, and how far it simply reflects a 'rational' preference for protection, as in the game analysed above. Sadly, there can be little doubt that failure to liberalize, particularly agricultural trade in recent negotiations was deliberate and calculated. The EU and the US knew perfectly well what they were doing. They decided that pandering to their agricultural, and other, lobbies offered more than going for serious trade liberalization.

This case is covered, in spirit at least, by the model. The pay-off from giving the farmers what they want plays the same role as an improvement in the terms of trade: it provides a perceived national benefit at the cost of external diseconomies. The Nash equilibrium is inefficient. Why in that case could a better negotiated settlement not be achieved? The answer is not obvious. The actual game played by trade negotiators is far more complicated than the simple classroom-style example given above. They are under pressure from numerous and diverse lobbyists. Their rewards may take the form of newspaper headlines more than precise mathematical pay-off functions. The game is not the static game of the example; so strategies may take the form of postponing the most rational moves until later; until a later that may never come. The players are diplomats and should not lose their tempers and become too emotionally involved if they feel betrayed; yet precisely that can happen. In short the assumption of rational actors may be incorrect. The political machine delivers what the political machine can deliver. In the light of that sad fact, the puzzle is often not why trade is not more free, but rather why protection is not more rife than it is.

## 11.4 Gains from Trade and Level Playing Fields

In the current debates concerned with globalization and the rules that do, and the rules that should, govern international trade, two opposing themes recur constantly. One of these themes is the old root idea that international trade based on comparative advantage is gainful to all parties. Many politicians support free trade, not because they have studied economics, but because it sounds like a good idea. Could that be explained by the presence of the word 'free' in the title of unrestricted trade? Possibly, but free love uses the same word, and typically commands little political support.

The other idea is that trade should be fair. Fairness is a powerful concept, but it is not easy to tie down. From a young age children acquire the notion that some things, indeed many things, are 'unfair'. The child's primitive definition of the 'fair' is usually quite a reasonable one, although its application can be imprecise and self-serving. Fairness equates to equal treatment. 'You let Tom do the same without punishing him, and now you punish me. *It's not fair*!'

While there may be a good deal more to fairness than simply equal treatment, equality of treatment is a good starting point. Most concepts of fairness include equality of treatment, often extending it, sometimes elaborately. Thus the man convicted in a court of law who complains that his sentence is far harsher than that imposed on the one guilty of the same offence who proceeded him, will probably not be satisfied with the explanation: 'He is a noble; you are a peasant.' Today at least, people feel that social rank is not a justified ground for unequal treatment. The idea of equal treatment for all is sometimes conveyed by the term 'a level playing field'. This term is intended to capture the idea that the field of play should not favour one side against the other. Whether a sloping field is exactly the essence of unfairness is arguable. In soccer, each team plays for the same time from each end, so the advantage of a slope should cancel out over the course of a match. If, however, the home team is accustomed to playing with or against a steep slope, that would give it an unfair advantage.

In the context of trade policy people frequently declare that they favour free trade, but that it needs a level playing field. What is intended typically is not well captured by a reference to the field of play. The idea is better described by saying that everyone should play by the same rules. And this is where objections to trade competition locate themselves. Protection is proposed because the others are cheating. They are protecting their own markets; they are subsidizing their exports; they are dumping output on our market; they employ child labour; they pay their workers starvation wages; their environmental regulations are non-existent; their exporter is a monopolist; their political system is oppressive. That kind of list conflates issues between which it is important to distinguish; even if the distinction is not always as clear as one might wish.

To put it simply, the question is whether trade rules should confine themselves to the economic, or whether they can properly extend their reach to politics, the environment, and human rights. In writing the last sentence I can almost feel my readers wincing in reaction to its naivety. Where is there a clear distinction between the economic and the non-economic? Take environmental harm as an example. That is certainly an economic issue, but how far should trade rules take it into account?

The Nobel-Prize-winner, Joseph Stiglitz (interviewed in *The Independent* newspaper, 20 February 2006), has suggested that an action against the US might be pursued in the WTO. The claim would be that, by refusing to apply any controls on, or charges for, greenhouse-gas emissions, the US is effectively conferring a trade advantage on its polluting sectors. Probably this provocative idea is not intended seriously. Lawyers defending the US side would certainly argue that competitors of the US in international markets are either exempted from emission controls by the Kyoto Treaty, or have not signed that treaty, or have signed the Kyoto Treaty, but will fail to meet its targets.

Leaving aside the practicality of the Stiglitz proposal, it illustrates perfectly a standard and recurring theme in arguments about free trade. Professor Stiglitz is seriously worried about global warming. He sees the harm caused by the emission of greenhouse gases, the result of human activity, as proceeding far more rapidly than had previously been supposed. And he is scared. The trouble is that the Bush administration takes the exact opposite view. It believes global warming to be not much of a problem, and it maintains that 'technology' will be able to take care of it. Professor Stiglitz is looking for leverage: a means of forcing the US to behave as it is disinclined to do.

US greenhouse emissions are an environmental bad; an environmental bad of a particular kind; they are a global environmental externality. This serious problem demands a response. Let us leave the George Bush view aside for the moment, and accept that global warming is indeed a grave problem. Then three questions follow:

1. Can anything be done about it?
2. What interventions, specifically international agreements, can best ameliorate the problem?
3. Should trade rules and environmental regulation be interlinked; as when environmental misbehaviour is made a legitimate reason for trade-restricting protection?

Notice that the third question invokes the level-playing-field issue. One could logically maintain that environmental misbehaviour is harmful and should be punished, without being considerably concerned by its trade-distorting consequences. And perhaps that is more or less the position of Joseph Stiglitz. He is greatly concerned by US environmental pollution, and he sees in WTO rules the possibility of punishing the US for the actions of which he disapproves. But that is not to say that it is genuinely US trade-distorting advantages that fuel his concern. His worries are directly about environmental pollution, and trade rules are an instrument for pursuing that concern.

## 11.5 Trade as a Lever

What is the correct scope of international trade agreements? To focus this question, consider two strongly opposing views.

1. Trade is a complex field, involving the balancing and offsetting of numerous considerations and interests. For that reason trade is inescapably a political issue. However there is nothing to be gained by allowing non-trade politics to enter into trade regulation. That only clouds and complicates already-tangled problems. Leave human rights, child labour, and the environment out of trade regulation. Deal with those questions separately.
2. In our awful imperfect world, in which many nations do many dreadful things all the time, there are few weapons that can be applied to pressure the wrongdoers into improving their behaviour. The UN is fairly ineffective; military interventions rarely work, and can do great harm; international treaties are hard to negotiate successfully, and are often honoured in the breach. Trade access, on the other hand, is a big stick. To gain that—membership of the WTO, the EU, or a bilateral trade agreement—countries will do all kinds of things to 'clean up their act'.

The second of the above views defines what may be called 'trade as a lever'. The idea is that trade access may be used as a reward for correct behaviour, or may be made a condition for gaining that access. If trade is taken to mean the same as EU membership (which would be a gross oversimplification), then that provides an ideal example of how trade can be a strong lever. Countries eager to join the EU, such as the Baltic States, or Turkey, have been willing to undertake, or to contemplate undertaking, radical reforms. These have ranged from improvements in human rights, and the independence of the judiciary, to the abolition of capital punishment. It cannot be imagined that such changes would have been initiated, and so rapidly, without the carrot of EU membership making itself felt.

An extreme instance of trade as a lever is provided by the example of trade sanctions. These have usually been implemented by the UN, which has the power to require every member state not to trade with a country. The Arab states' trade boycott of Israel is an example not involving the UN. The aim of a trade boycott is never to affect the trade policies of the country affected. Hence for South Africa it was the apartheid policies of the government to which exception was taken, and for a human rights reason, not because it was asserted that oppresson of the black population promoted South African exports.

The effectiveness of trade boycotts must always be questioned. One often hears it said that trade sanctions brought apartheid to an end. This is not true. When they operated it was estimated that the inconvenience to which they subjected the economy was similar to a 7 per cent worsening in its terms of trade: painful certainly but not critical. What finally killed apartheid was that international capital came to feel that the system was unsustainable, and capital flooded out of the country. An attempt was made to soften apartheid, including the release of Nelson Mandela from jail, and the rest, as the saying has it, is history.

Sanctions were notoriously unsuccessful in the case of Iraq in the last years of Saddam Hussein's rule. As in every other case, sanctions motivate unscrupulous operators to find ways around them. Sanctions were good for Saddam, as his corrupt thugs controlled much of the operation of the system. He could use it to harvest rents, and to reward the loyal. Also TV pictures of children dying in hospital, supposedly because of sanctions, were good propaganda; never mind that medicines were never embargoed.

The other possibility that the second view can implement is the level playing field discussed above. The issues here are not ones that abstract economic theory can readily resolve, as the following model indicates. Imagine, to keep the argument simple, that trade negotiations involve only two parties. The negotiations cover two vectors, each of which will take a specific value in each country. These are $x$ and $y$. The vector $x$ shows the conduct of the country concerned in matters that directly influence trade: tariffs, quotas, subsidies, etc. The vector $y$ includes other items. That the boundary between items that belong in $x$ and those that belong in $y$ is a difficult one to draw will be plain to the reader. That point is developed in the next section. Assume for the moment that the distinction is clear and unambiguous.

In that case, economic theory is bound to say that the interested parties will do better to negotiate the values of both vectors simultaneously. Trivially that approach dominates a restriction that requires negotiation of each type of vector separately. Separate negotiation is after all a special case of unrestricted negotiation. Suppose, for example, that one of the parties is already a free-trader and has nothing to offer the other on the trade front. It is however seriously damaging the environment. The other country does no harm to the environment, but its markets are highly protected. There is scope here for a mutually beneficial deal, where one country constrains its environmental damage in return for tariff cuts from its partner.

Despite this last compelling point, there is a good case for keeping trade rules largely restricted to trade. We have to close negotiations between a large number of countries, several of which may veto any agreement, and where the negotiations are enormously complicated. All parties are manoeuvring to gain advantage, and sometimes trying to confuse things to that end. The complete structure of the game is not common knowledge. Sometimes, as in many other aspects of life, it pays to keep things simple.

## 11.6 What is a Commodity?

The next argument springs from the above discussion of the scope of trade rules. Should international trade law allow countries to exclude imports produced by child labourers? This is just one instance of protection on what one might call moral grounds. Environmental degradation is a similar case. To see how these cases break new ground in economic theory, consider how they would be treated by the competitive trade theory of Chapter 5. As was explained there, competitive trade theory is created by applying the Arrow-Debreu general equilibrium model to the case of international exchange. It is true that the trade models are trivial baby applications of the general equilibrium model, yet they are general equilibrium modelling all the same.

In the Arrow-Debreu model there is a given finite listing of all goods. There is no distinction in the list between inputs and outputs. That distinction applies only to individual agents, consumers or firms, and depends upon the sign of their net supplies. For a firm, if net supply is positive, this is an output for the firm concerned; if net supply is negative, this is an input for the firm concerned. It is perfectly possible for an output to be defined in part by the inputs used to produce it. Organic vegetables are not always distinguishable, even in a laboratory, from their non-organic versions. Yet some consumers demand them. Two T-shirts, one produced by child labourers, the other not, may be indistinguishable by inspection. Yet some consumers would not wish to purchase the version produced using the labour of children. So is it legitimate to prohibit the import of T-shirts from a country known to use, or suspected of using, child labour?

The most superficial answer to the question is to say that this is a problem of information. As long as there is clear labelling of products, let the consumer be sovereign. But countries have their own laws against child labour, and these laws do not prohibit the sale of products produced in this manner, and not correctly labelled. They do not allow the consumer to decide. If a country can prohibit child labour on its own territory, why should it not prohibit the import of products produced by child labour abroad?

The important point is that issues such as the above must be dealt with by the WTO within a framework of well-formulated laws. One reason for this is that protectionism is to be found everywhere. So countries are often looking for excuses to restrict imports, and allowing them to make use of all kinds of 'extraneous' reasons for doing that is dangerous. Also self-serving assertions need to be checked properly. So a claim that child labour is employed must be subject to strict legal testing. Otherwise, we are left with the outrageous propaganda recently disseminated by the EU, against Asian shoe manufacturers, not for any genuine respect for free trade, but to protect its own failing shoemakers.

Child labour looks like an easy case, because nearly everyone agrees that it is undesirable. That said, simply exterminating it, with no other interventions, may do little good. If children are sent to work because their families are desperately poor, they are unlikely to attend school when their jobs are extinguished. More probably they will seek some other employment, less exposed to official regulation: sorting rubbish, or in the worst case child prostitution.

What if a country pays really low wages? Is a T-shirt produced by low-wage workers a different product, as a T-shirt produced by child labour is a different product? In its devious justification of its protection against shoe imports from Asia, the EU looked at production costs in Brazil, where wages are considerably higher. Without having the nerve to announce the view explicitly, that low wages are a justification for protection, the EU was in effect walking that destructive path.

The growing use of arguments about the environment, about labour standards, about human rights, in the trade context, are not evidence that politicians care more and more about these matters. Rather there is a worrying slide towards protectionism in major trading nations, and people are looking for good excuses. If protectionism wins it will be a tragedy, and the blame will lie with large country politicians. Ideally the trade leverage delivered by those countries will improve policies in poor countries, and in a real, not a nominal, direction.

# 12

# Trade, and Growth, and Catching-Up

## 12.1 Models of Trade and Growth

The theoretical literature on trade and growth is mostly of recent vintage, but already quite extensive. Aghion and Howitt (1998: ch. 11) provide a good overview, and allow the reader to obtain a sense of the shape and variety of the modelling approaches that have been adopted. It is no slight on the various authors who have contributed to this body of research to say that this work makes uncomfortable reading. There is a makeshift feel to these models, and it is clear that the authors cannot agree, even in broad terms, as to what are the crucial ways in which trade and growth are related to each other. This is unsurprising if one considers that both growth, and trade, have been modelled independently, according to their particular requirements. The danger is that a model that combines the two will be a kind of scissors-and-paste job, and the two parts will not fit together convincingly. A good example of this problem is provided by the question of how trade, on the one hand, and the international transmission of knowledge, sometimes called knowledge spillovers, on the other, are related. A close association is sometimes simply assumed, although Keller (1996) questions the connection. Why should that be the case? And even if it is the case, how is the scale of trade related to the scale of knowledge spillovers? Certainly the Japanese imported a few British motor cycles in the 1950s, and from these vehicles they might be said to have learnt how to make a motor cycle; or chiefly how not to make one. Yet to argue that this is trade carrying knowledge on its back is far-fetched.

In defence of the disordered literature on trade and growth it could be claimed that the connections are inescapably multifarious and complex, and that any model that claims to have the single and unique answer is bound to be wrong. While that is true, another problem is more wide-reaching: it concerns the elusive connection between good economic performance and the rate of growth. In the past an economy was often seen to perform well, or badly, according to how close it is to its production possibility frontier (PPF). Then distortions and bad policies pushed the economy inside its PPF, and inefficiency in that sense defined poor performance. Today that way of looking at things is unusual. The old idea of an efficient economy has been replaced by a fast-growing economy. Good policies it seems promote growth, bad policies produce low growth. It is the case that some of the world's most unsatisfactory economies are growing slowly; for instance the Arab world and the SSA region, as discussed in Chapter 7 above. However to insist too rigidly on an identification of good performance with rapid growth runs the risk of muddling hopelessly short-run and long-run issues. If an economy is a mess, with everything in sight bad, and in need of reform; then one would hope that its situation might be rapidly improved. In that case the rate of growth of the country concerned, for a while at least, should be high. To jump from that conclusion to say that well-run economies will grow rapidly in long-run steady state, is a large leap. The US economy is often taken to be one that performs well, is efficient if one likes. For two decades from 1970, however, at which time its productivity was the highest in the world, its rate of growth was slow in comparison with Europe or Japan.

The tangled problems thrown up by the difference between the long run and the short run make themselves felt when empirical studies are considered. It is a familiar fact, if that term is permitted, that countries more open to trade tend to grow faster than do countries more closed to trade. Winters (2004) in a survey of various studies shows that trade liberalization contributes positively to economic performance. This conclusion is consistent with the possibility that any boost to growth that results from the liberalization of trade will not be permanent. It may well be that trade liberalization moves the economy to a higher level of efficiency and economic effectiveness. It would do that if high-cost outputs are replaced by imports, and the resources released are better employed. This process is evidently difficult, and Winters concludes that it is especially in combination with parallel reforms that trade liberalization delivers its best results.

Many questions arise from conclusions built on cross-section regressions. The measurement of openness is inevitably unclear and subjective. Frequently used measures include the ratio of imports to national product, and indices of average tariff levels. The first measure is biased in favour of measuring small countries as more open, since these have to trade extensively, simply because they are small. The second measure runs into formidable index-number problems. Suppose, to take an absurdly simple example, that there are two countries, A and B. Country A is completely free-trading; nothing is protected. Country B is the same, except that it prohibits the import of champagne, interpreted here as an infinite tariff on that bubbly product. Country B's average unweighted tariff level is infinite; its trade-weighted tariff level is zero. More than one solution to this problem has been proposed, but none can be completely satisfactory, because there can never be a perfect path through the index-number jungle.

Yet another issue for the interpretation of cross-section growth regressions arises from precisely the long-run short-run distinction. Winters (2004) has already noted that the effect of trade reforms on growth may be positive in the short run, but ultimately transitory. The countries of the modern world that are growing significantly are divided to a great extent between those with a long-established economic development, such as Britain, and countries in rapid transitions, such as China. These bursts of growth in the early stage of development depend heavily on the reallocation of labour from low-productivity activities to far more productive activities. That can only continue so long as there is a pool of underemployed labour in agriculture, or elsewhere, to feed the rapidly growing sector. The Korean economic miracle, as discussed by Lucas (1993), presents a somewhat different picture, as the transition in this case is attributed to a big-push increase in education. No matter, in either case the growth miracle is not sustainable in the long run.

How the combination of countries in something like steady state, and countries enjoying short-run growth bursts, affects cross-section comparisons of growth and openness, depends upon how openness as measured varies with country, and the countries' various stages of development, or non-development. Both Japan and Korea were highly, albeit selectively, protectionist in the early stages of their growth miracles. Subsequently they became more open, without ever graduating to the status of model free-traders. The 'Asian Tigers'—Hong Kong, South Korea, Singapore, and Taiwan—provide more clear-cut examples of rapid growth based upon great openness. These countries illustrate another problem with the growth–openness relation. When their rapid growth was initiated they had little to protect; no substantial steel sectors for instance. For a careful treatment of trade-growth empirics, with due awareness of the short-run longer-run distinction, see Dollar and Kraay (2004).

## 12.2 Leading Modelling Issues

As has been noted above, one reason why the literature on trade and growth is elaborate and unfocused is that the influences involved are inescapably multifarious and complex. The following are leading questions that need to be answered, and have been answered in different ways, by various authors in this field.

1. Are goods easily substitutable, or substitutable only with difficulty?
2. Is technical progress exogenous or endogenous?
3. How internationally mobile is capital?
4. How mobile between firms (countries) is technical knowledge?

The first question pinpoints a fundamental difference between two ways of viewing the gains from trade. The classical view, examined in Chapter 5, says that the gains from trade arise from differences between countries that exchange goods that are identical, regardless of the country of origin. So food produced in a labour-rich, capital-scarce country, tastes the same at the dining table as does food produced in a labour-scarce, capital-rich country. It is the cost of products that matters, not their precise specification. As a consequence, the same goods of different national origin are perfect substitutes.

Two developments over the last three decades have taken the emphasis away from the perfect substitutes assumption. First, the introduction of imperfect competition models into international trade theory have admitted the consideration of the close but imperfect substitutes that underlie monopolistic competition. These developments permit distinct views of how trade is gainful. They do not contradict the classic comparative advantage idea; rather they complement it. And in the rarified arena of the academic classroom they can be demonstrated with models of trade between identical countries, where of necessity there can be no differences of comparative advantage. According to these new views, trade can be gainful if it increases the size of markets, allowing greater exploitation of economies of scale; or if it increases the variety of products available, allowing increased benefits to consumers and producers. These benefits arise, either because buyers are better able to purchase a variety closer to an ideal target specification, or because buyers purchase all types available, and enjoy variety by the simultaneous use of all types. The open Romer-style model demonstrated below shows how an imperfect-substitutes model can analyse the connection between trade and growth.

The second of our questions above asks whether technical progress is exogenous or endogenous. In the context of international trade and growth that question answers itself. Technical progress must be endogenous; that is, influenced by some variable features of the economies concerned, if trade is to influence it, and through it to affect the rate of growth. Baldwin, Forslid and Ottaviano (2003) argue that the effect of incremental growth on trade is the crucial question. There remains unresolved another question. Endogenous growth is like a disease, although in this instance a good one, not a bad. The incidence of almost any disease is endogenous, in the sense that its frequency and severity depends upon a huge number of variable features of the population under consideration. Among these features, to name but a few, are standards of hygiene; climate; the presence of disease-carrying vectors, such as insects; and in some cases social practices, including sexual behaviour. Notice that the items in the above list, or in any list, differ in the following regard. Some are outside individual control: short of migrating, an individual cannot affect his climate. Others are subject to individual choice: hygiene and sexual behaviour are clearly of that type. Remember that disease and economic growth are somewhat similar, but of opposite sign when they are evaluated. Disease is bad; economic growth is good. Then, in so far as they are subject to influence by behaviour at the individual level, desirable behaviour with regard to disease reduces it, or the risk of it; while desirable behaviour with regard to growth increases it. Now the analogy of disease allows us to depict simply one of the central ideas of modern endogenous growth theory. Some diseases can be avoided by the appropriate individual behaviour. Drinking only boiled water protects the individual against waterborne disease. Others require collective actions. Some precautions against a disease may be ineffective unless generally adopted. Then viewed at the individual level, the intervention appears to be ineffective. Yet, when widely adopted, the disease is controlled, for a reason that seems to the individual to be exogenous—an external gift. This description is parallel to the Romer (1986) model of exogenous growth. There are diminishing returns at the level of the individual enterprise; yet for the economy as a whole, due to the external benefits of capital accumulation, there may be increasing returns, or ideally constant returns. Diminishing returns to capital are absent at the aggregate level, allowing for the possibility of an AK-Model, where the aggregate rate of growth is determined solely by the rate of accumulation of capital. For endogenous growth theory it is helpful to introduce some terminology to distinguish between two cases. In the first, technical progress is endogenous for the economy as a whole, but it is not driven by individual maximizing decisions, because an individual agent is not rewarded for the social benefits of any progress that results from his actions. This is exactly the position with Arrow (1962), the famous learning by doing model, Romer (1986) is similar. We call this *accidental endogenous growth*. In another polar case, technical progress is generated by profit-maximizing entrepreneurs, who capture all the social benefits of their production. It may be however that the individual entrepreneur reaps some of the benefits of his activity but not all, as when a patent operates but only for a limited period. In either case producing technical progress becomes quite similar to producing bread for sale. We call this *profit-driven endogenous growth*.

The final two questions on our list are similar in their content. They concern the freedom of international mobility; in one case for capital; in the other case for technical knowledge. We have encountered both instances in earlier chapters. The classic HOS model can be interpreted as assuming that capital is perfectly immobile internationally, while technical knowledge is perfectly mobile. These stark and simple assumptions might be rejected. In the Krugman-Wood model, capital is freely mobile; but then another factor is introduced, skilled labour, so that we are left with two immobile factors. What matters here is what the various possible assumptions imply for economic growth. In the BMS model of Chapter 4, capital immobility, where a poor country is concerned, slows down GDP growth, but accelerates the rate of growth of national income. In the Ventura model, increased capital mobility, for a country diversified in its trade at least, has no effect on the rate of growth of national income. In more than one of the models examined below in this chapter; as with the endogenous-growth model with perfect goods aggregation; the issue of whether technological knowledge is transferable, is of great importance for the model's implications. In an ideal case, technological knowledge becomes a product, produced nationally, and traded internationally, just like any other good. In Chapters 8 and 10 above we have seen the serious and weighty reasons why factors, and also technology, may move internationally only with considerable difficulty. Whatever the precise answers to our questions, they will have considerable implications for economic growth, and how it varies with circumstance and location. On the interconnections between factor and knowledge mobility, and trade, see Ben-David and Loewy (2000).

The field-of-growth theory is now extremely extensive, and it has been well surveyed elsewhere. For that reason, no comprehensive survey will be attempted here. The following sections make use of various models, some established, some new, according to convenience. The dual aims of the exercise are to throw some light on that elusive relationship between growth and trade; and also to treat profit-driven endogenous growth in a way more fitting to a globalized economy, so that R & D workers in different countries are affected by similar work elsewhere. This leads to a theory of *catching-up*, meaning that the country behind gains an advantage from the contributions of leaders. This is a radically different idea from the convergence of different countries discussed in Chapter 4, because there the theory is of isolated countries growing according to their own local conditions. It will be seen that the precise way in which catching-up is modelled makes a large difference, and that supports the view that the theory of catch-up mechanisms deserves more research attention than it has so far received.

## 12.3 A Simple Learning-by-Doing Model

As already explained in Chapter 1, this author generally prefers models that locate endogenous technical change in profit-driven activity much the same as standard production. There are inescapable differences, because knowledge is a public good; but even so, the profit-lead approach is more satisfying than what we have called the sticky-tape approach. The latter makes technical progress an incidental external consequence of normal economic activity, and that may well happen. As has been remarked, however, with the sticky-tape approach almost anything can be demonstrated. That said, our first model is of the sticky-tape variety. It is assumed that a certain economic activity, in this case the employment of skilled labour in the high-tech sector, generates increased productivity, for that factor in the same sector, as an incidental externality. This means that the said increase in productivity has no influence on the employment of skilled labour in the high-tech sector, as from the social point of view it should. One could interpret the model as embodying learning by doing. The more time in total that skilled labour works in the high-tech sector, the more skilled and productive it becomes. If that sector consists of many small firms, and labour can move freely between employers, it is realistic to assume that firms will ignore the social benefits of their employment decisions, because they will not be rewarded for the higher productivity that they create.

The framework is standard HOS of the Krugman-Wood style. We add the condition that skilled labour working in the high-tech sector augments the productivity of all labour employed in that sector. Thus the production functions are:

$$y^l = f^l\left[\ell_s^l, \ell_u^l\right] \tag{1}$$

$$y^h = f^h\left[\theta \cdot \ell_s^h, \ell_u^h\right] \tag{2}$$

$$\frac{d\theta}{dt} = \alpha \cdot \ell_s^h \tag{3}$$

Here $\theta$ is a parameter that measures the additional productivity of skilled labour in the high-tech sector induced by learning by doing. If we compare two small countries, one better endowed with skilled labour, the latter has more labour employed in the high-tech sector, even at a time when they share the same values of $\theta$. The growth of $\theta$ through (3) then accentuates any initial difference and forces the faster-growing country into high-tech specialization. After that returns to efficiency units of skilled labour decline, but total earning of skilled people may grow. Notice that we can encounter an immiserizing growth outcome here, see Bhagwati (1958). The slow-growing country may do better than the fast-growing country.

In this model, what is the consequence of a partial freeing of trade, starting from a situation in which trade is restricted, as by a tariff? Following the idea already exploited in Chapter 6, we answer this question from a North-South model, where relative to the South, the North is better endowed with skilled labour. Opening up trade a bit raises the relative price of the high-tech good in the North, and lowers it in the South. In the North the high-tech sector expands and the level of $\ell_s^h$ increases. This must be so, because the low-tech sector contracts and also the higher cost of skilled labour will cause both sectors to substitute against that input. Now it follows directly from equation (3), that the rate of growth in the North will increase. In the South the story is a mirror image of the story for the North. The level of $\ell_s^h$ decreases and the rate of growth falls.

This little model invites ridicule. It is a trivial extension of a standard HOS model. Its findings follow simply from the manner in which its global endogenous growth has been written into it. Increased trade lowers growth in the South for a reason that is much the same as the Johnson paradox examined in Chapter 5. Because the market does not take into account the external benefit, in the form of increased productivity, that flows from the employment of skilled labour in the high-tech sector, these are distorted economies. More trade for the South increases the distortion, just as more capital did in the Johnson-paradox model. It contracts the sector that should be expanded. This is the true reason why more trade is harmful to the South. It is the case that the opening up of trade lowers growth in the South, just as it increases it in the North. Remember however that growth is neither the objective nor the measure of economic welfare. Growth can even be immiserizing. Should that be the case, an export tax is an optimal intervention. The optimal response to the distortion noted above is an appropriate subsidy on the employment of skilled labour in the high-tech sector.

While the present model is absurdly simple and unsophisticated, its tells us something that must surely be far more general. It can never be a universal conclusion that freerer trade increases the rate of growth for all participating, or even for any of the participating, countries. It depends on the structure of the case at hand. This confirms a conclusion of Grossman and Helpman (1991: ch. 6), also for small countries, though with a different model. As these authors write (p. 152):

> Does trade promote innovation in our model of the small economy? The answer is, 'It depends'. When trade causes resources to be released from the manufacturing sectors, which then find their way into research labs, the rate of innovation rises. But when the sectors that expand in response to trading opportunities compete with research labs for factor inputs, international integration may retard growth.

Also, we are reminded once again that growth is not the correct measure of good economic performance.

## 12.4 Variety, Quality, Growth, and Trade

In the learning-by-doing model of the previous section, the gains from trade arise from differences in relative factor endowments, and the effect of trade on growth is via the learning-by-doing externality. A number of related models take a different approach. They identify technical change with an increased variety of inputs. The opening of this type of model to international exchange is a leading part of the analysis of trade and growth in Aghion and Howitt (1998: ch. 11), and also in Feenstra (2004: ch. 10), and in Grossman and Helpman (1991).

Here we follow Aghion and Howitt (1998: chs. 1 and 11). The most convenient specification involves a continuum of firms on [0, $A$] each producing its own type of intermediate input to the production of a single homogeneous consumption output. The aggregate production function is:

$$Y = L_1^{1-\alpha} \int_0^A x(i)^\alpha \, di \tag{4}$$

With the increased-variety model there are some basic differences of view as to how the model should be specified. Each firm is an imperfect competitor in the sale of its own particular intermediate input, which can be produced for sale at constant marginal cost, sometimes taken to be zero. When this marginal cost is positive it is removed in effect from the main model. On one interpretation, adopted by Aghion and Howitt, the intermediate services are produced from capital, freely available at a fixed cost. The costs of increasing $A$, that is the cost of adding new designs, are kept within the model as a labour cost. These strange, and somewhat forced, assumptions are made for good reason. They are needed for indefinite economic growth to be possible, at a rate not predetermined by the growth of the effective labour supply.

The assumption of zero marginal cost is natural if the product is the use of a patented method. In any case all variety models have to deal somehow with the fundamental difficulty that growth must be made possible in a model with a finite (often constant) labour supply. That means that the production of a growing menu of varieties must not demand ever greater labour inputs. Otherwise we would arrive at a Harrod-style conclusion; long-run growth would be determined entirely by the natural rate of growth, the sum of labour-force growth and labour-augmenting technical progress. It is not only current production costs that are subject to this problem. There must be some kind of overhead costs, for otherwise variety would be increased without limit. For this last problem different authors, and the same authors at different times, have produced different solutions. Grossman and Helpman (1989) accepted the Harrod-style property, with growth eventually limited by effective labour supply. On the other hand, Grossman and Helpman (1991) and Romer (1990) assume that there is an overhead cost, but that its level is inversely proportional to the extent of variety. Finally, Aghion and Howitt (1998) assume that overhead costs are once-for-all expenditures, with the consequence that labour is only employed in conjunction with varieties, as $L_1$ in (4), and in doing R & D to increase varieties. The labour cost of research is represented by the next equation:

$$\frac{1}{A}\frac{dA}{dt} = \delta L_2 \tag{5}$$

When the R & D has been done, the firm has an indefinite monopoly on the application of that knowledge: hence the monopolistic competition. There is free entry, in the sense that anyone can establish a new type if they pay the R & D cost, so that net profit from that activity is zero.

The levels of the $x$ values in (4) are determined by the profit maximization and a free-entry condition that says that working profit is just equal to the overhead cost of entry. The level of saving is decided by the requirement that growth maximizes a Ramsey integral of discounted utility:

$$\int_0^\infty c(t)^{1-\varepsilon} e^{-\rho t} dt \tag{6}$$

Noticed that despite the maximization of (6), this is not a fully optimal growth model. The integral (6) is maximized subject to the constraint that the provision of intermediate goods services will be decided by the equilibrium conditions of free-entry imperfect competition.

In the symmetrical case, where all $x$-values are equal, the allocation of labour between $L_1$ and $L_2$ is given by the equal-marginal-productivity condition:

$$(1-\alpha)L_1^{-\alpha}Ax^{\alpha} = \delta A P_A \tag{7}$$

where $P_A$ is the price of increasing the flow of designs; that is the market valuation of $\frac{dA}{dt}$. From (7):

$$(1-\alpha)L_1^{-\alpha}x^{\alpha} = \delta P_A \tag{8}$$

What will be the precise value of $P_A$ depends upon how completely those responsible for innovation can capture the economic value of their contribution to knowledge. If the value is captured in full, we get an important reference case.

The details of the computations required to calculate the values are laid out in Aghion and Howitt (1998: ch. 1), and Feenstra (2004: ch. 10). Aghion and Howitt (1998: 39) derive the steady-state growth rate as:

$$g = \frac{\alpha\delta\overline{L} - \rho}{\alpha + \epsilon} \tag{9}$$

where $\overline{L}$ is the total labour supply in the economy. When two identical countries are introduced to each other, and they trade, the increased growth that results is due simply to the fact that the integrated economy that results is twice as large as either of the original economies that were combined by trade. The property that large economies grow faster than small economies appears strange in this context, as cross-section empirics do not reveal this feature. One needs to be rather careful, however, in moving from abstract theoretical models to the empirics here. The theory shows that moving from autarky to free trade increases growth. When we look at countries that trade extensively, the comparison is not the same. In general small countries trade more heavily than do large countries; and were they not to do so, it is plausible to believe that they would grow slowly.

The production function (4) has an advantage over the similar specification where there is a finite number of products, and the integral in (4) is replaced by a sum. This is that $A$ is not an integer, and technical progress can be a smooth continuous process, without the need for discrete leaps as new varieties are added. That is an analytical convenience, and not closer to reality than the discrete version, as arguably technical innovations do in fact arrive as discrete jumps. Another issue of realism is posed by the increased-variety description itself. Obviously technical innovations are not all of one kind. And sometimes progress does involve the addition of a new product to the list of existing products. The telephone-answering machine would be such a case. In other instances, however, a new product assigns an existing product to the scrap heap, as happened eventually with the automobile and the horse cart. This way of viewing things corresponds to the quality-ladder model; see Grossman and Helpman (1991: ch. 4).

Contrary to what the increased-variety models seem to tell us, different countries frequently do not back off into their own special products, but rather compete head-to-head over the sale of the same product. This happens with steel, for example, and it seems that something close to it is seen with small family cars. The various versions of these for sale in world markets have converged hugely over the last few decades. Vast sums of money are spent on advertising, to persuade the consumer that one version is better than others, and this advertising usually associates the said car with glamorous locations, and not infrequently with glamorous girls. Advertising like this always tells us that the products concerned do not differ as fundamentally as the advertisers would like us to believe. As Chamberlin (1933) noted long ago, imperfect competitors strive actively to increase the degree of product differentiation, because it lowers the elasticity of demand for their particular products.

Head-to-head competition over sales of the same product have been extensively analysed in terms of the Cournot-Nash model of oligopoly equilibrium. See in this connection Brander and Krugman (1983), and Krugman (1979) and (1981). For a general trade-theory perspective, various models can be combined at will: quality, variety, and Cournot-Nash. Where endogenous growth is concerned, however, the theory would become impossibly cumbersome if it united different production specifications, and for this reason authors have selected one specification at a time.

## 12.5 A Model with Perfect Goods Aggregation

The next model is new. It is useful in particular for the examination of international trade in knowledge. In that context, trade in goods is best set aside, and to that end there is one tradeable good. We can imagine that different countries swap this good in different versions. The special assumption is that the relative prices of different versions are always the same. This is a standard one-consumer Ramsey economy that maximizes:

$$\int_0^{\infty} U\,[c(t)]\,e^{-rt}dt \tag{10}$$

Each country has a fixed endowment of unskilled and skilled labour, normalized to 1 and $\ell_0$. The tradeable good is produced under constant returns using both types of labour, just as in the usual Heckscher-Ohlin model. A research sector is skilled-labour intensive. It produces nation-specific Hicks-neutral technical progress in the tradeables sector.

The small-country balanced-payments planning solution maximizes (10) subject to:

$$c(t) = A \cdot f^t \left\{\ell^s(t), \ell^u(t)\right\} \tag{11}$$

And:

$$\frac{dA}{dt} = f^r \left\{\ell_0 - \ell^s(t), 1 - \ell^u(t)\right\} \tag{12}$$

Optimizing with respect to $\ell^s(t)$ and $\ell^u(t)$, gives a reduced form:

$$c(t) = H\left[A, \frac{dA}{dt}\right] \tag{13}$$

where $H\,[\,]$ is homogeneous of degree 1, and the form of $H\,[\cdot,\cdot]$ depends upon $\ell_0$. The higher is $\ell_0$, the higher is $c(t)$ for given $A$ and $\frac{dA}{dt}$ values.

Given $A$, the higher is its value, the less is $c(t)$ reduced by a particular increase in $\frac{dA}{dt}$. As in standard HOS analysis, the better endowed is the country with skilled labour, the lower the cost in terms of consumption of given level of $\frac{dA}{dt}$.

So we maximize:

$$\int_0^\infty U\left\{H\left[A, \frac{dA}{dt}\right]\right\} e^{-rt}dt \tag{14}$$

and the Euler equation gives:

$$\frac{d}{dt}\left\{uH_2e^{-rt}\right\} = uH_1e^{-rt} \tag{15}$$

where $u$ is the marginal utility of consumption, and subscripts to $H$ denote partial differentiation. From (15):

$$\frac{du}{dt}H_2e^{-rt} + u\frac{dH_2}{dt}e^{-rt} - ruH_2e^{-rt} = uH_1e^{-rt} \tag{16}$$

$$-\frac{1}{u}\frac{du}{dt} = \frac{1}{H_2}\frac{dH_2}{dt} - \frac{H_1}{H_2} - r \tag{17}$$

The growth of consumption satisfies the Ramsey rule with the usual marginal productivity of capital term replaced by:

$$\frac{1}{H_2}\frac{dH_2}{dt} - \frac{H_1}{H_2} = \frac{\frac{dH_2}{dt} - H_1}{H_2} \tag{18}$$

In a steady state at growth rate $g$ and a constant elasticity utility function $U = \frac{1}{\alpha}C^\alpha$ (18) becomes:

$$(1-\alpha)g = -\frac{H_1}{H_2} - r \tag{19}$$

If we compare two countries identical except for their factor endowments when they have equal levels of $A$, the country with more skilled labour will have a higher value of $-\frac{H_1}{H_2}$. A good endowment of skilled labour lets a country grow faster. Now imagine that the growth of $A$ is an internationally traded service. Prior to trade this service is cheaper in the skill-abundant country, which is why it grows faster. With complete free trade, the relative price of consumption and technical change are equalized across countries. The skill-abundant country does more R & D, but trade in the good and the R & D service equalizes growth rates.

## 12.6 Comparative Advantage in doing R & D

The next model is somewhat similar to the increased-variety model examined above. It is simpler in structure, mainly because it has no imperfect competition and no increasing returns. Consumption is:

$$C = AL_1^{\beta} \tag{20}$$

with $0 < \beta < 1$. The country's technical level is measured by $A$ on $[0, \infty)$, and $L_1$ is the labour allocated to producing consumption. The total labour force is 1. Then $1 - L_1$ workers are employed in the R & D sector. The dynamic equation for $A$ is:

$$\frac{dA}{dt} = \mu A(1 - L_1) \tag{21}$$

with $\mu$ a positive constant. This is similar to equation (5) above, but notice that although $A$ is used in each case to denote the technical level achieved, the precise meaning of $A$ is different in the two cases. All output is consumed and the planner maximizes:

$$\int_0^{\infty} C^{\alpha} e^{-\delta t} dt = \int_0^{\infty} A^{\alpha} L_1^{\alpha\beta} e^{-\delta t} dt \tag{22}$$

subject to (21).

We choose a constant value of $L_1$ to solve this problem. With $L_1$ constant the growth rate of $A$ is $\mu(1 - L_1)$, and $C$ is given by:

$$C = A_0 e^{\mu(1-L_1)t} L_1^{\beta} \tag{23}$$

and the maximand becomes:

$$\int_0^{\infty} A_0^{\alpha} L_1^{\alpha\beta} e^{[\alpha\mu(1-L_1)-\delta]t} dt = -\frac{A_0^{\alpha} L_1^{\alpha\beta}}{\alpha\mu(1 - L_1) - \delta} \tag{24}$$

provided that $\delta > \alpha\mu(1 - L_1)$, which is required for the convergence of the integral (22) in any case.

The maximization of (24) requires:

$$\frac{\alpha\beta L^{\alpha\beta-1}[\delta - \alpha\mu(1 - L_1)] - \alpha\mu L_1^{\alpha\beta}}{[\delta - \alpha\mu(1 - L_1)]^2} = 0 \tag{25}$$

This implies:

$$\beta[\delta - \alpha\mu(1 - L_1)] - \mu L_1 = 0 \tag{26}$$

$$L_1 = \frac{\beta(\delta - \alpha\mu)}{\mu(1 - \alpha\beta)} \tag{27}$$

We always have $\delta > \alpha\mu(1 - L_1)$. Then (27) must give a positive $L_1$, for if $L_1$ were to be zero then this last condition would say $\delta > \alpha\mu$, when (25) would make $L_1$ positive. So only $L_1$ positive is consistent. The formula (25) only applies if it gives a value of $L_1 \preceq 1$. If $\mu$ is sufficiently small $L_1$ as given in (25) will exceed 1. In such a case $L_1 = 1$ will be the model solution, and from (21) $\frac{dA}{dt}$ will be zero, and there will be no growth.

**Theorem 12.1:** *If $\mu$ is sufficiently small, there will be no economic growth. The values of $\mu$ which allow for economic growth are those such that $\beta\delta < \mu$.*

**Proof:** From (27) it will be seen that $L_1 < 1$ only if $\beta(\delta - \alpha\mu) < \mu(1 - \alpha\beta)$. This simplifies to $\beta\delta < \mu$, as required. □

To see the effect of the value of $\mu$ on growth differentiate (27) with respect to $\mu$ to obtain:

$$\frac{dL_1}{d\mu} = \frac{-\alpha\beta\mu^2(1 - \alpha\beta) - \beta(\delta - \alpha\mu)(1 - \alpha\beta)}{[\mu(1 - \alpha\beta)]^2} \tag{28}$$

The sign of (28) is given by:

$$-\alpha\mu^2 - (\delta - \alpha\mu) = \alpha\mu(1 - \mu) - \delta \tag{29}$$

So the effect of an increase in $\mu$ on $L_1$, and hence on growth, is ambiguous. Notice however, from (21), that $\mu$ is the rate of growth divided by the share of labour employed in the R & D sector. For that reason it should be small, certainly less than 1. So perhaps the most plausible case is $\frac{dL_1}{d\mu} < 0$. That means that a higher $\mu$ entails a higher rate of growth.

The most important point to understand is how it could be possible for an increase in $\mu$, that is an increase in labour productivity in the R & D sector, to lower the rate of growth. What is happening is similar to an income effect. A higher $\mu$ points to less current consumption and more future consumption, because growing consumption has become relatively cheaper. This is the substitution effect. At the same time, a higher $\mu$ points to more consumption at all times. This is the income effect. And more consumption early on means lower growth. This is why the ambiguity above arises. In the following section we focus on the more intuitive case, in which the substitution effect dominates outcome, in which case a higher $\mu$ raises the rate of economic growth.

## 12.7 International Exchange of Ideas

The above simple model of endogenous growth can be used to demonstrate the effect of the international exchange of ideas. Specifically we ignore trade in goods as such, and just take into account the point that research in one country can be influenced by the state of research in other countries.

In this context a critical question is whether it is an advantage or a disadvantage to be behind with respect to R & D. Being behind can be given more than one interpretation. It could mean having a lower level of labour productivity in research; that is a lower level of $\mu$. It could also mean having attained a lower level of technical efficiency in producing consumption; that is a lower level of $A$. In either case a country can learn from the example of others, and in that connection learning must be more effective for those with less than average attainment. The champion does not learn from a beginner. It is questionable, however, that the performance of the champion will inevitably tend to slide downwards, because he has no example to help him to improve. That will be a feature of our modelling, but plainly its realism can be questioned.

Whatever the reaction of the reader to our internal debate concerning assumptions, the following point deserves emphasis. A large part of the convergence that the world has witnessed over the last decades, notably the movement of European economies up towards the US level of productivity, has taken the form of catching-up, of copying in some form US practice and technique. That is equivalent to saying that the US to some extent suffered from a first-mover disadvantage. In the next section it will be shown that the strength of the international copying effect need only be extremely weak to make eventual convergence, in the relevant sense, happen. Of course if the catching-up force is weak, convergence may require a huge amount of time.

## 12.8 A Second Model of R & D Comparative Advantage

Consumption satisfies:

$$\ln C = T + \ln \mu_c + \beta \ln L_c \tag{30}$$

where the country's technical level is measured by $T$ on $(-\infty, +\infty)$. The total labour force is 1 and $L_c$ is the labour allocated to producing consumption, with $0 < \beta < 1$ and $\mu_c$ a positive constant. Then $1 - L_c$ workers are employed in R & D. The dynamic equation for $T$ is:

$$\frac{dT}{dt} = \mu_r (1 - L_c) \tag{31}$$

with $\mu_r$ a positive constant. For convenience denote $\ln C$ by $c$, and $\ln L_c$ by $l$. Then (30) and (31) can be written:

$$c = T + \ln \mu_c + \beta l \tag{32}$$

$$\frac{dT}{dt} = \mu_r (1 - \exp\{l\}) \tag{33}$$

All output is consumed and the planner maximizes:

$$\int_0^\infty U[c]\, e^{-\delta t} dt \tag{34}$$

subject to (33). Notice that (34) is the usual optimal growth objective function. However $U$ [.] maps the log of consumption to utility.

The control variable is $l$, the state variable is $T$. The Hamiltonian with $p_0 = 1$ is:

$$U[T + \ln \mu_c + \beta l]\, e^{-\delta t} + p_1 \mu_r (1 - \exp\{l\}) \tag{35}$$

The maximization of (35) with respect to $l$ requires:

$$\beta U_1 [c]\, e^{-\delta t} - p_1 \mu_r \exp\{l\} = 0 \tag{36}$$

where subscripts denote differentiation.

The costate variable condition requires:

$$\frac{dp_1}{dt} = -U_1 [c]\, e^{-\delta t} \tag{37}$$

Differentiating (36) totally with respect to time and taking into account (37) gives:

$$\beta \frac{dU_1[c]}{dt} e^{-\delta t} - \delta\beta U_1 [c]\, e^{-\delta t} + U_1 [c]\, e^{-\delta t} \mu_r \exp\{l\} = p_1 \mu_r \exp\{l\} \frac{dl}{dt} \tag{38}$$

Or,

$$-\frac{\frac{dU_1[c]}{dt}}{U_1[c]} = \frac{\mu_r}{\beta} \exp\{l\} - \frac{p_1 \mu_r \exp\{l\} \frac{dl}{dt} e^{\delta t}}{\beta U_1 [c]} - \delta \tag{39}$$

Equation (39) is like a standard Ramsey necessary condition, except that the right-hand side is complicated, and $c$ is the logarithm of consumption. The country can jump directly to the steady state, as there is no sticky variable. So, with $\frac{dl}{dt} = 0$ (39) simplifies to:

$$-\frac{\frac{dU_1[c]}{dt}}{U_1\,[c]} = \frac{\mu_r}{\beta} L_c - \delta \tag{40}$$

Note that marginal utility and its rate of change are with respect to $c$, not $C$. Try a special case:

$$U\,[c] = \frac{1}{1-\eta} \exp\left\{(1-\eta)\,c\right\} \tag{41}$$

Equation (41) is just the constant elasticity of intertemporal substitution function popular with growth theorists, written so that utility is a function of $c$. The unrestrained use of this form has been criticized in Chapter 3, but here we just accept it. Denoting derivatives with respect to $c$ by subscripts:

$$U_1 = \exp\left\{(1-\eta)\,c\right\} \tag{42}$$

$$U_{11} = (1-\eta) \exp\left\{(1-\eta)\,c\right\} \tag{43}$$

So if $U$ is to be strictly concave in $c$ we must have $\eta > 1$.

Now (40) becomes:

$$-(1-\eta)\frac{dc}{dt} = \frac{\mu_r}{\beta} L_c - \delta \tag{44}$$

Or:

$$\frac{dc}{dt} = \frac{\frac{\mu_r}{\beta} - \delta}{\eta - 1} \tag{45}$$

This is the steady-state rate of growth. It is independent of $\mu_c$ although that parameter will influence the level of consumption.

## 12.9 Trade in the Second R & D Model

We take a simple view of international trade here. It allows an individual country to obtain the effect of more plants (a larger $n$) without having to pay all of $n_i.\alpha_i$. It achieves this by producing from a subset of plants and obtaining other intermediate inputs via trade. The final result is that trade raises the $\mu$ values, and the extent to which it does so more in one sector than the other depends upon the country's size and how trade gains are biased in favour of one sector or the other. Then from (45) we infer:

**Theorem 12.2:** *Trade increases the level of consumption if it increases $\mu_c$. Trade increases the steady-state growth rate of the economy if it increases $\mu_r$.*

**Proof:** By inspection of equation (45). □

To provide the intuition of this result it is only necessary to note that for trade to increase the growth rate it has to raise R & D efficiency. Increasing the efficiency of delivering current consumption is good but not growth-enhancing.

## 12.10 Catch-up Dynamics

Suppose that:

$$\frac{d\mu^j}{dt} = \zeta \left[ \frac{\sum_i \mu_i}{N} - \mu^j \right] \tag{46}$$

where $N$ is the number of countries. The idea is that the efficiency of R & D in any country increases (decreases) at a rate proportional to the extent to which it lags behind (is ahead of) the world average efficiency. Note that these efficiency dynamics are independent of the technical levels and growth dynamics of individual countries.

**Theorem 12.3:** *With the adjustment process shown in (46), all countries converge in growth rates, but not in levels (values of $A$).*

**Proof:** By inspection of equations (45) and (46). □

An alternative idea is that catching up applies directly to $A$. So replace (46) by:

$$\frac{dA^j}{dt} = \mu \left[ \frac{\sum_i A^i}{N} - A^j \right] \left(1 - L_1^j\right) \tag{47}$$

where, as with many imperfect competition models, the term $[\frac{\sum_i A^i}{N} - A^j]$ is treated by the planner as a function of time independent of the choice of $L_1$ at any time; this despite the fact that choices of $L_1$ do have a small effect on the said term. The implication of this last amendment on the maximization calculations above is the same as if $\mu$ were to be replaced by an arbitrary continuous function of time. This makes no other difference to the optimality conditions.

**Theorem 12.4:** *With the adjustment process shown in (47), and if growth increases with $\mu$, there is $\beta$-convergence, and all countries converge both in growth rates and in levels.*

**Proof:** From inspection of equation (47) it may be seen that a country with a low relative value of $A$ is the same as one with a high value of $\mu$. Then consumption will be growing faster, which is $\beta$-convergence.

Now take any time and two countries for which $A$ takes different values, denoted $A^+$ and $A^-$, with $A^+ > A^-$. Now the lower-$A$ country will grow faster, and this will remain the case until the two $A$ values reach equality, should that happen. With ln $[A^+ - A^-]$ declining and bounded below by a constant, it must either converge to that constant or to a limit greater than the constant. In either case we have in the limit two countries with a constant logarithmic distance between their $A$ values, and the country with the lower $A$ value always growing faster. Eventually the consumption of the country with the lower $A$ value will overtake the consumption of the higher-$A$-value country, and will continue to grow faster. However the logarithm of consumption is $\ln A + \beta \ln L_1$, so that the country with the low $A$ value can only have its consumption gaining without limit on the consumption of the country with the high $A$ value if $L_1$ increases without limit, which is impossible as $L_1$ is bounded above by 1. It follows that $\ln[A^+ - A^-]$ cannot converge to a limit and the proof is complete. □

## 12.11 Convergence and Catch-up Dynamics

It is received wisdom that with endogenous growth the $\beta$-convergence characteristic of the Ramsey-Solow model is no longer to be expected. We have arrived at a similar conclusion above. The level of consumption in (32) depends upon $\mu_c$, assumed to be the same for all cases, $L_c$, and $T$. The value of $L_c$ is chosen to optimize, but that choice is not affected by the level of $T$. It depends only upon parameters again common to all cases. The rate of growth is given by (45) and does not depend upon $T$. Therefore there is no tendency for convergence in levels and with shocks consumption levels will be an untrended random-walk process.

There is nothing mysterious about this conclusion. Unlike the Ramsey-Solow model, there is nothing in our endogenous-growth model to lend a comparative growth advantage to the poorer unit. The R & D sector is the growth driver, and its dynamic equation, repeated here for convenience as equation (48), is the same for all units, regardless of level.

$$\frac{dT}{dt} = \mu_r (1 - L_c) \tag{48}$$

At the cost of additional complications we can replace (46) by:

$$\frac{dT}{dt} = \mu_r [T] (1 - L_c) \tag{49}$$

where $\mu_r [\cdot]$ is a weakly monotonically decreasing function such that:

$$\mu_r [T] = \mu_0 \quad \text{for } T > T_0 \tag{50}$$

The last condition allows for a steady state when the technical level is sufficiently advanced, while at the same time giving a growth advantage to poorer units. It says that a given labour input to the R & D sector yields more growth when the technical level is low. The result is $\beta$-convergence but only for poor units. Rich units, $T > T_0$, do not necessarily converge.

## 12.12 Concluding Remarks

With several models new and old competing for attention, this chapter badly needs a summing-up and a pulling-together of ideas and key conclusions. The following points, either shown by our analysis or worth adding now, are of central importance.

1. The relation between trade and growth is fundamentally ambiguous. It is true that no country closed to trade has ever grown rapidly. For the efficient open economies of the economics classroom, however, more trade will, as always, boost some activities and shrink others. The boosted activities may be more growth-enhancing than those shrunk, or the opposite.
2. It is the case that any slow-growing economy is condemned to relative poverty in comparison with a faster-growing economy. Our analysis, however, has treated small economies facing fixed prices. In reality, prices will not be fixed in the long run. They will move, and typically so as to favour products with a lower tendency to grow. Again, growth is not the be-all and end-all of economic success.
3. Trade results in a country specializing to a greater extent in those activities in which it enjoys a comparative advantage. A country may enjoy a comparative advantage in growth, possibly simply on account of its factor endowments. Redding (2002) calls this dynamic

comparative advantage. When that happens, trade may lead to the country concerned growing faster by itself. But should R & D services be tradeable, then growth may be similar in all countries, while the country with the comparative advantage in growth will undertake more of the R & D work.

4. Even when R & D services are not tradeable, a country undertaking its own research can learn from other countries. When that happens, it is possible, although not certain, that the country with a lower technical level can advance more rapidly. We have called this catching-up, and the analysis indicates that catching-up, even when it is a weak force, can make a large difference in the long run. Exactly how the catch-up dynamics are specified affects the type of convergence that results.
5. It is not certain that the country with the lower technical level can take advantage of its backward situation to advance more rapidly. A low level may inhibit the process of learning by copying. With a lecture course on quantum mechanics, it will not be the student who knows the least physics who will benefit most. This is a kind of non-convexity, similar to those analysed in Chapter 3.

# 13

# Two Models of Growth and the Resource Curse

## 13.1 The Resource Curse

An influential study, Sachs and Warner (1997*a*), examines the relation between resource richness and economic growth. The authors use a wide cross-section of countries for the period 1970–89 and claim that resource richness is negatively correlated with economic growth. In this connection see also Auty (1990) and (2001), Gelb (1988), and Hausmann and Rigobon (2002). Sachs and Warner assert that the negative association between resource abundance and economic growth survives the inclusion of additional standard variables in the growth regression, in particular measures of trade openness and of the quality of bureaucratic administration. Several oil-rich Arab states are not included in the sample for lack of data, but their growth performance is even worse than included resource-rich nations. This is an ambitious study, well designed to provoke debate and speculation. It is not without problems. Some coefficients are barely significant. Also the trade-openness and bureaucratic-quality variables measure what is required fairly imperfectly. For instance, one oil-exporting state with no manufacturing to protect may well be 100 per cent open, where another resource-rich nation uses the revenue that its resource exports generate to subsidize local manufacturing, and is counted as less open.

These little points hardly matter as the grand question remains: why should growth be lower in resource-rich states? One known consequence of resource richness is the so-called *Dutch disease*. A resource discovery or boom increases the demand for domestic non-tradeable goods, either because the resource sector requires these goods, or from the income effect from local increased prosperity. Non-traded goods are then drawn out of the non-resource, traded-goods sector. This constitutes a disease only because structural adjustment is painful, particularly for factors that must move, or whose price falls. Notice that this account includes no obvious association with economic growth beyond a temporary decline that may follow from factor unemployment.

Sachs and Warner suggest that an endogenous-growth explanation may lie behind their findings. If resource sectors generate less endogenous growth than do manufacturing sectors, then resource abundance may depress growth simply by shrinking the alternative good-for-growth sectors. That cannot be the complete story for the sub-Saharan African (SSA) countries, as an SSA dummy variable is typically significantly negative in cross-section regressions, while these countries are resource-rich, or not, to a greatly variable extent. Robinson, Torvik and Verdier (2005) examine the negative political consequences of resource richness.

## 13.2 How To Model Growth

This chapter builds its arguments on the basis of a wholesale rejection of what may be called 'sticky-tape' endogenous growth models. The term 'sticky tape' means that productivity increases are added to (stuck onto) normal productive activities as an incidental externality. Of course external economies may be present. Arrow's pioneering learning-by-doing model, Arrow (1962), is sticky tape if one likes. It should be noted however, that the Arrow paper details how the externality arises, rather than just assuming it. And the assumption has empirical support at least for one sector, even if it is applied to the whole economy.

The general objection to sticky-tape modelling is that it allows too many conclusions to be given what looks like theoretical support. A temperance campaigner can attach productivity increases to economic activities not producing intoxicating liquors, and claim that the control of alcohol consumption will increase economic growth. His opponent can write down a model in which there are huge learning-by-doing economies in the brewing and distillation industries, and arrive at the opposite conclusion.

The alternative to sticky-tape modelling is shown by Grossman and Helpman (1991) in their important volume. They argue eloquently that growth generation should be modelled as the result of profit maximizing activities, like any other costly uses of scarce resources. Then commercial R & D, such as pharmaceutical research, becomes a sector of the economy, and profit maximization determines its scale, just as profit maximization decides the size of the bread-baking sector of the economy. A decentralized private-sector treatment of R & D comes up against the problem that the solution is crucially dependent on the extent and duration of the private capture of social benefits. Most technological progress is a public good, because the use of the knowledge by one party does not exclude its full use by others. In the present chapter we skate around this last problem by assuming that the R & D sector is a separate sector that separately employs resources; while determining the size of that sector by social optimization. This gives clear and definite conclusions. The reader may note immediately that the social optimization treatment is bound to produce more growth than will be achieved in a decentralized economy with imperfect private capture. Therefore if it can be shown that resource abundance may lower growth in our models, then it can certainly lower growth in an economy with partial private capture of the benefits of technological progress.

## 13.3 What is Resource Abundance?

Everyone would recognize some things as resource abundance. A country has huge gas or oil reserves, as with Russia or Saudi Arabia. In such a case the effect takes the form of a transfer to the country of the net international value of extracting and exporting the product concerned. This way of looking at the resource effect, as a gift to the trade balance, calls to mind the early consequence of the exploitation of UK North Sea oil in the 1970s. Hardly anyone in the UK at that time had any expertise in undersea oil extraction. What happened, to simplify the story, is that Texan experts were called in to get the oil out, and after they had done their work and taken their fees, the remaining surpluses were transferred to the UK government and private shareholders as a dollar credit. This is precisely the way in which resource abundance is treated in our Model 1.

That way of modelling resource richness fails to capture what might be called an export boom. Take the Kenya coffee boom (1976–9). This was the result of a dramatic increase in the world price of coffee as it impacted on a coffee-producing country. That could be viewed as a large improvement in the balance of trade, but to treat it that way would be to miss the impact of consequent relative price changes on the domestic economy. This was essentially a change in the terms of trade. With Model 2 we have two tradeable goods, so we can model resource abundance as an improvement in the terms of trade, and see how that affects the optimal rate of growth.

It is important to add that the Kenya coffee boom is not an ideal illustration of what Model 2 depicts, because this was a characteristically brief episode, while the model only analyses the steady-state effects of a permanent improvement in the terms of trade. The Sachs-Warner study only examines resource abundance as a fairly permanent feature of the countries concerned. Even so, the distinction between evident resource richness and the consequences of a favourable terms-of-trade change is by no means solid. To illustrate this last point consider the rich tar-sand deposits of Alberta in Canada. The existence of these deposits has been well known for at least a century, and the assessment of their scale requires no costly exploration, as they lie close to the surface. Yet they have never until recently been regarded as a huge resource for the Canadian economy. This is because they are costly to exploit. The oil has to be steamed or burnt out of the sand that holds it, and until the world oil price reaches about \$50 a barrel it is not economic to do that. Recently the world oil price has been well above \$50 a barrel, and if it stays there, as it may well do, Canada will become a major energy exporter and will experience a resource boom. It is already resource-rich on account of timber and minerals. If the new resource boom happens it will be the consequence of a terms-of-trade change: a huge rise in the price of oil relative to other goods.

## 13.4 Growth and Resources: Model 1

The model is kept as simple as possible so as to illustrate root ideas. Its form is similar in spirit to the model used by Forsyth and Kay (1981) to analyse the consequences of North Sea oil on UK manufacturing from the 1970s. There are two goods produced by labour, two types skilled and unskilled. One good 1 is tradeable, and good 2 is non-tradeable. Then:

$$1 = c^1\,[w^s, w^u] \tag{1}$$

$$q = c^2\,[w^s, w^u] \tag{2}$$

where the $c\,[\,]$ functions are unit-cost functions, and $w^i$ ($i = s$ or $u$) are the wage rates of the two types of labour in terms of the tradeable good. Finally $q$ is the *real exchange rate*, the price of the non-traded good in terms of the traded good.

Because this is a partly closed economy a large part is played by domestic demand. In the simple HOS model, world markets supply demand, at least for a small economy, whatever may be required. For that reason local demand does not affect the domestic production equilibrium. Assumptions on demand, such as the homotheticity assumption, only play the role of translating home comparative advantage into exports and imports in the natural and intuitive direction.

None of the above is true for the current model. The relative price $q$ is not given by international markets: it is a domestically determined variable. Now given supplies of the two kinds of labour, outputs depend upon the price $q$ just as with the Stolper-Samuelson effect. Then demand-supply equilibrium satisfies the following equations.

$$E_1[1,q,U] = S^1[q] + R \tag{3}$$

$$E_2[1,q,U] = S^2[q] \tag{4}$$

where partial derivatives of $E$ functions, that is subscripted $E$ values, give demands; while $S$ functions give supplies. The parameter $R$ shows the effect of a resource discovery, modelled here as a simple transfer across the balance of trade of a stock of the tradeable good. Differentiating (3) and (4) totally with respect to $R$ gives:

$$\left\{E_{12}[1,q,U] - S_1^1[q]\right\}\frac{dq}{dR} + E_{13}[1,q,U]\frac{dU}{dR} = 1 \tag{5}$$

$$\left\{E_{22}[1,q,U] - S_1^2[q]\right\}\frac{dq}{dR} + E_{23}[1,q,U]\frac{dU}{dR} = 0 \tag{6}$$

Or, in matrix form:

$$\begin{bmatrix} E_{12}[1,q,U] - S_1^1[q] & E_{13}[1,q,U] \\ E_{22}[1,q,U] - S_1^2[q] & E_{23}[1,q,U] \end{bmatrix}\begin{bmatrix} \frac{dq}{dR} \\ \frac{dU}{dR} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{7}$$

In a normal case, with no perverse income effects, an increase in $q$ causes the net demand for good 1 to increase, and the net demand for good 2 to decrease. Assuming both goods are consumed in positive quantities, an increase in utility requires expenditure to increase. We conclude that the sign pattern of the square matrix in (7) is:

$$\begin{bmatrix} + & + \\ - & + \end{bmatrix} \tag{8}$$

so that the determinant of the matrix is positive. Using Cranmer's rule to solve for the effect of small changes in $R$ gives:

$$\frac{dq}{dR} = \begin{vmatrix} 1 & + \\ 0 & + \end{vmatrix} \Big/ D > 0 \tag{9}$$

$$\frac{dU}{dR} = -\begin{vmatrix} + & 1 \\ - & 0 \end{vmatrix} \Big/ D > 0 \tag{10}$$

These are unsurprising conclusions. A larger resource endowment raises the real exchange rate, which is to say that it makes the non-tradeable good more expensive relative to tradeables. This is the classic Dutch-disease effect. The same larger resource endowment increases the utility of the single consumer, which seems odd in connection with the use of the term *disease*. Few diseases make their sufferers better off. There is no genuine paradox here. With more than one type of consumer some may be hurt even when there is in some sense general gain. Also the model has perfectly flexible prices. Hence the involuntary unemployment of some workers, often seen as a typical feature of Dutch-disease situations, cannot happen here.

## 13.5 Growth Consequences of Resource Discoveries

So far our model is entirely static and neither models nor refers to growth. However the addition of an R & D sector allows a growth effect to be included. And because a separate sector is taken into account this is not a case of sticky-tape modelling. We allow the labour inputs that are applied to direct production in the above model to be variable quantities, denoted $(l^s, l^u)$. Then from the solution of the model the level of current utility attained is the indirect function:

$$V\,[l^s, l^u, R] \tag{11}$$

which function increases with each of its arguments. It is the partial derivatives of this function that will provide us with insight into the growth effect of an increase in $R$. The derivatives:

$$\frac{\partial V\,[l^s, l^u, R]}{\partial l^s} \text{ and } \frac{\partial V\,[l^s, l^u, R]}{\partial l^u} \tag{12}$$

are the utility shadow prices of the two types of labour. Both are positive numbers because more of any type of labour means more output and higher utility. Less clear are the signs of the second-order cross-partial derivatives:

$$\frac{\partial^2 V[l^s, l^u, R]}{\partial l^s \partial R} \text{ and } \frac{\partial^2 V[l^s, l^u, R]}{\partial l^u \partial R} \tag{13}$$

These measure what an increase in resources $R$ does to the marginal utility valuations of the labour types. The answer depends upon the relative factor intensities of the tradeable and the non-tradeable sectors. Suppose for example that the tradeables sector is relatively intensive in the use of unskilled labour. An increase in $R$ raises $q$, that is lowers the relative price of the tradeable good. That via Stolper-Samuelson magnification lowers $w^u$ and raises $w^s$. These wage rates are in terms of the tradeable good. They are translated to marginal utility values when they are multiplied by the marginal utility of the tradeable good. As that marginal utility value is positive the cross-partial derivatives (13) will move in the same direction as the $w$ values.

Now take a simple short-run view of growth. The economy allocates its labour to maximize the sum of two indirect functions:

$$V[l^s, l^u, R] + G\left[L_0^s - l^s, L_0^u - l^u\right] \tag{14}$$

where $V[]$ is the current utility function defined above, $G[]$ is the value of growth-inducing activity, and $L_0^s$ and $L_0^u$ are the aggregate supplies of the two types of labour. Here $G[]$ will be assumed to be a concave function. An attractive consequence of representing the benefit of growth-inducing activity by going straight to the indirect function $G[]$ is that it allows for great breadth of interpretation. Thus $G[]$ may measure the total social benefit of extra growth, or it may measure the private benefit of growth-promoting activity, when some of the social gains are dissipated by external economies that private agents do not capture.

An undesirable feature of the particular form that has been chosen for the expression (14) may be that $R$ does not appear as an argument of $G[]$. It is possible that extra resources will affect the value produced by the allocation of given labour resources to growth enhancement. Note that $R$ is simply the transfer of tradeable resources to the home country, and that it has already been completely credited to current utility in the function $V[]$, so it is not obvious why it should appear in $G[]$ as well. In any case, the exclusion of $R$ from $G[]$ will only upset the following argument if more $R$ has a large positive effect on $G[]$, and it is not at all clear why that should be so.

Given the omission of $R$ from $G\,[\,]$ in (14), the effect of an increase in $R$ on growth can be analysed in a straightforward manner. The maximization of (14) with an interior solution requires:

$$V_1\,[l^s, l^u, R] - G_1\left[L_0^s - l^s, L_0^u - l^u\right] = 0 \tag{15}$$

and:

$$V_2\,[l^s, l^u, R] - G_2\left[L_0^s - l^s, L_0^u - l^u\right] = 0 \tag{16}$$

where subscripts denote partial differentiation. For a regular maximum we require that the Jacobean matrix of second-order derivatives has a particular sign pattern. Differentiating (15) and (16) totally with respect to $R$ gives:

$$\begin{bmatrix} V_{11} - G_{11} & V_{12} - G_{12} \\ V_{21} - G_{21} & V_{22} - G_{22} \end{bmatrix} \begin{bmatrix} \frac{dl^s}{dR} \\ \frac{dl^u}{dR} \end{bmatrix} = \begin{bmatrix} -V_{13} \\ -V_{23} \end{bmatrix} \tag{17}$$

where the arguments of functions have been dropped for the sake of clarity. Some things concerning the sign pattern of (17) can be established by the following arguments. First if $V\,[l^s, l^u, R] - G\left[L_0^s - l^s, L_0^u - l^u\right]$ is concave in the two $l$ values, then the determinant of the matrix on the left-hand side of (17) is positive definite. This implies:

$$V_{11} - G_{11} < 0 \tag{18}$$

and:

$$(V_{11} - G_{11})\,(V_{22} - G_{22}) - (V_{12} - G_{12})^2 > 0 \tag{19}$$

Next assume that $V_{12} - G_{12} = V_{21} - G_{21} > 0$. This says that the two types of labour are complements in the maximization of (14), so that more of one type of labour increases the marginal product of the other. That would not be the case if the two types of labour were near perfect substitutes. So this is an assumption, albeit a reasonable assumption. From the above discussion it follows that the sign pattern of (17) must be:

$$\begin{bmatrix} - & + \\ + & - \end{bmatrix} \begin{bmatrix} \frac{dl^s}{dR} \\ \frac{dl^u}{dR} \end{bmatrix} = \begin{bmatrix} - \\ + \end{bmatrix} \tag{20}$$

The determinant of the matrix is denoted $D$ again, but now $D$ has a new value. Using Cranmer's rule again gives:

$$\frac{dl^s}{dR} = \begin{vmatrix} - & + \\ + & - \end{vmatrix} \Big/ D \tag{21}$$

$$\frac{dl^u}{dR} = -\begin{vmatrix} + & - \\ - & + \end{vmatrix} \Big/ D \tag{22}$$

Without knowing the value of the various terms in (21) and (22), but just relying on the sign pattern, we cannot sign either $\frac{dl^s}{dR}$ or $\frac{dl^u}{dR}$. This is not so surprising. The effect of extra $R$ is to raise current utility, while it increases the opportunity cost of skilled labour, but decreases the opportunity cost of unskilled labour.

It is true that growth generation is assumed to be intensive in the use of skilled labour, so it might seem that the rise in the cost of skilled labour will be the more important effect, and that growth will definitely decline. We cannot argue that way however because the argument does not take into account the size of the changes, not just their signs. Thus if skilled labour becomes more expensive, but only slightly so, while unskilled labour becomes cheaper and greatly so, then growth may increase.

Our investigation contrasts with the informal argument provided by Sachs and Warner, but this is entirely explained by different modelling assumptions. Sachs and Warner say that extra resources increase the demand for non-tradeables, due to an income effect presumably, and that pulls non-tradeables away from growth-promoting activities, causing growth to decline. In our model non-tradeables are produced goods. Then an increase in their demand translates to higher demand for the non-producible factors that make them. With more than one non-produced factor the increase in demand leads to a rise in demand for one non-produced factor but a fall in the demand for the other, in standard Rybczynski style. That is why factor-intensity assumptions have to feature in our analysis where they do not feature in the Sachs-Warner argument.

## 13.6 Growth and Resources: Model 2

Now in contrast to Model 1, we move to a model in which all outputs are tradeable. This allows us to model terms-of-trade changes. There are two consumption goods produced by unskilled and skilled labour. The world price of the skill-intensive good 2 in terms of the other good 1 is $p$. Then:

$$1 = c^1 \left[w^s, w^u\right] \tag{23}$$

$$p = c^2 \left[w^s, w^u\right] \tag{24}$$

where the $c\,[\,]$ functions are unit-cost functions, and $w^i$ ($i = s$ or $u$) is the wage rate of the $i$th type of labour. The total supplies of the two types of labour are each normalized to 1. For R & D the two types of labour are used in fixed proportions, $a > 1$ units of skilled labour with 1 unit of unskilled. Then quantities $aL$ and $L$ of respectively skilled and unskilled labour are employed in the R & D sector. Therefore the full-employment conditions for the remaining HOS-style economy are:

$$\frac{\partial c^1\,[w^s, w^u]}{\partial w^s} y^1 + \frac{\partial c^2\,[w^s, w^u]}{\partial w^s} y^2 = A\,(1 - aL) \tag{25}$$

$$\frac{\partial c^1\,[w^s, w^u]}{\partial w^u} y^1 + \frac{\partial c^2\,[w^s, w^u]}{\partial w^u} y^2 = A\,(1 - L) \tag{26}$$

where $y^i$ is the output of sector $i$, and $A$ is a labour-augmenting efficiency factor for goods production.

It is easiest to ignore the details of the HOS equilibrium solution, and go straight to the indirect function which gives the maximized value of national output in terms of good 1. This function is:

$$AV\,[p, 1 - aL, 1 - L] \tag{27}$$

where $V\,[\,]$ increases with each of its three arguments. Then instantaneous utility $U$ is defined by:

$$AV\,[p, 1 - aL, 1 - L] = E\,[p, U] \tag{28}$$

If we think of a developing country with a comparative advantage in producing good 1, then an increase in $p$ corresponds to a negative terms-of-trade shock, and given $A$ will be associated with a fall in $U$. If $U$ is extracted from (28) as an explicit function it takes the form:

$$U\,[p, A\,(1 - aL)\,, A\,(1 - L)] \tag{29}$$

So far no restriction on the form of $U$ has been proposed. However to allow for steady-state growth $U$ should be homogeneous, so that (29) takes the form:

$$A^{\beta} U\,[p, 1 - aL, 1 - L] \tag{30}$$

where $0 < \beta < 1$.

The technical level is measured by $A$ on $[0, \infty)$. The dynamic equation for $A$ is:

$$\frac{dA}{dt} = \mu AL \tag{31}$$

where $\mu$ is a positive constant and $L$ is the size of labour teams in the proportions $(a, 1)$ employed in the R & D sector.

The planner maximizes:

$$\int_0^\infty U e^{-\delta t} dt = \int_0^\infty A^\beta U\,[p, 1 - aL, 1 - L]\, e^{-\delta t} dt \tag{32}$$

subject to (31).

We choose a constant value of $L$, the labour variable, to solve this problem. Then (32) becomes:

$$A_0^\beta U\,[p, 1 - aL, 1 - L] \int_0^\infty e^{[\beta\mu L - \delta]t} dt = -A_0^\beta \frac{U\,[p, 1 - aL, 1 - L]}{\beta\mu L - \delta} \tag{33}$$

provided that $\delta > \beta\mu L$, which is required for the convergence of the integral (32) in any case.

The maximization of (33) with respect to a positive value of $L$ requires:

$$\frac{[-aU_2 - U_3]\,[\delta - \beta\mu L] + U\beta\mu}{[\delta - \beta\mu L]^2} = 0 \tag{34}$$

where numerical subscripts denote partial differentiation, and arguments of functions have been omitted where there is no ambiguity.

This implies:

$$\frac{aU_2 + U_3}{U} = \frac{\beta\mu}{\delta - \beta\mu L} \tag{35}$$

How is the growth rate affected by a change in $p$, in particular by the fall in $p$ that would represent a resource boom? The opening up of a somewhat closed economy might also be represented by a fall in $p$. Notice first that the equation (35) may not admit of a solution for $L$ positive. Instead we may have:

$$\frac{aU_2 + U_3}{U} > \frac{\beta\mu}{\delta} \tag{36}$$

This means that, at least for some values of $p$, it is not optimal for the country to commit any resources to R & D, and its optimal growth rate will be zero. In the following discussion we ignore this case, but it is worth keeping in mind the possibility that a change in $p$ might trip the country from a positive- to a zero-growth-rate outcome.

The numerator of the left-hand side of (35):

$$aU_2 + U_3 \tag{37}$$

has a simple meaning. It is the shadow price in terms of utility of a labour team in the proportions demanded by the R & D sector. What does a fall in $p$ do to this shadow price? More than one effect is involved in the answer to this question.

1. An increase of labour in the proportion $(a,1)$ increases the proportion of labour that is skilled. This raises the share of output that is good 2 to a magnified extent. A fall in $p$ lowers the international value of this change, and for that reason lowers the shadow price under consideration.
2. In terms of utility the fall in $p$, which is the price of the imported good, raises the utility value of extra production, and for that reason raises the shadow price under consideration.

The conclusion must be that the sign of the change in (34) is ambiguous.

The denominator of the left-hand side of (35) is $U$. Notice that $U$ as defined by (29) includes the production and valuation effects of a change in $p$, as well as the direct utility effects (the Slutsky effect) of the change. In the case of interest a fall in $p$ increases $U$ because both the production and consumption effects are to the benefit of the home country. By itself this would lower the left-hand side of (35), and that would require $L$ to fall.

Therefore we conclude:

**Theorem 13.1:** *An improvement in the terms of trade lowers the rate of growth if it lowers the shadow price of a labour-supply vector* $(a, 1)$ *and if it also raises* $U$*. These conditions are sufficient but not necessary. An improvement in the terms of trade may lower the rate of growth even if the shadow price of the labour supply vector rises, or if utility falls.*

**Proof:** By inspection of equation (35). □

We now have part at least of a simple resource-curse type of argument. It applies most straightforwardly to the more compelling case in which a fall in $p$ does lower the shadow price of a labour-supply vector $(a, 1)$ and where it also lowers $U$. This is the case in which the production consequences of a terms-of-trade change dominate the consumption consequences. A coffee boom makes coffee drinking more costly for Kenyans, but this is a tiny effect in comparison with the production and wealth-increasing consequences. A fall in $p$ raises utility via the direct price (Slutsky) effect. A parallel point indicates that this direct price effect will be small.

The above argument has nothing to do with that sketched by Sachs and Warner, and it is entirely independent of it. Sachs and Warner refer to a model that includes a non-tradeable output. What happens in their case is that the positive income effect that follows from a resource discovery (or an improvement in the terms of trade) pulls the non-tradeable resource out of growth-increasing activities. That is similar to classic Dutch disease. In the model of this chaper all outputs are tradeable, so the Sachs-Warner effect cannot be present.

So what is the intuition behind our findings? That question is answered easily from the recognition that the model is an absurdly oversimplified steady-state model. Equation (11) confirms that the value of any plan is decided by the size of the base utility term, and also by the rate of growth. It is as if one were to visit a tree nursery to purchase a tree, with a fixed sum of money to spend, and with the objective of maximizing the integral of wood over the life of the tree. On sale are large slow-growing trees and small fast-growing trees. The optimal tree to purchase can be selected by comparing the various integrals available for the sum of money to be expended. What a terms-of-trade improvement does is to make trees that start large relatively cheaper. Then it is natural that lower-growth varieties are selected.

## 13.7 Concluding Remarks

Real exchange-rate theory was developed in the 1970s, and the fundamental ideas require no amendment today. This chapter adds a new angle. It uses a highly simplified Grossman-Helpman type of endogenous growth model to investigate the relationship between steady-state terms of trade and the rate of growth. The resulting analysis is complex, but in the most compelling case for a small developing country, improved terms of trade are associated with a lower growth rate.

Some people will doubt that an abstract mathematical model can throw light on the complex economics of a developing country. Where is the R & D sector of Chad? Of course R & D is only a title for growth-increasing activities, whatever they may be, and such always exist. A deeper problem may be that the model is an optimal growth model, and who supposes that any real country grows optimally? The standard argument says that there will be underinvestment in R & D because investors cannot capture the external benefits of their innovations. Oddly that problem may be less acute in poor simple economies than in large complex industrial economies. Be that as it may, this chapter proposes a start with the task of reconsidering the real exchange rate, and if it is at all successful it will not be the finish.

# 14

# Unequal Trade and Trade between the Unequal

## 14.1 Globalization and its Discontents

Winston Churchill said of democracy: 'the only argument for it is that everything else is much worse'. Exactly the same might be said of trade liberalization, and the structural reforms that accompany it when programmes promoted by the IMF, the World Bank, and many economists, are under consideration. The heading of this section is borrowed from a prominent book by the Nobel-Prize-winner Joseph Stiglitz, Stiglitz (2002). This chapter chooses to concentrate on the analytical content of this contribution to the literature, leaving aside the author's intemperate attack on IMF staff. Stiglitz argues that everything else is most certainly not worse than the rigid formulaic adjustment programmes applied by his *bete noire*, the IMF. His case is broad and far-reaching, and a full evaluation of it does not belong here. See Bliss (2003*b*) and Wolf (2004) for a more detailed discussion of some of the issues.

Stiglitz's arguments are part of a continuing debate concerning the correct roles of the IMF and the World Bank at the end of the second millennium, and the start of the third. In their original incarnations these institutions had well-defined and distinct job descriptions. The IMF was tasked with the support of a dollar-based, mainly fixed-exchange-rate, world exchange system. Its interventions in the affairs of member nations were intended to have the character of a hospital's emergency room: the provision of quick, powerful, and short-term remedial treatment, and the hard-currency loans needed to support such interventions. The International Bank for Reconstruction and Development, usually known as the World Bank, and in this chapter for convenience just the Bank, was charged with a different job. This was initially the financing and oversight of Europe's post-war reconstruction. From that role it graduated to become an agency for economic development in countries known variously as underdeveloped, developing, Third World, and the South. Not one of these terms is satisfactory, but everyone knows which are the countries to which the Bank extends its soft loans. Until recently at least, they are all little industrialized, and in terms of per capita income they are all poor.

The tidy demarcation between what is the IMF's business and what is that of the Bank has become greatly blurred over the years. First, the Bank has had to concern itself with exchange rates, specially with overvalued exchange rates. Asked to support a project to produce shoes for export, for instance, it has responded by saying that the project cannot succeed unless the domestic currency is devalued. What this means in effect is that the Bank takes the view that the dollar wages of shoe workers need to be lower if the shoes are to be exported competitively. Just as the Bank has crossed the border into IMF territory by taking positions concerning exchange rates; so the IMF has increasingly demanded extensive packages of economic reform as a condition of loans required to help countries cope with exchange-rate crises. IMF conditionality, as it is known, is a considerable extension of what was always part of its brief.

IMF conditionality has come in for increasing and noisier criticism. The annual meetings of the IMF and the Bank, where these institutions are joined by finance ministers of major countries, and other 'men in suits' have been targeted by anti-globalization demonstrators. A substantial body of writing, ranging from the popular to the academic, blames the IMF for increasing poverty, raising unemployment, harming the environment, destroying female empowerment, and nearly any other bad that can be brought to mind. Apart from the academic literature, evaluations of the IMF ranging from critical to outright hostile have come from *The Meltzer Report* of the US House of Representatives, Friends of the Earth, Oxfam, The Third World Network, and other NGOs. The IMF in its turn has responded actively, posting many notes and papers on its website.

## 14.2 The Washington Consensus

The IMF is only one party to the alleged conspiracy, although a leading weight is attributed to its malign role. The accused collectively go under the name of 'the Washington consensus'. This term is meant to capture the ideas widely accepted by the US and other rich-country governments; the Bank; the IMF, of course; and liberal (in the US sense of that term) economists. The Washington consensus is an oversimplified depiction of views that are frequently more complex and nuanced in detail. Yet for our present purposes a stereotypical description serves well. Like a cartoonist's depiction of a politician's face, what matters is not that it be a complete and faithful portrait; but rather that it should capture something that is present and essential in the character's appearance.

The Washington consensus holds that necessary, and usually sufficient, conditions for growth and development are:

1. Private markets largely free of government interference and distortions, and open to foreign trade at close-to external relative prices.
2. A public sector limited in size, because it does only those things that it can, and does, do well. This implies privatization of government activity that can be done more efficiently and appropriately by private firms. Power-generation, construction or transport projects, and telephone services, are typical examples of activities that are proposed for privatization, and where there are examples of apparently successful privatization experiments.
3. Macroeconomic management that is balanced and disciplined. Some think an independent central bank to be an important requirement here, although this is for the most part a recent fashion, and more time is needed to judge how crucial it may be. In any case, excessive government fiscal deficits, even when not financed by money creation, are certainly to be avoided; as are any policies that lead to rapid inflation.
4. Like trade in goods, capital movements should be as free as possible. Controls on capital movements, and dual exchange rates, are to be avoided. Easy capital movements are seen as a means of financing the borrowing that will be needed to finance the transition to a well-ordered economy, and for the investment that growth will require. The idea is that capital that can leave more freely will enter the country in greater abundance.
5. The external exchange rate should be stabilized at a rate appropriate to the economic condition of the country. A highly overvalued exchange rate, including one protected by capital controls, should be devalued. Hard-currency pegging can be a good route to stability, if the peg is chosen correctly.

Many of the anti-globalization demonstrators at international meetings would reject every part of the above list. But a good many of these would like to abolish capitalism. That experiment has been tried already, in the Soviet Union, in Communist China, and elsewhere, without producing even a modicum of prosperity for the masses, and with massive loss of human lives in more than one case. If the choice were simply between the Washington consensus and the wild dreams of street protesters, the consensus would be the only prudent choice.

Of course the choice of policies is broader than the black-and-white options just depicted. Surprisingly, given the harshness and free-market orientation of the consensus package, much of it is accepted to a great extent by a majority of economists, including Professor Stiglitz. The critical questions then concern, not the long-term objectives of policy design, but the messy how and when of transition and policy reform. Around these seemingly detailed issues the battles are fought.

## 14.3 The le Chatelier Principle and Reform Sequencing

We have already visited some of the problems that have to be worked out in Chapter 5 above. Recall that the section of that chapter concerned with reform sequencing found that for an economy distorted by inappropriate tariffs, freeing capital movements could be harmful. Is that result a challenge to the Washington consensus? The superficial answer says: no. If the entire Washington package had been implemented instantaneously, there would be no inappropriate tariffs, and therefore no possibility of harm. That approach to reform sequencing is sometimes called the 'Big-Bang' approach. Reforms should not be sequenced; they should all be implemented together as soon as possible. IMF structural reform programmes sometimes look Big Bang in style, and therein lie many of the typical problems that they bring with them, and also the basis of some of Stiglitz's criticisms. A Big-Bang approach is rather like getting a variable group of individuals to all walk at exactly the same pace. Often that proves impossible: the fit youngsters race ahead of the old people walking with the aid of sticks. The crucial point is that even when all policy changes can be implemented simultaneously, the effects of those policies do not all arrive at the same time. On the sequencing of structural reform, see McKinnon (1991), Rodrick (1996), Toye (1999), and Williamson (1994).

Moreover the static, one might say the comparative static, consequences of a policy shift are typically different from the short-run transition consequences. Indeed an old idea, the le Chatelier principle, borrowed by Paul Samuelson from thermodynamics, describes the situation. See Samuelson (1948). The principle says that when a given compressive force is applied to a balloon, then the rise in the internal pressure in that balloon will be greater when the temperature of the air inside the balloon is allowed to rise in response to the pressure (partial adjustment), than when the same air temperature is maintained throughout at the level of the external ambient temperature (long-term adjustment). Take as an example the effect on wages of a currency devaluation. If the external exchange rate is overvalued in the initial situation, a labour-intensive export sector may be too small in comparison with its scale in an ideal everything-well-adjusted equilibrium, and its real wage level too high. Following the devaluation, if every value could adjust immediately, the nominal wage would fall, and to a greater extent than the devaluation. This is the magnification effect shown above in Chapter 5. The fall in the real wage is accompanied by an increase in the return to capital in the export sector. Eventually capital in that sector will expand, offsetting some of the increase in its return, and the real wage will recover partially, without returning to its original level. We have here the le Chatelier principle again. The real wage adjusts further when capital inputs are maintained at constant levels, than it does when they are allowed to adjust.

To use this analysis for an evaluation of IMF conditionality, we need to be able to generate welfare evaluations of the different paths that would be entailed by different sequencing of the various adjustments involved. To have a specific case to examine, compare immediate devaluation to the long-run equilibrium level with a gradual decline in the exchange rate over time, with the same final destination. That means a higher real wage on average during the transition period, lower employment, and lower profitability in the export sector. The slower transition is inefficient, in the sense that with lump-sum transfers, it would be possible to make everyone better off, all workers and capitalists, by implementing the immediate adjustment of the exchange rate. That point is not of great interest, because in reality compensations are never lump sum, and even when they occur they are of extremely limited scope. Without compensating payments, the only way to arrive at a welfare evaluation is to use explicit weights to make the gains and losses of different parties comparable between one group and another.

This last conclusion is of the greatest significance. Discussions of macroeconomic adjustment, including IMF programmes, sometimes leave the impression that the issues are technical and bureaucratic; that is, not value-laden and political. This is never the case. In the simple example just discussed, rapid adjustment is good for capitalists, and bad for workers, whether they take wage cuts and stay employed, or hold out for a higher wage and become unemployed. The simple example carries a message that is completely general. Rapid macroeconomic adjustment is like free trade: it is more efficient, but it involves conflict. Without side-payment compensations, a useful fiction for the economics classroom, but unrelated to reality, rapid adjustment tends to favour one group over another; and in many cases the gainer is capital.

Toye (1999) depicts the problem of reform sequencing as inescapably a political problem. This is because political obstructions are rather like large humps in a road. If the programme is not planned and steered skilfully, any one of these can throw the vehicle off the road, and destroy the entire enterprise. Toye also proposes a new and valuable way of viewing reform sequencing. When the success of one reform depends upon the earlier successful completion of another reform, the situation is comparable to the planning of a complex enterprise, such as a large construction project, where sequencing is important. If the roof beams are not in place, the roof cladding cannot be fitted. Critical path analysis is the method for designing these interdependent sequence projects, and Toye's idea is to apply the critical path method to the design of economic reform. On the design of economic reform in India, see Joshi and Little (1997).

Consider another example, highly relevant to real-life structural adjustment programmes. We start with a seriously unbalanced government budget, with expenditure far in excess of income. Let everyone agree, for the sake of this discussion, that the initial position is unsustainable, and that something must be done. It is an implication of simple arithmetic that either government expenditure must fall, or income from taxation must increase. Just looking at the budgetary imbalance, these are alternative approaches, and either will do. The welfare implications of the two methods are very different, however. Drastic cutting of government expenditure hits hard those who have enjoyed its benefits. The precise implications depend on which government expenditures are curtailed. These have typically been social expenditures that target the poor, medical clinics, schools, etc.

To understand why rapid cutbacks in government expenditure on the poor feature frequently in IMF-supervised structural adjustment programmes, it is necessary to appreciate the nature of the relation between the IMF and its clients. Two points deserve emphasis. First the IMF can only advise; it cannot compel. So when adjustment programmes are strongly biased against the poor, this owes quite a lot to the preferences of the country seeking assistance. The second point to note is that the volume of IMF lending today is small in comparison with the total volume of international lending from private capital markets. That does not mean that IMF decisions are unimportant. On the contrary, when they judge the soundness of lending to the country concerned, most other lenders take their cue from the IMF. For that reason, a 'bill of good health' from the IMF, meaning that rapid reform is underway, is a certificate of creditworthiness that opens world capital markets for the country concerned.

## 14.4 Improving Structural Adjustment Programmes

Structural adjustment is somewhat like a weight-reduction exercise. The pace of the programme is one of its most important features. Crash dieting, meaning a massive reduction of energy intake, aiming for a large concentrated burst of weight loss, is painful and usually ineffective. By contrast, extremely slow programmes can be ineffective as well. They are too tolerant of slackness, as weak implementation of the programme makes only a small difference. The temptation to postpone weight loss has to be countered by strict monitoring, and this is difficult with a very slow programme. One might say that it is irrational for the hungry individual today to push the job of losing weight onto the same individual tomorrow. True, but such irrationality is common. And when the individual is a government, the continuity of agency into the future is likely to be absent.

This argument helps to explain why international agencies often prefer rapid adjustment to slow adjustment, even when great costs attach to rapid adjustment. A standard problem with rapid adjustment is that only things that can be adjusted rapidly permit of rapid adjustment. To illustrate this point, take the correction of severe budgetary imbalance discussed in the previous section. We saw that either expenditure cuts, or tax increases, could take care of the problem. Sometimes, however, it is easier to adjust expenditure quickly than to adjust taxation equally rapidly. It may be, for example, that expenditure is at the discretion of the government, whereas tax changes require new legislation, and that may be slow or uncertain. Even given the tax legislation, the extra revenue that will be generated is far from certain.

Whether we examine an advanced industrial country, or a poor developing country, it is seldom the case that the total effect of government activity—both its taxation and its expenditure—is redistributive in favour of the poor. While the rich should pay more taxes (not even this is certain) they typically derive larger benefits from the government. They often make greater use of subsidized infrastructure, such as highways, and they participate far more than the poor in higher education, and similar activities that are provided at far less than their cost. From this argument it does not follow that cutbacks in government expenditure will harm the rich and not the poor. Expenditure must fall far more than taxation to rebalance the budget, and often the cuts in expenditure that are selected are precisely those that are particularly valuable to the poor. This is because the poor are often under-represented in concrete politics. They are poorly organized; they face discrimination in the design of the political system; they cannot pay for expensive campaigns. They are also frequently insecure in their employment situations. So when macroeconomic retrenchment creates large-scale unemployment, as has happened on numerous occasions, it is often the poor who suffer more than others.

Our list of Washington Consensus objectives includes the removal of capital controls and the liberalization of capital movements. There is no more unpopular item in that list, and the rejection of it is a leading part of the case argued against the IMF by Stiglitz. See in this connection Stiglitz (2004). None of our analysis above really engages with the issues involved in their most relevant form. The reason is that our analysis has been static, whereas the most important controversy turns on dynamics. Open capital markets, to put it quite simply, create instability. They lead to countries borrowing 'hot money'; that is, short-term borrowing that finances long-term illiquid projects. When the capital flees, the borrowing country faces a crisis. The IMF may come in and put out the fire; but as with most fire-fighters, they cause extensive water damage. The late James Tobin proposed his famous Tobin tax as a mechanism to counteract the instability inherent in excessive capital mobility. The administrative feasibility of this proposal was always questionable. Even leaving that aside, a trouble with this neat idea is that the level of the tax that would seriously inhibit excess capital mobility would discourage valuable capital movements.

Notice that our discussion of reform sequencing in Chapter 5 does not provide a good model for addressing the questions now under examination. That model says that free capital movement can be welfare-reducing in an economy distorted by tariff protection. That is purely static analysis: it says nothing about instability. What would a better dynamic analysis conclude? Really the basic points are quite elementary. The IMF must be wrong when it encourages countries to free up capital movements, without requiring of them a correct and conservative programme of debt management; one that pays particular attention to the term structure of loans in comparison with term structure of income flows.

Anyone told to finance the purchase of a house, by borrowing from a bank at an attractively low interest rate, when the bank could call in the loan at any time, would be exceedingly badly advised. A loss of confidence in the housing market could lead the bank to call in the loan, and leave the borrower to find whatever funds he could, at whatever price, or lose his home. The analogy to the foreign borrowing of developing countries is pertinent. On occasion this has been imprudent, both with regards to level and structure. As with a householder, if borrowing has been excessive, perhaps to finance an unsustainable budget deficit, then painful retrenchment is unavoidable. In these circumstances it is tempting, but seriously wrong, to say that the capital market is not to blame when borrowers abuse it. A loan transaction always has two sides: borrower and lender, and both should behave responsibly. When a firearm is used to kill an innocent party, we rightly ask why the gun-smith felt it appropriate to sell the criminal the gun. The IMF and private lenders' clubs have often attempted to make poor-country borrowers bear all the cost of imprudent borrowing. This is unjust, surely. But it also reflects an inefficient design of the principal-agent problem embodied in an imperfect-information, incentive-driven, borrowing arrangement.

What applies to the level of borrowing applies equally to the term structure of loans. Too often governments and private borrowers have chosen to finance their cash needs with 'hot money'; the short-dated liquid capital of the house-finance example above. With anxiety about a nation's prospects, hot money in terms of the annual interest rate demanded, is often cheaper money. But it comes at a price, because the capital market is then a revolving door, and what comes in can easily leave. An additional problem takes the form of imprudent exchange-rate speculation. Here is a typical story. A country pegs its currency to a hard alternative, say the US dollar. Then the US dollar becomes seriously overvalued, perhaps because of US fiscal policy. The home country can devalue, but only with a large loss of credibility. Instead it hangs in with the existing peg, arguing that the dollar must depreciate soon. Meanwhile it borrows short-term dollars, believing that its loans can be refinanced cheaply after the dollar has fallen in value. The dollar does not fall as it should, confidence in the home country ebbs away, and the whole house of cards comes crashing down. The Asian crisis of 1997 fits this description to some extent.

Where a country rapidly removes strong capital controls, it will often be the case that capital will be locked inside the country, capital that would prefer to be elsewhere, for whatever reason. It follows that the removal of capital controls is sometimes followed by domestic capital flight. If domestic capital and foreign capital were to be perfect substitutes, this would be no problem. The national portfolio would be rebalanced, with improved diversification, and the cost of capital would not increase. That is improbable. On account of insider information, domestic capital can be supplied more cheaply, because more efficiently. Less than optimal diversification is the cost of the gain in cheapness. Sometimes capital flight avoids domestic taxation, but that is only a problem when foreign capital cannot be taxed.

## 14.5 The Control of Inflation

As part of the discussion above we saw that the pace of reform is a critical question. In many cases a slower implementation of the measures that are essential has distinctly different consequences from a more rapid application of the same changes, and in several cases of interest the slower pace is more benign for the poor. The same argument noted that the slower pace is not always feasible, or at least that it is not without problems. Nowhere is it more apparent that a slow adaptation may not be feasible than when the control of rapid inflation is the objective of policy. Over the last few decades several countries have moved from rapid, or near-hyper, inflation to moderate and controlled inflation. In not one case known to the author has that transition taken the form of a slow dignified movement from high to low inflation, not accompanied by serious disruption of production, and a large increase in unemployment.

Some of the rapid transitions from high to low inflation have been the result of, or at least have been associated with, IMF interventions and adjustment programmes. The countries on the receiving end of those interventions have often been advised to tighten monetary policy sharply, forcing up interest rates, protecting the exchange rate, and inevitably causing bankruptcies and job losses. There can be no question that the disruptive effects of those shock interventions are dreadful to experience, and awful to observe. They are the target of some of Joseph Stiglitz's strongest invective, directed at the IMF and its staff.

That such policy interventions are unattractive and difficult to defend is beyond doubt. Yet the critical question must be: what is the alternative? Stiglitz advocates subtle carefully nuanced slow-acting policies, each one precisely tailored to the particular situation of the country that needs to adjust. Nothing must be brutal and overdrastic, nothing ever standard and formulaic. It is a necessary truth that, in an ideal form, such delicate and customized policy interventions will be better than harsh and standardized adjustments. In life, however, we often have simple standardized rules for the good reason that they work better than more complicated regulation. Simple tax systems can be preferable to delicately designed complex systems. The latter produce huge problems when they are exploited by skilful tax avoiders, and generate distortions as a consequence. It is distinctly possible that the IMF has overdone standardization in its interventions, but some degree of standardization is unavoidable.

An excessive emphasis on the harm that structural adjustment programmes do to the poor and vulnerable, takes attention away from the fact that severe macroeconomic imbalance, given that it must eventually be redressed somehow, is itself severely costly to the poor. This is clearest where the control of rampant inflation is under consideration, but is not confined to that problem. Restrained and moderate macroeconomic policies, in the spirit of the Washington consensus, are good for economic growth, good for high employment, and essential for effective health and education programmes. Governments that choose not to follow such virtuous principles, must share responsibility for the subsequent burdens that fall upon their poor citizens.

## 14.6 Globalization and Inequality

Does free trade lower or raise inequality? What appears to be a reasonable question is in fact no such thing. First, it is extremely improbable that the answer to that enquiry could ever be an unqualified yes, or no. A worldwide movement towards more open trade, and one proceeding at a rapid pace, will certainly benefit many and harm a significant number. All major historical changes do the same thing. Win-win situations are the product of fertile imaginations, whether in the Bank or elsewhere; they are hardly ever observed in reality. Once that is conceded, the question of implications for inequality becomes sensitive to how inequality is measured. See Theil and Seale (1994).

It is no paradox to one schooled in the algebra of measurement that income inequality may increase in every country in the world, while world inequality declines. All that is needed for that result is that incomes for a significant weight of individuals in relatively poor countries should be growing more rapidly than world incomes in general. And that is exactly what has been happening in the last twenty years. The rapid growth of China and India in particular, even though it has been associated with increased inequality within those countries, has greatly increased the incomes of many millions of people who were initially poor by world standards. Using standard measures of world inequality, such as the Gini coefficient, these changes register as a reduction in world inequality.

One might say that a reduction in world inequality is poor comfort to someone starving in Chad. That is a highly relevant observation for two reasons. Firstly, it reminds us that aggregate measures of inequality are just that, aggregates. They cannot by their nature represent all the detailed features that may be going on within the hugely complex developments for individual incomes over the billions of individuals who make up the world's population. Secondly, the reference to Chad reminds us that China and India, despite their evident poverty, are not the world's poorest countries in per capita terms. The very poorest countries are to be found particularly in sub-Saharan Africa, our SSA region again. In many of those countries per capita incomes have been stagnant at best, often falling. Probably inequality of the falling income levels has been increasing in many of those countries as well, although concrete evidence on such detail is lacking. It may be that in the tragic situation that has afflicted Zimbabwe, income inequality has declined, as the economic condition of the entire population takes a nosedive. In any case, where world income distribution is concerned, what happens in the SSA region is relatively unimportant. Only about 10 per cent of the world's population lives in SSA, so that developments there are usually not given sufficient weight. That comment applies to the Gini coefficient. One can design measures of inequality that give any weight desired to the situation of the poorest, but these would have to be powerfully directed to the low incomes of the poorest to make the failure of SSA countries register on a world inequality index.

Setting aside inequality indices particularly weighted towards the very poorest, it seems that trade has been good for equality. It has proved to be a powerful instrument for raising incomes rapidly in some highly populated poor countries. And if we focus on the world's poorest countries, in particular the many in SSA, then their economic stagnation cannot be explained to any great extent by their having opened their markets to freer trade. As we saw in Chapter 7, the limited participation of the SSA countries in international trade, especially when primary products are excluded, is a most striking feature. The leading problem for the SSA countries, and the same can be said of the Arab world, is not that they went to the globalized trade party, and drank too much. The chief problem is that they did not attend that party, and failed to enjoy the benefits that other developing countries have derived from it.

There remains the question of whether more open trade, particularly in labour-intensive manufactures, and in other labour-intensive activities, such as call centres, has greatly increased inequality in rich developed countries. There is clear evidence that income inequality has risen over time in most industrial countries. See Glyn (2004). This trend is variable to a surprising degree. It is most marked in the USA and in Britain, less apparent in mainland Europe.

In theory, meaning by that term that precise theoretical economic models can demonstrate the result, easier international trade could explain the observation. In detail, however, the trade-theoretic model is unconvincing. It does not do well in replicating the timing of the change in inequality. That has been proceeding for twenty-five years at least, while the large upsurge in manufacturing imports into the US is concentrated in a shorter period towards the end of that quarter-century. Also, there is another factor to be taken into account. Technical change in the US, and elsewhere, has been strongly labour-saving, meaning here unskilled-labour-saving. Motor vehicles are manufactured using robots, cash is dispensed by ATM machines, effecting payment for purchases is rapid and easy with the help of bar-codes and electronics, etc. These changes have demanded that the labour force increases its skills and education. Those who have failed to do that, who offer the labour market 'a pair of hands', have suffered a declining standard of living, or unemployment, depending in part upon the institutions of the country in which they live. So the provision of indefinite unemployment benefit, as in France, leads to much long-term unemployment. The absence of that benefit in the US produces the working poor; as well as drug-dealing, crime, and consequent imprisonment, as a kind of alternative social security system.

A simple neoclassical model may help to throw light on apparent trends in the world income distribution. The idea is that many of the world's poorest have enjoyed increases in real incomes recently, particularly those in countries that have participated in growing world trade. Equally the rich have been doing well, with huge increases in company profitability. In contrast to the above, the middle of the income distribution has been having a tough time. In the USA real hourly wages have hardly increased. The rise in real household incomes is accounted for by increased participation: longer hours and greatly increased female participation in the labour market.

To model this crudely, imagine a huge overarching world production function that unites world capital with world labour. The factor inputs only include those of countries that join in the globalized world economy. Now China and India particularly decide to enter the world market. They bring to the world production function some capital, but far more labour. The wage rate falls and the rate of profit rises. For those workers already in the world economy this development brings a loss. For owners of capital the same development brings gains. The poorest, however, are represented by the labour of China and India. They gain, because even the reduced wages provided by the world economy are higher than the wages they used to earn.

This model is crudely oversimplified, no doubt. There is no overarching world production function, if only because factors are not allocated perfectly for worldwide efficiency. Yet the model may capture some crucial features of the recent increase in participation in world exchange. How does this essentially one-sector model relate to the multisector trade models of Chapters 5 and 6 above? The differences are not profound. Take a basic HOS model with just capital and one type of labour. With no trade, the wage rate is low in the capital-poor South, and high in the capital-rich North. On average the return to capital is low, as most of it is located in the labour-scarce North. Now free up world trade completely. Wages rise in the South and fall in the North. The return to the bulk of capital rises. It is not greatly different from the one-sector picture, and not so greatly different from reality.

## 14.7 Trends in World Inequality

World inequality is not a value that can be measured in any simple direct sense. At best it can be estimated, and what is estimated is the level of some index of inequality. To the extent that there are several possible indices that might be used, that index will be arbitrary. And even if there is complete consensus as to what index is appropriate, its measurement is certain to be rough and approximate. The data available is seriously incomplete, and even when the numbers required can be obtained, they may be inaccurate.

In addition to these considerable difficulties, the whole exercise throws up technical and philosophical questions that are not easily answered, Why is inequality an issue? Even if it is clear that there is a problem when a community suffers from inequality, what defines a community? Is the aggregate of all humanity a community? The spirit of modern thinking, and of the UN Charter, says yes. But precisely what inequality is deleterious to the well-being of the aggregate of all humanity? Above we have mentioned the growing inequality of earnings in the USA. Is that a cause for any concern whatsoever for the vast mass of humanity that are far poorer than even the worst-off in the US? That increase in US inequality registers on all standard indices of inequality. The Atkinson index (see below) is most flexible when it comes to giving little importance to inequality between the rich.

Other questions are conceptual and do not admit of purely technical answers. For instance, is inequality about uneven consumption of privately purchased market goods, or does it include also public services and collective goods? Some crude measures of global inequality reduce it solely to the inequality between nations attributable to variations in national real income per capita. Not all national income is consumed by means of private purchases. Therefore such a measure shows no inequality between the residents of two nations, identical in all respects, including national income per head, except that one nation spends hugely on defence, where the other does not. Should we say that one of these nations happens to take out the benefit of its income to a far greater extent in security, while the other eats better, and this is not a matter that need concern the measurement of inequality?

Another issue finds a majority of economists on one side, but does not command agreement from all commentators. Should incomes be measured in terms of purchasing-power, or should they include a partial, or in an extreme case a complete, weighting for value at official exchange rates. Large differences rest on this distinction. Generally speaking, price levels converted to hard currencies, such as the US dollar, are low in poor countries. There are various reasons for this. One is the Balassa-Samuelson effect. Non-traded goods, especially labour-intensive goods and services, are cheap in poor countries, simply because dollar wages are low in poor countries. These cost differences affect even apparently tradeable goods.

A kilogramme of rice costs far less in dollars in Bangladesh than it would in the US. This is because one never buys strictly just rice: it always comes packaged with the transport and retailing services that bring it to the store or to the market.

Income inequality across the world is far larger when income is measured in US dollars than when income is measured in terms of purchasing-power. Since we are interested in inequalities in economic welfare, it seems natural to make comparisons in terms of what people can purchase with their incomes; which is why economists prefer it. Purchasing-power, however, involves deflating nominal income by a price index that measures what a basket of goods will cost the consumer. That exercise comes inescapably up against the classic index-number problem. If we know exactly what consumers in Bangladesh buy, we can cost it in dollars in that country, and cost the same basket in the US. Suppose that the basket costs five times as much in the US. Then we can multiply Bangladesh incomes by five to obtain a purchasing-power comparison of incomes in the two countries. We could have done an exactly parallel exercise based on the US consumer's basket. Suppose that basket costs half as much in Bangladesh as it does in the US. That would not be surprising, because some things consumed in the US are little consumed in Bangladesh, partly because the latter country is poor, but also because those items are particularly expensive in Bangladesh.

Now should we multiply Bangladesh dollar incomes by five, or by two; or even by three and one half, averaging the two estimates? There is no right answer to this question. As we are particularly interested in the welfare of the poor, it may seem right to base the measure of real income on what the poor consume, not on what is consumed in some other country. We have to use prices in other countries, however, to make any comparison, and to suppose that US consumers are buying mainly rice, even at a considerably higher price than the Bangladesh consumers, is to overestimate the welfare of US consumers, who in fact buy heating oil, medical care, and expensive housing.

The ideal method for addressing the index-number issue when computing measures of purchasing-power is to use a chain index approach, similar to that which is employed when computing cost-of-living indices through time. For countries the technique is not as straightforward as when it is applied to time series. Countries can be arranged along a line according to dollar income per capita, or using some other measure. Yet while time nearly always brings continuity, because huge immediate changes are unusual, dollar-income contiguity, or geographical proximity, do not entail closeness in consumption patterns. Some imprecision in estimates of purchasing-power can be taken for granted, but it is notable that even apparently small differences in the estimates can have large implications for the measurement of inequality. Estimates of China's per capita income in purchasing-power terms range from one-tenth of the US level to one-eighth. Why worry? China is clearly extremely poor in comparison with the USA. For an estimate of global inequality, say one using the Gini coefficient, the implied difference is large.

Finally, the measurement of inequality can be extended beyond income or direct consumption, to include life expectancy, infant mortality, health, education, etc. Sutcliffe (2004) provides a good account of the possibilities. One option is to look at inequality using the UN Human Development Index (HDI). Fortunately, the pictures that emerge from these various techniques are fairly in accord. SSA is always revealed as the world's poorest and most worrying region. The rich countries are rich, however measured. The economic miracle in China during the last two decades is confirmed by the various measures.

## 14.8 What the Evidence Shows

The evidence available is extensive and various. It covers different collections of countries, over different time periods, and using different measures of inequality. Sala-i-Martin (2002) and Sutcliffe (2004) provide broad useful surveys. The longer the time period required, the smaller will be the number of countries included. Angus Madison (see Madison (2003)) offers the longest and most solid time series, and other researchers have often drawn on his data. See also Berry, Bourguignon and Morrisson (1983), Chen and Ravallion (2001).

A critical question is whether inequality is to be measured between countries, or between what purports to be individuals. Observing the development over time of the per capita incomes of countries is equivalent to looking at the economic convergence that is the subject of Chapter 4. The difference between inequality among countries and across individuals is well illustrated by the case of China. In a cross-country comparison, China is just one observation. For inequality between individuals it provides over one billion people, and their incomes vary significantly. As China has been growing rapidly for the last two decades, the weight that its population carries will make a large difference.

An obvious point is that useful data for the income of every individual in the world is not even remotely available. Researchers usually work from estimates of shares of incomes for quintiles of the population, attributing to all individuals within the same quintile the identical income level. It is pretty rough and ready, but as with many a statistical exercise, really strong features will show through, and the rest is unimportant. If everyone agrees that it is raining, then it is raining. If opinions differ, who knows? But certainly there is not a downpour. On different concepts of inequality, see Ravallion (2004).

Generally speaking, as noted above, there is considerable consensus among different measures and measurers concerning the last two decades of the twentieth century. Bourguignon and Morrisson (2002) find no increase in inequality but little reduction. Sala-i-Martin (2002) reports great reductions in inequality using several measures. All agree that China has played a huge role in whatever has happened. All confirm that the SSA region has done badly, and that its low share of world population explains why inequality may have fallen despite poor African performance.

Milanovich (2002) and (2004) provides a distinctly different view by making use exclusively of data from household consumption surveys. That avoids the imputation of incomes to individuals based on averages for groups. On the other hand, it makes household consumption the sole measure of welfare. Consider China yet again. All agree that it has been growing at an exceptionally rapid rate in recent years. Equally it is clear that the Chinese save, even if only private saving is accounted, at an unusually high rate. That means that the household consumption of a hugely important slice of world humanity significantly underestimates its income. If only current consumption generates Chinese welfare, why do the Chinese save? The low-saving residents of the USA are the mirror image of the Chinese. Arguably their high consumption levels significantly overestimate their true welfare. It is plain that the conceptual and philosophical questions raised by the Milanovich approach are quite weighty.

## 14.9 Inequality among the Rich and the Atkinson Index

Many measures of inequality are mechanical in the sense that they propose an index of the dispersion of incomes, when any dispersion inevitably contributes to the index. That is the case with the Gini coefficient, as it is with the variance of log incomes. We may prefer a measure that allows the insertion of value judgements concerning what inequalities we care about, and how much we care. A measure of inequality that achieves that feature is the Atkinson index, Atkinson (1970).

Let $Y_i$ be the income of those in the $i$th income range, and $\theta_i$ be the proportion of the population in the $i$th range. There are $n$ ranges altogether. The mean income is $\overline{Y}$. The Atkinson measure of inequality is:

$$1 - \left[ \sum_{i=1}^{n} \theta_i \left( \frac{Y_i}{\overline{\overline{Y}}} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} \tag{1}$$

See Atkinson (1970, and 1975: 48–9). The index has a simple economic interpretation. It measures how far total income could be reduced, while holding the value:

$$\left[ \sum_{i=1}^{n} \theta_i \left( \frac{Y_i}{\overline{\overline{Y}}} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} \tag{2}$$

constant, if the income were distributed equally across all members of the population. An unequal distribution ‘wastes’ income by giving it to richer individuals, when its marginal valuation is higher among the poor. If the Atkinson index is 0.8, it would only need 80 per cent of the original total income to give an equal valuation to the income distribution, when the new distribution becomes one of complete equality.

Inspection of the function (2) shows that it belongs to the constant-elasticity-of-substitution family. From that comes a feature of the valuation that deserves emphasis. Suppose that the valuation, and hence the Atkinson index, is neutral with regard to the following transaction. From $m$ individuals earning \$100,000 per annum we take \$1,000, in return for which we are able to give \$3,000 per annum to $m$ individuals earning \$150,000 per annum. If we are indifferent concerning that transaction, then we must be indifferent concerning the following. From $m$ individuals earning \$400 per annum we take \$4, in return for which we are able to give \$12 per annum to $m$ individuals earning \$600 per annum.

The two cases can be perceived as very different. People earning \$100,000 per annum are not rich, but they are comfortably off. So reducing their incomes by 1 per cent is not a terrible harm. True, if that makes the same number of even richer individuals 2 per cent better off, we might not get very excited. In a more outlandish example, if Bill Gates gives Rupert Murdoch one million dollars, should that transfer register at all as a reduction in world inequality?

By contrast, however, people earning only \$400 per annum are seriously poor. They have only a little more than the \$1 per day level that is widely used as a definition of absolute poverty. To reduce the incomes of these people by 1 per cent is no small thing; it is a disaster. It is true that the transfer by assumption is giving an extra \$12 per annum to people who earn less than \$2 per day. One could argue the case for the acceptability of either transaction at length. What seems hard to accept, however, is that the constant-elasticity-of-substitution specification should be allowed to settle the matter. Transactions examined above do turn on the elasticity of substitution. If it is high, we are willing to accept large transfers between different income earners, if the terms on offer are sufficiently favourable. If it is low the opposite is true.

In Chapter 3 above we saw that assuming the elasticity of intertemporal substitution (the EIS) to be a constant is a seriously limiting assumption. For example, a varying EIS can yield infinitely many steady states in the Diamond model. There the argument took the form of proposing that the EIS might increase with wealth. The same concept can be applied to the Atkinson index.[1] We can allow the ease with which transfers from poorer to richer can be accepted, or the reduction in a higher income that offsets an increase in a lower income, to vary with the levels of income at which the said transfers are implemented. As with the EIS, the idea will be to make such transfers easier between the rich than between the poor.

The variable elasticity-of-substitution Atkinson index (the extended Atkinson index) awaits empirical implementation. A problem will be that it offers a great increase in flexibility for the measure of inequality. We saw in Chapter 3 how a function can be tailored to produce a particular outcome; in that case a continuum of Diamond-model solutions. With the extended Atkinson index that might be less of a problem, as it is not proposed to tailor the function to obtain a particular result. Assuming an increasing elasticity with income, the measure would only show in summary fashion what an inspection of income shares by deciles already reveals roughly; namely how much inequality is divided between the various levels of income.

Sala-i-Martin (2002) reports trends in an Atkinson index for the world population showing declining inequality, for the same reason as his other measures, importantly the explosion of economic growth in China.

[1] I owe the observation that my work on the EIS can be applied to the measurement of inequality to Anthony Atkinson.

## 14.10 Trade and Inequality

Two widespread presumptions are both incorrect. One holds that increased globalization over the last two decades has been accompanied by a great increase in inequality in the world. We have seen that the evidence does not support this opinion. The second idea is that the increase in inequality that is supposed to have occurred is a result of the rise in international trade, particularly the increased involvement in trade of countries that previously had little involvement. This view sits oddly with the observation that those initially poor countries that have been growing rapidly, notably China, have been doing so because of their increasing involvement in international trade. On the other hand the countries that have been stagnating over the same period, or even going backwards, as in parts of the SSA region, have extremely low involvement in international export markets. This is particularly the case if primary-product exports are excluded from consideration. Primary exports are important, and they have been growing in some cases. This is not a recent development, however, and it has not saved the region from the lowest growth of per capita incomes of any region in the world. Of course globalization means more than just non-primary trade. Capital markets have not been good for Africa. African countries have allowed heavy borrowing of funds that have been wasted at best, and corruptly misappropriated at worst.

If the world has been becoming somewhat less unequal, it is not thanks to SSA, or the Arab world. Indeed Sala-i-Martin (2002) argues convincingly that SSA may eventually be the cause of a reversal of the trend towards equality. If large poor countries, China and India, converge towards OECD levels of per capita income, they will cease to be motors of declining inequality. In that case what then happens to inequality will depend upon the performance of the remaining poor countries. If that is largely SSA, with a higher population weight by that time, and continuing negative growth rates, then world inequality will be rising. If that happens it will not be because of globalization and free trade.

Because trade lends itself well to economic analysis, it is tempting for trade theorists to give to trade the leading role in explaining trends in equality and inequality across the world. That approach is particularly appealing to anti-globalization lobbyists, because it makes capitalism in its globalized form the accused, required to defend itself. Our world is more complex and disordered than that story allows. Take SSA yet again. Its trade involvement has been low, and what involvement it has experienced has been heavily biased towards primary product exports, with the numerous problems that this implies. Yet top of a list of explanations for Africa's acute problems would be AIDS, corruption, and war. That list is in alphabetical order, as there is no need here to decide how these afflictions rank in importance; and the answer in any case would vary from country to country.

# 15

# Conclusions and Unresolved Issues

Gossen's law says that the bad money drives out the good. This means that if silver and gold coinage circulate together, and the metallic value of one, say silver, is lower, then gold coinage will disappear from circulation. No-one owning a gold coin will spend it for less than its underlying value. Goldsmiths will be offering more than face value in terms of silver to buy gold coins, to melt them down. There should be a name for a Gossen-like law of arguments. The good arguments are subtle and solid, and they need hard work to appreciate. The bad arguments are simplistic and slogan-like, but they satisfy the lazy or preoccupied listener. The bad arguments drive out the good.

Too often where trade and economic policy is concerned, the bad arguments win. Globalization harms the poor. Free trade requires a level playing field. The WTO is undemocratic. Allowing developing countries to protect their home markets will foster their development. Each of these propositions has an immediate appeal, and each would require a seminar to evaluate and qualify it. Yet all are hugely popular with people who will never accept an argument containing the phrase: on the one hand, on the other hand.

Looking at this overweight volume, an impatient person might ask to be told, in just a few words, what explains inequality; and for nations, what trade has to do with it. Such an individual is unlikely to be satisfied with the answer that it mostly depends on history and culture, in enormously complicated ways, and that trade, while it is associated with prosperity, is as much an effect as a cause. Critics of economics sometimes claim that one can build an economic model to demonstrate anything. This is itself a half truth. Within the ground rules of economic modelling which the profession in general accepts, there are conclusions that cannot be derived. The scope for 'valid' results, however, is extremely broad, and that is why the discipline demands judgement as well as technique. Sometimes empirical investigation gives a more definite picture than does abstract modelling by itself. We saw that in Chapter 14. In theory globalization might improve equality at a world level, or might make it much worse. In fact, in recent years, the latter has not occurred. The data shows clearly why that is. As is often the case, the answer questions the question. Once it is understood that world inequality measures are dominated by a few large countries, and are sensitive to how inequality is defined, it will be seen that global inequality is a less pressing concern than the condition of the poorest countries that are failing to share in worldwide growth.

Given the extensive literature, and the numerous empirical investigations, it might seem that the question of what causes rapid economic growth is more or less answered. That is far from the case, but it is surely true that cross-country regressions have provided about as much insight as they ever will. Their limitations have been examined, and will not be repeated here. What research methods can be employed in their place? It is possible that the entire project, the search for a philosopher's stone of world economic development, is misconceived. It may be that there is more than one way of doing economic development well. The Chinese economic miracle, recently inaugurated, seems to indicate that this may be so, but it is far too early to reach any definite conclusion.

The field-of-growth theory itself has been thrown into uncertainty by the innovations associated with the idea of endogenous growth. Impressive theoretical models have been developed, but empirical studies by Mankiw, Romer, and Weil (1992) and Jones (1995) call into question the importance of the root idea. It is embarrassing to have to admit that economists are still far from clear about why some countries grow rapidly, where others fail to do the same.

One theme that recurs many times in the previous chapters is that economic theory can often throw light on what is happening in the contemporary world, but that it frequently happens that the most simple models prove to be inadequate. Our world is complex to an extent that we can barely think about. We are forced to express our thoughts in terms of categories—labour, trade, inequality, etc.—that only weakly represent the world. And our questions often impose a straightjacket on the answer. Is globalization good for the poor? Which globalization? Which poor, where, and how? One strong suggestion that emerges from the foregoing arguments is that globalization and trade are not the right things to be looking at if we want to understand the condition of the poorest, and the slowest growing. To huge extent these are to be found in sub-Saharan

Africa, a region that trades to a low extent, and mostly in the wrong things from the point of view of poverty reduction and economic growth.

This volume has promoted the concept of an economic environment, and without any quantitative measures of its importance, the reader will be left in no doubt that the author believes that this feature may be of the greatest importance in explaining relative economic performances. The idea of an economic environment is similar to North's concept of institutions. While the latter idea has rightly enjoyed great influence, it suffers from a lack of definiteness. In particular it is not, to borrow some currently fashionable terminology, well micro-founded. In other words we need more than the correlation between certain institutions and economic performance. We need to know precisely why this form of government, or that religion, inhibits the natural tendency of human beings to found businesses and accumulate wealth. This volume has certainly not plugged that gap. It has at best pointed the way. So that is our first field for future research: a better institutionalism is needed. Similar in certain respects to institutions is the concept of Social Capital; see Fine and Green (2000). This is mainly a different variety of institutionalism, bringing in such ideas as investment in building social networks and establishing relationships involving trust. All this is likely to prove important for future research. In order to deliver, these concepts need to be translated into concrete empirical predictions that can be tested. Big ideas must lead on to little experiments, and they often do that; so there is every reason to expect that institutions, or economic environments; the precise term matters little, will have much to contribute to our understanding of economic development in the future.

A deeper understanding of economic environments should help with one of the leading puzzles of world economics. Many models of classic trade theory, and not a few discussions of international economic relations, assume that technology, or production functions as it is sometimes called, are common knowledge. In fact, productivity, even when it is measured as total factor productivity, varies hugely across the world. Hall and Jones (1999) examines the observed variations in labour productivity and conclude that they cannot be explained by other inputs, such as capital. Does the explanation for the large variation in productivity lie in the quality of inputs? Or is it to be found in different production functions? Is there a clear distinction between these two last possibilities? This is not an easy field in which to undertake research. The concepts can prove to be slippery, and the data that would ideally be required is largely unavailable. Yet the importance of the problem is undeniable.

In the proceeding pages we have touched several times on the design of reformist policy changes. This type of question was involved in the discussions of reform sequencing, and it is central to Joseph Stiglitz's confrontation with the IMF. The question awaits a clear and useful answer. It is plain that simple formulaic rules will always fall short of the ideal. On the other hand, ultra-subtle interventions requiring massive quantities of information, and blind to the constraints of incentive compatabilty, will not be the answer. The debate concerning economic adjustment soon migrates to an argument about the reform of the international economy's great pillars, the IMF and the World Bank. Some have argued that the IMF should return to an earlier role as a short-term emergency lender, and leave long-term adjustment problems to the World Bank. That line of reasoning involves defining the roles of the two pillars on a distinction without a difference. Large-scale short-term loans, with no collateral, always and everywhere involve the assessment of longer-term viability, and this is recognized in the IMF charter. Neither does some sniping at critics of the IMF serve to determine how international economic governance should be organized, nor does it imply that deep reform of these pillars is not needed. The truth is that the entire structure was created for a world radically different from that which now exists. For this reason reform might be of great value, and its design is another large unanswered question. And that question dissolves into two parts: what do we want; and can we get there? Any reform that does not involve a new treaty can only be of limited effectiveness. However negotiating a new treaty is likely to be extremely difficult if not impossible. The current troubles of international trade negotiations support this view. The creation of the WTO shows that the seemingly impossible is sometimes feasible. In that case, however, US leadership, and the cooperative line taken by other major economic powers, were critical for the success of the project. Already the world situation is bleaker. It may be necessary to explore the internal reforms that the IMF and the Bank can implement within the existing treaty structure. If so, what those reforms should be constitutes a large outstanding research area. On governance reform see Wolf (2004).

In writing this book I have frequently found myself thinking: give me five or ten more years, and everything will become clear and more secure. Of course I understand that this is an illusion. As time moves forward, the past comes into sharp focus, but the future remains cloudy. There will never be a last and final book on Trade, Growth, and Inequality. What you have here is at best a partial report on the current state of play.

# References


Abrego, Lisandro and Whalley, John (2005), 'Decompositional Analysis Using Numerical Equilibrium Models', in Timothy Kehoe, T. N. Srinivasan, and John Whalley (eds.), *Frontiers in Applied Equilibrium Modelling: In Honor of Herbert Scarf* (Cambridge: Cambridge University Press).

Acemoglu, Daron (1995), 'Reward Structures and the Allocation of Talent', *European Economic Review*, 39: 17–33.

—— and Robinson, James A. (2005), *Economic Origins of Dictatorship and Democracy* (Cambridge: Cambridge University Press).

Aghion, Phillipe and Howitt, Peter (1998), *Endogenous Growth Theory* (Cambridge, MA: MIT Press).

Aidt, T. S. (2003), 'Economic Analysis of Corruption: A Survey', *Economic Journal*, 113: 632–52.

Akerlof, George (1970), 'The Market for "Lemons": Quality Uncertainty and the Market Mechanism', *Quarterly Journal of Economics*, (August).

Armington, P. (1969), 'A Theory of Demand for Products Distinguished by Place of Production', *IMF Staff Papers*, 16: 159–76.

Arrow, Kenneth J. (1962), 'The Economic Consequences of Learning by Doing', *Review of Economic Studies*, 29/3: 155–73.

—— Solow, Robert, Chenery, Hollis, and Minhas, Bagish (1961), 'Capital-Labor Substitution and Economic Efficiency', *Review of Economics and Statistics*.

Atkinson, Anthony B. (1970), 'On the Measurement of Inequality', *Journal of Economic Theory*, 2: 244–63.

—— (1975), *The Economics of Inequality* (Oxford: Clarendon Press).

—— Bourguignon, F., and Morrisson, F. (1992), *Empirical Studies of Earnings Mobility*, DELTA Discussion Paper.

—— Rainwater, L., and Smeeding, T. (1995), *Income Distribution in OECD Countries* (Paris: OECD).

Auty, R. M. (1990), *Resource-Based Industrialization: Sowing the Oil in Eight Developing Countries* (New York: Oxford University Press).

—— (2001), 'The Political Economy of Resource-Driven Growth', *European Economic Review*, 45/4–6: 839–46.

Azariadis, Costas (1996), 'The Economics of Poverty Traps: Part One', *Journal of Economic Growth*, 449–86.

Balassa, Bela (1967), 'Trade Creation and Trade Diversion in the European Common Market', *Economic Journal*, 77/305 (March): 1–21.

Baldwin, Richard, Forslid, Rikard, Martin, Phillipe, Ottaviano, Gianmarco (2003), 'Incremental Trade and Endogenous Growth: A q-Theory Approach', NBER Working Paper 6477.

Bardhan, Pranab (1997), 'Corruption and Development: A Review of Issues', *Journal of Economic Literature*, 35/3: 1320–46.

Barham, V., Boadway, R., Marchand, M., and Petsieau, P. (1995), 'Education and the Poverty Trap', *European Economic Review*, 39/7: 1257–75.

Barro, Robert J. (1991), 'Economic Growth in a Cross-section of Countries', *Quarterly Journal of Economics* (May): 407–43.

—— (1997), *Determinants of Economic Growth* (Cambridge, MA: MIT Press).

—— Mankiw, J. G., and Sala-i-Martin, X. (1995), 'Capital Mobility and Economic Growth', *American Economic Review*, 85/1 (March): 103–15.

—— and McCleary, Rachel M. (2003), 'Religion and Economic Growth', Harvard University.

—— and Sala-i-Martin, X. (1992), 'Convergence', *Journal of Political Economy*, 100: 223–51.

—— —— (1995), *Economic Growth* (Boston, MA: McGraw-Hill).

Baumol, William J. (1986), 'Productivity, Growth, Convergence and Welfare: What do the Long-run Data Show?', *American Economic Review*, 76/5: 1072–85.

Becker, G. S. (1971), *The Economics of Discrimination*, 2nd edn. (Chicago, IL: University of Chicago Press).

Bell, Clive (1988), 'Credit Markets and Interlinked Transactions', ch. 16 in Hollis Chenery and T. N. Srinivasan (eds.), *Handbook of Development Economics*, Vol. 1 (Amsterdam: North-Holland).

Ben-David, D. and Loewy, M. B. (2000), 'Knowledge Dissemination, Capital Accumulation, Trade, and Endogenous Growth', *Oxford Economic Papers*, 52/4: 637–50.

Berry, A., Bourguignon, F., and Morrisson, C. (1983), 'Changes in the World Distribution of Income Between 1950 and 1977', *The Economic Journal* (June): 331–50.

Bevan, David, Collier, Paul, and Gunning, J. W. (1992), 'Consequences of External Shocks in African-Type Economies', in C. Milner and J. A. J. Rayner (eds.), *Policy Adjustment in Africa* (London: Macmillan).

Bhagwati, Jagdish N. (1958), 'Immiserizing Growth: A Geometrical Note', *Review of Economic Studies*, 25: 201–5.

—— (1982), 'Directly Unproductive, Profit-Seeking (DUP) Activities', *Journal of Political Economy*, 90/3: 988–1002.

—— (2002), *Free Trade Today* (Princeton, NJ: Princeton University Press).

—— Brecher, R. A., and Srinivasan, T. N. (1984), 'DUP Activities and Economic Theory', in D. C. Colander (ed.), *Neoclassical Political Economy* (Cambridge: Ballinger).

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

Bliss, C. (1995*a*), 'Capital Mobility, Convergence Clubs and Long-run Economic Growth', Oxford: Nuffield College Discussion Paper No. 100.

—— (1995*b*), 'The Kuznets Curve Revisited', in N. S. S. Narayana and A. Sen (eds.), *Poverty, Environment and Economic Development* (Bangalore: Interline Publishers).

—— (2002), 'The Stationary Distribution of Wealth with Random Shocks'. Working paper available online at <http://www.nuffield.ox.ac.uk/economics/papers/2002/w6/statDist.pdf>.

—— (2003*a*) 'A Specific Factors Model with Historical Application', *Review of International Economics*, 11/2 (May): 268–78.

—— (2003*b*), Review of *Joseph Stiglitz and the World Bank: The Rebel Within*, Ha-Joon Chang, *Economic Journal*, 113/488: F361-F382.

—— (2004), 'Koopmans Recursive Preferences and Income Convergence', *Journal of Economic Theory*, 117/1 (July): 124–39.

—— (2008), 'The Elasticity of Intemporal Substitution', *New Palgrave Dictionary of Economics*, 2nd Edition, forthcoming.

—— and di Tella, Rafael (1997), 'Does Competition Kill Corruption?', *Journal of Political Economy*, 105/5 (October): 1001–23.

—— and Nalebuff, Barry (1984), 'Dragon-slaying and Ballroom Dancing', *Journal of Public Economics*, (December).

—— and Stern, N. H. (1978), 'Productivity Wages and Nutrition, 1 The Theory; 2. Some Observations', *Journal of Development Economics*, Vol. 5 Issue 4, December: 331–62.

Bloom, David E. and Sachs, Jeffrey E. (1998), 'Geography, Demography and Economic Growth in Africa', *Brookings Papers in Economic Activity*, 98/2: 207–95.

Bourgignon, Francois and Morrisson, Christian (1989), *External Trade and Income Distribution* (Paris: OECD).

—— —— (2002), 'Inequality Among World Citizens: 1820-1992', *American Economic Review*, September 727–44.

Brander, Richard and Krugman, Paul (1983), 'A Reciprocal Dumping Model of International Trade', *Journal of International Economics*, 15: 313–21.

Calomiris, Charles W. (2005), 'Capital Flows, Financial Crises, and Public Policy', in Michael M. Weinstein (ed.), *Globalization: What's New?*, Council of Foreign Relations (New York: Columbia University Press).

Chamberlin, Edward H. (1933), *A Theory of Monopolistic Competition* (Cambridge, MA).

Chen, S. and Ravallion, M. (2001), 'How did the World's Poorest Fare in the 1990s?', *Review of Income and Wealth* Vol. 47, No. 3: 283–300.

Chipman, John (1966), 'A Survey of International Trade Theory: Part 3, The modern theory', *Econometrica*, 34/1: 18–76.

Collier, Paul (2003), 'Primary Commodity Dependence and Africa's Future', in Boris Plekovic and Nicholas Stern (eds.), *The New Reform Agenda*, Annual World

Bank Conference on Development Economics, The World Bank and Oxford University Press.

Collier, Paul and Hoeffler, Anke (1998), 'On the Economic Causes of Civil War', *Oxford Economic Papers*, 50/4: 563–73.

Corden, W. M., and Neary, J. P. (1983), 'Booming Sector and De-industrialisation in a Small Open Economy', *Economic Journal*, 92: 825–48.

Dasgupta, P. (1993), *An Inquiry into Well-Being and Destitution* (Oxford: Clarendon Press).

Dal Bo', Ernesto and dal Bo', Pedro (2004), 'Workers, Warriors and Criminals: Social Conflict in General Equilibrium', unpublished.

David, Paul A. (1975), *Technical Choice, Innovation and Economic Growth* (Cambridge: Cambridge University Press).

Deininger, K. and Squire, L. (1996), 'A New Data Set Measuring Income Inequality', *World Bank Economic Review*, 10: 565–91.

De La Croix, David and Michel, Phillippe (2002), *A Theory of Economic Growth: Dynamics and Policy in Overlapping Generations* (Cambridge: Cambridge University Press).

De Long, Bradford J. (1988), 'Productivity, Growth, Convergence and Welfare: A Comment', *American Economic Review*, 78/5: 1138–54.

De Soto, Hernando (2000), *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else* (New York: Random Books).

Devarajan, Shantayranan and Robinson, Sherman (2005), 'The Influence of Computable General Equilibrium Models on Policy', in Timothy Kehoe, T. N. Srinivasan, and John Whalley (eds.), *Frontiers in Applied Equilibrium Modelling: In Honor of Herbert Scarf* (Cambridge: Cambridge University Press).

Diamond, Peter (1965), 'National Debt in a Neoclassical Growth Model', *American Economic Review*, 55: 1126–50.

Diewert, W. E. (1982), 'Duality approaches to Microeconomic Theory', ch. 12 in Kenneth J. Arrow and Michael D. Intriligator (eds.), *Handbook of Mathematical Economics*, Volume II (Amsterdam: North-Holland).

Dixit, Avinash K. (2004), *Lawlessness and Economics* (Princeton, NJ: Princeton University Press).

—— and Norman, V. (1980), *The Theory of International Trade: A Dual Equilibrium Approach* (Cambridge: Cambridge University Press).

Djajic, Slobodan and Milbourne, Ross (1988), 'A General Equilibrium Model of Guest-Worker Migration', *Journal of Development Economics*, 25: 335–51.

Dollar, David (2005), 'Globalization, Poverty and Inequality', in Michael M. Weinstein (ed.), *Globalization: What's New?*, Council of Foreign Relations (New York: Columbia University Press).

—— and Kraay, A. (2004), 'Trade, Growth and Poverty', *Economic Journal*, 114/493: F22–F49.

Dornbusch, Rudiger (1980), *Open Economy Macroeconomics* (New York: Basic Books).

Durlauf, S. N. and Quah, D. (1999), 'The New Empirics of Economic Growth', in J. B. Taylor and M. Woodford (eds.), *Handbook of Macroeconomics*, Vol. 1A (Amsterdam: North-Holland Elsevier Science), 231–304.

Easterly, William (2002), *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics* (Cambridge, MA: MIT Press).

—— (2005), 'The Rich have Markets, the Poor have Bureaucrats', in Michael M. Weinstein (ed.), *Globalization: What's New?*, Council of Foreign Relations (New York: Columbia University Press).

Eggleston, H. G. (1963), *Convexity* (Cambridge: Cambridge University Press).

Erikson, Robert and Goldthorpe, John (2002), 'Intergenerational Inequality: A Sociological Perspective', *Journal of Economic Perspectives*, Vol. 16, No. 3, August: 31–44.

Ethier, Wilfred (1984), 'Higher Dimensional Issues in Trade Theory', in R. W. Jones and P. B. Kenen (eds.), *Handbook of International Economics*, Volume I (Amsterdam: North-Holland).

—— (1986), 'Illegal Immigration', *American Economic Review*, 76: 258–62.

Feenstra, Robert C. (2004), *Advanced International Trade* (Princeton, NJ: Princeton University Press).

Fine, Ben and Green, Francis (2000), 'Economics, Social Capital, and the Colonization of Social Sciences', in Stephen Baron, John Field, and Tom Schuller (eds.), *Social Capital: Critical Perspectives* (Oxford: Oxford University Press).

Forsyth, P. J. and Kay, J. A. (1981), 'Oil Revenue and Manufacturing Output', *Fiscal Studies*, Vol. 1 (July): 1–28.

Frankel, Jeffrey A. (2005), 'The Environment and Economic Globalization', in Michael M. Weinstein (ed.), *Globalization: What's New?*, Council of Foreign Relations (New York: Columbia University Press).

Friedman, M. (1992), 'Do Old Fallacies Ever Die?', *Journal of Economic Literature*, 30 (December): 2129–32.

Gallup, John Luke, Mellinger, Andrew D., and Sachs, Jeffrey D. (1998), *Geography and Economic Development*, NBER Working Paper No. W6849.

Gambetta, Diego (1993), *The Sicilian Mafia: The Business of Protection* (Cambridge, MA: Harvard University Press).

Garcia-Penelosa, Cecilia (1995), *Essays on Human Capital, R&D and Skill Differentials*. D.Phil. dissertation, Oxford University.

Gelb, Alan H. (1988), *Windfall Gains: Blessing or Curse*? (New York: Oxford University Press).

Geweke, John (1984), 'Inference and Causality in Economic Time Series Models', ch. 19 in Zvi Griliches and Michael Intriligator (eds.), *Handbook of Econometrics*, Volume 2 (Amsterdam: Elsevier Science Publishers).

Glyn, Andrew (2004), 'The Assessment: How Far has Globalization Gone?', *Oxford Review of Economic Policy*, 20/1 (Spring): 1–14.

Glyn, Andrew (2006), *Capitalism Unleashed: Finance, Globalization, and Welfare* (Oxford: Oxford University Press).

Goldberger, A. S. and Manski, C. (1995), 'Review Article: *The Bell Curve* by Herrnstein and Murray', *Journal of Economic Literature*, 33/2: 762–76.

Grief, Avner (1993), 'Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition', *American Economic Review*, 83: 525–48.

Grossman, G. and Helpman, E. (1989), 'Comparative Advantage and Long-Run Growth', *American Economic Review*, 80: 796–815.

____ ____ (1991), *Innovation and Growth in the Global Economy* (Cambridge, MA: MIT Press).

____ ____ (1994) 'Endogenous Innovation in the Theory of Growth', *Journal of Economic Perspectives* (winter): 23–44.

____ ____ (1995), 'The Politics of Free-Trade Agreements', *American Economic Review*, 85: 667–90.

Hall, Robert E. and Jones, Charles I. (1999), 'Why do some Countries Produce so much more Output per Worker than Others?', *Quarterly Journal of Economics*, 114: 83–116.

Harris, J. and Todaro, M. (1970), 'Migration, Unemployment & Development: A Two-Sector Analysis', *American Economic Review*, 60/1 (March): 126–42.

Hausmann, R. and Rigobon, R. (2002), 'An Alternative Interpretation of the "Resource Curse": Theory and Policy Implications'. NBER Working Papers No. 9294.

Henderson, Vernon (2002), 'Urbanization in Developing Countries', *World Bank Research Observer*, 12/1: 89–112.

Herrnstein, R. J. and Murray, C. (1994), *The Bell Curve: Intelligence and Class Structure in American Life* (New York: Free Press).

Hinkle, Lawrence E. and Montiel, Peter J. (1999), *Exchange Rate Misalignment: Concepts and Measurement for Developing Countries*, A World Bank Research Publication (Oxford: Oxford University Press).

Humphreys, McCartan (2000), 'Essays on Trade and Poverty', M. Phil thesis, Oxford University.

Irwin, Douglas A. (2005), 'Trade and Globalization', in Michael M. Weinstein (ed.), *Globalization: What's New?*, Council of Foreign Relations (New York: Columbia University Press).

Jain, Arvind K. (2001), 'Corruption: A Review', *Journal of Economic Surveys*, 1, February, 71–.

Johnson, Harry (1967), 'The Possibility of Income Losses from Increased Efficiency with Capital Accumulation in the Presence of Tariffs', *Economic Journal*, 77: 151–4.

Jones, Charles I. (1995), 'Time Series Tests of Endogenous Growth Models', *Quarterly Journal of Economics*, 110: 495–525.

Jones, R. W. (1971), 'A Three-Factor Model in Theory, Trade and History', in J. N. Bhagwati *et als*. (eds.), *Trade, Balance of Payments and Growth* (Amsterdam: North-Holland).

—— and Easton, Stephen T. (1983), 'Factor Intensities and Factor Substitution in General Equilibrium', *Journal of International Economics*, 15: 65–99.

—— and Neary, J. P. (1984), 'Positive Theory of International Trade', in R. W. Jones and P. B. Kenen (eds), *Handbook of International Economics*, Volume I (Amsterdam: North-Holland).

Joshi, V. and Little, I. M. D. (1997), *India's Economic Reforms 1991–2001* (Delhi: Oxford University Press).

Kaldor, Nicholas (1960), 'Economic Aspects of Advertising', reprinted in *Essays on Value and Distribution* (London: Duckworth).

Kamin, L. J. (1995), Review of *The Bell Curve*, *Scientific American* (February).

Kedouri, Elie (1992), *Politics in the Middle East* (Oxford: Oxford University Press).

Kehoe, Timothy (2005), 'An Exploration of Applied General Equilibrium Models on the Impact of Nafta', in Timothy Kehoe, T. N. Srinivasan, and John Whalley (eds.), *Frontiers in Applied Equilibrium Modelling: In Honor of Herbert Scarf* (Cambridge: Cambridge University Press).

Keller, W. (1996), 'Are International R&D Spillovers Trade Related? Analyzing Spillovers Among Randomly Matched Trade Partners'. Unpublished MS.

Ketter, Vincent (2004), *Small Firms: Red Tape*, Business and Transport Section, House of Commons Library, Research Paper 04/52, June.

Kreuger, Anne (1974), 'The Political Economy of the Rent-Seeking Society', *American Economic Review*, 64: 291–303.

Krugman, Paul (1979), 'Increasing Returns, Monopolistic Competition and International Trade', *Journal of International Economics*, 9: 467–76.

—— (1981), 'Intraindustry Specialization and the Gains from Trade', *Journal of Political Economy*, 89: 959–73.

—— and Lawrence, Robert (1993), 'Trade, Jobs, and Wages,' NBER Working Papers 4478, National Bureau of Economic Research.

—— and Venables, Anthony, with Masa Fujita (1999), *The Spatial Economy: Cities, Regions, and International Trade* (Cambridge, MA: MIT Press).

Kuznets, S. (1955), 'Economic Growth and Income Inequality', *American Economic Review*, 45/1 (March): 1–28.

Kuznets, S. (1963), 'Quantitative Aspects of the Economic Growth of Nations VIII', *Economic Development and Cultural Change*, 2/2: 1–80.

—— (1971), *Economic Growth of Nations: Total Output and Production Structure* (Cambridge, MA: Harvard University Press).

—— (1973), 'Modern Economic Growth Findings and Reflections', *American Economic Review*, 63/3 (June): 247–58.

Landes, David S. (1998), *The Wealth and Poverty of Nations: Why Some are Rich and Some So Poor* (New York: Norton).

Lau, L. J. (1976), 'A Characterization of the Normalized Restricted Profit Function', *Journal of Economic Theory*, 12: 131–63.

Little, Ian M. D. (1982), *Economic Development: Theory, Policy and International Relations* (New York: Basic Books).

—— and Mirrlees, J. A. (1972), *Manual of Industrial Project Analysis in Developing Countries*, Organization for Economic Co-operation and Development.

Lipsey, R. G. (1957), 'The Theory of Customs Unions: Trade Diversion and Welfare', *Economica*, New Series, 24/93: 40–6.

Lucas, Robert E. Jnr. (1993), 'Making a Miracle', *Econometrica*, 61/2 (March): 251–72.

—— (2002), *Lectures on Economic Growth* (Cambridge, MA: Harvard University Press).

—— and Stokey, N. (1984), 'Optimal Growth with Many Consumers', *Journal of Economic Theory*, 32: 139–71.

Maddison, A. (2003), *The World Economy: Historical Statistics*, Development Centre Studies (Paris: OECD).

Mankiw, N. Gregory, Romer, David, and Weil, N. (1992), 'A Contribution to the Empirics of Economic Growth', *Quarterly Journal of Economics*, 107/2 (May): 407–37.

Massey, Douglas S. and Parrado, Emilio A. (1998), 'International Migration and Business Formation in Mexico', *Social Science Quarterly*, 79: 1–20.

Masters, William A. and McMillan, Margaret S. (2000), 'Climate and Scale in Economic Growth', *The Centre for Study of African Economies Working Paper Series*, Paper 123.

McKinnon, R. I. (1991), *The Order of Economic Liberalization: Financial Control in the Transition to a Market Economy* (Baltimore: Johns Hopkins University Press).

Meade, James (1955), *Trade and Welfare* (London: Oxford University Press).

Mesnard, Alice (2004), 'Temporary Migration and Capital Market Imperfections', *Oxford Economic Papers*, 56: 242–62.

Milanovich, Branko (2002), 'True World Inequality 1988 and 1993: First Calculations based on Household Surveys Alone', *Economic Journal*, 112 (January), 476: 511–92.

—— (2004), *Worlds Apart: Global and International Inequality 1950–2000* (Princeton, NJ: Princeton University Press).

Mirrlees, J. A. (1975), 'A Pure Theory of Underdeveloped Economies', in L. Reynolds (ed.), *Agriculture in Development Theory* (New Haven, CN: Yale University Press).

Myrdal, Gunnar (1968), *Asian Drama: An Inquiry into the Poverty of Nations* (Harmondsworth: Penguin).

Newbery, David and Stiglitz, Joseph (1984), 'A Pareto Inferior Trade', *Review of Economic Studies*, 1–12.

Ng, Francis and Yeats, Alexander J. (2000), 'On Recent Trade Performance of Sub-Saharan African Countries: Cause for Hope of More of the Same?', *Africa Region Working Paper* No. 7, The World Bank.

—— (2002), 'What Can Africa Expect from its Traditional Exports?', *Africa Region Working Paper* No. 26, The World Bank.

Nikaido, Hukukane (1972), 'Relative Shares and Factor Price Equalization', *Journal of International Economics*, 2: 257–63.

North, Douglas C. (1990), *Insitutional Change and Economic Performance* (Cambridge: Cambridge University Press).

—— (2005), *Understanding the Process of Economic Change* (Princeton, NJ: Princeton University Press).

Plekovic, Boris and Stern, Nicholas (eds.) (2003), *The New Reform Agenda*, Annual World Bank Conference on Development Economics, The World Bank and Oxford University Press.

Quah, D. (1997), 'Empirics of Growth and Distribution: Polarization, Stratification and Convergence Clubs', *Journal of Economic Growth*, 2/1: 27–59.

Ramsey, Frank (1928), 'A Mathematical Theory of Saving', *The Economic Journal*, 38/152 (December): 543–59.

Ravallion, Martin (2004), 'Competing Concepts of Inequality in the Globalization Debate', World Bank.

Redding, Stephen (2002), 'Specialization Dynamics', *Journal of International Economics*, 58: 299–344.

—— and Venables, Anthony J. (2002), 'Explaining Cross-Country Export Performance: International Linkages and Internal Geography', *CEP Discussion Papers* 0549, Centre for Economic Performance, London School of Economics.

Richardson, Craig J. (2004), *The Collapse of Zimbabwe in the Wake of the 2000–2003 Land Reforms* (Lewiston, NY: Edwin Mellen Press).

Robinson, James A., Torvik, Ragnar, and Verdier, Thierry (2005), 'Political Foundations of the Resource Curse'. Paper prepared for a conference in honour of Pranab Bardhan.

Rodrick, D. (1996), 'Understanding Economic Policy Reform', *Journal of Economic Literature*, 24/1: 0–41.

Romer, P. M. (1986), 'Increasing Returns and Long-Run Growth', *Journal of Political Economy*, 94/5: 1002–37.

Romer, P. M. (1990), 'Endogenous Technical Change', *Journal of Political Economy*, 98: S71–102.

Rose-Ackerman, Susan (1978), *Corruption: A Study in Political Economy* (New York: Academic Press).

Ruffin, Roy J. (1981), 'Trade and Factor Movements with Three Factors and Two Goods', *Economics Letters*, 7: 177–82.

Ruffin, Roy J. and Jones, Ronald (1977), 'Protection and Real Wages: The Neoclassical Ambiguity', *Journal of Economic Theory*, Vol. 14, 2 (April): 337–48.

Rybczynski, T. M. (1955), 'Factor Endowments and Relative Commodity Prices', *Economica,* 22: 336–41.

Sachs, Jeffrey D. (2005), 'Globalization and Patterns of Economic Growth', in Michael M. Weinstein (ed.), *Globalization: What's New?*, Council of Foreign Relations (New York: Columbia University Press).

____ and Warner, Andrew (1997*a*), 'Natural Resource Abundance and Economic Growth', NBER Working Paper No. 5398.

____ ____ (1997*b*), 'Sources of Slow Growth in African Economies', *Journal of African Economies*, 6: 335–76.

Sala-i-Martin, Xavier (2002), 'The Disturbing "Rise" of Global Income Inequality', NBER Working Paper No. 8904.

Samuelson, Paul A. (1948), *Foundations of Economic Analysis* (Cambridge, MA: Harvard University Press).

____ (1971), 'Ohlin was Right', *Scandinavian Journal of Economics*, 73/4, 1071: 365–84.

Schelling, Thomas (1960), *The Strategy of Conflict* (Cambridge, MA: Harvard University Press).

Schliefer, Andrei and Vishny, Robert W. (1993), 'Corruption', *Quarterly Journal of Ecnomics*, 108 (August): 599–617.

____ ____ (1998), *The Grabbing Hand: Government Pathologies and their Cures* (Cambridge MA: Harvard University Press).

Selten, Rheinhardt (1975), 'Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games', *International Journal of Game Theory*, 4/1: 25–55, reprinted in H. W. Kuhn (ed.), *Classics in Game Theory* (Princeton, NJ: Princeton University Press, 1997).

Sen, Amartya K. (1966), 'Peasants and Dualism with or without Surplus Labour', *Journal of Political Economy*, Vol. 74: 425–50.

Solow, R. M. (1956), 'A Contribution to the Theory of Economic Growth', *Quarterly Journal of Economics*, 70: 65–94.

____ (1970), *Growth Theory: An Exposition* (Oxford: Clarendon Press).

Stewart, Dugald (1793), *Account of the Life and Writings of Adam Smith LL. D.*, Transactions of the Royal Society of Edinburgh.

Stiglitz, J. E. (1969), 'Distribution of Income and Wealth among Individuals', *Econometrica*, 37/3 (July): 382–97.

____ (2002), *Globalization and its Discontents* (London: Allen Lane).

____ (2004), 'Capital-Market Liberalization, Globalization, and the IMF', *Oxford Review of Economic Policy*, 20/1 (spring): 57–71.

____ (2005), 'The Overselling of Globalization', in Michael M. Weinstein (ed.), *Globalization: What's New?*, Council of Foreign Relations (New York: Columbia University Press).

—— and Weiss, A. (1981), 'Credit Rationing in Markets with Imperfect Information', *American Economic Review*, 71: 393–410.

Stolper, W. F. and Samuelson, P. A. (1941–2), 'Protection and Real Wages', *Review of Economic Studies*, 9: 58–73.

Strotz, R. H. (1955), 'Myopia and Inconsistency in Dynamic Utility Maximization', *Review of Economic Studies*, 23/3, 62: 165–80.

Sutcliffe, Bob (2004), 'World Inequality and Globalization', *Oxford Review of Economic Policy*, 20/1 (spring): 15–37.

Tawney, R. H. (1938), *Religion and the Rise of Capitalism* (Harmondsworth: Penguin).

Temin, Peter (1966), 'Labour Scarcity and the Problem of American Industrial Efficiency in the 1850s', *The Journal of Economic History*, Vol. 26, No. 3: 277–98.

Theil, H. and Seale, J. L. (1994), 'The Geographic Distribution of World Income, 1950–1990', *De Economist*, 4.

Todaro, M. P. (1969), 'A Model of Labor Migration and Urban Employment in Less Developed Countries', *American Economic Review*, 59: 138–48.

Toye, John (1999), 'The Sequencing of Structural Adjustment Programmes: What are the Issues?', available online at <http://www.ictsd.org/html/Toye.pdf>.

Tullock, G. (1967), 'The Welfare Costs of Tariffs, Monopolies and Theft', *Western Economic Journal*, 5: 224–32.

United Nations (2003), *The Arab Human Development Report*, Geneva. Available online at <http://www.rbas.undp.org/ahdr2003.shtml>.

Vanek, Jaroslav (1965), *General Equilibrium of International Discrimination: The Case of Customs Unions* (Cambridge, MA: Harvard University Press).

Ventura, J. (1997), 'Growth and Interdependence', *Quarterly Journal of Economics* (February): 57–84.

Viner, Jacob (1950), *The Customs Union Issue* (New York: Carnegie Foundation for International Peace).

Weber, Max (1930), *The Protestant Ethic and the Spirit of Capitalism* (London: Allen and Unwin).

Willenbockel, Dirk (2006), 'Structural Effects of a Real Exchange Rate Revaluation in China: A CGE Assessment', Munich, Germany.

Williamson, Jeffrey G. (1988), 'Migration and Urbanization', ch. 11 in Hollis Chenery and T. N. Srinivasan (eds.), *Handbook of Development Economics*, Vol. 1 (Amsterdam: North-Holland).

Williamson, J. (ed.) (1994), *The Political Economy of Policy Reform* (Washington, DC: Institute for International Economics).

Winters, L. A. (2004), 'Trade Liberalization and Economic Performance: An Overview', *Economic Journal*, 114/493: F4–F21.

Wolf, Martin (2004), 'Globalization and Global Economic Governance', *Oxford Review of Economic Policy*, 20/1 (spring).

Wood, Adrian (1994), *North-South Trade, Employment and Inequality: Changing Fortunes in a Skill-Driven World* (Oxford: Clarendon Press).

Wood, Adrian (2003), 'Could Africa be more like America?', in Plekovic and Stern 2003.

—— and Jordan, Kate (2000), 'Why does Zimbabwe Export Manufactures and Uganda Not?', *Journal of Development Studies*, 37/2: 91–116.

—— and Mayer, Jorge (2001), 'Africa's Export Structure in a Comparative Perspective', *Cambridge Journal of Economics*, 25/3: 369–94.

Woodland, A. D. (1974), 'Demand Conditions in International Trade Theory', *Australian Economic Papers*, 13: 209–24.

World Bank (2002), *Rural-Urban Linkages and Interactions: Policy Implications for Development Planning and Poverty Reduction* (Washington, DC: World Bank).

# Index