TOWARDS THE MICROFOUNDATIONS OF FINANCE AND GROWTH

Alex William Trew

**A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews**

2007

Towards the Microfoundations of Finance and Growth

Alex William Trew

Submitted for the degree of

Doctor of Philosophy (Economics)

at the University of St. Andrews

04 May 2007

I, Alex Trew, hereby certify that this thesis, which is approximately 50,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date: 04 May 2007. Signature of candidate: ____________________.

I was admitted as a research student in September 2004 and as a candidate for the degree of PhD Economics in September 2005; the higher study for which this is a record was carried out in the University of St. Andrews between 2004 and 2007.

Date: 04 May 2007. Signature of candidate: ____________________.

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD Economics in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: 04 May 2007. Signature of supervisor: ____________________.

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration.

Date: 04 May 2007. Signature of candidate: ____________________.

Address for correspondence:

School of Economics & Finance
University of St. Andrews
St. Andrews, Fife, UK
KY16 9AL

Email: alex.trew@dunelm.org.uk
Web: www.st-andrews.ac.uk/cdma/a.trew.html

For my parents

Abstract

We take a critical view of the standard approach to finance and growth. The mapping between the theory and empirics is shown to be poorly understood, and this is traced to deficiencies in our understanding of the microeconomics at play. By looking at both primary and secondary historical evidence we argue that issues of aggregation are critical, and that spatial factors are also prevalent. Further, we suggest that these disaggregated elements can change over the course of an industrial revolution. A model in the spirit of standard finance and growth theories is extended to consider these further effects, and we calibrate the model to data on historical growth paths.

In order to advance our understanding of the microeconomic factors that cause the observed phenomena in the finance-growth nexus, we develop a general equilibrium theory of financial intermediation in which exchange costs are endogenously determined by technologies, endowments and preferences. We suggest that incomplete contracts might be central to these phenomena. We link this framework to an understanding of power and political economy in a setting with heterogeneous agents. We develop these results numerically, showing a number of interesting interactions between markets, exchange costs and institutions in economies with different levels of wealth.

The model of endogenous exchange costs can be thought of in terms of the findings coming out of our historical analysis. We outline in some detail the further steps that need to be taken before we can speak of *the* microfoundations of finance and growth with any confidence. First, a fully dynamic model of markets and coalitions must be embedded within a story of economic growth that can match the dynamic observations. Second, we must develop our conception of incomplete contracting and the link with institutions and political economy. The thesis thus opens a number of interesting avenues for future research.

JEL Classifications: D51; L14; O16; O40; N23.

Keywords: Finance and growth; economic history; institutions; exchange costs; incomplete contracts.

Acknowledgements

All aspects of this thesis have benefited from the patient and generous guidance of my supervisors, Charles Nolan and Gary Shea. Gary has shown me that history can indeed be the laboratory of the social sciences. His willingness to meet the strident generalisations of growth theory with hard historical fact seems to make him a very rare form of economic historian. The research project forming the basis of his *Handbook* has made this thesis strive to address the important issues: The feeling of purpose throughout this thesis is down to him. Charles has been an exceptional supervisor. His intellectual and personal guidance has, over three short years, made me see economic research as an exciting and honourable profession. Mostly, his intention that a student should immediately begin working on research topics, finding important questions, forming new ideas and writing papers, has been inspiring. A few weeks after arriving I muttered the intention that, "Well, I'm not looking to invent a new growth theory." His response, "Why the hell not?", set the tone for the remainder of my time as a student and, hopefully, for the rest of my career.

The Centre for Dynamic Macroeconomic Analysis provided financial support for the first half of the period in which this thesis was written. I am grateful to the University of St. Andrews for providing the initial finance for this Centre. Being surrounded by a group of ambitious faculty and students, all of whom have been intent on pursuing interesting macroeconomic agendas, has very often contributed to any progress that I have made. Comments made on this work at internal PhD presentations and Brown Bags were always thought-provoking and helpful. In addition, my fellow CDMA students have made my time at St. Andrews far more interesting and enjoyable than I could have anticipated. The School of Economics & Finance has been an excellent place to write this thesis. The academic and administrative staff have been welcoming, helpful and understanding at all points.

The Shell Foundation provided generous funding for the second half of the period in which this thesis was written. I am indebted to Kurt Hoffman and Marc Lopatin for their openness to my initial advances, and for their consequent support and ambition for a larger project of which this thesis is now part.

I am grateful to Gary Shea for his permission to replicate in Chapter 3 parts of his original material from Shea (2007) (particularly, Figures 3.2–3.4). That Chapter also benefited from discussions with Kurt Hoffman, Andrew Vivian and Robert E. Wright. Parts of Chapter 4 are joint work with Charles Nolan (specifically, Sections 4.3 and 4.7). For discussions on aspects of that Chapter we thank participants at workshops in Edinburgh and St. Andrews. Particularly, Vladislav Damjanovic, Oliver Kirchkamp, John Hardman Moore and József Sákovics. Discussions with Helmut Rainer were especially helpful. The usual disclaimer applies to *all* of these acknowledgements.

I owe the greatest debt to my family and close friends. My parents and my sister have been unbelievably supportive. Their acceptance of my career choices, and their tolerance of my frustrations, which have too often been directed at them, has meant more than I can say. Chloë has been there for me when I needed her with love and understanding. The belief shown in me by my family and close friends has been a great source of strength.

This thesis was typeset in LaTeX 2_ε.

Contents

List of Figures

List of Tables

Introduction

What is the relationship between the financial development of an economy and its sustained level of economic growth? This is an old question, going back at least to the nineteenth century. It is also an important question. The financial market sits between savers and investors. It exists within an institutional framework as ordered by the legislative and political environment. Financial intermediaries, banks and stock markets facilitate insurance against risk, finance entrepreneurial activity, evaluate and monitor investments and become experts in the gathering and dissemination of costly information, to name just a few of their functions. Imperfections in financial markets, and the parts that institutions have to play in determining these imperfections, are thus central to economic activity.

The flurry of activity in finance and growth research over the last fifteen years has covered a great deal of empirical and theoretical ground. We can now be confident in stating that there does exist a positive correlation between financial depth and economic growth. Consequently, financial development has been accepted, in some academic and policy circles, as one of the factors considered to play a determining role in the level of sustained economic growth. But the literature is still pushing to be taken seriously in an academic sense. The best textbook on economic growth, Barro and Sala-i-Martin (2004), neglects the subject almost entirely. For, while a correlation exists, we do not yet have a solid understanding of the relationship on a more fundamental level.

The lack of consensus is exacerbated by the nature of the debate. There is little dialogue between opposing schools but, as is often the case, a compromise might be found. It could be argued that while proponents have been too quick to roll-out policy prescriptions for engendering industrial take-off in developing countries, paying too little attention to the underlying theoretical mechanisms and their quantitative significance, opponents have been too unwilling to take the consistent empirical and theoretical findings as at all meaningful.

This thesis intends to begin to address a number of the problems in the orthodox approach to finance and growth. We concentrate on an applied theoretical perspective, but also look at some of the foundations of current empirical work. We intend to make a step towards a deeper, more qualified understanding of the mechanics of finance and growth. In the process, it is hoped that we can both contribute to our understanding of any finance-growth mechanism and demonstrate to sceptics that the question is worth taking seriously.

The Scope of this Thesis

We focus on the connection between finance and growth in the context of the development of pre-industrial economies. As such, we abstract from short-run fluctuations, concentrating instead on the conditions of finance and growth in the long-run. Also, we leave aside the importance of financial institutions in transition economies. Given the already-present wealth of empirical evidence on the subject, we spend most of our analysis developing a theoretical understanding of finance and growth. However, the empirical and historical analyses that we do undertake are intended to directly feed into the theoretical steps that we take.

Despite this focus, throughout we use the word 'finance' in the same general sense in which it is interpreted in the literature: Everything from the microeconomic, contractual relationships between financial institutions and the agents (both debtors and creditors) demanding their services to macroeconomic aggregates such as ratios of financial debt to national output. Our primary aim is to shed light on the ways in which finance, in these senses, can have an impact upon the entry into sustained high growth of a poor, pre-industrial economy. This is why we focus on the historical record of industrialisation in market economies.

Outline of the Thesis

The thesis is structured as follows. Chapter 1 introduces the literature on finance and growth from three perspectives: Empirics; theory; and history. We develop a critique of the current understanding of the relationship between finance and growth. Chapter 2 extends this survey by developing a parsimonious model that can capture the salient mechanisms of prevailing theories of finance and growth. This model is then calibrated to historical data. We draw some conclusions about quantitative implications of work of this kind. We develop more general conclusions by extending a standard quality-ladders endogenous growth model to include similar finance and growth mechanisms. Common themes from these models are assessed in the light of our critique from Chapter 1.

Chapter 3 then takes a step towards a richer understanding of the dynamic and disaggregated elements to the finance-growth nexus. This is principally motivated by an investigation of both primary and secondary historical sources. We take the lessons from this analysis and form a model of finance and growth which extends that in Chapter 2. We find a number of more subtle, microeconomic issues at play. As such we need to construct, from microeconomic foundations, an understanding of the causes and effects of the more nuanced aspects of the finance and growth connection. Chapter 4 thus develops a new framework for understanding the existence and size of exchange costs and coalitions. Chapter 5 begins to tie together the findings from each of these Chapters and suggests directions for future work.

Chapter 1

Finance and Growth: A Critical Survey*

The literature on finance and economic growth has experienced a renaissance in the last fifteen years. The construction of a large World Bank dataset covering the second half of the twentieth century facilitated a large number of cross-country studies. While most of this work supports the hypothesis that finance plays a determining factor in economic growth, there have been one or two voices urging a more cautious interpretation of the data.

At the same time as creating new opportunities for research, the database has engendered a, perhaps excessive, focus on cross-sectional results based on financial depth alone. Recognising this, some economic historians have begun constructing datasets to reveal the time-series experience of countries going through a period of industrial and financial revolution. However, the time-series data remain somewhat sparse and, in general, the implications of the literature in terms of growth and transition *over time* have been largely neglected.

The theory of finance and growth has been developed, almost in parallel to

*Part of this chapter has been released as Trew (2006a), and published as Trew (2006b).

the cross-section empirics, to explain why finance may cause growth. It has been demonstrated that, in a comparative sense, financial institutions can play a role in the level of sustained growth. There is here, however, no clear *quantitative* lesson to be drawn from the existing literature; yet modern macroeconomic theory is judged largely against its ability to be calibrated by and replicate data in a consistent way. Many theoretical considerations of the finance-growth nexus do not rigorously confront theory with data.

This Chapter serves to motivate and frame the remainder of the thesis. We provide a general survey, but concentrate on those aspects of the literature which are not covered in orthodox surveys of the subject, such as Levine (2005). We argue that growth theory and growth empirics have become disconnected, especially in relation to the question of finance and growth; in an important sense, they answer different questions. We take up some of these issues in Chapter 2. In addition, we demonstrate that both theory and empirics can learn from cliometric evidence and, in general, a greater appreciation of the historical context. We begin to study the historical context in more detail in Chapter 3. The conclusions we draw from this work feed into an analysis of the microeconomics of finance and growth in Chapters 4 and 5.

The Chapter is organised as follows: Section 1.1 surveys the current state of knowledge in empirical, theoretical and historical terms. Section 1.2 considers the potential for future research on finance and growth to be more fully integrated across empirical-theoretical-cliometric lines. Section 1.3 concludes with a summary of our main findings.

1.1 Existing Literature

1.1.1 Contemporary Empirics

Empirical exploration of the relationship between economic growth and financial structure has been widespread, beginning with the ground-breaking work of Goldsmith (1969) and pioneered in recent years by Ross Levine, with much of his work summarised in Demirgüç-Kunt and Levine (2001) and Levine (2005).

Goldsmith (1969, *Preface*) saw the need to "relate certain measurable characteristics of financial structure to a quantitative expression of economic growth." His novelty was in considering what *constitutes* financial structure (how to define it), in collating a wide range of data from a number of countries across time and in attempting to present quantitative evidence for the idea that financial structure matters. Goldsmith was important for taking the first step and exposing areas of weakness in understanding and empirical inadequacies, but could draw no robust conclusions from his analysis.

The core measure of a country's financial development used by Goldsmith is the financial interrelations ratio (FIR): The value of all financial instruments outstanding divided by the value of national wealth. On cross-country evidence from mainly developed countries it is shown that there is a 'loose and irregular' association between financial development and higher per capita income (in his regression analysis, $R^2 = 0.19$). On the relationship between financial development and growth the conclusions are less clear: "Even in the long run there is no close correlation between the rate of growth of real national product and of FIR." (*op. cit.*, p. 378.) Goldsmith finds some evidence, therefore, but is unable to draw conclusions on causality, partly, he argues, because there is not sufficient data to show it. McKinnon (1973) also embarks upon an empirical investigation, and again finds mixed evidence on causality.

King and Levine (1993a,b) were the first to demonstrate the potential for large panel datasets to make rigorous the finance and growth debate. They found not only a consistent contemporaneous relationship between aggregate measures of financial depth and growth, but also a strong *predictive* component. They argued that current financial depth can predict economic growth over the consequent ten to thirty years, concluding that "better financial systems stimulate faster productivity growth and growth in per capita output by funneling society's resources to promising productivity-enhancing endeavours." (King and Levine, 1993b, p. 540.)

Beck et al. (1999) introduced the large World Bank database that grew out of the early King and Levine work. This, in turn, made possible most of the empirical research to which we refer in this thesis. The obvious endogeneity problems inherent in finance and growth regressions were considered in Levine et al. (2000): The positive causal link from finance to growth was found to be robust to various instruments. The longer dataset forms the backbone of Demirgüç-Kunt and Levine (2001). The book asks the questions first posed by Goldsmith: Does financial structure change as countries develop; does financial development influence economic growth; and does financial structure (bank- or market-based) influence economic growth? The findings are relatively consistent. It is shown again that financial development does accelerate growth, that an economy generally becomes more market-based as it grows, and that growth does not seem to be significantly different depending on whether a country is predominantly market- or bank-based. Rather, they argue that institutions such as the legal system are crucial in sustaining growth. The importance of property rights and contract enforcement, as substitutes for social capital, is underlined (*op. cit.* p. 12.),

> ...policymakers may achieve greater returns by focusing less on the extent to which their country is bank-based or market-based and more on legal, regulatory, and policy reforms that boost the functioning of

markets *and* banks.

These studies, and many others besides, find empirical support for the argument that finance leads growth, in some sense.

There have been some questions raised by, among others, Driffill (2003) about the interpretation of empirical results like those outlined above. These concerns have been omitted from surveys such as Levine (2005), so we pay special attention to them here. Indeed, on further analysis, we suggest that the impact of data limitations may be more acute than is generally believed. It will be argued that these limitations have the tendency to exaggerate the role of finance in determining economic growth.

The World Bank dataset, which currently[1] covers the period 1960-2005 for 192 countries, is not complete for all countries. For example, consider a preferred measure of financial intermediation, the private credit-to-GDP ratio: The mean number of observations per country is 24.1. As such, a typical cross-sectional estimation necessarily involves an average of financial variables over a long period, often over the whole sample period. Some empirical work has begun to use panel data (for example Levine et al., 2000), but the longitudinal scope of the panels used is limited (in the case of Levine et al., the panel consists of five seven-year averages over the period 1960-95) and this depth comes at the cost of cross-sectional breadth (Levine et al. include only 71 countries from a total sample of nearly 150 at that time). Not only is it hard to think about finance leading growth when the averaging time periods are so long, but such a trade-off also leaves estimations open to selection bias issues.

A number of specific results, obtained using this dataset, have been called into question. Driffill (2003) and Manning (2003) argue that the results in Levine and

[1]The dataset is available from http://econ.worldbank.org/programs/finance/. We refer to the March 14 2005 revision throughout this thesis. A September 2006 update brings the dataset up to 2005, but makes no other substantial revision.

Zervos (1998) have implausible implications for the effect of financial development on growth. Specifically, the results suggest a one-percent per annum increase in growth rates could be obtained if developing countries increased the level of financial development to those of more successful countries. He demonstrates that a number of results hinge on the inclusion of outliers, while the inclusion of regional dummies, especially those for the Asian Tigers, also renders coefficients on financial development insignificant. Driffill goes on to consider the robustness of the work on industry-level data of Rajan and Zingales (1998). It is shown again that the positive effect of financial development on growth is contingent upon the specification employed, particularly that including broad regional dummies tends to neutralise the significance of financial variables. Driffill concludes that the positive results on data over this period were likely driven by the growth of the Asian Tigers, growth which is more naturally attributed to other factors (on this see, among others, Young, 1995; Rodrik, 1996; Landes, 1998).

With this in mind, it should be noted that the Levine et al. (2000) dataset ends in 1995, before the Asian financial crisis; a period of economic downturn preceded by deepening financial markets. Financial depth, as measured by Levine et al.'s preferred indicator of financial depth (the ratio of private credit-to-GDP), increased significantly in the Asian Tiger countries over the period 1992-1998: In China by 30.9%; Hong Kong, 40.3%; Indonesia, 21.9%; the Korean Republic, 53.8%; Malaysia, 45.7%; the Philippines, 142.9%; and Thailand, 89.3%. The year 1998 saw a reduction in GDP in all of these countries except China (the respective percent changes in real GDP per capita were: 5.35; -8.21; -11.6; -9.08; -0.62; -4.06; -10.74).[2] While it is, of course, not possible to draw any hard conclusions from such analysis, if Driffill (2003) is correct in suggesting that most of the significance of financial variables is driven by the growth experience of the Asian Tigers, then this episode

[2]This data is from the Penn World Table, see Heston et al. (2002).

calls for a more refined classification of financial depth. A measure which controls for both institutional and regulatory factors that might determine the efficacy of financial deepening in spurring growth may obtain very different results.[3] Additionally, a distinction in growth regressions between foreign and domestic providers of finance may provide more qualified results.

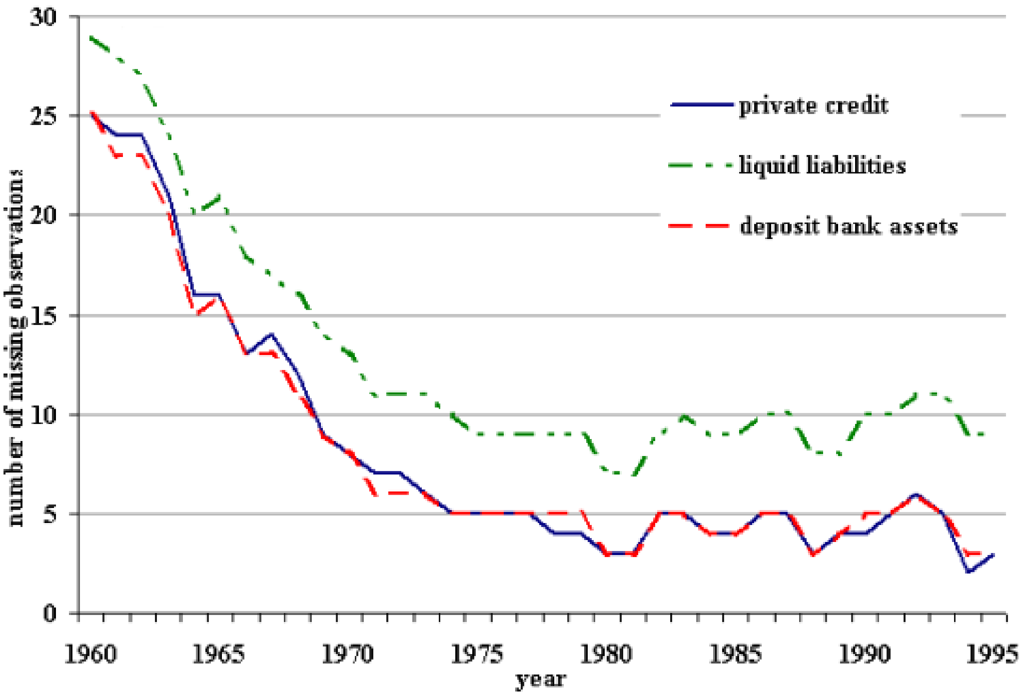
Aghion et al. (2005) use the same 1960-95 dataset as Levine et al. (2000). They also include the same 71 countries despite using the dataset in a purely cross-sectional way (employing an average of the private credit variable over the entire thirty-five-year period) to demonstrate the positive effect of financial development on convergence. It is possible that their results would be very different if we re-estimated on the whole sample, increasing both the number of countries and the endpoint to include properly the Asian financial crisis. While Aghion et al. (2005) do test for some geographical differences, they do not test specifically for the East Asian bloc.

The potential significance of selection bias issues is here even more important since Aghion et al. take an average for their financial proxy over the whole sample period. Countries with available data are more likely to have converged (for example the sample includes only 11 of 54 African countries) and countries with sparse data are generally those that were poor in 1960, such that available data tends to be at the end of the sample, as can be seen in Figure 1.1. Data are given in Appendix Table B.1.

Added to this, the trend of financial development as measured by the credit-to-GDP ratio is itself rising over time across countries. This can be seen by inspection of the data: The credit-to-GDP ratio trends upwards in around 55 of the 71 countries in the sample. As such, the measure of financial development for countries who were relatively poor in 1960 and so with data for only the later years, is biased

[3] There has recently been a move to consider institutional and legal issues, see particularly Levine (2004) and Beck and Levine (2005).

Figure 1.1: Distribution of Missing Observations for the Finance to GDP Ratios in the Levine et al. (2000) Dataset



upwards relative to a rich country with data for every year. It should be noted that the upward trend is not specific to the credit-to-GDP ratio; two of the three alternative proxies used in Aghion et al. (2005, Table 4), trend upwards. The third, a ratio of commercial bank to central bank assets, is relatively stable for most countries over the period, and this is the one proxy for which the coefficient on financial development is insignificant (this variable has been dropped from the Beck et al. (2000) dataset in later revisions).

The combination of these factors – the long average, the data sparsity, the sample selection bias and the upward trend in the financial development indicator selected – means that those countries that did converge have, as a result of the methods used, necessarily had a higher measure of financial development over the period. This would explain, at least partly, why the results in Aghion et al. (2005) are so robust to alternate specifications. This critique is, unfortunately, not specific to the Aghion et al. paper; see Beck et al. (2004, p.9): "...we sometimes use data averaged over the period 1960-1999, and sometimes we use data over the period 1980-2000...".

It should be clear that one ought not to be overly reliant on either purely cross-sectional empirics or limited panel datasets. Driffill (2003) suggests a greater emphasis on long-run, historical time-series. He stresses in particular the importance of comparing countries at similar stages of development in order that more robust conclusions might be drawn.

The overall message from contemporary empirical research on finance and growth is indicative but problematic; time-series evidence must be consolidated in order that we can speak with confidence on the relation between financial institutions and growth *within* a country over a period of transition.

1.1.2 Theories of Finance and Growth

Following the pioneering contributions of Lucas (1988), Romer (1990), Grossman and Helpman (1991) and Aghion and Howitt (1992), endogenous growth theory has been at the heart of most theoretical attempts to understand the mechanisms of long-run growth. The finance-growth nexus is no exception. Theories have generally differed along three aspects: The type of endogenous growth; the finance mechanism (the sphere in which constraints impact); and the treatment of asymmetric information (within that sphere of financial influence). Using this structure, Table 1.1 outlines the main features of some of the most influential finance and endogenous growth models. It should be clear by inspection that, regardless of the source of endogenous growth, the main feature determining growth in most models is some financial constraint on the acquisition of either knowledge via education or technology via entrepreneurship.

Financial intermediation in most models takes the form of a perfectly competi-

Table 1.1: Core Features of Some Finance and Growth Models

	source of endogenous growth	**finance mechanism**	**information problem**
BS91	production externalities	insurance market and entrepreneurship	exogenous liquidity shock
S92	production externalities	capital market	exogenous productivity shock
KL93b	vertical innovation	entrepreneurial funding, heterogeneous agents	adverse selection (screening)
BC96	production externalities	contract or screen heterogeneous entrepreneurs	adverse selection (ration or screen)
FM96	horizontal innovation	funding and monitoring entrepreneurship	moral hazard (effort aversion)
BH98	horizontal innovation	entrepreneurship, project appraisal, risk diversification	moral hazard (deceit)
GK98	human capital accumulation	credit markets vs. intergenerational altruism	none
M03	vertical inn., capital accumulation	entrepreneurship, screening	moral hazard (effort aversion)
AHM05	vertical innovation	entrepreneurship, credit constraints	moral hazard (deceit)
BBC05	production externalities	entrepreneurship, markets and banks	adv. selection and moral hazard

N.B. BS91 is Bencivenga and Smith (1991); S92 is Saint-Paul (1992); KL93b is King and Levine (1993b); BC96 is Bose and Cothren (1996); FM96 is de la Fuente and Marín (1996); BH98 is Blackburn and Hung (1998); GK98 is de Gregorio and Kim (1998); M03 is Morales (2003); AHM05 is Aghion et al. (2005); BBC05 is Blackburn et al. (2005).

tive system of banking or financial intermediation. Some consider a role for stock markets, but often only as a choice between mutually-exclusive banks and markets (such as Greenwood and Smith, 1997). Blackburn et al. (2005) is an interesting exception, and considers both the joint-determination *and co-existence* of banks and stock markets as determined by state-dependent moral hazard conditions. In these models there is feedback from growth in the economy to the determination of optimal financial structure, be it based on banking alone or on a mixture of banks and markets. As an economy develops so it can afford those financial structures that better facilitate faster growth (Greenwood and Smith, *op. cit.*).

A related result is obtained by Greenwood and Jovanovic (1990), who extend Townsend (1978) to consider the consequences of the exogenous costliness of financial intermediation on market formation, inequality and economic growth. Intermediaries both provide information for the assessment of investment projects and facilitate diversification of risk. Something like Kuznets's (1955) inequality hypothesis is captured: In the early stages of development, financial markets are small and growth is low; as financial markets develop, so inequality increases as only the relatively wealthy can access their benefits; as the economy grows further, so financial constraints diminish and the relatively poor begin to benefit from greater financial sophistication, reducing inequality again. This is something also found by, among others, Aghion and Bolton (1997).

In a related literature, multiple equilibria can emerge as a result of countries with limited financial sectors caught in a low-growth trap. Saint-Paul (1992) is a further approach to the modelling of stock markets, wherein stock markets that facilitate international risk sharing enable specialisation in technologies and higher growth. There is thus a low- (high-)growth equilibrium associated with low (high) financial development, capturing the idea of different take-off points for industrialising nations. In Bose and Cothren (1996) a bank lends to an entrepreneur of unknown quality and selects by either designing a separating menu contract (where this is possible) or by implementing a costly screening technology, or by a mix of the two. They show that in the early stages of financial development, a fall in the screening cost can actually be growth-*reducing* because of the interaction between dependence on rationing or screening. In concurrence with Schumpeter (1934), the financial sector needs to reach 'critical mass' before advances in financial sophistication will improve growth.

Risk sharing also motivates the work of Acemoglu and Zillibotti (1997). An OLG economy is considered in which fixed costs present a barrier to large investments with high return. Minimum investment requirements mean that agents cannot always insure against the risk involved in investing in high-return projects. A series of positive shocks can cause financial development and economic development to take-off. Take-off occurs in all economies eventually, but at different stages as determined by the fortune of agents in the economy. Credit constraints also impact on growth and convergence in the model of Aghion et al. (2005).

A paper in which there is no explicit information problem is that of de Gregorio and Kim (1998). In an economy comprised of agents with different levels of ability, a first-best optimal level of investment in high-ability agents requires fully functioning capital markets. It is shown that in the absence of such markets, intergenerational altruism can obtain only a second-best equilibrium.

The majority of works outlined in Table 1.1 treat the financial sector as static[4] (where the nature of the relationship between agents and intermediaries does not change endogenously over time; again, Blackburn et al., 2005, is a notable exception), with the degree of efficiency of identifying/screening/funding/monitoring suitable debtors determining the costs of financial intermediation and so the level of economic growth. The static nature of this relationship requires that the level of financial depth remains constant as the economy grows. This is clearly at odds with the lessons drawn from the empirical literature, where at the least we know that financial depth *does* change over the course of long-run period of sustained growth. We develop in Chapter 2 a better understanding of this proposition.

Most papers consider some form of entrepreneur who cannot or will not personally fund a project either because it is too large or too risky. Increasing the ease with which entrepreneurs can obtain funds thus increases the rate of technological progress and so the rate of economic growth. Others consider a role for the ac-

[4]This concept of a static intermediation relationship echoes that of Auerbach and Siddiki (2004).

cumulation of education or human capital as entering directly into the production function; the efficiency with which this process is facilitated by screening or funding agents thus has an effect on economic growth. Further papers look at the effect of credit constraints on entrepreneurship and, again, the growth consequences of higher efficiency in financial intermediation are, for all intents and purposes, equivalent.

The major differences between these models largely revolve around the treatment of asymmetric information. In a few (such as King and Levine, 1993b, and Bose and Cothren, 1996) the information problem is relatively straightforward, wherein asymmetric information plays a role in pre-contracting, i.e., where there is adverse selection, and intermediaries are endowed with the ability to screen heterogeneous agents. Agent behaviour post-contracting in these models is not subject to asymmetric information. In others (such as de la Fuente and Marín, 1996; Blackburn and Hung, 1998; Morales, 2003; Aghion et al., 2005) there is a post-contract incentive for agents to shirk or deceive because of, respectively, an aversion to effort or an ability to hide research outcomes. Such moral hazard issues thus bring the modelling of static intermediation closer to reality, but often simply add another wedge between agents and firms, scaling up intermediation costs and so, *ceteris paribus*, scaling down balanced growth rates.

The implications for policy in these models are, in general, limited to advocating liberalised financial markets and efforts to increase the efficiency of banks and markets while providing the institutional support required to diminish the costs of moral hazard and enforce contracts. The book by Rajan and Zingales (2004) is a prominent example of the sorts of policy prescriptions derived from this literature. In terms of theory, King and Levine (1993b) show that a simple tax on income from financial intermediation will have a monotonic effect on the level of intermediation and so on growth. An interesting result is that of Morales (2003), where effort-averse entrepreneurs with limited liability can be influenced by being subject to bank monitoring. It is shown that, under certain conditions, it is possible that a research subsidy (one direct to the entrepreneur) will accentuate the moral hazard problem and actually reduce growth. It is suggested, therefore, that policy used to stimulate growth should concentrate on financial intermediation and that the optimal tax on research can be non-zero. Aside from that paper, we must be careful to take policy implications from models in which the absence of *any* governance is Pareto optimal. The lessons on appropriate institutions drawn from, among other things, the work of Demirgüç-Kunt and Levine (2001) have, as yet, not impacted significantly on theories of finance and growth.

In short, the theory reviewed briefly here suggests that greater financial efficiency (be it in providing insurance, pooling resources, screening entrepreneurs or monitoring borrowers) reduces the disincentive to, or increases the feasibility of, entrepreneurship or the accumulation of human capital, thus increasing the rate of technological progress and consequently also the long-run growth rate of the economy. In general, a country with higher financial efficiency, by way of lower information asymmetries or better financial technologies, will have a higher long-run economic growth rate. The relationship is monotonic and (normally) continuous; larger differences in efficiency obtain larger differences in long-run growth.

A key component missing from most of these models is a consideration of their *quantitative* implications. However, as we demonstrate in Chapter 2, even a simple numerical representation of the models outlined here must be treated with circumspection. The models described here are, for the most part, essentially cross-section theories: Changing parameter values results in comparative balanced growth paths for separate economies, not the movement within a single economy from one growth path to another. Or, rather, we could think of these models as reflecting changes within a country that exhibits no transitional dynamics.

Without doubt, the development of a theory that can easily be interpreted numerically will be a topic for future work. Doing so would help in three ways: It would aid an understanding of the principle mechanics behind most finance and growth theories; it would create the potential for testing against data; and, it would enable the implications of the theory to be placed in an historical context. In this regard, the recent work of Townsend and Ueda (2006) is a major advance. They build upon the work of Greenwood and Jovanovic (1990), solving the model analytically for transition growth paths. Again, financial structures are formed on the basis of exogenous fixed and marginal costs to using financial services. These costs, at the heart of the theory, are not explored. More significantly, the analytical model is calibrated to micro-level data for the Thai economy. They compare the quantitative implications of the model to data from Thailand for the period 1976-1997. There is here a direct connection between rich microeconomic data and the testing of a microfounded dynamic general equilibrium model. This exercise is successful in capturing a number of the observed dynamic interactions between finance, inequality and growth.

The intention of this thesis, however, is to consider the finance-growth relation in the context of industrial take-off. Further, we wish to view financial conditions as a *product* of their environment, not as exogenous constraints on borrowing, investment or exchange. As such, we need to look more closely at the evidence on financial intermediation and growth before proceeding to construct a model that might capture them.

We develop in Chapter 2 a representation of prevailing theoretical mechanisms that can be easily calibrated to historical data. In order to see the relevance of such an exercise we first must place the theoretical implications into a more general context. We discuss in Section 1.2 the coherence of the finance and growth literature as a whole. To facilitate this discussion, we first need to consider briefly the question of finance and growth in the light of the existing historical literature.

1.1.3 Historical Evidence

Historical and cliometric research can have a lot to add to our understanding of any subject. If our goal is to answer questions about the necessary preconditions for developing countries to enter a sustained period of higher growth, then cross-section analyses of developed countries or theories based on a static intermediation problem can only aid us in a limited way. It is clear that the onus on establishing more rigorous empirical evidence will lead to much fruitful historical research, and a few papers have already begun in this direction. A consideration of the relationship between finance and growth in countries going through a period of transition might tell us more about the dynamics at play. Were there important changes, for example, in the way in which agents raised finance over time? Did legislation inhibit the emergence of the financial sector? Why, and how, do different financial structures emerge? Why did the UK industrialise first, despite not being the first to develop a sophisticated banking system? Are there any cliometric tests which we could impose on theoretical models of finance and growth?

Rousseau and Sylla (2005) combine a long historical US-wide dataset with contemporary dynamic econometric techniques. Their dataset is comprised of aggregate data covering the initial emergence of the financial structure we see today, over the period 1790-1850. They argue that initial financial developments "placed the United States of the early 19th century on a trajectory of economic growth higher than that of other nations... The US financial system did (and does) what a modern financial system is supposed to do, namely mobilize and efficiently allocate capital, and provide opportunities for risk management" (Rousseau and Sylla, 2005, p. 21). Additional moves to present the empirics of finance and growth in an historical context include Rousseau and Wachtel (1998) and Wright (2004).

Bordo and Rousseau (2006) follow Rousseau and Sylla (2005) and embark upon a long-run analysis of the finance-growth link, and move to consider what they term 'deeper fundamentals'. Considering a number of case-study countries, again on a aggregate basis, they add parameters for legal origin (intended to capture a country's inherent attitude to property and contract rights), the political environment and other factors into regressions on finance and growth. Conclusions from this analysis are not clear since, "...there remains a substantial component of financial development that is correlated with growth and yet not related to these measures of deeper fundamentals." (p.26).

We wish for empirical analyses of both contemporary and historical data to enter into decisions made about the nature of a stylised theory of finance and growth. The historical research, out of practical necessity, considers *aggregate* financial depth. But the finding that financial depth may have led periods of sustained growth in a number of countries does not mean that increasing financial efficiency by cutting down on moral hazard and adverse selection will do so also.[5] This distinction is central to the disconnection between the theory and empirics of finance and growth. The theory motivates financial frictions and their effect on growth based on a microeconomic information gap between savers and investors. Empirical work looks for macroeconomic relations between aggregate financial depth and aggregate growth. If we are to attempt to consider messages emanating out of each approach as a single literature, we must specify then the relationship between financial efficiency and financial depth, something which is largely absent from the finance and growth literature. A deeper historical analysis may provide the key, however.

Cliometric analyses of the type outlined above cannot, by themselves, support theories based around information and the efficiency of intermediation. To do

[5] A novel and highly interesting exception, slightly out of place in this analysis, is that of Jayaratne and Strahan (1996), which demonstrates that bank liberalisation in the US increased economic growth for efficiency, rather than depth, reasons.

so would need a detailed consideration of the ways in which banks and markets emerged over time: An analysis of the role of asymmetric information and entrepreneurship in forming the financial structures observed around the period of industrial revolution, and not just of aggregate measures of financial depth. Large, national financial institutions did not appear overnight but were the response to economic incentives that emerged over time, building the financial structures we see today from informal coalitions of agents that saw the initial incentives to act as intermediaries.

Some, more normative, historical research has begun to approach these considerations in the context of finance and growth. Wright (2002) provides some evidence to support the mechanisms through which financial institutions can facilitate economic growth by compensating for asymmetric information conditions, and so backs-up both the empirical evidence in favour of the finance-led growth hypothesis and the dominant theoretical models. As Wright (p.212) notes, "Problems of information asymmetry, namely adverse selection, moral hazard, and the principal-agent problem, collude to limit effective lending." The author suggests that the early US financial system was in fact much more effective than previously believed, and invokes Adam Smith as being among the first to describe the ways in which banks spur growth by addressing information problems. The central part that asymmetric information plays in determining the efficacy of financial institutions in engendering sustained levels of high economic growth is the message of this work.

But such analysis does not get to the question of whether such financial structures emerge as a result of economic necessity or whether economic growth, and industrial take-off, can actually be forestalled by an inadequate financial system with the implication that an exogenous improvement in the financial environment will facilitate takeoff. As we shall see in more detail in Chapter 3, the broader historical consensus on UK growth is that the role of finance in *determining* industrial development was at best limited. Cottrell (1980), Harris (2000), Shea (2007), and others, cite both the ease with which a firm could find initial finance and the ubiquity of profit-ploughback as a means of expansion. It is also shown that a great deal of early financial intermediation was decentralised, where often the regional manufacturing industries opted for local finance and not the use of the London capital markets. Depicting only the growth of a national financial system thus masks a great deal of complexity and dynamism regarding the relationship between entrepreneurs and financial intermediaries. We leave a fuller analysis of these important issues to Chapter 3. For the moment, we discuss in Section 1.2 the congruence of the various aspects of finance and growth described above. This serves to motivate the remainder of this thesis.

1.2 Discussion

We have seen that the finance-growth nexus operates along at least four dimensions: The size of the financial sector as a proportion of the economy (financial depth); the effect of institutions and regulations on the efficiency with which financial services are provided (financial efficiency); the nature and extent of asymmetric information (both moral hazard and adverse selection); and the extent of disaggregation. In addition, each of these evolves over time. Viewed from this perspective, the literatures surveyed in this Chapter typically address the finance-growth nexus in an incomplete way.

It has been shown that applied econometric work considers financial depth while holding efficiency to be exogenous. By contrast, most theories consider financial efficiency holding depth to be constant. This clear disconnection has significant implications for the reconciliation of applied and theoretical work: Applied (theoretical) research of this sort cannot without qualification be held to support theoretical (applied) conclusions.

Quite apart from the mapping between the theory and empirics of finance and growth, there has been little in the way of historical motivation or cliometric testing in the standard approach to most theoretical modelling. Most theories consider asymmetric information and financial efficiency but pay little heed to questions of disaggregation; there is no scope for sub-national coalitions of financial intermediaries to provide services more efficiently.

The time-series historical analyses based on financial depth have demonstrated clear and consistent results that supplement what was learned from the cross-sectional research. This literature tells us that the level of financial depth *does* change over time, and that theories of static intermediation thus miss an important element of the story which robust econometric analyses suggest is so important. But in terms of *understanding* if and why financial depth leads economic growth we need to understand the reasons why financial structures emerge. Economic theory justifies the existence of banks, or more generally coalitions of agents who provide finance, by appealing to microeconomic conditions such as asymmetric information and risk aversion. It is necessary, if we are to endogenise the financial development of an economy going through industrial transition, to consider historical experience in this light.

Instead of appealing to such historical beacons, decisions about modelling are typically driven by cross-section evidence in a way that limits the time-series implications of the theory, i.e., there are few meaningful transitional dynamics. As we show in Chapter 2, a major drawback with present theory is that, even if we can back-out some indicative path for financial efficiency, this transition would be exogenous (i.e., the time-series properties would be imposed to fit the data). We must look properly at the historical record for clues to understanding the mechanics behind the degree of asymmetric information, and in this way develop a theory of endogenous growth and *endogenous financial development*. We begin to do this in Chapter 3.

The historical evidence presented in Section 1.1 suggests that the relationship between agents and financial intermediaries is dynamic and disaggregated, and where the level of financial depth is not constant. The mechanics of theories such as King and Levine (1993b) appear, prima facie, to be difficult to square with with the historical evidence presented here. The numerical implications of these models raise further questions since they are also difficult to reconcile with the historical literature. In general, the static approach to modelling finance and growth in which financial depth is not endogenously changing and where aggregative factors are not considered, is thus inappropriate and the implications of the models described in this survey must be considered with caution.

The problem is in establishing the quantitative significance of each of these aspects. If we can work towards uncovering the richness of the dynamic interplay between asymmetric information, financial structure, financial depth and economic growth, during the transition to an industrial economy, then we would have a new, more historically congruent, micro-founded, theory of finance and growth. Townsend and Ueda (2006) is an interesting start to work in this direction, matching rich, disaggregated data on Thailand to a theory of inequality, financial deepening and growth. But a step further than this would be to consider a long time series of a country incorporating the pre-industrial, take-off, and post-industrial phases of development with an understanding of the morphing financial depth, financial efficiency, asymmetric information and aggregation issues all coming into play.

1.3 Concluding Remarks

Applied and theoretical research on any question in economics cannot be considered in isolation from each other. We have argued that the theoretical, contemporary econometric and historical literatures on the finance-growth nexus are if not contradictory then at best simply disconnected. An attempt at reconciliation will need to move beyond the concentration on contemporary econometrics, beyond the assumption of static information asymmetry, and beyond the conception of aggregate variables alone.

These criticisms apply equally to empirical, theoretical and historical research. Future work will thus need to identify the key features of the interaction between finance and growth over continuous periods, such as the industrial revolution. The historical literature surveyed briefly here strongly suggests that current theories of finance and growth do not depict adequately the experiences of countries going through industrial revolution. A potentially more fruitful avenue for research will be to establish the historical experience of industrialisation, asymmetric information and intermediation, and then construct a growth theory founded in microeconomics that more faithfully reflects it. Understanding the relationship between increasing financial depth and evolving conditions of asymmetric information through a period of industrial revolution is required as a first step.

Chapter 2

Quantitative Implications of Finance and Growth Theory*

As described in Chapter 1, the connection between finance and economic growth has been the subject of increasing attention over recent years. The majority of this attention has been on its empirical aspect, however. Though the implications of the empirical results cannot be taken without qualification, the core messages emanating from this research have been consistent and forceful. It has been shown that the extent of financial development in an economy, at least in terms of depth, is related to the level of sustained economic growth.

Work such as King and Levine (1993a) has gone so far as to suggest that we can draw predictions about the rate of economic growth over ten to thirty years based on the extent of financial depth. By their nature, these sorts of implications, and those of succeeding papers, are open to further validation: In short, are they realistic? Driffill (2003) suggests that they are not, partly because of the strength of outliers (specifically, the Asian 'tiger' economies) in driving the results, and partly because of the simple implausibility of the predictions.

*Part of this chapter has been released as Trew (2006a).

The contemporary empirical work has more recently been augmented by longer time-series analysis such as the cliometric work of Rousseau and Sylla (2005) and Bordo and Rousseau (2006). The preliminary findings go some way to supporting those of the cross-section work on the latter half of the twentieth century. Specifically, they find a relationship over time between the level of aggregate financial depth and the transition to a higher level of sustained economic growth. Questions of causality, interaction with legal origin and the effect on technological progress are still open.

These and other issues were explored further in Chapter 1 so we leave them to one side here. Instead, having noted the capacity to hold empirical implications to account in this way, we begin to address the absence in most of the theoretical literature of any quantitative, and therefore testable, implications. Despite the near-ubiquity of calibration to data in the macroeconomic literature, growth theory remains largely untested in this respect. Indeed, the absence of numerical application is seen as one of the key weaknesses of the new growth theory which grew out of the work of Romer (1986, 1990) and Lucas (1988). A strong critique of the applied relevance of the endogenous growth literature is Parente (2001). It is the intention of this Chapter to make a first attempt at confronting extant finance and endogenous growth theory with data for historical growth paths.

Quantitative Finance and Growth

At the same time as finding relationships between financial development and economic growth, work such as King and Levine (1993b) seeks to bolster empirical findings with the development of theory relating financial matters to the determinants of growth. The tacit implication is that the empirical findings are supported by, and support, the theoretical results. But while the empirics typically consider cross-section regressions on aggregate financial depth, the theories, with some exceptions, relate measures of financial efficiency to economic growth. In these theories, there is assumed to be some wedge between savings and investments which acts to reduce the rate of technological progress or human capital accumulation by dampening entrepreneurial or educational activities. In order to consider results from each approach as a single body of research, the connection between these frictions and the level of aggregate depth then needs to be made explicit. But not only is this connection typically absent, each approach implicitly answers a different question. Indeed, it will be argued that, for a wide class of theories, balanced growth actually implies that the economy needs to obtain a constant level of financial depth. In general, then, the theory cannot without qualification be held to support the numerical implications of empirical work.

The empirical literature has moved towards increasingly rich analyses. Work such as Beck et al. (2005) has looked at the differing impact of financial development on firm size and growth. A number of papers have also begun to consider factors which determine the efficacy of finance in influencing growth. Beck and Levine (2005) and Bordo and Rousseau (2006) are examples in the context of legal origin.

Work such as Aghion et al. (2005), however, has continued to focus on the juxtaposition of cross-sectional econometrics with theory without explicitly stating the connection between them. There is, in this approach, no calibration of theoretical results to empirical facts, and no attempted simulation of time paths or cross sections found in the data. This sort of interdependence at the frontier of research between theory and evidence is the lifeblood of contemporary macroeconomics, and could yet be so in the economics of growth.

Outline of this Chapter

This chapter attempts to reduce the key mechanisms at work in many theories of finance and growth to a single model that can be calibrated to data. Given the lack of the sorts of historical data required, it is our intention to minimise the number of free parameters in the model. This allows us to interpret observed growth rates in terms of *implied* historical financial efficiency. The simplicity of the resulting model reflects the stripped-down nature of our approach. Nonetheless, we demonstrate the validity of this simple model by developing a less reductive endogenous growth model with quality ladders. This second model also serves to reinforce our argument that the theoretical and applied approaches to finance and growth are, in some senses, disconnected.

The intention is to develop quantitative implications which are transparent enough to allow an interpretation in terms of historical growth paths and cross-section results, though the focus will be on the former. We thus concentrate on the calibration of the simple model. To the author's knowledge, this is the first attempt to calibrate prevailing finance and growth theories to match quantitatively either historical time-series or cross-section empirics.

We first provide in Section 2.1 a brief recap of key theoretical contributions, and argue that they can be considered in the context of a small number of core mechanisms. We hope to keep any overlap with Chapter 1 to a minimum. We then develop in Section 2.2 our parsimonious model. 2.2.1 presents a version in the manner of King and Levine (1993b). This initial model with adverse selection is calibrated to an historical time-series of growth in the UK. We draw implications for historical and cross-sectional financial efficiency in this context. Section 2.2.2 extends the model to include moral hazard considerations, and again develops quantitative implications. Section 2.3 presents a more involved endogenous growth model with quality ladders, and suggest that our simple model is a parsimonious one. Section 2.4 concludes with our main findings.

2.1 The Prevailing Mechanics of Finance and Growth Theory

The intention then, in this Chapter, is to reduce the most commonly cited finance and growth theories down to their core, laying bare the central mechanisms through which finance is said to influence the rate of economic growth. We do not wish to go as far as Pagano (1993) in simplifying the finance-growth link. In that model, the wedge between investment and savings is simply posited. Instead, we wish to model, at least following the spirit of prevailing finance and growth theories, some level of detail in the microeconomic roots of financial frictions. We do not derive an empirically motivated theory; we intend only to reflect the state of prevailing thought in finance and growth theory. The representative theory will then be considered in the light of its implications for time-series growth paths and for cross-section comparative statics. This allows us to turn back to the more detailed evidence on the realities of finance and growth in Chapter 3.

Capasso (2004) and Levine (2005) go through in some detail the nature of a good deal of the theoretical literature. While a number of economists have long questioned any suggested role of financial matters in determining growth, the time-series evidence has shown that the two are at least positively related. Of course, in an Arrow-Debreu world there is no necessity for financial intermediation, and so no possible link between financial matters and the performance of the economy. In a microeconomic sense, in order to motivate the existence of specialised financial structures, we need to introduce frictions between those who save and those who wish to invest. These frictions can arise because of at least incomplete but also perhaps asymmetrical information. The symmetrical incompleteness of information is not often held to be the only motivating factor, though Saint-Paul (1992) shows that it is, of course, possible in the presence of economy-wide uncertainty. Most theories consider asymmetrically incomplete information to be the underlying cause of financial structures. This line of thought has been pursued by Capasso (2004), among others. Capasso demonstrates the similarities in many of the finance and growth theories that incorporate asymmetric information.

Given the focus on asymmetric information, we are left with choices regarding the source of economic growth, be it via exogenous changes in TFP or endogenous growth, and the nature of the asymmetry, be it adverse selection and/or moral hazard. In the main, the source of economic growth has followed the trend set by the new growth literature and included either production externalities or imperfectly competitive intermediate sectors which compete to increase either the quality or quantity of intermediate goods. The majority of those concerned with finance place at the heart of the growth mechanic some form of entrepreneurship in the pursuit of technological advance, though work such as de Gregorio and Kim (1998) does consider the accumulation of human capital.

As shown in Capasso (2004), the nature of the information asymmetry, whether *ex ante* or *ex post* private information, is an important determinant of the resulting financial structures. King and Levine (1993b) is the original finance and growth model with adverse selection, in which entrepreneurs are screened by a financial intermediary to determine their quality. The screening is costly, and the chosen entrepreneurs develop better quality intermediate goods with some known probability. In Bose and Cothren (1996), banks can choose potential creditors with either a costly screening technology or by designing a separating contract, or by a mix of the two.

Models that motivate financial structures by the presence of moral hazard include Blackburn and Hung (1998) and Morales (2003). In each there is a post-contract incentive for agents to either deceive or shirk. The latter is an interesting extension to Howitt and Aghion (1998), which includes a role for taxation in achieving an optimal level of research activity. This is one of the very few theories of finance and growth in which zero taxation is not Pareto optimal.

Most models specify the financial intermediation condition to be a static relationship between lenders and borrowers. Blackburn et al. (2005), however, combines problems of both moral hazard and adverse selection in a model that can allow for the co-existence of both stock markets and banks. That paper, along with Greenwood and Smith (1997), allows for the nature of financial structures to change as the economy develops. Blackburn et al. considers the joint-determination and co-existence of debt and equity finance. In these sorts of model, where there exists a decision over the nature of finance and so a relationship between economic development and optimal financing method, the financial intermediation condition is, in a sense, 'dynamic'.

We can draw together some key aspects of the mechanisms underlying predominant finance and growth theories. Something akin to entrepreneurship drives the accumulation of either human capital or better technologies. The efficiency with which the motive to innovate or accumulate translates into actual growth is determined by the ease with which entrepreneurs can obtain finance for their risky projects. With asymmetric information in the financial sector, this efficiency is dependent upon the sophistication of the financial technologies, and on the extent of moral hazard or asymmetric information. Additionally, as the economy becomes richer, so it can afford those financial structures that better facilitate higher economic growth.

A recent, and pathbreaking, analysis is that of Townsend and Ueda (2006), which builds on the theory of Greenwood and Jovanovic (1990) to include transitional behaviour. They develop a dynamic general equilibrium model of an economy with evolving levels of financial depth and economic inequality. Financial structures exist because of the imposition of fixed and marginal costs to exchange; i.e., the information problem is not explicit. Still, the model presented in Section 2.2 of this Chapter demonstrates why the approach of Townsend and Ueda is so superior to that which has gone before it. To reiterate, the purpose of this Chapter is to demonstrate the dangers of drawing conclusions from arguments which omit the sort of direct connection between data and theory propounded by Townsend and Ueda. A theory of economic growth and financial efficiency, however it is motivated, cannot be held to support, and nor can it be supported by, empirical relationships between measures of aggregate financial depth and economic growth.

In Section 2.2 we develop a model in the spirit of King and Levine (1993b) which links financial matters to economic growth. The model includes: A role for entrepreneurship in the accumulation of human capital, as facilitated by the existence of an intermediary-banking sector; asymmetric information between lenders and entrepreneurs in the form of moral hazard and/or adverse selection; and an interaction between economic growth and financial development. We invoke a model in which entrepreneurs wish to obtain finance for investment in their own human capital, rather than for an addition to technology. It will be clear that for our purposes the difference of each approach is minimal. The model is then developed numerically, with quantitative implications that can be compared with both cross-section empirics and time-series data.

The largest omission from this simple model is a role for the co-evolution of stock markets alongside banks. We leave this aside in order to keep the model and its calibration simple. This omission is not necessarily to the detriment of our conclusions. Indeed, as suggested by Blackburn et al. (2005, p. 145), "our analysis invites one to think of debt and equity as complementary, rather than substitute, means of corporate finance." This echoes the previously noted findings of Demirgüç-Kunt and Levine (2001). A further, related, aspect is the existence of a 'critical mass' effect in financial development, of which Saint-Paul (*op. cit.*) and

Bose and Cothren (*op. cit.*) are examples. We view these aspects as important, but, again, ultimately not significant for the numerical simulations we propose.

2.2 A Parsimonious Model of Finance and Growth

The purpose of this section is to outline an endogenous growth model that can capture the principle mechanics of significant theoretical works. It also reflects in part the historical debate on the nature of asymmetric information, both in terms of adverse selection and an extension to include moral hazard. We calibrate the model to historical data for the UK and so trace out the implied 'transition path' for financial efficiency over the period of the industrial revolution.

We will take the financial intermediation relationship to be static, i.e., the way in which intermediaries and agents interact does not explicitly change over time. In addition, we assume that there are no arbitrary credit constraints so that the causes of friction are entirely informational. With a suitable model we can thus use numerical methods to compare quantitatively the implications of such models for time-series growth with the historical pattern of industrial finance and growth.

The mechanism by which finance affects long-run growth follows the trend suggested by the theories discussed in Section 2.1: Ever since King and Levine (1993b), the majority of theories linking finance to growth revolve around entrepreneurship and either human capital accumulation or technological progress. We adopt that perspective also.

2.2.1 Financial Intermediation and Growth

In the model of King and Levine (1993b) intermediaries are effectively venture capitalists that have the technology necessary to screen potential entrepreneurs who are then employed and given funds to run a research project. Prospective entrepreneurs differ by type in their ability to carry through a research project. The fruit of such labour is an addition to the stock of knowledge (specifically, via a quality-ladders setup *à la* Grossman and Helpman, 1991). The screening cost is fixed and the intermediary consequently knows with certainty the ability of the applicant. There is no costly effort (so no moral hazard), and the intermediary market is perfectly competitive. Reductions in the cost of screening or in the tax on intermediary profits thus increase the efficiency of the financial sector, increase the rate of technological progress and so increase the rate of long-run growth.

Outline

In this model firms demand physical capital and human capital. We have a continuum of agents in each household of total mass one, and a random distribution of type within each. If we assume a large number of households then in the aggregate we can work with the average distribution of type within a given household. So, on average, a proportion φ_1 has no ability to acquire human capital whatsoever, a proportion φ_2 has low ability Λ' and the remainder, proportion $\varphi_3 = 1 - \varphi_1 - \varphi_2$ has high ability $\Lambda > \Lambda'$. It is important that able agents do not know their own level of ability, only that they have some. If agents knew their level of ability, given that the screening technology of the intermediary identifies ability with precision, and given also that agents know this, there would be no reason for those with less than high-ability to apply.

Agents with no ability take household responsibility for selling physical capital to firms. Only agents with high ability have the potential to develop human capital. All agents with nonzero ability apply to a financial intermediary to be screened. Those that are rejected do not contribute to household income. Those that are accepted are consequently funded by the intermediary to acquire education or conduct research, becoming human capital with fixed probability β. In the case of education this might reflect the likelihood of not dropping-out; in the case of research this might reflect the probability of useful innovation. Either way, we obtain the same result. Those that fail to develop human capital contribute nothing to household income, those that do develop human capital are consequently employed by firms and enter the production function as human capital. In the event that the agent succeeds in acquiring human capital it is the researcher that owns the human capital, paying a proportion t of income from human capital to intermediaries. The intermediary thus sets t to maximise expected profits.

Firms

Firms use human capital, H, and physical capital, K, as inputs to the production process, $Y_t = AK_t^{\alpha}H_t^{1-\alpha}$.[6] Each firm maximises profits, $\pi_t = Y_t - rK_t - hH_t$, where each takes the rates of return on physical capital, r, and human capital, h as given: $r = \alpha(Y_t/K_t)$ and $h = (1-\alpha)(Y_t/H_t)$. We can use equation for h to obtain the firm's demand for human capital, $H_t = [(1-\alpha)Y_t]/h$, which, upon substitution into the production function, obtains a form of the familiar Ak endogenous growth setup,

$$Y_t = \left[A \left(\frac{(1-\alpha)}{h} \right)^{1-\alpha} \right]^{\frac{1}{\alpha}} K_t. \tag{2.1}$$

Once we have found a relationship between the rates of return on human and physical capital, we can treat this model as one in which externalities to production are just enough to generate constant returns and 'Ak' growth. Following Barro and Sala-i-Martin (2004), we can then think of K as something like a proxy for a composite capital variable. In doing so, we assume that the $H : K$ ratio is constant. As in most of the extant theory, transitional dynamics will not exist here.

[6] In order to more faithfully reflect King and Levine (1993b), we might have thought of entrepreneurs adding to the stock of knowledge via a compound coefficient of technological progress of the form $Y_t = A\tilde{A}_t^{1-\alpha}K_t^{\alpha}$. This difference would not matter for the purposes of our simple numerical simulation.

Intermediaries

The intermediary incurs the cost $f(H) > 0$ to screen agents for ability and funds successful applicants to acquire human capital at cost $x(H) > 0$. Note that these costs are not invariant to the level of human capital, and we make the assumption that $f' > 0$ and $x' > 0$, i.e., that the costs of intermediation are proportional to the size of the demand for human capital. So both the outlay required to fund the acquisition of human capital, x, and the cost of screening candidate acquirers of human capital, f, is increasing in the level of human capital – a reasonable assumption if we imagine that the higher the level of human capital aspired to, the more costly it is to both fund and identify suitably able agents. This ensures that the costs of intermediation do not become insignificant over time as a proportion of the size of the economy.

We demonstrate in Section 2.3 that there is an analogous requirement for balanced growth in a quality-ladders setup so it may be a general result that for balanced growth in these simple economies with static financial intermediation we require that the size of the financial sector is constant over time. This explains why, even though (because of data limitations) econometric analyses consider largely financial depth, most theory considers financial efficiency; within an endogenous growth framework it becomes difficult to solve analytically for balanced growth when the size of the financial sector relative to the economy is changing over time. In an economy going through industrial transition, the size of the financial sector does change significantly. It may be that to reconcile these facts, i.e., for both the balanced growth rate and the level of financial depth to change endogenously over a period of industrial takeoff, we require the financial condition to be dynamic, instead of the static relationship depicted below.

We also require that it is not feasible for households to fund the amount $x(H)$ from their own resources. For a given agent, expected intermediary profits will be the probability-weighted incomes and expenditures. The probability that an agent who applies will be of low ability is $\varphi_2/(1-\varphi_1)$, in which case only the screening cost is expended. The probability of successfully developing human capital from high-ability agents and thus obtaining a rent from him is $\beta(1-\varphi_1-\varphi_2)/(1-\varphi_2)$. If we assume competition then the expected intermediary profit is zero,

$$\begin{aligned} E(\pi) = {} & \beta\left(\frac{1-\varphi_1-\varphi_2}{1-\varphi_1}\right)[thH - x(H) - f(H)] + \\ & + (1-\beta)\left(\frac{1-\varphi_1-\varphi_2}{1-\varphi_1}\right)[-x(H) - f(H)] + \\ & + \left(\frac{\varphi_2}{1-\varphi_1}\right)[-f(H)] = 0. \end{aligned} \tag{2.2}$$

Let us specify $x(H) = \eta_x hH$ and $f(H) = \eta_f hH$, where $\eta_x > 0$ and $\eta_f > 0$ are the cost parameters of intermediation, then we obtain a simple expression for the fee charged by the intermediary.[7] On the face of it, this restriction is not an unreasonable one: Both the costs *and revenue* of financial intermediation are proportional to agents' income from human capital, hH. We thus have a form of static financial intermediation. This specification allows us easily to derive balanced growth rates and it reflects the manner in which financial intermediation is modelled in the literature. However, as we will see in Chapter 3, such an assumption is quite likely to be inappropriate. The intention of this Chapter is to test the numerical implications of extant theories of finance and growth, so we proceed for now with static intermediation.

Using $x(H) = \eta_x hH$ and $f(H) = \eta_f hH$ with equation (2.2) we obtain,

$$t^* = \frac{1}{\beta}\left\{\eta_x + \left[\frac{1-\varphi_1}{1-\varphi_1-\varphi_2}\right]\eta_f\right\}. \tag{2.3}$$

[7]It is possible to generalise this functional form but the consequences for the model are not significant. Specifically, we can specify $y(H) = \eta_y H$, wherein we obtain $t^*(h)$ and a less simple form for the interest rate. We prefer the simpler form here since the problem becomes intractable when we consider the full model with moral hazard.

Equation (2.3) is increasing in the costs of financial intermediation, η_f and η_x, and in the share of low ability agents, φ_2, and decreasing in both the probability of human capital creation, β and the share of high ability agents, φ_3.

Households

The cost t^*hH is borne by consuming households. The household receives income from physical and human capital, however, at the rates r and h respectively. Using equation (2.3), the household budget constraint will thus be the familiar $c_t + \dot{k}_t = rk + \tau(1-t^*)hH$. We mirror King and Levine here by incorporating a tax on income from innovation, where $1-\tau$ is the tax rate applied to household income from human capital. Households maximise the discounted present value of future consumption,

$$\max_{c_t} U = \int_0^\infty e^{-\rho t} u(c_t) dt, \tag{2.4}$$

where $u(c_t)$ is the instantaneous utility function. Assume CES preferences of the form $u(c_t) = (c_t^{1-\theta} - 1)/(1-\theta)$. The Hamiltonian is,

$$\mathcal{H} = e^{-\rho t}\left(\frac{(c_t)^{1-\theta}-1}{1-\theta}\right) + \lambda\left[rk + \tau(1-t^*)hH - c_t\right], \tag{2.5}$$

given initial capital stocks. We also have the transversality conditions, $\lim_{t\to\infty} e^{-\rho t}\lambda_t k_t = 0$. The first order conditions are,

$$\frac{\partial \mathcal{H}}{\partial c_t} = e^{-\rho t} {c_t}^{-\theta} - \lambda = 0; \tag{2.6}$$

$$\frac{\partial \mathcal{H}}{\partial k_t} = \lambda r = -\dot{\lambda}; \tag{2.7}$$

$$\frac{\partial \frac{\partial \mathcal{H}}{\partial c_t}}{\partial t} = -\rho e^{-\rho t}(c_t)^{-\theta} - \theta e^{-\rho t}(c_t)^{-\theta-1}\dot{c}_t = \dot{\lambda}. \tag{2.8}$$

From the first order conditions we obtain the standard Euler equation governing the growth rate of consumption, $\dot{c}_t/c_t = \theta^{-1}(r-\rho)$.

Equilibrium Growth

In this setup, we mirror the literature by making no distinction between equilibrium and balanced-path growth. Hence, using the Euler equation, in equilibrium we require that the net return on capital is equal to the net return on human capital, i.e., that $r = \tau(1-t^*)h$.[8] From the production function, equation (2.1) we have the following expression for r,

$$r = \left[A \left(\frac{(1-\alpha)}{h} \right)^{1-\alpha} \right]^{\frac{1}{\alpha}}. \tag{2.9}$$

By the equilibrium financial intermediation condition, $h = r/[\tau(1-t^*)]$, we may solve for r from,

$$r = A[\tau(1-\alpha)(1-t^*)]^{1-\alpha}. \tag{2.10}$$

Hence, we have a simple closed-form solution for the equilibrium growth rate,

$$\gamma = \frac{1}{\theta} \left\{ A[\tau(1-\alpha)(1-t^*)]^{1-\alpha} - \rho \right\}. \tag{2.11}$$

An increase in the efficiency of financial intermediation, by reducing η_f or η_x *ceteris paribus* results in an increase in the equilibrium growth rate by reducing the cost of intermediation, t^*. So there is simply a wedge in between what firms pay for human capital and what agents receive, where the significance of this wedge reflects the efficiency of financial intermediation. This is the main theoretical result of King and Levine (1993b). Inasmuch as we can call exogenous changes in η_f changes in financial efficiency over time within a country we can now calibrate this model and consider its quantitative implications for historical growth.

[8]This is akin to the argument in Tsiddon (1992): "I assume that each financial intermediary can provide a risk-free return to lenders that is equal to or greater than the risk-free rate of return individuals can earn in the market for physical capital. Competition guarantees that each financial intermediary has zero profit." p. 305

Calibration

Using data from Crafts and Harley (1992) for the level of industrial production in the UK through the industrial revolution we can, with reasonable parameter values, trace back the implied efficiency of financial intermediation in this setup. We use here the 'revised best guess' (Crafts and Harley, 1992, Table A3.I) for the industrial production series. This is a standard reference for such data, and it shows a similar pattern to that in Bairoch (1982). The advantage of the Crafts and Harley dataset is that they provide annual values. We do not extend the data to the current day since the composition of output changed significantly, with a decreasing proportion of industrial production towards the end of the twentieth century. Figure 2.1 shows the path of the trend growth rate. We report the growth rate of HP-filtered series with both $\lambda = 100$ since this is annual data and with $\lambda = 10,000$ to show the general movement in growth. All these data are reproduced in Appendix Table B.2.

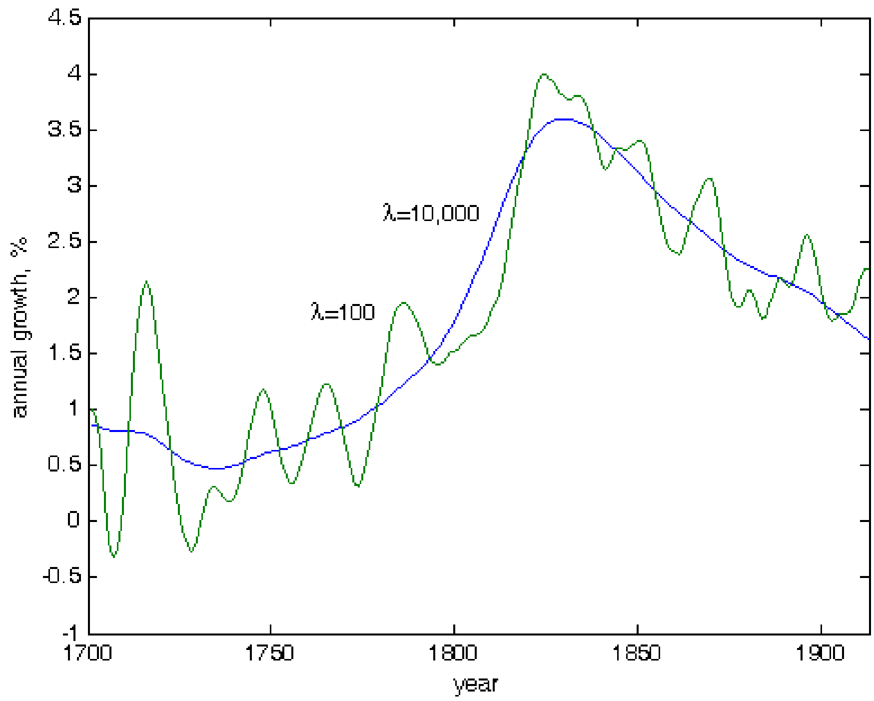
Whichever of the two weights we use in the filter the implications are the same. Trend growth in the UK increased through the early periods of industrial revolution and decreased slightly after around 1825. This is a relatively typical pattern for countries going through industrialisation during the eighteenth and nineteenth century, and the only exception seems to be the US.

We can use the growth trend with $\lambda = 10,000$ to find the implied value of η_f, *ceteris paribus*, by using the estimate of trend growth, $\hat{\gamma}$, and equations (2.3) and (2.11),

$$\hat{\eta}_f = \left[\frac{1-\varphi_1-\varphi_2}{1-\varphi_1}\right]\left\{\beta\left[1-\left(\frac{\theta\hat{\gamma}+\rho}{A}\right)^{\frac{1}{1-\alpha}}\frac{1}{\tau(1-\alpha)}\right]-\eta_x\right\}. \tag{2.12}$$

We must restrict some parameters for calibration purposes. For $\hat{\eta}_f > 0$ over the range of obvserved trend growth values we require $A[(1-\eta_x/\beta)(1-\alpha)]^{1-\alpha} > \hat{\bar{\gamma}}\theta + \rho$,

Figure 2.1: Trend UK Growth of Industrial Production



where $\hat{\bar{\gamma}}$ is the maximum growth rate observed over the sample. Bearing this in mind we use the parameter values given in Table 2.1 in simulations of this model. MATLAB code is reproduced in Appendix Section A.1. [9]

Numerical Implications

So we have inferred the historical *level* of financial efficiency, in the King-Levine mould, as that depicted in Figure 2.2. Here η_f is the implied proportion of intermediary costs that go to screening agents. We use this single measure to reflect implied changes in the efficiency of financial intermediation.

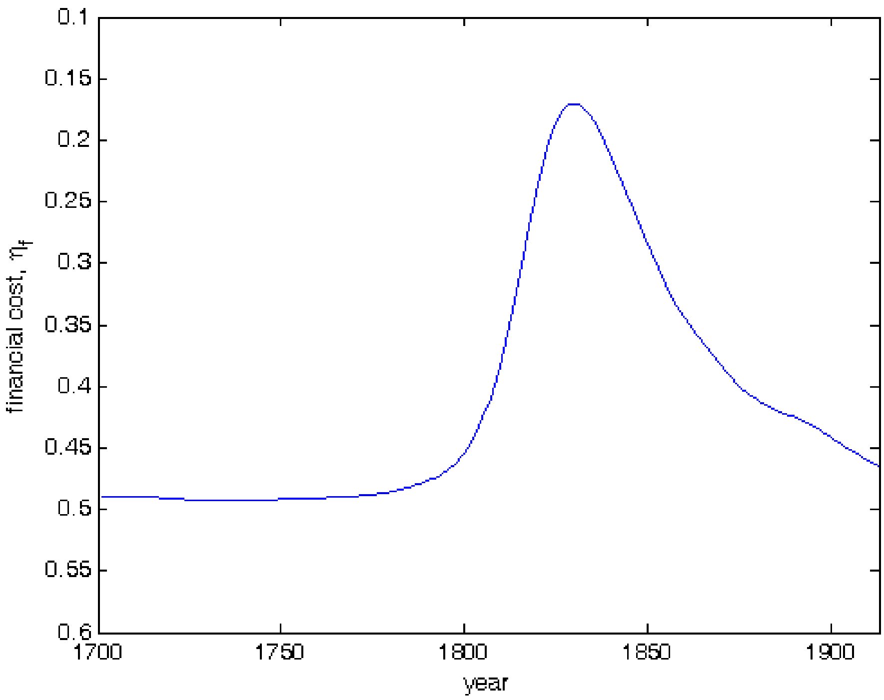
Here we have inverted the y-axis in order to reflect more clearly the movement of

[9] These results are considerably robust to changes in parameters, so long as we obtain positive values for η_f. Without significantly altering the *shape* of the implied efficiency data over time we can vary α between 0.1 and 0.9, φ_1 and φ_2 can take any value so long as $\varphi_1 + \varphi_2 < 1$, β can take any value in the interval $(0, 1)$, A can be varied widely (while satisfying the inequality for $\eta_f > 0$), θ can be varied (again, so long as we satisfy the inequality), and η_x can take any positive value below unity.

Table 2.1: Parameter Values for the Benchmark Model

capital share	α	0.67
subjective discount rate	ρ	0.02
elasticity of substitution	θ	5
human capital probability	β	0.75
proportion of zero ability	φ_1	0.4
proportion of low ability	φ_2	0.2
coefficient of technological progress	A	30
scale parameter on financial investment	η_x	0.01
tax parameter	τ	1

Figure 2.2: Historical Financial Efficiency in the UK



implied financial *efficiency* over the period. Plainly there is a great deal of implied movement in the parameter over time. We see an initially low level of financial efficiency at the beginning of the industrial take-off and a peak of financial efficiency at around 1830. The parameter then falls monotonically over the remainder of the sample. The path of financial efficiency mirrors the path of the trend growth rate and so the implied drop in financial efficiency reflects the fall in observed trend growth.

It should be noted also that we are only changing one variable; the implied path of financial efficiency would be affected if we were to account for the technological revolution by exogenously increasing A over the period. Nonetheless, the general shape of the path would remain. The implication of this finance and growth model is thus that the level of financial efficiency was, at the start of the industrial revolution, (relatively) low. Financial efficiency then increased up until around 1830 before dropping again, almost to pre-industrial levels.[10] This observation is an important one, and might be tested using appropriate historical data. In a sense it is surprising to consider a rapid rise and then decline of financial efficiency, and so this suggests that theories of finance and growth, in the main, cannot acceptably account for the dynamics of industrial take-off. Again, this question cannot move beyond speculation without further research.

In addition to these 'time-series' implications, we can consider the above results in the context of a cross-section analysis. Indeed, this might be the only acceptable implication since there are no dynamic elements to the transition from one level of financial efficiency in the model above; in essence, different levels of financial efficiency imply different balanced growth paths in different economies. Given the above calibration, we see that at low levels, differences in financial efficiency mean larger difference in growth. Changing η_f from 0.5 to 0.49 obtains growth path difference of 0.5 percentage points. Over higher levels of financial efficiency, however, to obtain a difference of 0.5 percentage points, η_f must change from, for example, 0.25 to 0.2.[11] The implication is that countries with low growth levels should exhibit lower variations in financial efficiency than those with high growth

[10]The process could equally well be applied to any other country with sufficient data; the growth patterns of many industrialised countries have mirrored this peak shape, see Bairoch (1982).

[11]This result holds up to a wide range of different calibrations, as detailed in footnote 9 above.

levels. At present, there are no empirical works that could validate this implication on a cross-country basis. The work of Jayaratne and Strahan (1996) considers the effect of bank liberalisation on economic growth in states of the US. They find a positive causal effect on growth via an increase in financial efficiency, rather than through increased financial depth. Perhaps partly because the focus of the article is elsewhere, they do not find the sort of nonlinearity implied by the model presented above.

2.2.2 Financial Intermediation, Moral Hazard and Growth

The degree to which the model presented in Section 2.2.1 reflects the state of the literature is limited. Most work on financial intermediation and growth considers some further role for asymmetric information, be it imperfect screening, costly state-verification or effort-aversion and costly monitoring. Here we will take the latter approach, along the lines of de la Fuente and Marín (1996) and Morales (2003).

In the King and Levine (1993b) model, the entrepreneur knows that there is no difference in his income between success and failure, i.e., he is fully insured, and yet he still supplies effort in the management of a research project, the success rate of which he has no influence over. In the modification presented in Section 2.2.1, the agent is not fully insured (though of course the clan of which he is part is fully diversified) but, still, he has no influence over the probability of acquiring human capital. It is likely, however, that there would be a relationship between the effort the agent puts into acquiring human capital and the likelihood, β, of it occurring. If agents are averse to effort there emerges a role for intermediaries in implementing a costly monitoring technology, where 'monitoring' is hereafter synonymous with 'controlling', to increase their expected income by forcing an increase in β.

Morales (2003) considers an endogenous growth model with financial imperfections but makes the probability of innovation endogenous. Researchers in Morales' model, analogous to the entrepreneurs of King and Levine, dislike effort and have limited liability, i.e., they pay back a certain amount less than their monopoly profits from starting up in the intermediate sector in the case of success, but do not suffer relative to their initial wealth in the case of failure. So there is a level of effort that the entrepreneur will provide given his preferences over effort. The intermediary then has the ability to *monitor* the entrepreneur and force him to increase effort, a mechanism used in a number of papers (*inter alia* de la Fuente and Marín, 1996; Blackburn and Hung, 1998). In the model presented here an agent with high ability is funded and acquires human capital with probability β, paying back an amount to the intermediary in the case of success and nothing otherwise.

A simple approach is to assume that agents are averse to effort and that a monitoring technology is required to increase effort. Effort in this model is reflected in the probability of a good agent becoming human capital, so an increase in effort is the same as an increase in β, though not one-for-one. We could endogenise the quality of intermediary screening but for now we leave the simple case where the agent and intermediary can only influence the probability of becoming human capital.[12] So, post-screening, the agent is faced with the following expected profit condition,

$$\beta(1 - t^*)hH - D(\beta), \tag{2.13}$$

where effort aversion enters as $D(\beta) = (hH\beta^2)/(2\kappa)$, which is an increasing and convex function of β, and also increasing in the level of human capital. The parameter $\kappa > 0$ reflects the agents' effort aversion, i.e., high κ suggests a low aversion to effort. These assumptions might be justified on two counts: The marginal effect of an increase in effort on the likelihood of success is decreasing in the probability of success; and the higher the level of human capital to which an agent aspires, the

[12] Trew (2004) considers such an extension in a model that is closer to King and Levine (1993b).

more difficult it is to succeed and so the higher the cost of increasing β. We also abstract from taxation in this version, so $\tau = 1$. The agent thus chooses his level of effort to maximise his private return, given t^* and κ,

$$\beta_0 = \kappa(1 - t^*), \tag{2.14}$$

which is, importantly, invariant to h.

We thus have a minimum effort level in the absence of monitoring equal to β_0. An intermediary can spend resources on 'monitoring' the agent in order to force his effort level higher. The cost of increasing effort is a function $M(\beta - \beta_0) = [hH(\beta - \beta_0)^2]/2s$ of the difference between the desired β and the minimum, β_0, where $s > 0$ is, again, some scale parameter that influences the cost of monitoring and we again assume that the cost of monitoring is increasing in the level of human capital. So s is some indication of the sophistication of financial intermediaries in mitigating the costs of moral hazard; the higher is s, the less costly is moral hazard. These simplifying assumptions are necessary for both β_0 and β^* to be invariant to H so they could be modified, but the algebra would not permit a simple closed form solution for growth rates. We define this function to be convex in the difference between desired and minimum effort levels (the convexity here is a consequence of convexity in effort-aversion). As such, the intermediary's expected profit considers this additional cost,

$$\begin{aligned} E(\pi) &= \beta\left(\frac{1-\varphi_1-\varphi_2}{1-\varphi_1}\right)\left\{thH - x(H) - f(H) - [H(\beta-\beta_0)^2]/2s\right\} + \\ &+ (1-\beta)\left(\frac{1-\varphi_1-\varphi_2}{1-\varphi_1}\right)[-x(H) - f(H)] + \\ &+ \left(\frac{\varphi_2}{1-\varphi_1}\right)[-f(H)] = 0. \end{aligned} \tag{2.15}$$

So the intermediary now maximises expected profits with respect to both β and

t. If we again specify $x(H) = \eta_x hH$ and $f(H) = \eta_f hH$, then the optimal β for a given t is the positive solution to,

$$\left\{3\beta^2 - 4\beta\kappa(1-t) + [\kappa(1-t)]^2\right\}/2s - t = 0, \tag{2.16}$$

which is,

$$\beta^* = \frac{2}{3}\kappa(1-t) + \frac{1}{3}\left\{[\kappa(1-t)]^2 + 6st\right\}^{\frac{1}{2}}. \tag{2.17}$$

It is easy to see from equation (2.17) that both an increase in the efficiency of monitoring (increasing the scale parameter s) and a lower aversion to effort (higher κ) results in a higher optimal effort.

Substituting the expression for the optimal β into the expected profit function, equation (2.15), and setting expected profits equal to zero, it follows that the optimal levy on agents acquiring human capital, t^*, is the solution to,

$$\left\{\frac{2}{3}[\kappa(1-t)] + \left[\frac{(\kappa(1-t))^2}{9} + \frac{2}{3}st\right]^{\frac{1}{2}}\right\}\left\{2st - \left[\frac{(\kappa(1-t))^2}{9} + \frac{2}{3}st\right] + \right. \tag{2.18}$$
$$\left. -\frac{2}{3}\kappa(1-t)\left[\frac{(\kappa(1-t))^2}{9} + \frac{2}{3}st\right]^{\frac{1}{2}} + \frac{1}{9}(\kappa(1-t))^2\right\} = 2s\left[\eta_x + \left(\frac{\varphi_2}{1-\varphi_1}\right)\eta_f\right].$$

We can now find the growth rate of the economy, as before, as a function of the financial intermediary conditions. It should be clear that parameterisation will not be as simple as in the case without moral hazard since here we require both $t^* \in (0,1)$ and $0 < \beta_0 < \beta^* \in (0,1]$, but there is a range of parameters for which we obtain sensible results.

Again we have $r = (1-t^*)h$[13] so the level of growth in the economy is equation (2.11). The effect of parameter variations on growth are the opposite of the

[13]This condition does not change from the model without moral hazard. Decisions over screening and effort are made within the cohort of agents, before employment as human capital, so do not affect the conditions for dynamic optimisation.

effect on the optimal t. For a reasonable range of parameters[14] it can be shown that the optimal financial intermediary cost, t^*, is decreasing in the efficiency of monitoring technology, s, and increasing the degree of effort aversion (decreasing in κ). Financial efficiency also has the expected effect, with t^* increasing in both η_f and η_x.

So we have a model of endogenous growth which incorporates both the role of financial efficiency, along the lines of King and Levine (1993b), and a facility to reflect the degree of moral hazard faced, in the spirit of Morales (2003). It is, therefore, possible to consider the results from section 2.2.1 in the light of changing moral hazard conditions over time.

We can also present an analogous result to that in Figure 2.2, with combinations of financial efficiency and moral hazard required to obtain the observed UK growth path through the industrial revolution. For each year we have an estimate of trend growth, $\hat{\gamma}$, from which we can infer, from equation (2.11), the implied estimate for t^*,

$$\hat{t}^* = 1 - \left(\frac{\theta\hat{\gamma} + \rho}{A}\right)^{\frac{1}{1-\alpha}} \frac{1}{1-\alpha}. \tag{2.19}$$

Again, there are restrictions on parameters in order that $\hat{t}^*$ is in the unit interval. Specifically this requires that $0 < \theta\hat{\gamma} + \rho < A(1-\alpha)^{1-\alpha}$, so we choose parameter values that satisfy this given the range of growth rates over the period 1701-1913.[15] Using parameter values given in Table 2.2, this inequality is satisfied for the entire sample. The fact that we can use the same parameters for both models demonstrates that the model without moral hazard is nested within the extended model presented here, and also that both are not overly sensitive to parameter variations.

Having identified $\hat{t}^*$ we can find combinations of financial costs, η_f, and s that

[14]In these experiments the benchmark parameterisation is the same as that given in Table 2.1, and in addition $\kappa = 1$, $\eta_f = 0.1$ and $s = 10$. We vary one parameter holding the others constant in order to infer the partial influences.

[15]The minimum growth rate, using the same procedure as in section 2.2.1, is 0.47% and the maximum is 3.60%

Table 2.2: Benchmark Calibration for the Model with Moral Hazard

capital share	α	0.67
subjective discount rate	ρ	0.02
elasticity of substitution	θ	5
proportion of zero ability	φ_1	0.4
proportion of low ability	φ_2	0.2
coefficient of technological progress	A	30
scale parameter on financial investment	η_x	0.01
scale parameter on screening	η_f	0.1
effort aversion parameter	κ	1
monitoring cost parameter	s	10

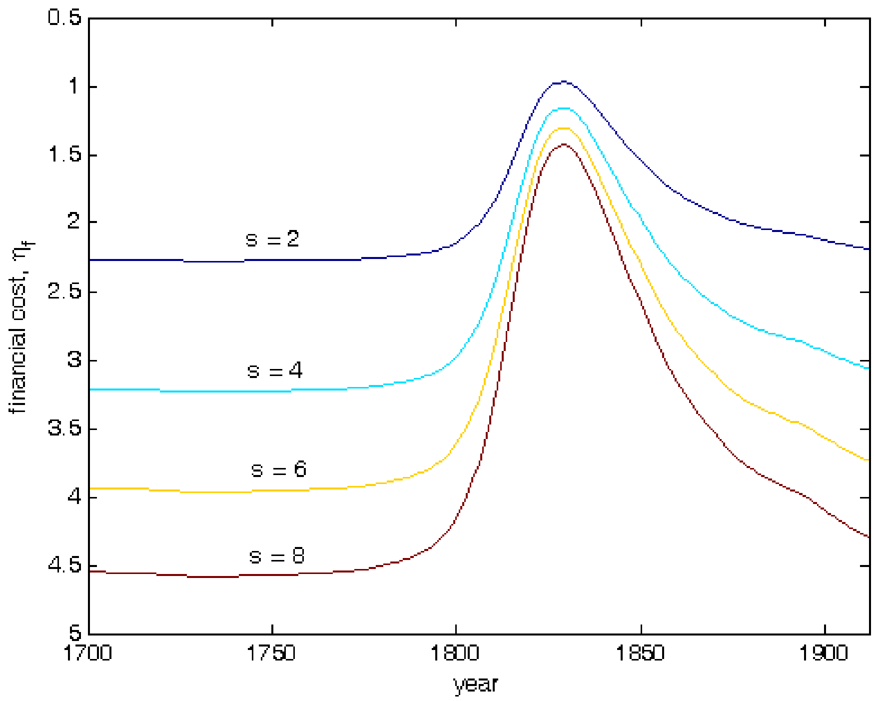
obtain this growth rate using equation (2.18), and thus combinations of financial efficiency and moral hazard that replicate the industrial revolution in the UK. Figure 2.3 depicts this relationship, where we again simply reverse the z-axis to give an impression of financial efficiency *à la* King and Levine (1993b). As anticipated, improving moral hazard conditions (increasing s) means that, *ceteris paribus*, a given level of growth can be obtained with lower financial efficiency.

Figure 2.3 gives combinations of financial efficiency and moral hazard that result in our estimated growth rate. MATLAB code is given in Appendix Section A.1. We can see that either high financial efficiency and high moral hazard costs or low financial efficiency and low moral hazard costs obtain the same growth rate, as in Morales (2003). We can imagine a cross-section of the figure as being equivalent to Figure 2.2.[16] So changes in the conditions of moral hazard affect the level of financial efficiency required to obtain a given growth rate.

The relation between growth and financial efficiency is monotonic but the degree of variation is clearly dependent on the relationship between the rate of growth and moral hazard conditions. Choosing a cross section at $s = 2$ suggests a high level

[16]We cannot think about a single cross-section as representing it perfectly since, in that model, β is fixed whereas here it is endogenous and so changing over the period, but the general pattern is consistent.

Figure 2.3: Industrial Growth, Moral Hazard and Financial Efficiency



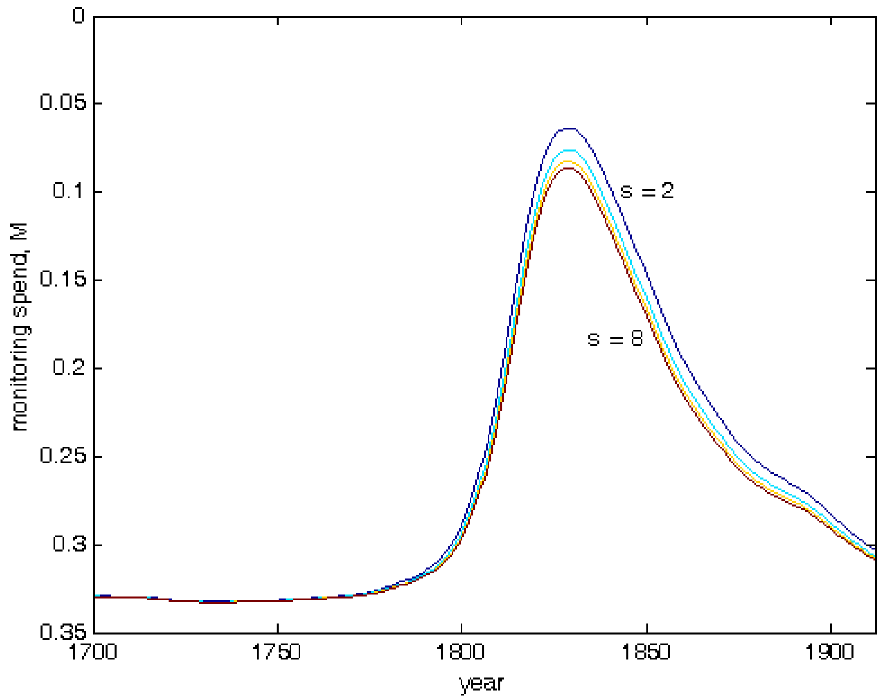
of financial efficiency throughout the period, while at $s = 8$ we see a (relatively) dramatic variation in the level of financial efficiency. So while $\partial\eta_f/\partial s < 0$ holds at all points we have in addition,

$$\left.\frac{\partial\eta_f}{\partial s}\right|_{\gamma=\overline{\gamma}} > \left.\frac{\partial\eta_f}{\partial s}\right|_{\gamma=\underline{\gamma}} \quad \text{where } \overline{\gamma} > \underline{\gamma}. \tag{2.20}$$

In other words, the partial effect of s on η_f is more negative (closer to zero) when growth is low (high). So the most unusual implication of the figure is that as moral hazard conditions deteriorate, so financial efficiency needs to vary less in order to obtain equal changes in the growth rate. It will help our understanding of the model if we consider the intermediary's total spend on monitoring.

Interestingly, the total spend on monitoring, $M(\beta^* - \beta_0)$, is almost invariant to s, as shown in Figure 2.4. The minimum effort level, β_0, is invariant to s but the optimal effort level, β^* is of course increasing in s while the overall monitoring cost simultaneously declines. So increases in the efficiency with which intermediaries can monitor agents endogenously decreases the level of moral hazard in intermediation without intermediaries actually spending significantly different amounts on monitoring agents.

Figure 2.4: Industrial Growth and Spend on Monitoring



The reason why total monitoring spend falls as growth increases is because of the implied increase in financial efficiency, decreasing the optimal intermediation levy, t^*, and so increasing the minimum effort level, β_0. This causes less emphasis to be placed on the the effect of monitoring on growth. As such, we observe that at high growth rates, when the minimum effort level is high and with roughly constant and, most importantly, low spend on monitoring, the effect of changing moral hazard conditions is less since it enters directly into the monitoring decision. The transmission from moral hazard to growth thus follows: When moral hazard conditions mean that the level of effort in the absence of monitoring is low, the total spend on monitoring is high and so, in such cases, the effect of changing moral hazard conditions affects growth more severely.

This sort of result, where moral hazard and economic growth conditions interact, is akin to that in Greenwood and Smith (1997) and Blackburn et al. (2005) and, in the simple form presented here, could be tested empirically with appropriate cliometric evidence.

The cross-section implication is more subtle than that in Section 2.2.1. The general result holds, that countries with low levels of growth require lower differences in financial efficiency to obtain the same changes as countries with high levels of growth. The effect of lower moral hazard is to mitigate this nonlinearity. In a sense, we have the same result: Greater financial sophistication, either in terms of the extent of moral hazard or in terms of financial efficiency, results in a lower effect on growth of improving the level financial efficiency.

2.3 Finance and Endogenous Growth with Quality Ladders

We wish to demonstrate that our findings from the simple Ak growth model might be extended to a more involved endogenous growth setup. We introduce a finance and endogenous growth model with quality ladders that reflects, to some degree, the theoretical mechanisms outlined above. We wish to identify the nature of financial conditions required to obtain balanced growth. In so doing, we wish to see whether our proposition that static intermediation and balanced growth together imply constant financial depth holds in a more general endogenous growth setting. Developing this model also brings home the difficulties we would face in calibrating such a finance and growth theory to data. We first lay-out a more general model without financial frictions, and then turn in Subsection 2.3.3 to our extension.

The following is a representation of the Barro and Sala-i-Martin (2004) endogenous growth model with quality ladders, which itself is based on the models of Grossman and Helpman (1991) and Aghion and Howitt (1992). The model is based around three sectors: a competitive sector of final-good producers; a monopolistic sector producing excludable intermediate goods necessary in the production of final goods; and a research sector producing designs for intermediate good production. In contrast to the horizontal endogenous growth models, where technological progress manifests in an increasing number of intermediate goods à la Romer (1990), here we have a fixed number, N, of intermediate goods that increase in quality. Certain conditions result in the holder of the highest technology in a given good being the pure monopoly producer of that good and this monopoly profit provides the incentive for research and so growth in the model.

2.3.1 Behaviour of Firms

Levels of Quality in the Production Technology

Final good production of good j by firm i requires labour and the input intermediate good j,

$$Y_i = AL_i^{1-\alpha} \sum_{j=1}^{N} \left(\tilde{X}_{ij} \right)^{\alpha}, \tag{2.21}$$

where $0 < \alpha < 1$ and $\tilde{X}_{ij}$ is the *quality-adjusted* intermediate good,

$$\tilde{X}_{ij} = \sum_{k=0}^{\kappa_j} \left(q^k X_{ijk} \right), \tag{2.22}$$

where quality is indexed by k and where κ_j is the highest quality rung in sector j. Increases in the quality of good j are by $q > 1$, where $q^0 = 1$. Intermediate good X_{ij} has unit marginal cost.[17] The total output of intermediate good j is thus the

[17]This is standard where output is a function of labour alone. If output $q = f(L)$ and price is

sum of output at each quality rung. A good j at stage $k = 2$ is a perfect substitute for the same good j at stage $k = 1$, and is preferable in cost terms. Where only the highest technology good is produced, the total quality-adjusted output of good j is $\tilde{X}_{ij} = q^{\kappa_j} X_{ij\kappa_j}$, and so output is $Y_i = AL_i^{1-\alpha}(q^{\kappa_j} X_{ij\kappa_j})^{\alpha}$.

The marginal product of an intermediate good is,

$$\frac{\partial Y_i}{\partial X_{ij\kappa_j}} = \alpha A L_i^{1-\alpha} q^{\kappa_j \alpha} (X_{ij\kappa_j})^{\alpha-1}. \tag{2.23}$$

If the leading-edge good is priced at $P_{j\kappa_j}$, the final-good competitors maximise,

$$\text{profit} = Y_i - wL_i - P_{j\kappa_j} X_{ij\kappa_j}, \tag{2.24}$$

so,

$$P_{j\kappa_j} = \alpha A (L_i)^{1-\alpha} q^{\kappa_j \alpha} (X_{ij\kappa_j})^{\alpha-1}. \tag{2.25}$$

The intermediate-producer's demand function is thus,

$$X_{ij\kappa_j} = L \left[A\alpha q^{\kappa_j \alpha} / P_{j\kappa_j} \right]^{\frac{1}{1-\alpha}}. \tag{2.26}$$

The monopolistic intermediary maximises profit, $P_{j\kappa_j} X_{ij\kappa_j} - X_{ij\kappa_j}$ based on this demand function,

$$\pi_{j\kappa_j} = (P_{j\kappa_j} - 1) L \left[A\alpha q^{\kappa_j \alpha} / P_{j\kappa_j} \right]^{\frac{1}{1-\alpha}}; \tag{2.27}$$

$$\frac{\partial \pi_{j\kappa_j}}{\partial P_{j\kappa_j}} = \frac{\alpha}{\alpha - 1} (P_{j\kappa_j})^{(2-\alpha)/(\alpha-1)} - \frac{1}{\alpha - 1} (P_{j\kappa_j})^{1/(1-\alpha)} = 0;$$

$$P_{j\kappa_j} = \frac{1}{\alpha}; \qquad \forall_j. \tag{2.28}$$

equal to marginal cost, $P = \omega / f'$, where ω is the nominal wage. If we divide both sides by P and define $w = \omega / P$ as the real wage, we obtain that the marginal cost (in real terms) is $1 = w / f'$.

So the monopoly price thus exceeds unity, since $\alpha < 1$.

The aggregate output of the j^{th} intermediate good is thus,

$$X_{j\kappa_j} = L\left[A\alpha^2 q^{\alpha\kappa_j}\right]^{\frac{1}{1-\alpha}}, \tag{2.29}$$

which is increasing over time with κ_j, unlike the models of increasing product variety.

We can now write the production function as,

$$Y_i = AL_i^{1-\alpha}\sum_{j=1}^{N}\left[q^{\alpha\kappa_j}(X_{ij\kappa_j})^{\alpha}\right], \tag{2.30}$$

which, with equation (2.29), becomes,

$$\begin{aligned} Y_i &= AL_i^{1-\alpha}\sum_{j=1}^{N}\left[q^{\alpha\kappa_j}\left(A\alpha^2 q^{\alpha\kappa_j}\right)^{\frac{1}{1-\alpha}}\right]^{\alpha} \\ &= A^{\frac{1}{1-\alpha}}L\alpha^{\frac{2\alpha}{1-\alpha}}\sum_{j=1}^{N}\left[q^{\frac{\alpha^2\kappa_j}{1-\alpha}}\right] \\ Y &= A^{\frac{1}{1-\alpha}}L\alpha^{\frac{2\alpha}{1-\alpha}}Q, \end{aligned} \tag{2.31}$$

where Q is the aggregate quality index,

$$Q = \sum_{j=1}^{N} q^{\frac{\alpha\kappa_j}{1-\alpha}}. \tag{2.32}$$

Aggregate intermediate output, $\sum_{j=1}^{N} X_{j\kappa_j}$, is,

$$X = A^{\frac{1}{1-\alpha}}\alpha^{\frac{2}{1-\alpha}}LQ. \tag{2.33}$$

We can see from (2.33) that since L and N are constant, the driver of growth in intermediate output is the changing level of the cutting-edge technology in each sector, κ_j. It is also clear that both final output and intermediate production is simply proportional to Q. Finding the growth rate of the quality index thus enables us to find the balanced growth rate of the economy.

The Flow of Monopoly Profit

Profit for the intermediate monopolistic producer is, as above, $\pi_{j\kappa_j} = (P_{j\kappa_j} - 1)X_{j\kappa_j}$. So,

$$\pi_{j\kappa_j} = \left(\frac{1-\alpha}{\alpha}\right) L[A\alpha^2 q^{\alpha\kappa_j}]^{\frac{1}{1-\alpha}}. \tag{2.34}$$

This profit accrues for only as long as the incumbent holds the cutting edge technology, κ_j. When a new technology, $\kappa_j + 1$ is invented, the incumbent is displaced, i.e. the profits accrue over the interval $T_{j\kappa_j} = t_{\kappa_j+1} - t_{\kappa_j}$. The value to a firm of innovation is, with a constant interest rate r, thus,

$$\begin{aligned} V_{j\kappa_j} &= \int_0^{T_{j\kappa_j}} \pi_{j\kappa_j} e^{-rt} dt \\ &= \pi_{j\kappa_j} \left[-\frac{1}{r} e^{-rt}\right]_0^{T_{j\kappa_j}} \\ V_{j\kappa_j} &= \frac{\pi_{j\kappa_j}}{r}\left[1 - e^{-rT_{j\kappa_j}}\right]. \end{aligned} \tag{2.35}$$

So the value to a firm of acquiring the cutting-edge technology increases in the sophistication of the good (κ_j), in the profit acquired while incumbent ($\pi_{j\kappa_j}$), and decreases in the interest rate (r).

Duration of Monopoly Profit

The monopoly profit, $\pi_{j\kappa_j}$, is known so to determine the incentive to innovate, $V_{j\kappa_j}$, we must find $T_{j\kappa_j}$, the duration over which profits accrue.

If $Z_{j\kappa_j}$ denotes the flow of resources (in units of Y) expended by innovators (researchers) in sector j when the current state of knowledge is κ_j, increasing $Z_{j\kappa_j}$ increases the probability of a research breakthrough, but breakthroughs become harder as κ_j grows. If $p_{j\kappa_j}$ is the probability per unit of time of a breakthrough, we can specify,

$$p_{j\kappa_j} = Z_{j\kappa_j}\phi(\kappa_j), \qquad \phi' < 0. \tag{2.36}$$

Equation (2.36) is a Poisson process wherein, for a given κ_j, the probability of successful innovation per unit of time depends upon current R&D effort only, i.e. we assume that individual sectors are small and that the probability of research success across sectors in independent.

The weak Law of Large Numbers implies that microeconomic 'jumpiness' resulting from individual sectors moving up the quality ladders will not transmit to macroeconomic outcomes, and instead we obtain a smooth and stable growth path.

If we define $G(\tau)$ as the CDF for $T_{j\kappa_j}$, i.e. the probability that $T_{j\kappa_j} \leq \tau$, then the change in $G(\tau)$ with respect to τ is the probability that innovation occurs at time τ. For innovation to happen at time τ we require, first, that it has not happened previously, with probability $1 - G(\tau)$, and, second, that it does happen in time τ, with probability $p_{j\kappa_j}$. The probability is thus,

$$\frac{dG(\tau)}{d\tau} = [1 - G(\tau)]p_{j\kappa_j}. \tag{2.37}$$

Assuming $p_{j\kappa_j}$ is constant over time, i.e. research effort, $Z_{j\kappa_j}$ is constant within sectors and between innovations,

$$\frac{dG(\tau)}{d\tau} + G(\tau)p_{j\kappa_j} = p_{j\kappa_j}, \tag{2.38}$$

so,

$$\begin{aligned} G(\tau) &= e^{-\int_0^\tau p_{j\kappa_j} dt} \left(A + \int_0^\tau p_{j\kappa_j} e^{\int_0^\tau p_{j\kappa_j} dt} dt \right) \\ &= e^{-p_{j\kappa_j}\tau} \left(A + \int_0^\tau p_{j\kappa_j} e^{p_{j\kappa_j}\tau} dt \right) \\ &= e^{-p_{j\kappa_j}\tau} \left(A + [e^{p_{j\kappa_j}\tau}]_0^\tau \right) \\ G(\tau) &= e^{-p_{j\kappa_j}} A + 1 - e^{-p_{j\kappa_j}\tau}. \end{aligned} \tag{2.39}$$

Since as $\tau \to \infty$, $G(\tau) \to 1$, it must be that $A = 0$, so,

$$G(\tau) = 1 - e^{-p_{j\kappa_j}\tau}. \tag{2.40}$$

The probability distribution function is, therefore,

$$g(\tau) = G'(\tau) = p_{j\kappa_j} e^{-p_{j\kappa_j}\tau}. \tag{2.41}$$

We now have the present value of profits from innovation *and* the PDF for $T_{j\kappa_j}$. We can thus substitute for $T_{j\kappa_j}$ into equation (2.35),

$$V_{j\kappa_j} = \frac{\pi_{j\kappa_j}[1 - e^{-rT_{j\kappa_j}}]}{r}, \tag{2.42}$$

and take expectations, letting $E[T_{j\kappa_j}] = \tau$,

$$E[V_{j\kappa_j}] = E\left[\frac{\pi_{j\kappa_j}}{r}(1 - e^{-r\tau})\right]. \tag{2.43}$$

Combined with the probability of innovation at a point in time, we have that the expected flow of profit from being at the cutting-edge of a given sector is,

$$\begin{aligned}
E[V_{j\kappa_j}] &= \frac{\pi_{j\kappa_j}}{r}\int_0^\infty (1-e^{-r\tau})p_{j\kappa_j}e^{-p_{j\kappa_j}\tau}dt \\
&= \frac{\pi_{j\kappa_j}p_{j\kappa_j}}{r}\int_0^\infty e^{-p_{j\kappa_j}\tau} - e^{-(r+p_{j\kappa_j})\tau}dt \\
&= \frac{\pi_{j\kappa_j}p_{j\kappa_j}}{r}\left[-\frac{e^{-p_{j\kappa_j}\tau}}{p_{j\kappa_j}} + \frac{e^{-(r+p_{j\kappa_j})\tau}}{r+p_{j\kappa_j}}\right]_0^\infty \\
&= \frac{\pi_{j\kappa_j}}{r} - \frac{\pi_{j\kappa_j}p_{j\kappa_j}}{r(r+p_{j\kappa_j})} \\
E[V_{j\kappa_j}] &= \frac{\pi_{j\kappa_j}}{r+p_{j\kappa_j}}. \qquad (2.44)
\end{aligned}$$

Using equation (2.34) we obtain,

$$E[V_{j\kappa_j}] = \left\{\left(\frac{1-\alpha}{\alpha}\right)L[A\alpha^2 q^{\alpha\kappa_j}]^{\frac{1}{1-\alpha}}\right\}/(r+p_{j\kappa_j}), \qquad (2.45)$$

which is random, since the duration of incumbency as technological leader is itself random.

There is additional uncertainty, however, since innovators cannot be sure that a research project will succeed – they must select an optimal level of effort.

Determination of Research Effort

The incentive to innovate, $E[V_{j\kappa_j}]$ affects the effort level of researchers, $Z_{j\kappa_j}$, and so the probability of innovation, p. We assume that researchers care only about $E[V_{j\kappa_j}]$, not about the randomness of its return.[18] The expected reward for pursuing

[18] Barro and Sala-i-Martin (2004) assume that research projects are carried out by 'syndicates' to such a degree that even if innovators are individually risk-averse, sufficient diversification will make the body of researchers act risk-neutrally.

the $\kappa_j + 1^{\text{th}}$ innovation is $p_{j\kappa_j} E[V_j, k_j + 1]$. The expected net return from innovation is thus $\Pi_{j\kappa_j} = p_{j\kappa_j} E[V_{j\kappa_j}] - Z_{j\kappa_j}$, which, using equations (2.36) and (2.45), can be expressed as,

$$\Pi_{j\kappa_j} = Z_{j\kappa_j} \left[\phi(\kappa_j) L A^{\frac{1}{1-\alpha}} \left(\frac{1-\alpha}{\alpha} \right) \alpha^{\frac{2}{1-\alpha}} \frac{q^{\frac{(\kappa_j+1)\alpha}{(1-\alpha)}}}{(r + p_{j\kappa_j})} - 1 \right]. \tag{2.46}$$

Assuming free entry into the research sector, if $Z_{j\kappa_j} > 0$ then $\Pi_{j\kappa_j} = 0$ must hold. So the free entry condition requires,

$$r + p_{j\kappa_j} = \phi(\kappa_j) L A^{\frac{1}{1-\alpha}} \left(\frac{1-\alpha}{\alpha} \right) \alpha^{\frac{2}{1-\alpha}} q^{\frac{(\kappa_j+1)\alpha}{(1-\alpha)}}. \tag{2.47}$$

We can see from equation (2.47) that the probability of innovation for a given rate of interest depends on κ_j. The expected flow of income from research success, $E[V_{j\kappa_j}]$, is increasing in κ_j while the resources required to maintain a given probability of innovation is also increasing in κ_j, because we have defined $\phi'(\kappa_j) < 0$. The dominating term will determine the total effect of κ_j on $p_{j\kappa_j}$. If the first effect dominates the latter, the rate of technological progress will grow with κ_j, and we have some form of increasing returns in the R&D sector. If the second effect dominates the former, the overall growth rate will fall over time as the technological frontier advances. If the two forces offset one another, the growth path is balanced and we have a steady-state growth path that can be compared with other models of endogenous growth.

To obtain the balanced growth path, we thus specify,

$$\phi(\kappa_j) = \frac{1}{\zeta} q^{-\frac{(\kappa_j+1)\alpha}{1-\alpha}}, \tag{2.48}$$

where $\zeta > 0$ is the 'cost' of research. The effects of ϕ and $E[V]$ thus offset,

$$r + p = \frac{L}{\zeta} A^{\frac{1}{1-\alpha}} \left(\frac{1-\alpha}{\alpha} \right) \alpha^{\frac{2}{1-\alpha}}, \tag{2.49}$$

where $p = p_{j\kappa_j}$ is invariant to j. As such, if r is constant over time, so is p.

The resources allocated to research, $Z_{j\kappa_j} = p/\phi(\kappa_j)$, are thus,

$$Z_{j\kappa_j} = q^{\frac{(\kappa_j+1)\alpha}{1-\alpha}} \left[LA^{\frac{1}{1-\alpha}} \left(\frac{1-\alpha}{\alpha} \right) \alpha^{\frac{2}{1-\alpha}} - r\zeta \right], \tag{2.50}$$

i.e. research input increases as sectors become more advanced, but only enough to offset the higher effort required to maintain an invariant p.

Aggregate R&D spending is thus,

$$Z = \sum_{j=1}^{N} Z_{j\kappa_j} = Qq^{\frac{\alpha}{1-\alpha}} \left[LA^{\frac{1}{1-\alpha}} \left(\frac{1-\alpha}{\alpha} \right) \alpha^{\frac{2}{1-\alpha}} - r\zeta \right]. \tag{2.51}$$

The Behaviour of the Aggregate Quality Index

Aggregate output, Y, aggregate resources to R&D, Z, and aggregate spending on intermediate good, X, are all constant multiples of Q. All will thus grow at the same rate as Q,

$$\gamma_Q = \gamma_X = \gamma_Y = \gamma_Z. \tag{2.52}$$

Recalling the equation for the aggregate quality index, $Q = \sum_{j=1} Nq^{(\kappa_j \alpha)/(1-\alpha)}$, we can see that if innovation in sector j does occur, the growth in quality will be,

$$\frac{q^{\frac{(\kappa_j+1)\alpha}{1-\alpha}} - q^{\frac{\kappa_j \alpha}{1-\alpha}}}{q^{\frac{\kappa_j \alpha}{1-\alpha}}} = q^{\frac{\alpha}{1-\alpha}} - 1. \tag{2.53}$$

Since this is invariant to sector, and given that the probability of innovation is p, the expected growth of the quality index per unit of time is thus,

$$E\left[\frac{\Delta Q}{Q}\right] = p\left(q^{\frac{\alpha}{1-\alpha}} - 1\right), \tag{2.54}$$

so,

$$\gamma_Q = \left[\frac{L}{\zeta}A^{\frac{1}{1-\alpha}}\left(\frac{1-\alpha}{\alpha}\right)\alpha^{\frac{2}{1-\alpha}} - r\right]\left(q^{\frac{\alpha}{1-\alpha}} - 1\right). \tag{2.55}$$

To derive the growth rate of the economy, therefore, we must find the equilibrium interest rate.

The Market Value of Firms

Only leading-edge goods are produced. Substituting equation (2.49) into equation (2.45) we obtain,

$$E[V_{j\kappa_j}] = \zeta q^{\frac{\kappa_j \alpha}{(1-\alpha)}}. \tag{2.56}$$

Aggregating over N sectors,

$$V = \zeta \sum_{j=1}^{N} q^{\frac{\kappa_j \alpha}{1-\alpha}} = \zeta Q. \tag{2.57}$$

The total market value of firms is thus a constant multiple of, and so grows at the same rate as, Q.

2.3.2 Households and Market Equilibrium

Individual households maximise utility,

$$U = \int_0^\infty \frac{c^{1-\theta} - 1}{1-\theta} e^{-\rho t} dt, \tag{2.58}$$

where c is per capita consumption, population growth $n = 0$, θ is the constant intertemporal elasticity of substitution and ρ is the subjective discount rate. The

Euler equation defines the growth rate of consumption,

$$\gamma_c = \frac{\dot{c}}{c} = \frac{r-\rho}{\theta}. \tag{2.59}$$

The economy-wide resource constraint is,

$$C = Y - X - Z. \tag{2.60}$$

By substituting for Y, X and Z from equations (2.31), (2.33) and (2.51) respectively we can see that consumption is also a constant multiple of Q,

$$C = \left[(1-\alpha^2)A^{\frac{1}{1-\alpha}}L\alpha^{\frac{2\alpha}{1-\alpha}} - \zeta p q^{\frac{\alpha}{1-\alpha}}\right]Q. \tag{2.61}$$

From the Euler equation, $r = \theta\gamma_c + \rho = \theta\gamma + \rho$. Substituting for the interest rate into the expression for the growth rate of the quality index, equation (2.55), and letting $\gamma = \gamma_Q$,

$$\gamma = \frac{\left[q^{\frac{\alpha}{1-\alpha}}\right]\left[(L/\zeta)A^{\frac{1}{1-\alpha}}\left(\frac{1-\alpha}{\alpha}\right)\alpha^{\frac{2\alpha}{1-\alpha}} - \rho\right]}{1+\theta\left[q^{\frac{\alpha}{1-\alpha}}-1\right]}. \tag{2.62}$$

The interest rate is,

$$r = \frac{\rho + \theta\left[q^{\frac{\alpha}{1-\alpha}}\right]\left[(L/\zeta)A^{\frac{1}{1-\alpha}}\left(\frac{1-\alpha}{\alpha}\right)\alpha^{\frac{2\alpha}{1-\alpha}}\right]}{1+\theta\left[q^{\frac{\alpha}{1-\alpha}}-1\right]}. \tag{2.63}$$

And, from $\gamma_Q = p\left[q^{\frac{\alpha}{1-\alpha}}-1\right]$, the equilibrium probability of successful innovation is,

$$p = \frac{\left[(L/\zeta)A^{\frac{1}{1-\alpha}}\left(\frac{1-\alpha}{\alpha}\right)\alpha^{\frac{2\alpha}{1-\alpha}} - \rho\right]}{1+\theta\left[q^{\frac{\alpha}{1-\alpha}}-1\right]}. \tag{2.64}$$

There are no transitional dynamics, and it can be shown that no other equilibria are possible. As with a model of expanding product varieties, growth is decreasing in ρ and θ and increasing in the technological coefficient A and in the inverse of the cost of research, $(1/\zeta)$. The growth rate in this model is also increasing in the innovative step, q.

2.3.3 Static Financial Intermediation with Quality Ladders

The following is an exposition of financial intermediation in the spirit of King and Levine. The financial intermediary here is, again, effectively a venture capitalist, operating in two stages. A given entrepreneur can be one of two types, either capable or incapable of managing a research project. First, the intermediary possesses a screening technology that allows it to select with certainty only those entrepreneurs[19] capable of managing a given research topic and second, having determined quality, funds the research project which has an uncertain outcome. If the research project succeeds, the entrepreneur patents the new design, sets up the new market leader and so acquires monopoly profits less an amount he is contracted to pay back to the intermediary. This is thus a departure from the King and Levine model, wherein the intermediary retains the rights to monopoly profits.

The justification derives from the required structure of the financial intermediary sector. King and Levine cite "competition among such organizations" (p. 518). Assuming that there is freedom of entry into the intermediation sector, however, the presence of intermediary monopoly profits is non-sustainable. Competition for entrepreneurs would drive intermediaries to offer the best feasible contract to the entrepreneur, that which generate zero expected profits for the intermediary. In a world of perfect competition however, there can be no co-ordination between these intermediaries. There is a large literature on venture capitalism, for example

[19] King and Levine use the term 'entrepreneur' to describe the agents that intrinsically possess the capacity to manage a team of researchers in pursuit of a research goal and who are employed at a fixed rate and have no influence over the fixed probability of innovation. This model actually comes closer to the real definition of an entrepreneur, since he selects his effort level, does the research and sets up the business, acquiring the monopoly profits net of fees to the intermediary. We refer to entrepreneurs and researchers interchangeably here.

Gehrig and Stenbacka (2003), which shows that such competition will always generate regular cyclicality in screening activity as a consequence of a (privately) rejected entrepreneur being able to apply a second time, to a different venture capitalist. There are, consequently, 'pool-worsening' effects following the screening period and pool improvements following the period of inactivity, a result was first noted by Broecker (1990). A monopolistically competitive intermediary who can contract entrepreneurs to being screened only once, or who publishes records of those rejected, satisfies the three main functions of evaluating projects, pooling funds, and diversifying risk, but more importantly this assumption also obtains the constant rate of screening denoted by King and Levine's equation (1).

If we imagine an intermediary, active in sector j and looking for an entrepreneur to reach technology level $\kappa_j + 1$, it must select an entrepreneur capable of managing the project.[20] If, among the pool of potential entrepreneurs a proportion μ is capable of managing the research project then we have a Binomial distribution in the probability of finding a capable entrepreneur.

The moment-generating function of Binomial distributions[21] shows that mean number of successes $m' = Np$ where N is the number of Bernoulli trials and p is the probability that the result of each Bernoulli trial is "true". It follows that on average the entrepreneur must screen $1/\mu$ potential entrepreneurs to find a single

[20]This ordering, wherein intermediary approaches entrepreneur, is non-trivial. Imagine a potential entrepreneur approaches an intermediary for funding to research a certain technology. The potential entrepreneur knows with certainty that the intermediary can apply his screening technology to accurately determine his quality. Why then should an entrepreneur that knows he is incapable apply to the intermediary for funding, and so why do we not consider that the intermediary accepts everyone who applies, following the well-known result that the threat of perfect screening will deter bad applicants (see, for example, Laffont and Martimort (2002, pp. 121-30) for a description of costly state verification results)? The result is that he would not apply, and the intermediaries' screening problem would be simplified. As such, if we were to allow entrepreneurs to approach intermediaries with proposals, we must further require that the entrepreneurs themselves are not aware of their own quality or of their capacity to manage a given research project. This is not an entirely unreasonable assumption if we consider that the potential entrepreneur does not appreciate the effort required to complete a project, or has a distorted opinion of his own capacity to innovate.

[21]Papoulis (1984), p.154.

capable one. Given that the current state of knowledge in sector j is κ_j, the cost of screening a potential entrepreneur is $f_{j\kappa_j}$ units of labour at wage rate w, and the expected return to the intermediary of funding the capable entrepreneur is $E[\xi_{j\kappa_j}]$. In a monopolistically competitive environment, the expected profits must equal zero, so we have,

$$E[\xi_{j\kappa_j}] - \frac{1}{\mu} w f_{j\kappa_j} = 0, \tag{2.65}$$

from which we obtain the entrepreneurial selection condition, equation (1) in King and Levine,

$$\mu E[\xi_{j\kappa_j}] = w f_{j\kappa_j}. \tag{2.66}$$

Since all intermediaries and entrepreneurs are identical, this condition is invariant to both sector, j, and technological vanguard, κ_j. We also infer from equation (2.66) that if the cost of screening increases over time with technological progress, so must the expected return from research grow at the same rate (assuming a constant wage and talent pool).

In this model, the financial intermediary does not have the right to own the patent afterwards, it is the researcher owns the design and becomes the new intermediate-good producer, paying only an amount, χ, to the intermediary in the case of innovative success, and nothing otherwise.

The probability that a research project is successful in innovating is $p_{j\kappa_j}$ and the cost of funding the research project to reach ladder point κ_j in sector j is $x_{j\kappa_j} > 1$[22] units of labour, at wage rate w. The expected net return to the intermediary from a rated entrepreneur is thus,

$$E[\xi_{j\kappa_j}] = p_{j\kappa_j}\chi - w x_{j\kappa_j}, \tag{2.67}$$

[22]Imposing $x_{j\kappa_j} > 1$ enforces the need for the entrepreneur to seek intermediary funding, i.e. a single entrepreneur cannot take on the research project alone and needs the intermediary to employ others to help him.

which on average must be positive, but only to the extent that it matches average screening costs. Upon substitution of equation (2.67) into equation (2.66) we obtain the following expression for the return to the intermediary in the case of successful innovation,

$$\chi = \frac{w[\mu x_{j\kappa_j} + f_{j\kappa_j}]}{\mu p_{j\kappa_j}}. \tag{2.68}$$

The rebate to the financial intermediary in the case of research success is thus decreasing in the quality of the talent pool μ, increasing in both those factors which raise the costs of finding capable researchers, i.e. in the cost of screening, $f_{j\kappa_j}$ and the wage rate, w, and in those factors which raise the cost of financing research, w and $x_{j\kappa_j}$. A higher probability of research success (higher $p_{j\kappa_j}$) also reduces the optimal rebate to intermediation, because it increases the expected net return by equation (2.67).

If $Z_{j\kappa_j}$ is the flow of resources (in units of Y) expended by the researcher in sector j when the current state of knowledge is κ_j, the researcher maximises expected profits,

$$E[\Pi] = p_{j\kappa_j}\left\{E[V_{j\kappa_j}] - \chi\right\} - Z_{j\kappa_j}, \tag{2.69}$$

which leads to the determination of the probability of innovation as before. We can use equation (2.68), with equations (2.36) and (2.45), to give the expected net return from innovation, $\Pi_{j\kappa_j} = p_{j\kappa_j}E[V_{j\kappa_j}] - Z_{j\kappa_j}$,

$$\Pi_{j\kappa_j} = Z_{j\kappa_j}\left\{\phi(\kappa_j)\left[LA^{\frac{1}{1-\alpha}}\left(\frac{1-\alpha}{\alpha}\right)\alpha^{\frac{2}{1-\alpha}}\frac{q^{\frac{(\kappa_j+1)\alpha}{(1-\alpha)}}}{(r+p_{j\kappa_j})} - \frac{w[\mu x + f]}{\mu p_{j\kappa_j}}\right] - 1\right\}. \tag{2.70}$$

Again assuming free entry into the research sector, if $Z_{j\kappa_j} > 0$ then $\Pi_{j\kappa_j} = 0$ must hold. So the free entry condition requires,

$$\mu p_{j\kappa_j}^2 + \left\{ \mu r - \phi(\kappa_j)\mu \left[LA^{\frac{1}{1-\alpha}} \left(\frac{1-\alpha}{\alpha} \right) \alpha^{\frac{2}{1-\alpha}} q^{\frac{(\kappa_j+1)\alpha}{(1-\alpha)}} \right] + \right.$$
$$\left. + w\phi(\kappa_j)[\mu x + f] \right\} p_{j\kappa_j} + \phi(\kappa_j) w[\mu x + f] r = 0. \tag{2.71}$$

The equilibrium probability of innovation is thus the positive root of equation (2.71).

As previously, for a balanced growth path we need to consider the conflicting pressures of the increasing profit incentive and the increasing difficulty of innovation as κ_j grows. Here we also have the need to pay back an amount to the intermediary which, if fixed, will fall as a proportion of monopoly profit accruing. As such, we need to offset all three effects in order that a balanced growth path is retained,

$$\phi(\kappa_j) = \frac{1}{\zeta_\phi} q^{-\frac{(\kappa_j+1)\alpha}{1-\alpha}}; \tag{2.72}$$
$$f_{\kappa_j} = \zeta_f q^{\frac{(\kappa_j+1)\alpha}{1-\alpha}}; \tag{2.73}$$
$$x_{\kappa_j} = \zeta_x q^{\frac{(\kappa_j+1)\alpha}{1-\alpha}}, \tag{2.74}$$

where ζ_ϕ, ζ_x and ζ_f are the nonnegative costs of research effort, research funding and intermediary screening respectively. Although a departure from King and Levine, this modification is intuitively very reasonable. Equations (2.73) and (2.74) imply that as the technological frontier advances, so it becomes harder to select researchers capable of handling the project to design the next patent (f increases), and so the length of the research project itself, or the resources required for its completion, becomes longer (x increases). Taken together, Equations (2.73) and (2.74) imply that financial costs comprise a constant proportion of total output in balanced growth. With a constant probability of innovation, these conditions imply that financial depth, by a number of measures, will be constant also.

Equation (2.71) is now,

$$\mu p^2 + \left\{\mu r - \mu\left[(L/\zeta_\phi)A^{\frac{1}{1-\alpha}}\left(\frac{1-\alpha}{\alpha}\right)\alpha^{\frac{2}{1-\alpha}}\right] + \right.$$
$$\left. +(w/\zeta_\phi)[\mu\zeta_x + \zeta_f]\right\}p + (wr/\zeta_\phi)[\mu\zeta_x + \zeta_f] = 0. \quad (2.75)$$

So we have a probability of innovation that is invariant to time and sectors, $p = p_j\kappa_j, \forall j, \kappa_j$, which is the requirement for a stable growth path. As such, equations (2.73) and (2.74) require that as technology progresses, so the rebate to financial intermediation, $\chi = f(x_{j\kappa_j}, f_{j\kappa_j})$, increases. This increasing rebate thus accounts for a constant proportion of research income in the case of success, and allows the intermediary to maintain zero profit in the face of increasing intermediation costs, resulting at a macroeconomic level in a balanced growth path. Without this additional effect, as patents continue to be invented, and as profits from the monopoly on their design grow, a constant rebate to intermediation would make research exhibit increasing returns to scale, and an increasing growth rate over time. Moreover, this resolves the proposition that motivated this model: For balanced growth, the size of the financial sector must be a constant proportion of total output.

The solution to equation (2.71) is,

$$p = \frac{1}{2\mu}\{\mu[r - (L/\zeta_\phi)\Upsilon] + (w/\zeta_\phi)[\mu\zeta_x + \zeta_f]\}$$
$$\pm\sqrt{\frac{1}{2\mu}\{\mu[(L/\zeta_\phi)\Upsilon - r] - (w/\zeta_\phi)[\mu\zeta_x + \zeta_f]\}^2 - 2(wr/\zeta_\phi)[\mu\zeta_x + \zeta_f]}, \quad (2.76)$$

where $\Upsilon = A^{\frac{1}{1-\alpha}}\left(\frac{1-\alpha}{\alpha}\right)\alpha^{\frac{2}{1-\alpha}}$ which is invariant to ζ_ϕ, ζ_f and ζ_x. Equation (2.3.3) thus makes clear that the probability of innovation is constant. This is a result of the constant proportionality of financial depth resulting from equations (2.73) and (2.74).

2.4 Concluding Remarks

Can we reasonably expect to be able to explain large deviations in growth rates in terms of the movements in financial efficiency as described above? And can reasonable calibrations capture the cross-sectional variations suggested implied by the empirical literature? This is exactly what the prevailing theoretical finance and growth literature would suggest. While changes in the coefficient of technological progress over the period of industrial revolution will soak up much of the implied fall in efficiency after the peak in growth, implications for financial efficiency and asymmetric information of the representative model developed above would likely follow a broadly similar pattern.

One key factor missing from present theory is the relation to financial depth, and, in turn, the interaction between depth and efficiency at the microeconomic level. It is likely that further developing models which can account for changing conditions of both financial depth and financial efficiency will be more open to comparison with the data. The work of Townsend and Ueda (2006) suggests that this will, indeed, be the case. In addition, the development of testable quantitative implications of theories such as Blackburn et al. (2005) will shed greater light on the importance of the finance-growth nexus in a changing environment.

This Chapter has provided an initial foray into matching the theoretical mechanisms of finance and growth with their numerical counterparts. It has shown that numerical implications can be drawn from major theories of finance and growth, and that these can be considered in the context of an historical growth path. We have seen that these numerical implications are not only unrelated to the empirical research, but also that the two may in fact be contradictory. We developed a second model of endogenous growth in which financial efficiency enters into a quality-ladders set-up, and showed that our simple model is, for our purposes, acceptably parsimonious. This second model also underlines this disconnection between the theory and empirics of finance and growth.

But what do we really learn about finance and growth from even this extended model? In the context of what we found in Chapter 1, it is abundantly clear that both the theoretical mechanisms considered and the current data available are insufficient. In Chapter 3 we will focus on developing theory that exhibits more of a connection with the empirical concept of finance and growth. To do that, we first take a more detailed looked at historical evidence. Coupled with that, the construction of a dataset, of the level of detail approaching that in Townsend and Ueda (2006), will greatly contribute to our understanding of finance and growth. The work of Shea (2007), we will suggest, can play this part in future research. Mirroring the success of the quantitative macroeconomic literature in the economics of growth will likely improve our ability to account for the persisting global inequalities in levels of income.

Chapter 3

Endogenous Financial Development and Industrial Takeoff*

3.1 Introduction

As described in Chapter 1, a plethora of supportive empirical studies and a more sparse but equally supportive theoretical literature has engendered confidence in the belief that there is a causal relationship between financial development and the long run level of economic growth. Indeed, the World Bank, for one, now finds this relationship to be a near-given. For example, Beck and de la Torre (2006, p.1), in an introduction to the World Bank's current research priority, access to finance, remark that the "causal link running from financial depth to growth has been rather convincingly established." Demirgüç-Kunt (2006) is more circumspect, but nonetheless looks to lay-out the policies that hasten financial development in the pursuit of higher economic growth. The tacit message is that the finance-growth

*Part of this chapter has been released as Trew (2007).

nexus is close to being resolved.

We demonstrated in Chapters 1 and 2, however, that there are good reasons for not proclaiming an end to the debate. First, we have argued that empirical cross-section results are strongly biased toward accepting the finance-causes-growth hypothesis. Second, much of theory and econometrics considers the question in only its comparative sense; an understanding of transitional issues is commonly omitted. This must have implications for policy in regard to the implementation of any finance-led growth strategy within a country. Third, the connection between theoretical mechanisms and empirical results in finance and growth has, largely, been neglected. Specifically, the former typically looks at financial efficiency; the latter at financial depth. How are efficiency and depth related, and what might a theory calibrated to data tell us about the quantitative effect of finance on growth? Fourth, we do not have an understanding of the quantitative significance of any finance-growth mechanisms. We can build models like those in Chapter 2 but without an attempt to connect them with an empirical reality we are left without an understanding of the magnitude of any finance-growth link. Fifth, we have found indications that aggregation issues play a highly significant role in the relationship between transitional growth paths and financial matters. Where does the literature on endogenous financial coalitions and growth impact on contemporary finance and growth theory? Sixth, we have seen that a dynamic financial intermediation story can be relevant. Omitting dynamic elements means that the finance-causes-growth hypothesis is poorly-conceived as a policy tool when applied across heterogenous economies.

At the heart of this critique is an uncertainty regarding the mapping between cross-sectional results and the ability to explain time-series growth patterns. Further, there is no clear theoretical understanding of the mechanisms by which this mapping can be accounted for. How then can we view the relationship between financial depth, financial efficiency and industrial takeoff *within* a country? The remainder of this thesis will strive to address these questions.

As detailed above, in Subsection 1.1.3, recent cliometric research has begun to consider the long-run time-series relationships between measures of aggregate financial depth and long-run growth. Additional research on firm-level data has also been conducted. The findings from these studies suggest to us that the positive relationship between finance and growth is not entirely driven by selection bias in cross-sectional estimations. There *is* a positive correlation between financial depth and economic growth. But still, aggregative cliometrics and richer cross-sectional econometrics do not help us to address the more fundamental, theoretical, issues regarding aggregation and the evolution of financial conditions. Understanding the dynamic interplay between finance and growth in an economy as it goes through a period of transition from low to high growth will shed a great deal of light on the remaining questions.

We intend to do two things in this Chapter: First, given the remaining questions raised by our critique of empirical results and the paucity of any proper longitudinal understanding, we delve into an examination of the finance-growth relation based on historical evidence. This entails an analysis of historical accounts of the finance-growth nexus. We develop a number of key distinctions to be made in our understanding of finance and growth and introduce a new dataset of financial coalitions through the industrial revolution in the UK. Restricting our attention to this evidence means that our findings have a particular implication for countries going through industrialisation. The role of financial markets in industrially developed economies going through transition from non-capitalism to capitalism has been considered elsewhere, e.g., Colombo and Driffill (2003). Second, taking this historical analysis as a guide, we develop a new theory of finance and growth that can account for some of the dynamic, disaggregated elements found in our historical analysis.

The Chapter is organised as follows: Section 3.2 presents our historical analysis. Section 3.3 then constructs an endogenous growth model to account for the stylised facts on finance and growth from the historical analysis. We then calibrate the model to data for industrial growth paths and interpret these numerical results. Section 3.4 concludes the Chapter with a summary of our main findings.

3.2 An Historical Context for Finance and Growth

Economic historians have long considered the industrial revolution from a macroeconomic perspective. The prime example is Landes (1969). More recently, Mokyr (1990) has stressed many of the same issues. Such studies place the technological progress at the heart of the growth mechanic, in concurrence with endogenous growth theory (though the latter also places emphasis on the accumulation and technology of human capital). The importance of financial matters in determining the rate of economic growth has, not least in the economic history literature, taken a back seat. The proposition that financial constraints do not matter has commonly been regarded as a truism: In this view, political and economic incentives are such that impediments to the finance of entrepreneurship are at worst transitory. As we shall see below, however, there is a great deal of evidence that this proposition is very far from a truism and that, in fact, financial constraints can have a significant, if indirect, effect on industrial development. Reconciling these views requires us to take a more detailed look at the financial history, and this in turn requires more disaggregated and richer data.

The analysis of the historical evidence below will, no doubt, paint too many broad brush strokes for the liking of an economic historian. In looking to the historical record for answers on any topic of debate a multitude of conflicting pieces of evidence will always be found. Furthermore, we must, to some extent, generalise away from historical detail in favour of telling a more cogent macroeconomic story. This is necessary if we are to begin to place the theories of Chapter 1 into a more realistic setting. It is hoped that in making these compromises between generality and specifics we do not move too far away from reflecting what actually happened.

Subsection 3.2.1 considers some commonly-cited historical perspectives. Subsection 3.2.2 goes through some historical evidence on the role of finance in entrepreneurship and industry. In doing this, we develop the central distinction to be made between types of finance and the effect of constraints on industrial development. Subsection 3.2.3 develops this distinction by drawing on the historical record of industrial revolution in Europe. Subsection 3.2.4 introduces part of the new *Handbook* of 18th and early 19th century British corporate finance. This *Handbook* sheds further light on our historical analysis and enables us to draw some firmer conclusions. Subsection 3.2.5 begins to draw-out some of the lessons from our historical analysis. Subsection 3.2.6 then looks at these findings in the light of alternative policy environments, specifically the European experience.

3.2.1 Historical Perspectives

The motivation of finance and growth as a subject for debate is often lent weight via the views of various prominent figures in political economy. Naturally, these views have played a part in shaping the way in which modern economists think about the nexus.

Chief among the oft-cited critics of the view that finance leads growth is Robinson (1952, p.86), who famously wrote, "...where enterprise leads, finance follows." The impression given by this phrase belies a deeper, more qualified statement on the importance of financial constraints. Robinson (p. 87) distinguishes between finance as a determinant of enterprise-led growth and as a determinant of finance-led growth; she advocates the view that finance can constrain but only enterprise can cause growth:

> ... the supply of finance cannot be regarded as a rigid bottleneck limiting the rate of investment, but must be treated rather as an element in the general atmosphere encouraging or retarding accumulation.

Lucas (1988) is also frequently cited as a contemporary critic of the view that financial constraints play any role. But his concern that the part played by financial matters in determining economic growth might be "over-stressed" does not preclude the importance of financial 'institutions' *per se.* There is a distinction between money-neutrality and financial services that might reduce transactions costs. He writes (ibid., p.6),

> ... insofar as the development of financial institutions is a limiting factor in development more generally conceived I will be falsifying the picture, and I have no clear idea as to how badly.

The inadequacy of innovation in creating wealth, and the additional need for efficient financial systems, was observed by Bagehot (1873). His *Lombard Street* was among the first to suggest that a scarcity of finance, "no spare money for new and great undertakings" (ibid. para. I.6), can be an element in keeping poor countries poor. Further, it is argued that an ineffectual institutional environment can mean that in rich countries "the money is too scattered, and clings too close to the hands of the owners, to be often obtainable in large quantities for new purposes." He lauds the London money market of Lombard Street as an "efficient and instantly-ready organisation," (ibid. para. I.12). The perceived importance of efficient financial markets is in allowing those that require it to obtain capital from disparate sources at reasonable rates.

There are, for Schumpeter (1934), two agents of economic growth: The first, and better known, depicts innovation as a search for monopoly, or entrepreneurial, profit; the second stresses the importance of finance in *determining* the rate of economic growth, not in simply emerging as an albeit necessary sideshow to technologically-driven growth. The first channel has been adapted into growth theory generally, such as in Aghion and Howitt (1998). The second channel has latterly come to support proponents of the finance-causes-growth school. Schumpeter (1934, p. 74) wrote,

> He [the banker] stands between those who wish to form new combinations and the possessors of productive means. He is essentially a phenomenon of development... He makes possible the carrying out of new combinations, authorises people, in the name of society as it were, to form them. He is the ephor of the exchange economy.

The role of financial intermediaries, according to Schumpeter, is in allowing entrepreneurs to *be* entrepreneurs by mobilising scarce savings, evaluating research projects, managing risk, evaluating future cashflow and facilitating transactions. This is the familiar list of properties attributed by finance and growth researchers, such as Levine (2005). For Schumpeter (*op. cit.*, p. 77), an intermediary exists to mitigate the entanglement of the "entrepreneur's essential function... with other kinds of activity, which as a rule must be much more conspicuous than the essential one."

The idea that finance impacts on growth via inhibiting entrepreneurship, and so technological progress, has taken hold. Of course, finance is central for a healthy entrepreneurship and so constraints on finance can have an impact. But what would cause the banker to *not* stand as a conduit between savers and investors? In other words, why does finance not follow enterprise? Schumpeter does not help us on this: For him, the banker is a phenomenon of development. Robinson and Lucas, among others, are only suspicious of arguments based on the exogenous existence, exogenous persistence, and exogenous impact on entrepreneurship of imperfections in financial markets. How can we reconcile these views? We must consider why such constraints arise in the context of economic development, how they persist and in which economic arenas they act to dampen entrepreneurial spirit.

3.2.2 Historical Evidence

Among the major financial innovations of the industrial revolution in Britain was the creation of the limited liability joint-stock company in the middle of the nineteenth century. Before the Joint Stock Companies Act of 1844, those wishing to establish a joint stock company had to obtain consent from the crown or through Parliament. The Limited Liability Act of 1855 then allowed companies to be incorporated under the protection of limited liability for its investors. Robinson (1952) puts the invention of the joint-stock company on a par with that of the steam-engine. Its formal emergence in England was, according to Hunt (1935a), the outcome of economic necessity in the face of substantial legislative and judicial opposition. This resistance was born partly out of lingering memories of the 1720 South Sea Bubble. As *The Times* of 1833 had it: "if, as a *sleeping partner*, [an investor] chooses to be robbed, the public ought not to be robbed because he chooses to *sleep*."[23]

The belated emergence of joint-stock finance was mirrored by the relatively late growth of formal stock markets. But this apparent delay in the development of financial services did not prevent the emergence over the course of the eighteenth and early nineteenth century alternative, and efficient, methods of industrial finance. As Pollins (1954, p.230) succinctly puts it,

> One of the commonplaces of English economic history is the fact that

[23] Quoted in Hunt (1935b), p. 342.

> manufacturing industry has seldom made use of the [London] capital market machinery for the raising of its capital. In the early period of industrialization small manufacturers were able to make use of their own resources, or loans from friends or from banks; later they ploughed back their profits.

This view is one widely supported in the historical literature. Take Hudson (2002, p.267): "It has long been accepted that internal self-finance was *the* dominant form of industrial finance during the industrial revolution in England."

Critical of this perspective, Harris (2000, p.289) argues that while Britain was enjoying the world's first industrial revolution, the "formal legal framework of business organization remained in its preindustrial state." He argues that this relative backwardness retarded entrepreneurial growth. He notes also, however, how the business corporation evolved from its origination in the sixteenth century to legal acceptance in the first half of the nineteenth. Further, this development "*paralleled* the rise of capitalism, in its mercantilist and industrializing phases." (p. 290, my emphasis). He also points (p.127) to the number of alternative sources of finance in eighteenth century England: Short-term credit and long-term personal borrowing from "banks, merchants and kin" was commonplace. The movement towards the use of stock issue as a method of finance occurred only gradually towards the end of the century.

For Landes (1994, p. 641), "Once they [Europeans] caught the whiff of wealth in their sails, no change in government policy, no want of official support, was going to stop them." Cottrell (1980) writes that it is more generally accepted, that "savings within the economy were not inadequate to support industrialization." (p.5). He also argues that profit-ploughback was the principal method of financing early growth in manufacturing. Some cotton firms even began to borrow money from their employees before looking to banks for finance. A reliance on bank finance for long-term investments only occurred later, when profit margins fell. In other words, firms chose the method of finance to suit their current situation. Further, he argues that the Bubble Act and Usury laws did not constitute an institutional barrier to industrialisation. The later emergence of banking and stock markets reflected not the release of some legislative or institutional constraint but an acceleration of demand for them.

The order of financing method also fits in with the 'pecking-order theory' following Myers (1984). The pecking-order of finance runs as follows: Internal financing is always preferable, followed, if external finance is unavoidable, by debt then equity. So, agents were not constrained to use profit ploughback or limited local credit in the early stages of the industrial revolution: This was their *preference.*

These sources depict an historical record in which financial constraints did not have a direct negative effect on the pace of industrialisation. The oft-cited example of such constraints, the creation of the joint-stock company as an accessible legal entity, was not central to the finance of the early part of the industrial revolution. Further, the resistance to its entrance into legislation did not substantially inhibit the pace of industrial growth. From this standpoint, the pattern of industrial finance through Britain's history is the outcome of relatively unconstrained optimal financing behaviour by firms in response to their desire for expansion under varying economic conditions.

This we take as the story of the finance of industry, in general. Proponents of the enterprise-leads-finance school rightly say that inefficiencies in the finance of industry, *per se*, have not constrained the economic development of the UK. The same was true for continental Europe: Milward and Saul (1973) and Mathias and Postan (1978) tell broadly similar stories in regard to the finance of European industry. Earlier case studies of banking through industrialisation, such as Cameron (1967, 1972) also support these conclusions.

If we consider that most investments in early industry could be small and/or non-lumpy, then this consistent historical story is perhaps not so surprising: An individual entrepreneur, especially a good one, could find the little start-up capital required or use reinvested profits to expand as and when conditions allowed. Even what we might think of to be a large fixed cost in manufacturing, the factory premises constructed to house workers and machines, were often rented in arrears from more wealthy individuals, with multiple tenancy, subletting and power-sharing prevalent (see Hudson, 2002). This stands at odds with the entrepreneur-centred channel offered by standard explanations of the finance-growth link.

In order to see where constraints on finance can have a real effect, we need to make a distinction between the types of activity requiring finance, based on the proportion of the investment which is fixed. Problems in raising finance for investment largely occur where there is a large fixed cost element. Large-scale infrastructure projects are thus prime examples of the class of investments in which financial conditions can have a large effect on economic growth. Among contemporary analyses, Hulten (1996) and Calderón and Servén (2004) indicate what we would expect: An effective supply (i.e., one that works efficiently) of infrastructure can have a large effect on the economic development of an industry that surrounds it. So while finance does not directly constrain industry, it might inhibit the expansion of markets along both supply and demand lines via its impact on the growth of fixed-cost investments such as infrastructure.

In short, it is with regard to investments where fixed costs are high that the financial system has to work harder. Without either wealthy backers or efficient financial markets from which to obtain funds, an individual entrepreneur must either obtain finance from a wide range of sources or forsake the opportunity. Often even a serious backer could not provide all the start-up capital required, and joint-stock operations became necessary. Take, for example, Milward and Saul (1973, pp. 347–8) on the power of French joint-stock companies in comparison with the wealth of even the deepest of individual pockets,

> No matter how large the private fortune of even a family like the Rothschilds it could not hope to bear comparison with a capital fund which was to be built up by selling shares to the public in relatively small denominations and thereby mobilising the collected savings of France.

For investments with a large fixed cost component, all those aspects of the policy, legislative and institutional environment which, we hypothesise, might have an impact on growth via finance come into sharper relief. We need to explore these issues in greater depth.

3.2.3 Finance, Industrial Growth and Infrastructure

We have argued that financial constraints did not directly inhibit the pace of industrial growth. We suspect, however, that limitations on the availability of finance can have an effect on growth via their impact on projects where fixed costs are a large proportion of the investment.

For Bagehot (1873, para. I.6.), the absence of adequate financial institutions is directly related to difficulties in infrastructure development,

> A citizen of London in Queen Elizabeth's time could not have imagined our state of mind. He would have thought that it was of no use inventing railways (if he could have understood what a railway meant), for you would not have been able to collect the capital with which to make them.

How would we characterise an 'efficient' market for the finance of infrastructure? It must be one in which information flows freely between large numbers of savers and investors. Financial constraints deriving from imperfections in financial markets are, in part, a product of both information asymmetry and the politico-institutional environment. In theory, it has been shown that deviations from the complete financial markets of Arrow-Debreu can arise as a result of information problems. For Grossman and Stiglitz (1980), costs to the acquisition of information mean that information-efficient markets are impossible. In anticipation of Chapter 4, we might add to the costs of obtaining information the (related) costs of forming comprehensively complete contracts.

The purpose of this section is to investigate the realities of those information problems, which costs exist to cause them, and the ways in which they impacted upon optimal financial systems. We also look at the part played by the institutional and legislative environment in this context. We are able to see the effect that fixed costs have on the nature of the financial systems that emerge to cope. We characterise in detail, using both secondary and primary sources, the nature of the financial coalitions that emerged to supply infrastructure finance and how these typically evolved over time. In Section 3.3 we develop a model that can capture a number of the macroeconomic implications for finance and growth, leaving to Chapter 4 a deeper theoretical investigation into some of the microeconomic roots of our findings.

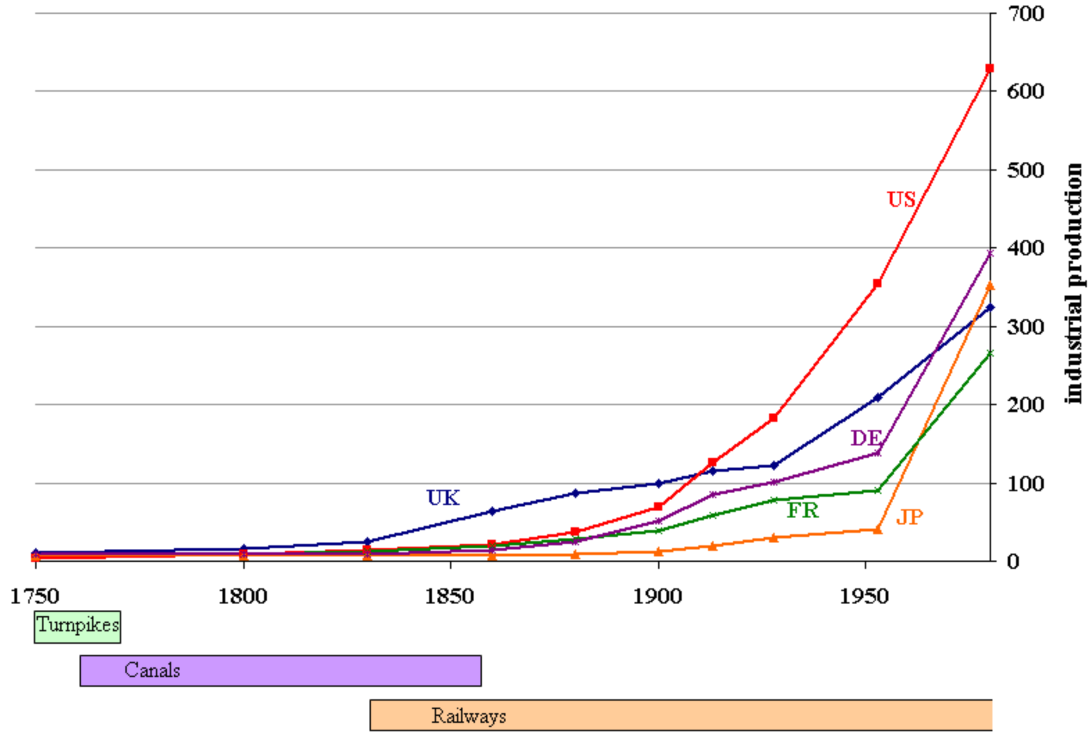
We focus on the finance of physical transport infrastructure for four reasons: They are a classic example of investment projects in which fixed costs are a large element; the development of an effective infrastructure is strongly correlated with industrial development; infrastructures are constructed using both public finance and private enterprise under public regulations; and, perhaps most importantly, there is a great deal of historical and cliometric evidence covering the finance of infrastructure through the period of industrial revolution. We must appreciate that there are limitations to this focus, however. As described below, much of the historical observations are based on the nature of financial coalitions that emerged to provide the *initial* finance of transport infrastructures. In the absence of data on the way that individual coalitions morph over the course of industrial revolution, we look primarily at the changing initial structures of different coalitions. A further drawback from concentrating on physical transport infrastructure is the potential bias towards finance from those that actually use the infrastructure. As users, firms with high freight costs will naturally benefit from financing infrastructure in their area. As investors, the firms also benefit from any return on the shares they purchase. We proceed with these reservations in mind, and draw special attention when we can to findings that partially address these qualms.

We can broadly date the phases of each of the main forms of transport infrastructure in the UK: In the 1750s and 1760s, there was a boom in the number of turnpike trusts; from the 1760s to the 1850s canals dominated, with a boom in 1792–3; and from the 1830s the 'Railway Age' commenced. The first railway 'mania', of 1843–47, saw capital expenditures on railways increase tenfold in the space of five years (see Kenwood, 1965).

For reference, we replicate in Figure 3.1 these phases of infrastructure development against time paths for the per capita volume of industrial production, using the data in Bairoch (1982). The data (reproduced in Appendix Table B.3) are normalised relative to 100 in the UK for the year 1900. We can see that the UK had an early lead in industrial takeoff which was sustained for over a century and a half.

The recent dispute between Landes (1994, 1995) and Crafts (1995a,b) demonstrates that there are no simple answers to the question 'Why Was England First?' Additionally, we have argued above that underlying shifts in technological progress, those which are subject to entrepreneurship, have not been significantly dampened by financial constraints. We are not suggesting a new 'single-cause' explanation. What we wish to understand, however, is whether financial constraints might have

Figure 3.1: Per Capita Industrial Production and UK Infrastructure, 1750–1980



played a part in limiting the growth rate through their effect on the development of infrastructure.

The early history of the development of physical infrastructure in Britain was one of private enterprise and local finance. Government practiced its regulatory role with a light-touch, enforcing only some standards on construction and maintenance. Except for major disputes over land use, Parliament did not meddle with the layout of infrastructure.

Along with the standardisation in the construction of infrastructure, the industrial revolution saw a gradual standardisation of the way in which infrastructure finance was regulated through Parliament. The Bubble Act of 1720 necessitated that joint-stock companies be authorised by royal charter. Thus, the construction of any piece of infrastructure required a Bill to be passed in Parliament. Further, from 1794, after the canal mania, requirements for an infrastructure Bill included the need to deposit three things: A map of landholdings in the vicinity of the project; reference books (linked to the map of landholdings) of landowners and occupiers as well as their support or opposition to the plan; and, a subscription list of proposed financial supporters. These deposits enabled Parliament to judge not only the likely success of the proposed project, but also to consider conflicting local interests before securing private property rights to a corporation.

Most of the evidence cited on the finance of infrastructure is based on analysis of these records. As such, it is an excellent source and we can read a great deal into them with confidence. Using these sources also means, however, that they often cover only the initial period of any piece of infrastructure and not its performance over time after its construction. We pay attention to dynamic analyses of infrastructure finance wherever possible.

The Emergence of Turnpikes in the UK

A surprising amount of evidence can be drawn upon to characterise the emergence of turnpike trusts in the early and middle parts of the eighteenth century. Prior to the enactment of legislation allowing the establishment of turnpike trusts, road maintenance was carried out by local parishes, funded by levying local taxes. From the beginning eighteenth century, however, Parliament could approve the establishment of individual turnpike trusts. The trusts could raise finance through issuing debt and levying tolls on road users. Bogart (2005) suggests that this institutional innovation brought forth a wider road transport revolution. It is beyond question that the construction of canals and railways aided industry by reducing the cost of transporting goods. Bogart shows that, in addition to waterways and railways, the levying of tolls by turnpikes did not increase freight charges, and may have even reduced them.

Buchanan (1986, p.227) notes that most turnpike companies "were run by men representative of the economic life of the area." His analysis sheds light on the existence of a significant regional capital market and demonstrates, using the Ward (1974) classification scheme, that investors were large in number and from a wide range of social strata. Further to this, Bogart (2006a) finds regional and network effects in the diffusion of turnpikes. Turnpike trusts were typically spatially concentrated in the vicinity of major economic centres.

Financing the UK Canal Network

Ward (1974) develops a highly detailed analysis of the finance of canal companies in England through the eighteenth century. A group of industrial and merchant leaders would see the 'economic' motive, in terms of the direct benefit to their business, of installing a canal in their vicinity. These promoters would then either call upon a financier, or become financiers themselves, to sell scrip and shares in a joint-stock canal company under a 'financial' motive of potential future returns on holding the issue. Ward shows that canals were generally financed by those local to the route of the canal. The financiers would tour local public houses, hold town meetings, coax relatives and friends, to convince local individuals to invest in the scheme. Those who invested were by no means uniformly wealthy. Ward classifies investors by occupation and social status, showing that the majority of investments came from local landowners, merchants, tradesmen, manufacturers and professionals – people whose wealth was not only relatively limited but also mostly tied up in their primary employment.

The spatial concentration was, particularly during and after investment booms, sometimes a restriction ordered by financiers wishing to avoid speculative investment. More often, however, the parochial nature of finance was a result of informational asymmetries: A local familiarity with market conditions, with local industry and an affinity with the canal promoters made it more easy to raise finance locally than on the London market. Ward puts the spatial concentration down to inequalities in the social and geographic distribution of capital. He notes (ibid., p.171–2) an unwillingness of London creditors to invest in regional infrastructure projects because "appropriate capital markets did not exist and the scale of investment was insufficient to support them." We prefer to interpret this observation in the following way: The costs of forming spatially concentrated coalitions of small investors was, under the institutional environs, less than the costs of inducing interest from those with more investment experience and deeper pockets. We come on to the possible reasons for this below. This pattern is seen across the country and throughout the century. Ward notes, however, that through the eighteenth and into the nineteenth century, centralised stock markets became more willing to support infrastructure projects; in the process, the problems of finance were gradually relaxed.

Turnbull (1987) finds not only a similar pattern of regional finance of canals but also a localised pattern in the *construction* of canals. The importance of an integrated, national system of waterways gave way to local and regional demands for routes unconnected to trunk lines. Canals were built as local enterprises first, and formed part of a national network only later. This was the outcome of market forces, and reflects the idea that the industrial revolution itself was regional. For Cottrell (1980, p.19), the "industrial revolution... was essentially a process of regional growth." Recent work, such as O'Brien (2006), continues a growing revisionist literature on the industrial revolution as a provincial phenomenon.

The Finance of Railways in the UK

There is a great deal of evidence that railways were financed using methods similar to those employed in turnpikes and canals, namely spatially concentrated coalitions of local, relatively modest investors. Hunt (1935b) suggests that the English railway network was established without the use of the London stock exchange. Pollins (1954, pp.230–1) describes the establishment of a typical railway during the first half of the nineteenth century:

> Some public-spirited men. . . would recognize the need for improved communication in their locality. They would subscribe among themselves to finance a survey, or would call public meetings in the locality to arouse support and obtain subscriptions. Later a definite route would be decided upon, public meetings held to sell scrip, and the committee of the company would appoint local agents to obtain subscriptions and take deposits. Those who took scrip would be asked to sign the subscription contract (or parliamentary undertaking) required by parliamentary Standing Orders, and an application would be made to parliament for an Act of incorporation.

Pollins suggests this process is repeated across the country in the finance of canals, tunnels, docks and railways. Again, it is observed that companies often reserved shareholdings for local landowners and occupants of towns along the route of the railway. He suggests there was also some element of learning to the emergence of financial coalitions: Before a new form of transport has been tried and tested, potential investors took more convincing that buying shares was worthwhile. Even once technologically proven, during boom-periods a proximity to the route of a proposed railway aided potential investors in deciding over which railway companies to support. Only after a few years of local finance did companies float on the stock exchange.

Broadbridge (1955) also finds similar results on the spatially concentrated nature of early railway capital. That paper also points to the later emergence of regional centres of finance, particularly in the North of England. A good deal of even Scottish railway stock was held in Lancashire. He tracks a gradual movement in the second half of the nineteenth century away from local subscription toward London. This did not happen just because the stock market was there, but because conditions and capital requirements changed: His evidence supports the "conventional view that railways were drawing their capital from ever-widening sources in the early 1840's, as compared with previous decades." (ibid., p.206).

To firm-up our understanding of the financial coalitions formed to finance infrastructure projects during the industrial revolution we can draw upon a new dataset of corporate finance built by Gary Shea. We summarise some of the current findings from Shea (2007) in the next section, and give a concrete example of how these sorts of findings can be drawn.

3.2.4 The Shea *Handbook*

We intend to give an indication of the sorts of data that can be drawn upon to make inferences about the finance of infrastructure. To do so, we present two case studies from Shea (2007). The full *Handbook*, once completed, will cover a number of case studies, for around 20 railways, 80 canals, 20 energy companies, and 40 others over the period 1760-1834.

We include here two examples of how such records, along with other materials, can be used to illustrate the successful projection of a local infrastructure project. There is nothing special about the Wigan Branch Railway, constructed in 1830. It was minor railway company, initially issuing 700 shares, which could be paid up to £100 each, but it was typical of about another 6 or 7 small railways built in Lancashire from the late 1820's and it was typical of how many other early British railways, canals and gas and water supply were built. In addition, we present some data one the larger St. Helens and Runcorn Gap Railway, which was first promoted in 1825 and constructed in the early 1830s. That railway company issued 1200 shares of up to £100 each.

Figure 3.2: Wigan Branch Railway: Geography

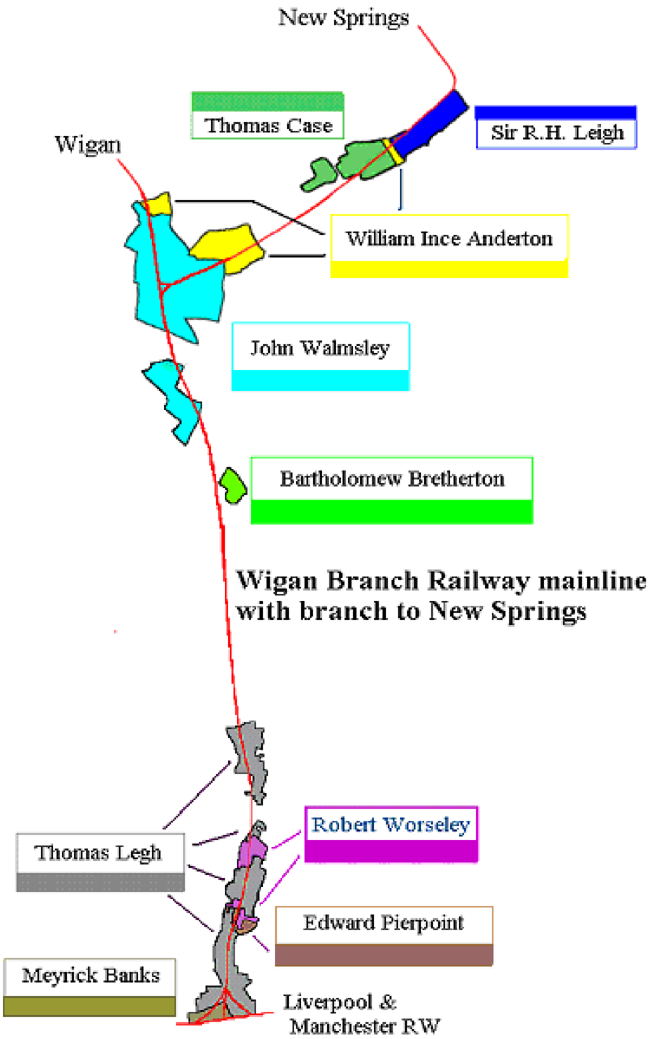


Figure 3.2 maps the route of the Wigan Branch, as well as noting those landowners along the route who also invested in the railway company. These landowners comprised almost 10% of total capital investment. Figure 3.3 gives a breakdown of investors under the Ward (1974) classification scheme. Figure 3.4 does the same for the St. Helens Railway.

A great deal of detail about the railways will, upon its completion, be contained in Shea (2007). For example, how were the railways promoted; what support or dissent was offered in Parliament and by private individuals; how was the final routes of these railways arrived at; and, what was their position within a wider

Figure 3.3: Wigan Branch Railway: Investors

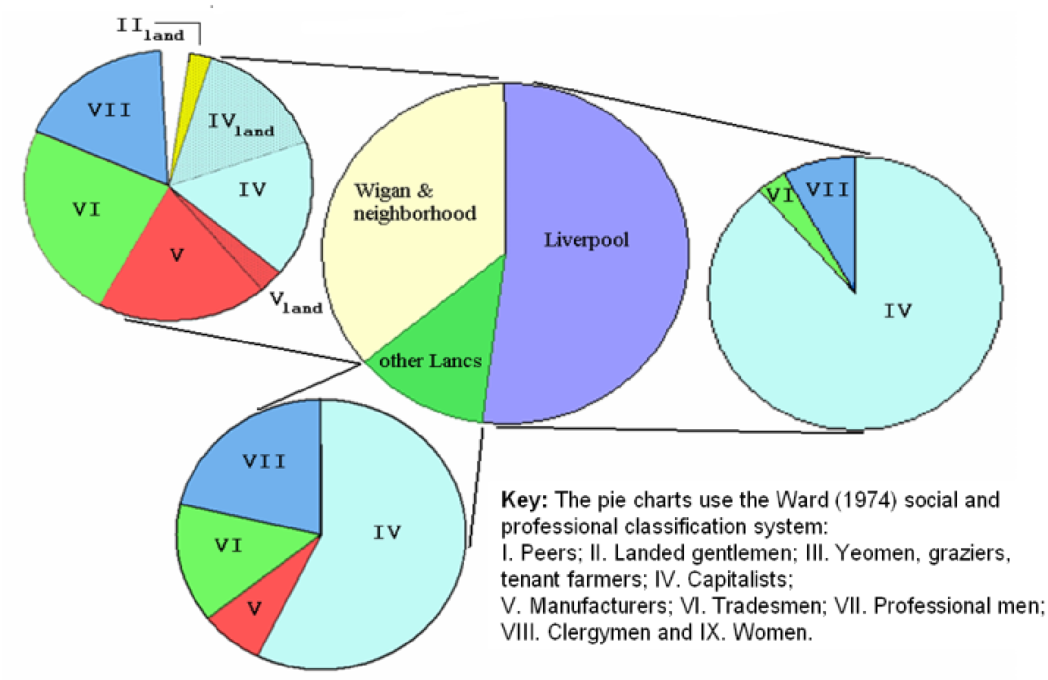
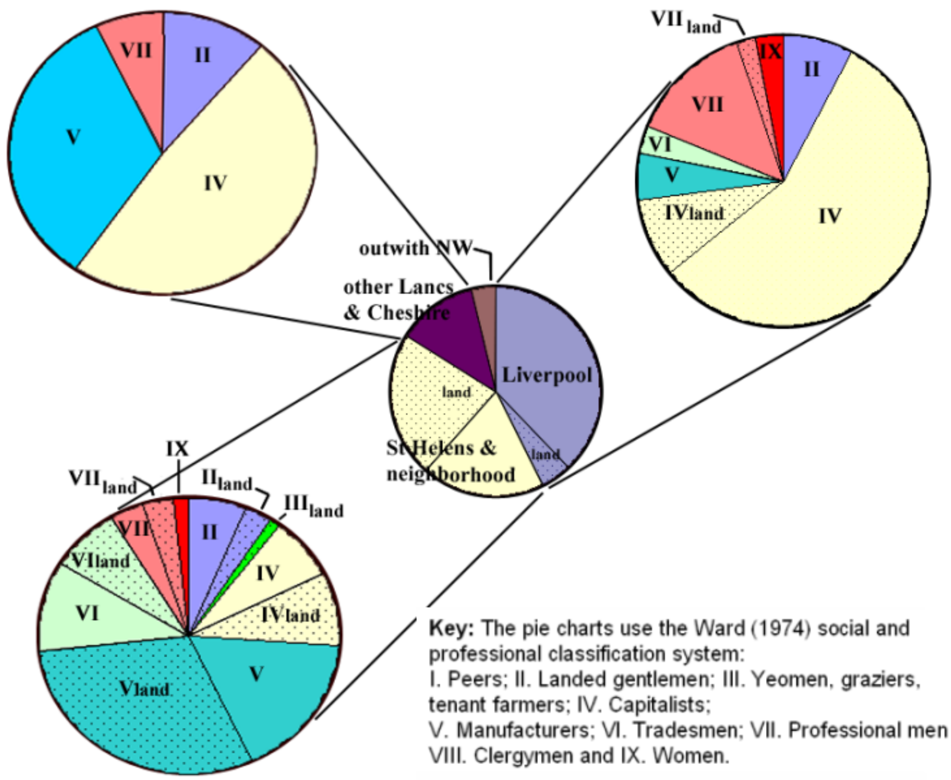


Figure 3.4: St. Helens and Runcorn Gap Railway: Investors



context of the construction of major trunk lines? This richness in the dataset will enable us to eventually develop a fundamental understanding of the evolution of railways through the eighteenth century. This will aid in the development of a dynamic theory of the microfoundations of finance and growth.

For the purposes of this thesis, however, we take only a few more salient points, as illustrative support for the historical arguments based on secondary evidence above. We can see clearly from these Figures, and from the dataset in Shea (2007), how investment was heavily financed by share issue, not debt acquisition. They suggests that shares were not purchased by a small number of very wealthy individuals, but by a large number of relatively modest local capitalists, manufacturers, tradesmen and professionals. It also clearly demonstrates how these investors were spatially concentrated around the location of the railway, and suggests a pattern for landowners along the path of a railway to also invest in it.

3.2.5 Regional Growth, Spatial Concentrations and Dynamic Aspects

We can summarise a few stylised facts from the above analysis in relatively short order. The finance of early infrastructures in Britain generally took the form of spatially concentrated coalitions of large numbers of relatively modest individuals. In the early stages, infrastructure development was also nationally disjointed. A national network emerged later, as the industrial revolution took hold. At the same time, centralised financial markets became the primary method of raising finance, and infrastructure projects could raise finance from the stock exchange or large banks.

How should we interpret both the nature of coalitions and their change over time? The scale and form of financial intermediation was the key decision in any entrepreneurs' attempt to finance an infrastructure project. We need to consider how the costs of intermediation are related to its scale; the role played by information asymmetries must be central.

In the early stages of development, the cost of obtaining external finance from central financial markets was greater than the cost of raising it from scratch from among a relatively modest local populace. Even though the process of raising local finance in this way was time-consuming and expensive, lesser information problems at the local level made local markets preferable to seeking finance in more distant ones. This must have something to do with the cost of forming effective contracts. For financial coalitions to form, members must be contracted into it: Payoffs, responsibilities, and other actions must be specified in an environment where there may be a great deal of uncertainty about the future. The costs of forming these contracts cannot always be considered to be negligible. And there is a consequent decision to be made about the *quality* of that contract (i.e., how well it specifies party behaviour, how many loopholes are left, how enforceable it is), given its cost. Moreover, there will be a degree of heterogeneity: Some agents are better informed, or better educated than others while the level of trust between agents can also vary. This heterogeneity can be related to the proximity of a potential investor to an investment project. In addition, the institutional environment can affect the costs of forming contracts (and so the choice of contract quality) by, among other things, establishing industry standards, supporting property rights, easing the gathering of information and enforcing written contracts. We come on to consider further these issues in Chapter 4.

The very existence of spatially concentrated financial coalitions suggests that the national financial market was not perfectly efficient in informational and contractual terms. In the absence of information and contracting problems at the national level, and given a conducive institutional environment, a national financial market would have been first-best efficient in terms of the supply of finance. Without such an efficient national financial market, financiers' only alternative was to build financial coalitions of local investors.

We cannot draw from this that the financial arrangements which emerged to finance infrastructure were not themselves *optimal*: Given the costs and benefits of obtaining information, and the costs and benefits of forming contracts, the financial structures observed through the industrial revolution might simply have equated marginal conditions in information gathering and contract writing. But this is not just a problem of private arrangements. It is clear that the costs of obtaining information and forming contracts can be subject to further institutional and legislative constraints. Moreover, the capacity of the public purse to mitigate information problems is itself not limitless: Optimal levels of public expenditure, and so optimal levels of taxation, follow. There is then, in theory at least, an optimal combination of private and public behaviour that combines to support the optimal, private, financial arrangement. We develop a theoretical framework for understanding these trade-offs in Chapters 4 and 5. For the purposes of the current Chapter, we abstract from formalising the existence of optimal information extraction and optimal contract formation. Our focus is, then, on the institutional environment as the source of inefficiencies (sub-optimal outcomes) in the finance of infrastructure. Even where the institutional and legislative environment deviates from that which was optimal, we consider that rational private arrangements to mitigate information problems were optimal responses to *both* the present institutional conditions and the private costs and benefits of obtaining information and writing contracts.

Despite the spatial concentration that characterised the early industrial period, there was an institutional and economic learning process. Over time, as infrastructures became larger, as industrial centres became more evident and as industrial development began to require more sophisticated external financing, the financial systems of London evolved into markets more capable of evaluating distant (and, increasingly, larger) infrastructure projects. Institutional and legislative changes played a part in this evolution. It has been argued that the government played a role in advancing property rights and encouraging the private provision of public goods. Bogart and Richardson (2006) introduce a large database on the passage of Acts of Parliament and show that the passage of Acts pertaining to enforcing property rights and encouraging public good provision were positively correlated with (and sometimes led) the provision of infrastructure and the rate of economic growth in the run-up to industrial revolution. That paper also draws attention to the unique position that Parliament held in acting as a forum for transforming the structure of landholding into its modern, capitalist form. As they note, in most other nations this transformation was delayed and, in France and Russia, sowed the seeds of revolution.

Through the middle of the nineteenth century, centralised and specialised financial services gained precedence and began to cater for the greater demands of larger infrastructure projects. The informational problems at the national level began to wane as the institutional framework for centralised finance developed, and as the returns on infrastructure investment became more reliable. Eventually, central financial markets were such that constructing local and regional coalitions of finance was the less efficient method of finance.

The pattern in provincial infrastructure development is mirrored by the provincial nature of the industrial revolution. Disjointed local and regional infrastructures supported a local and regional industrial growth that itself comprised, on the national level, a disjointed patchwork of regional economies. As the national economy emerged, so, in parallel, did both the national infrastructure and the national financial markets to finance it.

So there are three effects here: First, a learning process in national financial markets made them gradually more amenable to the finance of distant infrastructure projects; second, the development of political and institutional environment had an impact on the efficiency of centralised financial markets; and third, economic development and the growth of the stock of infrastructure made economic integration and market expansion an additional incentive to build and finance infrastructure nationally.

We develop in Section 3.3 a model of finance and growth that can capture a number of the stylised facts we have outlined here. Before that, we look in Subsection 3.2.6 at whether alternative policy environments, such as a bias toward public finance and public planning, had an effect on the development and finance of infrastructure in other countries going through industrial revolution.

3.2.6 Alternative Policy Environments

To appreciate properly the experience of British infrastructural development we must look at it in the light of a wider context. France is a good example of a very different approach to that in Britain. We might also look to the emerging body of cliometric research on the US and Canada to widen the application of our hypotheses: Wright (2002) draws attention to the importance of information problems in the emergence of the US financial system over the period 1780–1850; and, Sylla et al. (2006) track the integration of transatlantic capital markets through the first half of the nineteenth century. We focus on the Anglo-French contrast here, since it serves to motivate an understanding of the finance of infrastructure in what were, at the outset of the industrial revolution in Britain, otherwise relatively similar economies.

Broadly, the development of British infrastructure was one based on market forces; that of France was the outcome of public planning and a great deal public finance. The French industrial revolution occurred much later than in Britain, some argue that it began properly as late as the 1850s; this is affirmed by inspection of Figure 3.1.

Milward and Saul (1973) are among those that put the delay in French economic development down, in part, to the way in which infrastructure policy was formed. French governments of the late eighteenth and early nineteenth century initiated a publicly financed, centrally planned and tightly regulated system of canals that, it was intended, would serve all citizens at no charge. The Becquet plan of 1822 envisioned a public-private partnership: A rational (i.e., centrally planned) waterway system paid for by private capital. A group of civil engineers, the Corps des Ponts et Chaussées, was charged with setting and enforcing the regulations for a waterway network of sufficient quality. The plans did not come to full fruition. The routes which did get built quickly were those where local economic demands most greatly necessitated them.

Lévy-Leboyer (1978) notes that the centralised nature of infrastructural development in France extended beyond canals and covered also railways. There was, in addition, a great deal of overlap in the plans. The Corps, at first, viewed railroads as 'dry canals'. Smith (1990, p.665) writes that Becquey and the Corps were remembered for "committing their country to waterways on the eve of the railroad age."

Many of the new railway schemes were in direct competition to the previously planned canal networks. A solution based on complementarity between railways and waterways was found: Canals, it was thought, served best the transport of heavy goods while railways carried light goods and passengers. Indeed, plans for some stretches of infrastructure specified parallel lines of canal and railway (see Figure 2 in Smith, 1990).

The Legrand plan of 1838 began to map the national plan for canals to one for railways. A system of trunk lines emanating from Paris to each of the largest cities was envisaged, a so-called Legrand Star. Private companies were restricted from constructing major routes for fear that it would interfere with the greater plan. Up till 1837, only three private bids to construct railways were accepted. From 1865, smaller lines could, at the discretion of local authorities, be appended to the trunk lines as and when local economic conditions demanded.

By the middle of the nineteenth century, Napoleon III began to promote the private finance of a dominant railway infrastructure. Private infrastructure developments were still subject to the layout, location and specifications dictated by the Corps. Milward and Saul (1973, p. 336) note that government "beset railway building with so many safeguards as to delay its flourishing by a full decade." Again, those railway lines that were taken up were those in greatest demand by local industry.

The French experience of public planning can be contrasted with the experience of Germany. There, it is argued, the country was able to construct a railway infrastructure much more quickly because of less stringent requirements on standards and a more liberal approach to the granting of private enterprises. Further, the political fragmentation of mid-nineteenth century Germany allowed separate regions to go ahead with railway developments in opposition to more central directives: Regional political units could better coordinate infrastructural development.

In our discussions of previous sections we viewed the finance of infrastructure in Britain as somewhat inefficient. The experience of France suggests at least that the encouragement of private finance and a relatively laissez-faire approach to regulation was the right policy approach in terms of the development of infrastructures. We do not wish to imply (though it is by some) that the retardation of infrastructural development was the root of the delayed entrance of France into high growth paths, of course there can be many other reasons.

There is, however, further evidence on the relations between infrastructure development, policy issues and economic growth. Hulten (1996) draws attention to the importance of maintaining an *effective* stock of infrastructure. Using World Bank indices for road condition, locomotive availability, electricity loss and telephone faults, the paper finds a link between the effectiveness of infrastructure and the rate of economic growth. A large positive impact on growth of improving the effectiveness of a given stock of infrastructure is found, especially among poor countries. Bogart (2006b), using cross-country evidence for the nineteenth century, finds a strong and positive correlation between the level of investment in railways and the proportion of private ownership. Further, the rate of railway diffusion is positively related to the extent of private sector involvement.

So policy can have a role to play in the development of infrastructure, and this role can have implications for financial development. But we do not seek to draw firm conclusions on the role of policy in the finance of infrastructure. Counterfactuals can always be found to any argument based on case studies. The experience of Germany may be one such counterfactual. We rather think of policy, with apologies to Robinson (1952), as an element in the general atmosphere of encouraging or retarding the finance of effective infrastructures.

3.3 Some Analytics of Growth, Finance and Infrastructure

Given what we have learned from the historical analysis above, we now proceed to construct a theory of finance and growth that can account for a number of the disaggregated and dynamic aspects observed. This model is something of an intermediate step between the aggregative, cross-sectional finance and growth theories summarised in Chapter 2 and the fully microfounded models advanced in Chapters 4 and 5. We wish to capture the stylised facts from our historical analysis in a relatively transparent model of finance and growth that can be, broadly, matched to data. Our model is distinct from work such as Acemoglu and Zillibotti (1997) primarily because an economy in our model can be caught in a low-growth trap. In the paper of Acemoglu and Zillibotti, industrial takeoff is an inevitability. Though it might reflect on some levels the relationship between finance and growth in countries (or regions) that *do* industrialise, it does not help us to understand why some countries do and other countries do not enter periods of sustained high growth.

We first describe the structure of our model before presenting it in more formal terms. We have a closed economy with two major regions. Factors of production are capital and infrastructure. Following the discussion above, we also consider that this infrastructure is the product of private enterprise. A local supply of infrastructure benefits all those firms who pay for it.

Firms have no trouble raising capital; they can use profit-ploughback or sell claims on future profits. But infrastructure projects are subject to significant fixed costs; they cannot be funded by individual firms or individual agents alone. Entrepreneurs exist to see the demand for infrastructure and organise finance via financial intermediation to construct and lease infrastructure to firms. In this vein, we account for the interaction between infrastructure and financial intermediation by thinking of infrastructure as a direct input to production; the efficiency of financial intermediation then determines the costliness of raising finance for that infrastructure. Firm output is determined in part by the level of infrastructure that the firm is willing to pay for, given the costs of raising finance for that infrastructure.

In Chapter 4 we will consider in greater detail the interaction between the costliness of exchange and the institutional, economic and social environment in which exchange takes place. The purpose of the remainder of this Chapter, however, is to lay-out the implications of the discussion from Section 3.2 for a relatively standard theory of finance and growth. This then serves as a bridge between the critique of current theory in Chapters 1–2 and a closer look at the fundamentals driving these interactions in Chapter 4.

So, the employment of infrastructure as a factor of production incurs the costs of financial intermediation. The economy is populated by agents endowed with a money income each period. We could otherwise have thought of agents endowed with a unit of labour earning a money income by renting their labour to firms. When agents in the economy see the demand for infrastructure they can become entrepreneurs. Entrepreneurs use the services of a financial intermediary to raise finance for the construction and maintenance of an infrastructure. Entrepreneurs then rent that infrastructure to the firm.

We do not have to restrict the intermediary, or even the entrepreneur, to being external to the firm; we simply want to allow for an effect of intermediation costs on spatial decisions and growth. Of course, it is likely that specialised financial intermediation will emerge, so we naturally think of the existence of a market for financial intermediation. Additionally, it may be that the financial intermediary is also the entrepreneur supplying infrastructure to the firm. For clarity, we think of firms, entrepreneurs and intermediaries in isolation.

3.3.1 Finance, Productivity and Economic Integration

We found a clear and consistent pattern in the historical evidence discussed above: In the early stages of development financial coalitions, infrastructure and markets are, broadly, local; as the economy grows infrastructures grow, financial intermediation becomes more sophisticated and markets become more national. We wish to account for and understand these effects: Why do firms choose local markets at low levels of development? Might an economy be trapped in a spatially disparate, low-growth trap? What characterises the transition from small, spatially concentrated markets to large, national ones? What part does the efficiency of financial intermediation play? What government policies might instigate faster growth paths?

In a set-up with two regions, funds can be raised at the regional level, via regional financial intermediaries, or at the national level, using national financial intermediaries. We make two central assumptions: 1) A regional financial intermediary can only finance a regional infrastructure; and, 2) a firm can only operate at the scale of the infrastructure that it employs. The costs of financial intermediation are subject to scale effects. Specifically, we impose that financial intermediation at the national level can, *ceteris paribus*, be more costly.

Part of the incentive for firms to operate at the national level arises from *extensive* scale effects, that is, scale effects resulting from the linking of separate economies. For McDermott (2002, p. 373), extensive scale increases potential income partly by enhancing "the productivity of research and study". Rivera-Batiz and Romer (1991) takes a similar perspective in the context of an endogenous growth theory. In Alesina et al. (2000) productivity is directly related to economic integration via the imposition of iceberg costs in the trade of intermediate goods. In our economy, the integration of two regions into a national economy raises firm-level productivity and can obtain a higher growth rate.

Extensive scale also has a role in financial intermediary conditions. We consider two types of effect: Fixed information costs and learning costs. The historical analysis above has indicated that both firms and investors can have a preference for local finance where markets are small. We have argued that this is the result of a bias towards exchange and finance with those who are geographically closer which arises when the returns to centralised finance are outweighed by its costs. This is what we call the fixed information cost: It is easier for an entrepreneur, E, to convince an investor, I, to invest if I knows and trusts E personally; if E has a good local reputation; if I knows the local market conditions well; if institutional factors are conducive to I evaluating E; if if I can more easily monitor the activities of E. We posit that these scale effects are fixed; they hold true no matter what the market size. What is important, however, is that these information costs can depend on the legislative and institutional environment within which exchange takes place. We use this fixed scale effect to capture exogenous changes in the institutional environment.

We have also seen that a transition from spatially concentrated finance and markets to economy-wide finance and markets can occur. We have suggested that this is not entirely the result of exogenous changes in political or institutional factors. We can, in part, put it down to endogenous changes in financial intermediation conditions. We account for, in addition to the fixed information cost effect, a learning cost in financial intermediation. We have seen hints of this effect in the historical analysis. We can think of four separate effects, though there are, no doubt, more.

First, consider a demonstration effect: It is harder to raise finance for the first national turnpike/canal/railway than it is for the fiftieth because of the initial newness of a technology or as a result of the unwillingness to risk money when the demand is uncertain. As a new form infrastructure is tried and tested by those willing to make initial investments, and as the demand for said infrastructure is demonstrated, it becomes progressively easier to raise finance. Secondly, a scale effect in construction can be considered: The first national turnpike/canal/railway will be more expensive to build than the fiftieth. This can result from technological improvements in rolling-out an infrastructure as it is used more. Third, there will be a learning effect in the *technology* of financial intermediation at the national level. We have seen that sophisticated financial markets do only emerge gradually, and that this emergence often parallels the growth of the economy. Fourth, the demand for major infrastructure projects that require national financial intermediation only occurs over time, as the economy grows regionally and as regional infrastructure networks are contstructed.

Of course, we are abstracting from such market-effects at the regional level for simplicity. In practice, these learning effects will exist at the regional level. The difference is that at the regional level these effects can be greatly mitigated because of the local nature of the investors.

Intermediation at a regional level is thus optimal when learning and fixed costs are relatively high. As the regional economies grow so learning costs fall (with the building of a regional infrastructure) and national integration becomes more feasible. Likewise, if market-invariant information costs fall at a national level, we can move from an equilibrium with two regions to an equilibrium with one national economy. If the regional economy lies on the balanced growth path then over time the learning-cost barrier to integration will become insignificant. There will thus be three possible outcomes, dependent on both the extent of information costs of national intermediation and initial infrastructure conditions: Information costs remain high and regional economies are optimal in the long run; information costs are low, initial infrastructure is good and economy-wide output is optimal from the start; or information costs and initial infrastructure are such that we begin with regional economies and (via falling learning costs or exogenous institutional shifts) we move endogenously to a national economy.

3.3.2 Formal Model

We have two regions, A and B, in the same closed economy (i.e., there are no exchange rate complications here). In each region there are $\#F^A$ and $\#F^B$ firms, respectively, and $\#C^A$ and $\#C^B$ consumers, respectively. We normalise each of these cardinalities to unity (though below we leave them in entrepreneurship equations for generality). One region cannot trade capital or finance with another without financial intermediation and infrastructure on a *national* level. There are two possible situations at any given point in time: Either the economy operates as two separate regions with no trade in capital or finance between regions; or we have a combined national economy with a national output.

Figure 3.5 depicts the flow of resources between consuming agents, intermediaries, entrepreneurs and firms in our model. Agents maintain a money supply (by dint of a Central Bank) and are given an initial capital endowment. Firms demand capital and infrastructure. Agents can sell capital direct to firms but infrastructure is supplied by entrepreneurs who use financial intermediaries to raise the finance from agents. The costliness of intermediation drives a wedge between savings and investments that impacts upon what both firms and agents receive. Consumption optimisation by agents, combined with a specific production function for firms, generates an endogenous growth based on externalities in the manner of Rebelo (1991). This story abstracts from any distinctions over integration, so we now proceed to consider the model depicted in the Figure under both the regional and the national context.

Regional Growth

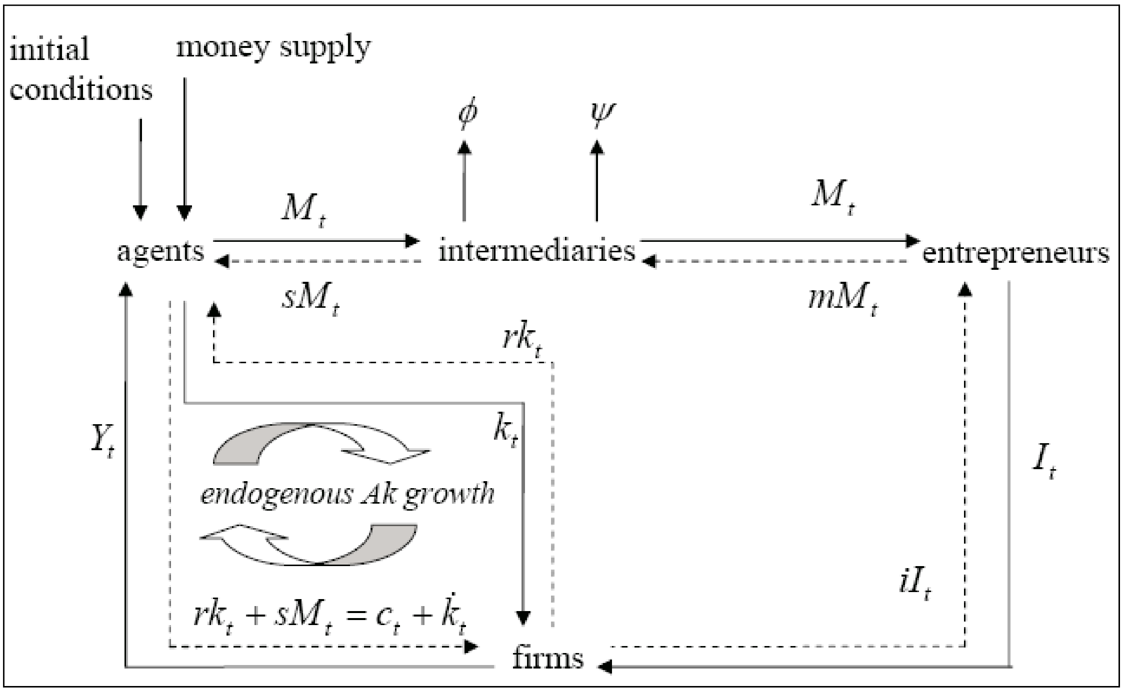
Two factors enter the production function: Capital, k_t, and infrastructure, I_t. If we assume both regions are specified identically and that initial values for capital and infrastructure, k_0 and I_0, are identical in both regions then we can find equilibrium growth for one region knowing that it is equal to growth in the other region. So we consider region A, in which firms produce a single good,

$$Y_t^A = A(k_t^A)^\alpha (I_t^A)^{1-\alpha}, \tag{3.1}$$

where $0 < \alpha < 1$. Each firm maximises profits,

$$\pi_t^A = Y_t^A - rk_t^A - iI_t^A, \tag{3.2}$$

Figure 3.5: Financial Intermediation and Growth Schematic



where each takes the rates of return on capital, r, and infrastructure, i, as given,

$$r = \alpha \frac{Y_t^A}{k_t^A}; \tag{3.3}$$

$$i = (1-\alpha) \frac{Y_t^A}{I_t^A}. \tag{3.4}$$

There is a market for the construction of infrastructure. Agents can recognise the demand for infrastructure by firms but cannot fund it themselves, they must obtain the services of financial intermediaries to raise the necessary capital. In so doing, they become entrepreneurs. Suppose that infrastructure is produced without capital, using money alone: $I_t = f(M_t)$. We could otherwise have thought of money raised to buy capital or labour for the construction of infrastructure. The entrepreneurs' incentive is the rent she can charge for firms' use of the infrastructure. Let us suppose that there are a large number, $\#E^A$, of agents who wish to become entrepreneurs. We normalise $\#E^A$ to unity; i.e., any agent can become an entrepreneur. Further let us suppose that there are no costs to *becoming* an entrepreneur. Perfectly competitive entrepreneurship will obtain,

$$\#F^A i I_t^A = mf(\#C^A M_t^A), \tag{3.5}$$

where m is the rate on finance supplied by the intermediary and M_t^A is the finance raised from each agent in region A. If we specify a simple linear production function for infrastructure, $I_t = M_t$, then $i = m$. This is the perfectly competitive outcome: If entrepreneurs could develop a monopolistic position with respect to firms or a monopsonistic position with respect to intermediaries then there would be a surplus to entrepreneurship in which $i > m$.

A financial intermediary exists to raise finance from agents and sell it to entrepreneurs. Any agent who sees the demand for financial intermediation can become an intermediary. Again, we are being purposefully loose about occupational choice here: An agent can become an entrepreneur or an intermediary or, indeed, can do both jobs. The mingling of occupations in this fashion is not contrary to reality, and so we do not require that agents formally choose between potential occupations.

We write the profit to an agent from financial intermediation at the *regional* level as,

$$\mathcal{F}^A = \#E^A(1-\psi)mM_t^A - \#C^A(1+\phi)sM_t^A, \tag{3.6}$$

where s is the private return that consuming agents obtain from selling finance to intermediaries.

Two costs are incurred by the intermediary. First, a cost ϕ of collecting finance from agents reflects the costs of communicating the worthiness of investment in terms of expected risk and return. We have seen from the historical analysis that this cost can be significant. Second, a cost ψ of distributing finance to entrepreneurs reflects the cost of evaluating and monitoring potential entrepreneurs. This follows on from the issues discussed in setting-up the models of Chapter 2. For the purposes of this analysis, we do not specify the sources of these costs analytically. We simply take the view, following both the critique of current methodology and the historical analysis above, that these costs exist and can be significant. We will take on these more fundamental issues later: Chapter 4 demonstrates that basing these costs in firm microeconomic foundations is not a straightforward matter. Our purpose here is to capture the broad implications of our critique of current finance and growth theory. As such, we first construct a model in which these costs are exogenous and think about their microeconomic roots later.

We take the view that the market for financial intermediation is also perfectly competitive: Given a large number of firms, agents and entrepreneurs, and given no fixed costs to becoming an intermediary, any profits from intermediation are competed away. From equation (3.6), with $\mathcal{F}^{\mathcal{A}} = 0$ under perfect competition, we have the following relationship between the rates of return on finance,

$$m = \frac{(1+\phi)}{(1-\psi)} s. \tag{3.7}$$

Equation (3.7) reflects the *wedge* between saving and investment: The more efficient the financial intermediation, the lower are the costs of collecting and disseminating finance, and the closer are the rates of return on saving and investment.

Substituting the demand function for infrastructure from equation (3.4) into the production function, and given $i = m$, we have,

$$Y_t^A = \left[A \left(\frac{(1-\alpha)(1-\psi)}{(1+\phi)s} \right)^{1-\alpha} \right]^{\frac{1}{\alpha}} k_t^A, \tag{3.8}$$

which is a simple form of Ak production which we know will generate endogenous growth. The relation to the parsimonious model of Chapter 2 is clear.

To close the model we specify conditions of consumer optimisation. Infinitely-lived consumers maximise their expected discounted income stream,

$$U = \int_0^\infty e^{-\rho t} u(c_t) dt, \tag{3.9}$$

where we define instantaneous utility as have constant elasticity of substitution,

$$u(c_t) = \frac{c_t^{1-\theta} - 1}{1-\theta}. \tag{3.10}$$

Agents maintain an idiosyncratic stock of finance that is controlled by, for example, a central bank. The central bank ensures that aggregate money supply is a constant proportion of aggregate output. The consumer chooses how much capital to sell to firms, how much finance to sell to intermediaries and how much to consume.

The Euler equation in consumption is obtained,

$$\frac{\dot{c}_t^A}{c_t^A} = \frac{1}{\theta}(r - \rho), \tag{3.11}$$

which is equal to the balanced growth rate of the economy, γ. From the production function, equation (3.8), we can derive r,

$$r = \left[A \left(\frac{(1-\alpha)(1-\psi)}{(1+\phi)s} \right)^{1-\alpha} \right]^{\frac{1}{\alpha}}. \tag{3.12}$$

Assume that a second market for finance exists that can give agents the same return as that on capital: Let us impose, equivalently, that agents are able to convert their money endowment into capital. If the return on finance is greater than the return on capital all finance will be sold to the intermediary. If the return to agents from selling finance to intermediaries is less than the return on capital, the finance could be sold directly to firms as capital. Competitive intermediation thus ensures $s = r$

and so, from equation (3.12),

$$s = r = A\left(\frac{(1-\alpha)(1-\psi)}{(1+\phi)}\right)^{1-\alpha}. \tag{3.13}$$

As such, from the Euler equation and this expression for the interest rate we have in both regions the growth rate,

$$\gamma^A = \gamma^B = \frac{1}{\theta}\left[A\left(\frac{(1-\alpha)(1-\psi)}{(1+\phi)}\right)^{1-\alpha} - \rho\right]. \tag{3.14}$$

National Growth

In the light of the historical evidence discussed above, it is reasonable to allow for the possibility that there are significant scale-effects in the costs of financial intermediation; historically we have seen an initial pattern of regional industrial takeoff in industry financed by local agents. Only once a local infrastructure is built and the regional economy becomes mature do financial intermediaries begin to operate on a country-wide basis. Additionally, information problems inhibit one region's ability to obtain finance from another, so there is a higher cost of coordinating investment on a national level relative to the regional level. So we might define the national financial intermediary conditions to be,

$$\mathcal{F}_t^* = (\#E^A + \#E^B)(1-\psi^*)i^* M_t - (\#C^A + \#C^B)(1+\phi^*)s^* M_t, \tag{3.15}$$

where $\psi^* = \Psi + \frac{2\omega}{I_t}$ and $\phi^* = \Phi + \frac{2\nu}{I_t}$, $\omega, \nu \geq 0$. The parameters Ψ and Φ reflect exogenous political and institutional factors. At early stages of development, or if fixed costs are always high, a financial intermediary incurs additional costs to operate at the national level, and to maintain zero-profit requires a higher return on finance sold to firms. The fixed cost premia, $\Psi - \psi$ and $\Phi - \phi$, reflect the difference in underlying efficiency of the regional cf. the national financial intermediary given the institutional environment. The non-fixed costs, parameterised by ω and ν, reflect the learning costs of establishing an infrastructure in order to operate an economy at the larger level. As regional markets grow, so a regional markets gradually mature. Proxying for the size of regional markets by the level of local infrastructure demand, $I_t^A = I_t^B = \frac{1}{2}I_t$, we effectively have that the cost raising finance to build a national infrastructure is decreasing in the size of regional markets. In the long-run, therefore, these costs become insignificant.

Given an extensive scale effect on productivity, we make the assumption that the the coefficient of technological progress at a national level is higher than that at the regional level. The national production function is thus,

$$Y_t^* = \bar{A}k_t^{\alpha} I_t^{1-\alpha}, \tag{3.16}$$

where $\bar{A} > A$, i.e. the incentive for agents to want to fund projects at a national level is the higher productivity of their capital and finance driven by the higher coefficient of technological progress, but this must be tempered by the cost of funding financial intermediation to facilitate that production. As in the case of the regions, we can find an analogous expression for the rate of interest to consumers on capital and finance in the case of integration,

$$r^* = \bar{A}\left(\frac{(1-\alpha)(1-\psi^*)}{(1+\phi^*)}\right)^{1-\alpha}. \tag{3.17}$$

It should be clear that r^* will not be constant so long as $\nu, \omega > 0$. We will still obtain a balanced growth path in the long run, but we approach it from below as $I_t \to \infty$. The long-run growth rate of the national economy is,

$$\gamma_{LR}^* = \frac{1}{\theta}\left[\bar{A}\left(\frac{(1-\alpha)(1-\Psi)}{(1+\Phi)}\right)^{1-\alpha} - \rho\right]. \tag{3.18}$$

The transition to this asymptotic growth rate follows,

$$\gamma_{SR}^{*} = \frac{1}{\theta}\left[\bar{A}\left(\frac{(1-\alpha)(1-\Psi-\frac{2\omega}{I_t})}{(1+\Phi+\frac{2\nu}{I_t})}\right)^{1-\alpha} - \rho\right]. \tag{3.19}$$

We will need to re-formulate this expression for any numerical simulation but, for the moment, the transitional growth dynamics should be clear: The rate of growth of consumption and infrastructure is inversely related to the *level* of infrastructure. The rate of change of economic growth is at first positive and reduces to zero as time goes to infinity: For an integrated national economy, $\lim_{t\to\infty}\gamma^{SR} = \gamma^{LR}$.

This growth rate will only be realised if the regional economies integrate. The rate of interest at the national level, equation (3.17), reflects the combination of increased productivity and increased cost of integrating the two regional economies. Integration thus takes place if $r^* > r$, where r is the rate of interest in the regional economies. This condition is satisfied where,

$$\frac{\bar{A}}{A} > \left(\frac{(1-\psi)(1+\Phi+\frac{2\nu}{I_t})}{(1-\Psi-\frac{2\omega}{I_t})(1+\phi)}\right)^{1-\alpha}. \tag{3.20}$$

Once this occurs, regional finance and capital supplies are combined and we have the national production function of equation (3.16) and no separate regional output, i.e. no agent would prefer to operate regionally when national output is possible. At the point where $r^* = r$ the growth rate at the national level is equal to that at the level of the regions. By equation (3.20), the feasibility of integrating is decreasing in both the relative additional costs of intermediating at a national level and the ratio of coefficients of technological progress.

The timing and transition to national integration here is, save exogenous institutional change, wholly endogenous to the model. This contrasts with work such as McDermott (2002) and Parente and Prescott (2005) where transition from one growth path to another is exogenously imposed.

Equilibria

There are three possible equilibria for the economy, dependent on both parameter values and the initial demand for infrastructure, I_0. We either have regional separation, national integration, or a transition from the former to the latter.

Given that over time the learning costs diminish into insignificance, the only thing that will prevent integration in the long-run are high fixed information costs relative to the productivity improvement, i.e. if,

$$\frac{\bar{A}}{A} \leq \left(\frac{(1-\psi)(1+\Phi)}{(1-\Psi)(1+\phi)} \right)^{1-\alpha}. \tag{3.21}$$

So it is possible that in the presence of either a low effect of integration on productivity ($\frac{\bar{A}}{A}$ is close to unity) or persistent high premia of pooling and coordinating savings over the larger economy (Φ and Ψ are significantly higher than ϕ and ψ) then we can be caught in a low growth trap. As such, there is, in this case, room for exogenous intervention to make integration feasible, i.e. we could mitigate information problems by for example legislating for contract rights. Government intervention to build a public infrastructure will have no effect on the feasibility of integration so long as equation (3.21) holds because they will not overcome information costs this way; this result thus falls nicely into the category of France vs. UK industrial growth with regard to the different attitudes to public infrastructure.

A second equilibrium will occur where initial infrastructure supply, I_0 is such that we begin with an integrated economy in the first instance, if,

$$\frac{\bar{A}}{A} > \left(\frac{(1-\psi)(1+\Phi+\frac{2\nu}{I_0})}{(1-\Psi-\frac{2\omega}{I_0})(1+\phi)} \right)^{1-\alpha}. \tag{3.22}$$

In this case, either a high productivity increase from integration or very low fixed information cost can mean that a low initial infrastructure supply and low learning cost effects (low ν and ω) could create conditions such that the economy is always integrated.

The most interesting case in terms of endogenous growth is the intermediate one, where the economy begins in its disintegrated form and endogenously integrates when conditions become right. This requires,

$$\left(\frac{(1-\psi)(1+\Phi)}{(1-\Psi)(1+\phi)}\right)^{1-\alpha} < \frac{\bar{A}}{A} \leq \left(\frac{(1-\psi)(1+\Phi+\frac{2\nu}{I_0})}{(1-\Psi-\frac{2\omega}{I_0})(1+\phi)}\right)^{1-\alpha}. \tag{3.23}$$

Of the three equilibria, this case perhaps comes closest to reflecting the actual pattern of industrial growth. In time zero, scale costs mean that it is optimal for financial intermediaries operate on a small scale, using local finance to fund the construction of a regional infrastructure. In this initial phase, growth is low. Over time, regional markets grow and a local infrastructure is constructed to support local output. This also lessens the cost of raising finance to build infrastructure and integrate at a national level. At a critical value of local market size we have national integration and a smooth takeoff in growth, approaching γ^* over time as the economy matures. We thus have an acceleration in industrial output growth as determined by endogenously improving conditions for financial intermediation. In this case there is room for exogenous action bring forward the takeoff point. The critical value of infrastructure, at which we integrate, is the positive root of,

$$(1-\Psi)(1+\Phi)I_t^2+\left[\left(\frac{\bar{A}}{A}\right)^{\frac{1}{1-\alpha}}\frac{(1+\phi)}{(1-\psi)}2\omega(1-\Psi)-(1+\Phi)2\omega-(1-\Psi)2\nu\right]I_t-4\omega\nu=0. \tag{3.24}$$

So there is a potential role for accelerating development by reducing the costs of information problems, as in the disintegrated equilibrium, but also here we can bring forward the point at which we integrate via the public funding of infrastructure technology and improving awareness. We can, therefore, draw comparisons here to the 'big-push' literature of the type espoused by Murphy et al. (1989); while on a collective basis agents could gain by integrating at the same time, individually they face financial intermediary costs related to the size of the entire economy, which they cannot significantly influence by their individual action.

3.3.3 Numerical Solutions

Consumption, money, capital and infrastructure all grow, in continuous time, at the rate $\gamma = \max\{\gamma^A, \gamma^*_{SR}\}$. For the purposes of a numerical extension we need to consider the growth rate of the economy in a discrete-time form, so $\gamma_h = (x_t - x_{t-h})/hx_{t-h}$ for all growth variables x in the economy where h is the length of each discrete time increment. So if we want to think of an annual growth rate while taking quarterly increments then we let $h = \frac{1}{4}$. In the limit as $h \to 0$ we have that $\gamma_h \to \gamma$. In the regional economy, and in the long-run integrated economy, the growth rate is constant. The transitional growth rate, equation (3.19), is dependent on the stock of infrastructure at time t, however. We can re-write the transitional growth rate as,

$$\frac{I_t - I_{t-h}}{hI_{t-h}} = \frac{1}{\theta}\left[\bar{A}\left(\frac{(1-\alpha)(1-\Psi-\frac{2\omega}{I_t})}{(1+\Phi+\frac{2\nu}{I_t})}\right)^{1-\alpha} - \rho\right]. \tag{3.25}$$

We can solve for I_t in terms of I_{t-h} and obtain a solution for the growth rate of the economy that can be solved numerically. (MATLAB code is reproduced in Appendix Section A.1.) Let $I_t = \mathcal{I}(I_{t-h})$ be the solution to,

$$\left(1+\Phi+\frac{2\nu}{I_t}\right)\left\{\bar{A}^{-1}\left[\rho+\theta\left(\frac{I_t - I_{t-h}}{hI_{t-h}}\right)\right]\right\}^{\frac{1}{1-\alpha}} = (1-\alpha)\left(1-\Psi-\frac{2\omega}{I_t}\right). \tag{3.26}$$

Of course, we need to check first that there is only one finite and real solution to equation (3.26).[24] Then we can write equation (3.19) as,

$$\gamma^*_{hSR} = \frac{1}{\theta}\left[\bar{A}\left(\frac{(1-\alpha)(1-\Psi-\frac{2\omega}{\mathcal{I}(I_{t-h})})}{(1+\Phi+\frac{2\nu}{\mathcal{I}(I_{t-h})})}\right)^{1-\alpha} - \rho\right]. \tag{3.27}$$

Table 3.1 gives a benchmark calibration with which we can demonstrate some of the growth dynamics. Figure 3.6 plots the course of the economy with an initial money stock of $M_0 = 20$ and initial capital stock of $k_0 = 40$. This economy is one which begins regionally separated and integrates as endogenous financial intermediation costs drop over time. As can be seen, the rate of growth of the regional economy is constant at around $\gamma^A = 0.054$. That of the integrated economy begins low and climbs to a long-run rate of around $\gamma^* = 0.064$. The thick black line indicates the equilibrium growth rate at any point in time, with national integration occurring at $t = 23$ and a smooth transition towards the long-run growth rate of the integrated economy.

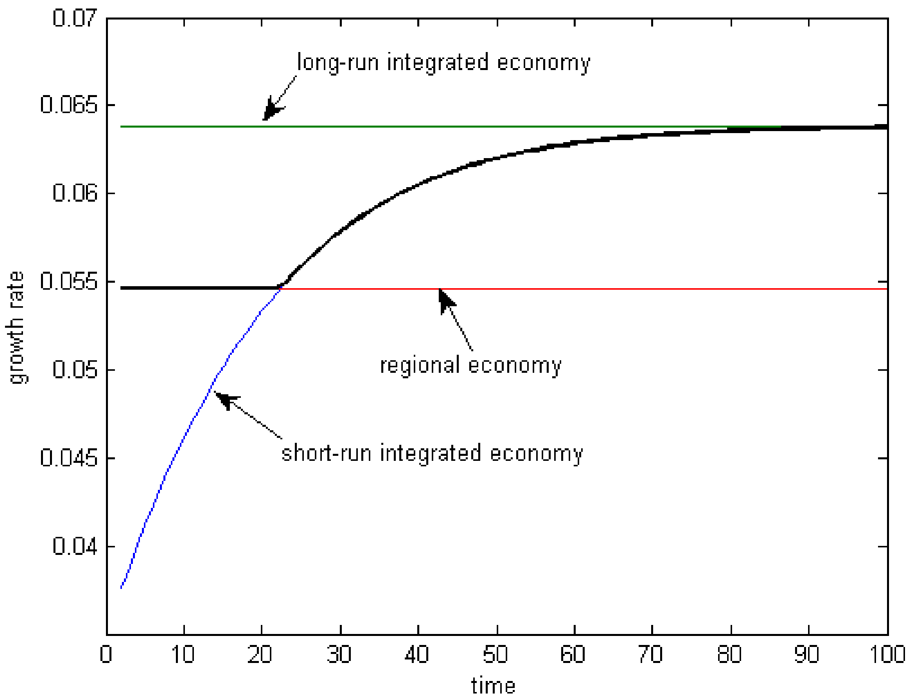
Table 3.1: Calibration for Finance, Integration and Growth

initial capital	k_0	40
initial finance	M_0	20
capital share	α	2/3
subjective discount rate	ρ	0.02
elasticity of substitution	θ	5
regional coefficient of technological progress	A	0.5
national coefficient of technological progress	$\bar{A}$	0.6
fixed cost parameter on regional intermediation	ψ	0.25
fixed cost parameter on regional intermediation	ϕ	0.25
fixed cost parameter on national intermediation	Ψ	0.3
fixed cost parameter on national intermediation	Φ	0.3
scale cost parameter on national intermediation	ν	5
scale cost parameter on national intermediation	ω	5

We can embark on a calibration of this model to the data of Bairoch (1982).

[24]This is generally the case. The exact solution method is straightforward, so we leave it to Appendix Section A.2.

Figure 3.6: Example Growth Path



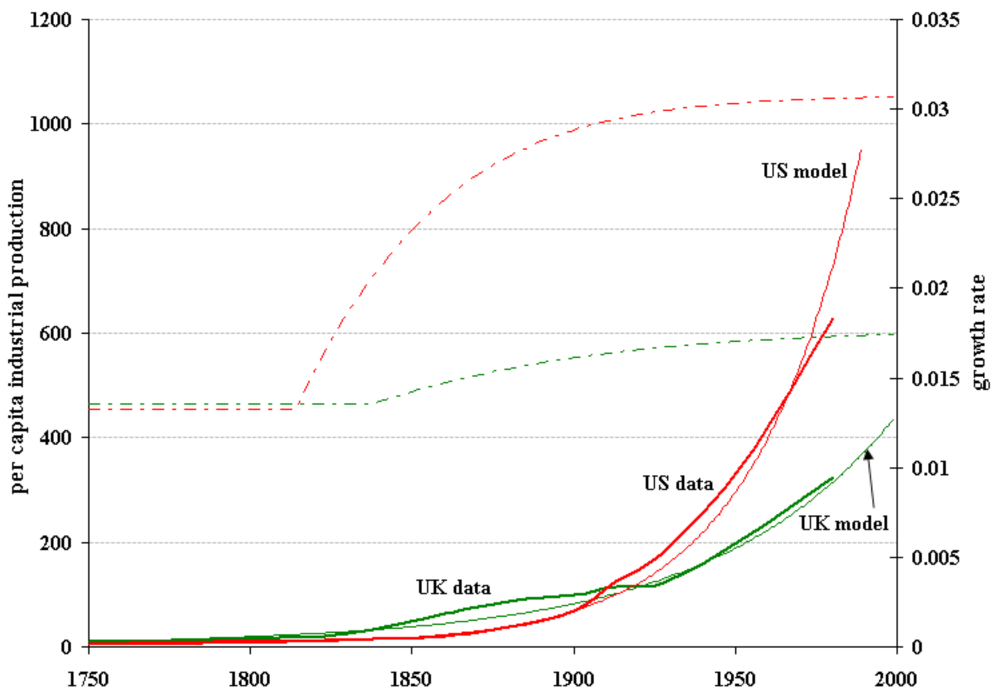
There are too many free parameters to be able to pin-down a full calibration based on the evolution of financial efficiency, national integration and technological progress. We can, however, more generally consider whether the sorts of growth path implied by our model can capture historical trends. Table 3.2 gives an indicative calibration to US and UK data. Figure 3.7 depicts the numerical results. There are three things that we wish to capture from the data: First, the different levels of income in 1750; second, the different rates of growth through the period of industrial revolution; and, third, the different levels of income at the end of the period. We do not make any mid-period exogenous changes to the financial intermediation conditions; all growth effects here are endogenous.

Of course, this is one of many stories that we can tell; a different parameterisation might match the data equally well. There are two main differences that cause the numerical growth path: First, the US has a higher level of productivity

Table 3.2: Calibration for UK and US Growth Paths

		UK	US
initial capital	k_0	10	4
initial finance	M_0	5	2
capital share	α	2/3	2/3
subjective discount rate	ρ	0.02	0.02
elasticity of substitution	θ	5	5
regional coefficient of technological progress	A	0.15	0.15
national coefficient of technological progress	$\bar{A}$	0.2	0.32
fixed cost parameter on regional intermediation	ψ	0.25	0.275
fixed cost parameter on regional intermediation	ϕ	0.25	0.275
fixed cost parameter on national intermediation	Ψ	0.35	0.35
fixed cost parameter on national intermediation	Φ	0.35	0.35
scale cost parameter on national intermediation	ν	2	1.25
scale cost parameter on national intermediation	ω	2	1.25

Figure 3.7: Calibrated US and UK Growth Paths



at the national level, $\bar{A}$; second, the learning element of national intermediation is, in the US, substantially more rapid, i.e., the ν and ω are lower. This means that, despite beginning with a lower initial capital endowment, the US economy catches up and overtakes the UK one. In the US, integration occurs earlier and proceeds to a higher long-run growth path at a faster rate. These distinctions appear to match the historical growth patterns relatively well.

Of course, this interpretation is based on a very loose calibration. If we had data on, say, the endogenous or exogenous evolution of financial efficiency throughout the industrial revolution then we could make substantially firmer conclusions. The emerging time-series cliometric analyses of industrial takeoff are likely to help us in this regard. Tieing these data with an analysis of the evolution of disaggregated coalitions would further restrict the range of possible calibrations. Moving towards the completion of rich datasets such as Shea (2007) is likely to prove invaluable in helping us to do this.

3.4 Concluding Remarks

We have demonstrated that by conceiving of finance and growth in a static and aggregative way we lose an understanding of the underlying transmission mechanisms. As a consequence, we are poorly armed to consider any policy implications. By looking at evidence from history, we have been able to identify a greater complexity in the transmission between entrepreneurial finance and industrial growth. Given an appreciation of the importance of fixed costs in the finance of any investment project, we identified the financial structures which supported the development of physical transport infrastructure.

Further to the historical analysis, we developed a theory of finance and growth that can capture some of the interrelations between scale, finance and infrastructural development. The quantitative aspects of our theory can broadly replicate some observed historical growth paths. The theory laid-out above has a number of shortcomings, however. In terms of matching the stylised facts of aggregate finance and growth, we do not observe an increase in financial depth over time. Moreover, the upward trend in the interest rate on capital does not reflect the historical record. There is obviously room for improvement in the theory. A number of the problems cannot be addressed in such a top-down model, and we shall see in the next Chapter that generating such disaggregated macroeconomic phenomena from microeconomic foundations is not straightforward. The model presented here, we hope, at least at least places stress on those aspects of the finance and growth mechanism that we believe are most important. The emergence of the literature on the industrial revolution as a regional phenomenon has the potential to greatly benefit theories such as that developed here, and leads us to the first potential extension to the theory we developed above. If we had data for regional and provincial estimates of economic growth, we could compare our eventual estimates from the *Handbook* on the emergence and gradual centralisation of financial coalitions. Our ability to account for industrial takeoff in a model of endogenous growth that can be calibrated to disaggregated data would promise an interesting avenue for future research. We could shed light on a number of important questions.

A second advance in the theory would be to make optimal government behaviour endogenous to the model. We can outline in broad terms how this might be done: In the early period of industrialisation, when scale costs in financial intermediation are very large, a government might be tempted to promote the accumulation of infrastructure by allocating resources to the amelioration of economy-wide financial constraints. Of course, there is a trade-off between the marginal positive impact of spending on mitigating information and contracting problems and the marginal negative impact of taxation on welfare. On the basis of this simple model, when coalitions are spatially concentrated the marginal impact of government spending in promoting centralised financial markets is minimal. The conclusion, then, is that in the early stages of growth, a government ought to behave with a relatively light touch, allowing private enterprise to form efficient financial structures that themselves mitigate information and contracting problems. The role of the government in the early phase of industrialisation seems to be the support of private enterprise by forming institutions that make it easier to write enforceable contracts. This has come out of both our historical analysis and the theoretical model we constructed to match our stylised facts.

We turn to a more formal analysis of these propositions in the next Chapter.

Chapter 4

Endogenous Exchange Costs in General Equilibrium

We have seen in Chapters 1–3 that the existence of financial market imperfections are central to the connection between finance and economic growth. The nature of these imperfections gives rise to spatial and dynamic elements in the existence of financial coalitions. In this chapter, we capture these imperfections at the microeconomic level by considering the endogenous determination of exchange costs. In Chapter 5 we place the findings of this analysis into the context of finance and growth, and identify a large and exciting agenda for establishing the microfoundations of finance and growth.

Exchange costs are central to a great deal of economic theory. Typically, however, they are an exogenously fixed component of a wider model. It is our view that such costs are more usefully thought of as the endogenous outcomes of individual behaviour. Indeed, as noted by Coase (1992, p. 716) in his Nobel lecture,

> a large part of what we think of as economic activity is designed to accomplish what high transaction costs would otherwise prevent.

A wide range of literatures have exchange costs at their heart. These are generally intended to reflect frictions caused at the microeconomic level: Think of the exchange costs invoked by, among others, Greenwood and Jovanovic (1990), to support the existence of financial structures; the trade friction models of Obstfeld and Rogoff (2000); the core-periphery work on globalisation by Krugman and Venables (1995); and the new institutional economics, surveyed in Williamson (2000), wholly founded on the existence of transactions costs. In addition, the work of Dixit (1996) on political economics is based on a transaction-cost politics perspective. Each of these contributions has demonstrated that exchange costs have a huge impact on theoretical outcomes. But in each of these contributions, those exchange costs are not explicitly related to the technologies of exchange.

Making Exchange Costs Endogenous

Why should we bother to endogenise exchange costs? Take, for example, the theoretical works on finance and development of Greenwood and Jovanovic (1990) and Townsend and Ueda (2006). The answer is clear: If, in the process of development, the conditions that give rise to markets change then the conditions that give rise to exchange costs also change. It may be that agents in economies of different levels of wealth allocate different portions of their endowments to the amelioration of exchange costs. Central to the nature of economic development, then, is a decision over amelioration which in turn determines exchange costs and, as a consequence, optimal diversification and market size. Considering development when the exchange cost is exogenously fixed detracts from the potential richness of any theory of finance and growth. Boyd and Prescott (1985) also note this in the context of models of financial coalitions: "[T]ransaction costs are assumed to exist and are not explicitly related to exchange technologies nor differentiated between types of trade," (p.212).

Agents in our economy wish to diversify against a stochastic technology shock.

The key decision that agents make regards the formulation of ex ante arrangements to ameliorate ex post exchange costs. If the ex ante arrangement is not perfectly "*effective*", then ex post exchange is, we impose, costly. The less effective the ex ante arrangement, the more costly is ex post exchange. In general, we see both private and public arrangements to ameliorate exchange costs. Therefore, in our model we distinguish between two forms of ex ante arrangements. Efforts such as organizational choices, private infrastructure or learning about property rights – activities specific to the transaction being made – are local and excludable. We label these activities as "*specific amelioration*". The legal enforcement of contracts, fiduciary duties, public infrastructures or competition policy also act to reduce exchange costs but are, by contrast, public goods; this we call "*general amelioration*". Each type of arrangement benefits the other: Without a public institution to support them, private arrangements are more costly; without private arrangements, there is little an institution can do.

It is the coordination of these ameliorating arrangements which can give rise to a richer framework of exchange costs. We are led to ask: How can institutional arrangements which ameliorate the costs of private exchange be supported by a decentralised economy populated by self-interested agents?

The result of this general-specific distinction in large economies is a coordination failure at the economy-wide level in the provision of general amelioration (i.e., in the existence of institutions). Voluntary contributions to the public good under perfect information are zero, and only specific amelioration emerges. In the presence of 'compulsion', a combination of specific and general amelioration arrangements is optimal. In an heterogeneous agent economy, however, we face political economy questions regarding power and influence of certain groups in the provision of such public goods. Within this simple general equilibrium framework, therefore, ineffective ex ante arrangements can account for the coalitional structures of exchange, the existence and optimality of institutions and an optimal taxation policy in which a public institution emerges to facilitate private exchange.

The Sources of Costly Exchange

North (1990, p.27) puts transactions costs down to "the costs of measuring the valuable attributes of what is being exchanged and the costs of protecting rights and policing and enforcing arrangements." Why do these costs exist ex post? And is there scope for agents to ameliorate them by allocating resources to some ex ante technology? The costs listed by North must, to some degree, be the result of an, at least partial, incompleteness of the ex ante contract. That is, contracts written ahead of the state of the world being fully known do not detail specific actions in every possible state of nature; there are, ex post, costs of measuring attributes and of enforcing, protecting and policing ex ante arrangements. But the exchange cost is not the cost of writing the partially complete contract, it must be its complement: Contracts do not simply emerge to cope with pre-existing exchange costs; the two concepts are irretrievably connected, in a sense two sides of the same coin.

This relates to what is, we believe, a general misconception of exchange costs as an exogenous fundamental to the general equilibrium framework. Townsend (1983b, p.259) is representative of the standard view of exchange costs as the "costs of bookkeeping, the cost of enforcement, the cost of monitoring…". The costs of exchange are, in this view, the costs of forming and applying effective contracts through both private behaviour and public institutions. But, taken literally, this implies that the greater the resources allocated to contracts, the greater are the exchange costs. Why then do agents carry out bookkeeping, etc., at all? There is not some fixed amount of bookkeeping, enforcement and monitoring that facilitates any given transaction. The exchange cost is the cost of *not* allocating resources to bookkeeping, etc. We prefer to think of agents choosing the extent to which such contractual procedures are carried out. Given preferences, technologies and endowments, then, exchange costs are an endogenous result of the marginal costs and benefits of carrying out the formation, writing and enforcement of complex contracts.

This derives from a trade-off underpinning the determination of exchange costs. A 'cost' of bookkeeping, etc., implies that forming contracts is a costly process. Further, agents can choose to not form a comprehensively complete contract if the costs are large relative to the gains from doing so. In making exchange costs endogenous, we are able to rationalise, in some senses, the existence of the *optimally [in]complete contracts subject to constraints* noted in Hart and Moore (1999). Boyd and Prescott (1985) stress the importance of contracting to the formation of intermediary-coalitions (under conditions of asymmetric information). In their interpretation, an intermediary can be thought of as announcing a complex system of group rules, to which potential coalition members can choose to attach themselves. Despite their complexity, however, these contracts are not costly to form in the Boyd-Prescott economy. And yet, following on from the remarks by Coase noted at the beginning of this Chapter, it seems desirable to have these costs reflected in the allocations that are feasible. Indeed, by making exchange costs endogenous within a numerically tractable model we are able to capture, in a quantitative way, the argument that resources directed at reducing transactions costs can comprise a large portion of economy activity.

Of course, there is more to exchange costs than incomplete contracts. In addition, the existence of physical barriers to exchange (such as inadequate transportation networks), human capital shortcomings (too few trained lawyers, for example) and problems in communication (such as a language barrier) can add to costs of physical exchange. None of these can necessarily be considered to constitute an incomplete ex ante contract, but they are problems which can be ameliorated with the allocation of resources. Furthermore, we are not offering here a microeconomic model linking contractual incompleteness with the existence of costly exchange. As such, we leave a fuller discussion of this incomplete contract interpretation to Section 4.7. For the moment, we concentrate on the central feature of our model: An agent not only chooses the market into which she will diversify but also the resources to be allocated to the amelioration of ex post exchange costs and, as such, the effectiveness of the ex ante arrangement.

Outline of this Chapter

The rest of this Chapter is set out as follows. In Section 4.1 we set out the basic structure of our economy. In Section 4.2 we examine cooperative equilibria, of a particular sort, and examine feasibility. We turn to the issue of decentralizing our cooperative solution in Section 4.3. In Sections 4.4 and 4.5 we begin to address some political economy issues. First, we discuss the feasibility of economy-wide institutions emerging in our simple set-up, and then in Section 4.5 we turn to the issue of heterogeneity. Section 4.6 discusses some detailed numerical simulations of our model that shed light on some important issues that we refer to. Section 4.7 discusses the model in the light of the incomplete contract literature and concludes with some implications for future work.

4.1 The Economy

Our basic set-up is inspired by Townsend (1978). A virtue of that work over a number of other transaction cost general equilibrium models is that it allows us to consider, in a tractable fashion, a non-convex technology of exchange that gives rise to specialisation. In our economy, there is a countable infinity of agents, $i \in I$. Each agent i receives an endowment, $e^i = \{k^i, \lambda^i\}$. Capital $k^i \in \Re_+$ can be used in conjunction with a stochastic linear production technology, $\lambda^i \in \Lambda \subseteq \Re_+$, which is i.i.d. across agents. Agent i produces consumption good, $\lambda^i y^i$, where y^i is the amount of capital used in production. We assume that all agents receive the same k.[25] Agents can also use their endowment to buy shares in a consumption bundle of other agents' output. All agents have the same (CRRA) utility function, $u^i(c)$, so we typically drop the superscript i. The probability of receiving some $\lambda \in \Lambda$ is $p(\lambda) \in (0,1)$, $\sum_{\lambda \in \Lambda} p(\lambda) = 1$. The average technology in the economy is then $\bar{\lambda} = \sum_{\lambda \in \Lambda} p(\lambda)\lambda$. Agents know $p(\lambda)$ for all $\lambda \in \Lambda$, so they also observe $\bar{\lambda}$. Let ω represent the state of nature, i.e., a list of λ^i for all $i \in I$ where the i^{th} element of ω, denoted ω^i, is λ^i. Let Ω be the set of all possible states of nature, and $p(\omega) = p(\omega^1) \times p(\omega^2) \times \cdots \times p(\omega^I)$ the probability of some $\omega \in \Omega$. An agent's expected utility is $V = \sum_{\omega \in \Omega} p(\omega) u(c(\omega))$.

Agents are risk-averse and so they have an incentive *ex ante* to invest in a diversified consumption bundle, a portfolio, if one is available. They can do so by forming markets in which agents meet to share risk before the realisation of ω. The diversification between agents is naturally subject to the formation of ex ante arrangements. In a world in which ex ante arrangements are what we term 'perfectly effective', the ex post exchange is free; there are no costs to the exchange and diversification takes place across all agents. Agents have formed securities that allow them to diversify against all idiosyncratic risk. If, however, ex ante arrangements are only partially effective there are expected to be some associated losses, which we think of as a simple 'ex post exchange cost'. In the event that ex ante arrangements do not specify actions for all $\omega \in \Omega$, *ex post* costs to exchange are greater than zero.

In the manner of Hart and Moore (2006), there can exist ex ante perfect com-

[25] An alternative procedure suggested to us was to dispense with the technology endowment and have stochastic endowments. Though technically possible, agents would have no ex ante resources with which to form costly arrangements. Ex post, then, there would be no sharing of endowments; risk sharing in our model is a costly process and requires that agents have some resources ex ante.

petition and ex post bilateral monopoly in exchange arrangements; so agents who form an ex ante arrangement under competitive conditions can be said to be *locked-in* to that arrangement ex post.[26] So agents choose to form an ex ante agreement that ties themselves to both ex post exchange *and* the cost of that exchange. The degree of effectiveness of an arrangement between agents is indexed on $\Pi = [0, 1]$. We let $\pi^{ij} \in \Pi$ be the effectiveness of some arrangement between two agents $i \in I$ and $j \in I$ with $j \neq i$. Of course, $\pi^{ij} = 1$ if $i = j$. Since the null arrangement is $\Pi^{null} = \{0\}$, partially effective and perfectly effective arrangements are denoted by $\pi \in \Pi \backslash \Pi^{null}$.

$\tau_s^i \in [0, 1]$ and $\tau_g^i \in [0, 1]$ are the proportions of the endowment agent i contributes to specific and general amelioration respectively, where $\tau_s^i + \tau_g^i \leq 1$. Let $\tau^i = \left\{\tau_s^i, \tau_g^i\right\}$.

π^{ij} is determined by the following mapping,

$$\pi^{ij} = F(S^{ij}, G),$$

where,

$$\begin{aligned} S^{ij} &\leq \mathcal{S}^{ij} = \frac{1}{2}\left(\tau_s^i + \tau_s^j\right) k; \\ G &\leq \mathcal{G} = \frac{1}{\# I} \sum_{i \in I} \tau_g^i k. \end{aligned}$$

G reflects the general ameliorating capital in the economy, that is public ameliorating arrangements, and $\mathcal{G}$ total ameliorating capital raised. S^{ij} reflects specific, excludable ameliorating capital between agents i and j. (We have let $S^{ij} \leq \mathcal{S}^{ij}$ and $G \leq \mathcal{G}$ to allow for the possibility of rents to intermediation or the provision of institutions, respectively.) $F(.,.)$ is continuous, increasing, strictly concave and

[26] And, indeed, in a general equilibrium sense it might be optimal for agents to choose to tie themselves down in this way, if there is some penalty against deviation.

separable in its arguments, mapping from ameliorating capital into arrangement effectiveness. We assume further Inada-type conditions. First, the limits of the function are defined as $F(0,0) = 0$ and $F(S^{ij}, G) \to 1$ as $\{S^{ij}, G\} \to \{\infty, \infty\}$. Let us define $F^n(S^{ij}, G)$ as the first derivative of the function F with respect to its n^{th} argument, x_n. So, second, first derivatives with respect to each variable approach $F^n(S^{ij}, G) \to \infty$ as $x_n \to 0$, and $F^n(S^{ij}, G) \to 0$ as $x_n \to \{\infty\}$, for $n \in \{1, 2\}$. If agents do not allocate resources to amelioration, the null arrangement obtains. If the F function varies across agents, then issues of power and political influence will come into sharper relief. For now, we assume homogeneity across agents in this respect. We outline some of these further issues in Sections 4.4 and 4.5.

We let α^{ij} be the cost of a single exchange between agent i and j,

$$\alpha^{ij} = \left(1 - \pi^{ij}\right) k. \tag{4.1}$$

So, the more effective the arrangement, the lower the exchange cost. Further, the exchange cost is zero when arrangements are perfectly effective ($\pi^{ij} = 1$), and exchange costs are equal to k under the null arrangement, ($\pi^{ij} = 0$). Therefore, if agents devote no resources to forming an arrangement, they cannot gain from diversification unless amelioration is costless.

4.1.1 Intermediation

A coalition is a set of agents $C \subseteq I$ with cardinality $\#C$. The set of agents with whom agent $i \in I$ exchanges directly is denoted N^i. A coalition C constitutes a market M if $\#C < \aleph_0 = \#I$ and for all $i \in C$, $N^i \subseteq C$ and where there is no coalition $B \subset C$ for which, for all $i \in B$, $N^i \subseteq B$. We can think about the size of the market M in terms of its cardinality, $\#M$. Where $\#M < \aleph_0$ we call M *finite-sized*. The set of all markets in the economy is $\mathcal{A}$, so $\bigcup_{M \in \mathcal{A}} M = I$.

Agent h is said to act as an intermediary if $N^i = h$ for $i \in M \backslash h$ and $N^h = M \backslash h$. So markets are disjoint; agents only exchange with one intermediary. Where H is the set of all intermediaries in the economy, let M^h be the market intermediated by agent $h \in H \subseteq I$. Agents exchange their endowment of capital on a one-for-one basis for shares in the portfolio put together by the intermediary, subject to their contribution to amelioration and their being able to meet their share of the ex post exchange cost.

We think of an intermediary as forming a single multilateral arrangement of effectiveness level $\pi^h \in \Pi$ that defines the exchange costs for transactions between each market participant and the intermediary, $\alpha^h = \left(1 - \pi^h\right) k$. In an homogeneous set-up, this generalisation loses us nothing. In some ways, we may think of an intermediary *specializing* by forming a multilateral arrangement. We denote specific ameliorating capital in a market M intermediated by agent h as S^h. For example, if,

$$S^h = g\left(\#M^h\right) \sum_{i \in M^h} \tau_s^i k,$$

then the function $g(\#M^h)$ determines the effect of increasing the number of agents in the arrangement on the aggregation of ameliorating capital. We could, for example, let $g\left(\#M\right) = \left(\#M\right)^{-1}$; this would correspond to taking ameliorating capital as the average contribution from agents in a market. This is what we (must) do in the general capital formulation, otherwise economy-wide amelioration would be priced at either infinity or zero, so we typically allow g to take this form for symmetry.[27] It follows that exchange costs for an agent in a market intermediated by agent h is $\alpha^h = \left(1 - \pi^h\right) k$ where $\pi^h = F(S^h, G)$ is the degree of effectiveness of the multilateral arrangement. To diversify in a market of size $\#M$ intermediated

[27]Perhaps, this is a natural modelling choice: A multilateral arrangement with some level of ameliorating capital per person has the same level of effectiveness as a bilateral arrangement with the same level of ameliorating capital per person. Otherwise, we could have let $g\left(\#M\right) = \left(\#M\right)^{-2}$; this would make larger arrangements (ones covering more agents) progressively harder to write. Letting $g\left(\#M\right) = 1$ would make larger arrangements progressively easier to write.

by agent h, each agent thus pays $2\alpha^h \frac{(\#M-1)}{\#M}$ in per capita bilateral exchange costs.

4.2 Cooperative equilibrium

Definition 4.1 *An allocation* $x^h = \{c^i, y^i, \tau^i; i \in M^h\}$ *is feasible for market* M^h *intermediated by agent h if, for all* $\omega \in \Omega$,

$$\sum_{i \in M^h} \left(1 - \tau_s^i - \tau_g^i\right) k - y^i \geq 2\alpha^h \left(\#M^h - 1\right); \tag{4.2}$$

$$\sum_{i \in M^h} \lambda^i(\omega) y^i \geq \sum_{i \in M^h} c^i(\omega); \tag{4.3}$$

$$G \leq \mathcal{G} = \frac{1}{\#I} \sum_{i \in I} \tau_g^i k; \tag{4.4}$$

$$S^h \leq \mathcal{S}^h = \frac{1}{\#M^h} \sum_{i \in M^h} \tau_s^i k; \tag{4.5}$$

$$\alpha^h = (1 - \pi^h) k; \tag{4.6}$$

$$\pi^h = F(S^h, G). \tag{4.7}$$

The *composition*, $\mathcal{C}(M^h)$, of a market intermediated by h is denoted by the n-tuple: $\mathcal{C}(M^h) = \{\#M^h; \{\tau^i\}_{i \in M^h}; S^{M^h}; \alpha^h; \pi^h\}$.

Before looking at a decentralised case, we first examine a cooperative version of the model where we designate agents at random to set up a market. We characterise the equilibrium of the model and then show, under certain conditions, that it is in the core.

The intermediary's problem can be characterised as follows in a way that ensures feasibility of the resulting allocation:

$$\max_{\mathcal{C}(M^h)} EU \left[\left\{ \left(1 - \tau_s^i - \tau_g^i\right) k - \left(\frac{\#M^h - 1}{\#M^h} \right) 2\alpha^h \right\} \left(\#M^h\right)^{-1} \sum_{i \in M^h} \lambda^i \right] |_{\{\tau^i\}_{i \in I}}. \tag{4.8}$$

The expected utility of an agent $i \in M^h$ is contingent on the behaviour of agents $i \in \mathcal{A} \backslash M^h$ because of the public-good nature of the general amelioration. If the market to which agent i belongs is finite-sized, its contributions to the institution will not cause changes in general amelioration unless agents in other markets contribute also. We assume that some commitment technology exists such that intermediaries across markets can cooperate. We deal with decentralization later.

Equation (4.8) may be written as,

$$E\left[u\left(c^i\right)\right] = \sum_{\omega\in\Omega} p\left(\omega\right) u \left\{ \left[\left(1 - \tau_s - \tau_g\right) k - \frac{2\alpha^h \left(\#M^h - 1\right)}{\#M^h} \right] \bar{\lambda}\left(\omega, M^h\right) \right\},$$

$$= \sum_{\omega\in\Omega} p\left(\omega\right) u \left\{ \left[\left(1 - \tau_s - \tau_g\right) k - 2\left[1 - F\left(S^h, G\right)\right] k \frac{\left(\#M^h - 1\right)}{\#M^h} \right] \bar{\lambda}\left(\omega, M^h\right) \right\}, \quad (4.9)$$

where $\bar{\lambda}\left(\omega, M^h\right)$ is the average technology in market M^h given the state of nature ω. Optimality in each market M^h with $\#M^h \geq 1$ requires,

$$2\left(\frac{\#M^h - 1}{\#M^h}\right) k \frac{\partial F\left(S^h, G\right)}{\partial S^h} = 1; \quad (4.10)$$

$$2\left(\frac{\#M^h - 1}{\#M^h}\right) k \frac{\partial F\left(S^h, G\right)}{\partial G} = 1. \quad (4.11)$$

The number of technologies, so long as there are two or more, does not matter; the optimal arrangement for a given market size will always have to satisfy these conditions. **Propositions 4.1–4.3** will help us characterise the core of this economy. Before that, **Lemma 4.1** will help us to establish the uniqueness of optimal outcomes.

Lemma 4.1 *The solution to equations (4.10) and (4.11) for a given market size* $\#M > 1$ *and endowment* k *is a unique pair of contributions,* $\tau^{\#M} = \left\{\tau_s^{\#M}, \tau_g^{\#M}\right\}$ *where* $\tau_s^{\#M} > 0$ *and* $\tau_g^{\#M} > 0$ *for* $\#M > 1$*. For* $\#M = 1$ *we have* $\tau_s^{\#M} = 0$ *and*

$\tau_g^{\#M} = 0$.

Proof. This is a proof by contradiction. Fix $\#M > 1$. Suppose for a given k there is a unique optimal specific contribution τ_s' but two optimal general contributions, τ_g' and $\tau_g'' \neq \tau_g'$. By the strict concavity of F in G, it cannot be the case that (4.10) and (4.11) are satisfied for both $\left\{\tau_s', \tau_g'\right\}$ and $\left\{\tau_s', \tau_g''\right\}$. Likewise for a unique τ_g' and two specific contributions τ_s' and $\tau_s'' \neq \tau_s'$. There is thus some unique pair of optimal contributions for any given market size $\tau^{\#M} = \left\{\tau_s^{\#M}, \tau_g^{\#M}\right\}$. From the properties of $F(.)$ we know that these unique solutions to equations (4.10) and (4.11) require $\tau_s^{\#M} > 0$ and $\tau_g^{\#M} > 0$. For $\#M = 1$, we require by equations (4.10) and (4.11) that $F^x = \infty$ for $x = \{s, g\}$, which happens only when $\tau_s^{\#M} = 0$ and $\tau_g^{\#M} = 0$. ∎

So, if an agent trades with at least one other agent, some amount of specific and general amelioration is jointly optimal. In the current set-up, all agents in all markets contribute the same. The relative efficacy of general and specific amelioration determines the equilibrium balance between each form of arrangement.

We wish to consider the behaviour of the economy as the level of wealth changes. Three propositions can be formalised.

Proposition 4.1 *For a given endowment, optimal contributions rise, ex ante arrangements become more effective and exchange costs fall as the market size increases.*

Proof. Since $2\left(\frac{\#M-1}{\#M}\right)$ is increasing in $\#M$, F_s' and F_g' must fall to satisfy optimality. So for a given k, $\tau_g^{\#M}$ and $\tau_s^{\#M}$ rise in $\#M$. As $\pi^h = F(S^h, G)$, the optimal degree of arrangement effectiveness increases in market size. Since $\alpha^h = \left(1 - \pi^h\right) k$, exchange costs decrease as the market size increases. ∎

So we know that as agents further diversify, ceteris paribus, ex ante ameliorating costs rise, arrangements become more effective and exchange costs fall. To consider the behaviour of the economy as the level of the endowment increases, we first consider the optimal choices over amelioration for a given market size.

Proposition 4.2 *For a given market size greater than one, optimal contributions can rise, ex ante arrangements become more effective and the exchange cost either rises or falls with the endowment.*

Proof. A corollary of **Lemma 4.1** is that for a given market size greater than one and given the level of the endowment, the optimizing choice of $\left\{S^h, G\right\}$ is unique and positive.

We can re-write the optimisation problem as,

$$\max_{S,G}\left\{k - S - G - \theta k[1 - F(S,G)]\right\}, \tag{4.12}$$

where $\theta = 2\left(\frac{\#M-1}{\#M}\right)$. The first-order conditions are,

$$\theta k \frac{\partial F}{\partial S} = 1 \tag{4.13}$$
$$\theta k \frac{\partial F}{\partial G} = 1 \tag{4.14}$$

It is clear by the concavity of F that S and G are increasing in k; but it is unclear whether this means that τ_s and τ_g are also increasing in k since we have that $\tau_s(k) = \frac{S(k)}{k}$. Under some further restrictions on F, let us assume that τ_s is increasing in k. We know that π is rising, that is, the optimal degree of arrangement effectiveness increases as the economy becomes richer. This means, however, that the effect on exchange cost, $\alpha = (1-\pi)\,k$ is ambiguous. We require $\frac{\partial \alpha}{\partial k} = 1 - \pi^*(k) - \frac{\partial \pi}{\partial k}k < 0$ for the case in which the exchange cost falls in the level of the endowment. We have that $\frac{\partial \alpha}{\partial k} = 1 - \pi - k\left[\frac{\partial F}{\partial S}\frac{\partial S}{\partial k} + \frac{\partial F}{\partial G}\frac{\partial G}{\partial k}\right]$. Using the marginal conditions, this reduces to $\theta\frac{\partial \alpha}{\partial k} = \theta(1-\pi) - \left[\tau_s + \tau_g + k\left(\frac{\partial \tau_s}{\partial k} + \frac{\partial \tau_g}{\partial k}\right)\right]$. Given our assumption on F, as k becomes large (and as π approaches unity) so exchange costs will fall in k. As k

approaches zero, however, though $\frac{\partial \tau_s}{\partial k}$ and $\frac{\partial \tau_g}{\partial k}$ may become very large we cannot without further restriction on F know that $\frac{\partial \alpha}{\partial k}$ will not be positive. It may be that exchange costs at first rise in the endowment, before falling. ■

Propositions 4.1 and **4.2** characterise the behaviour of the economy in only a partial equilibrium way; either as the market size changes holding endowment constant or as the endowment changes holding market size constant. We wish to be able to speak of the behaviour of the economy as the endowment increases given an optimal choice of market size.

We may characterise more formally the intermediary's problem. We know the optimal arrangement for every k and $\#M$. Preferences and scale effects in amelioration then determine the optimal level of diversification. The choice facing the intermediary is to optimise over market size, given knowledge of the optimal arrangement for each market size. Let Γ be the set of integers that maximise expected utility,

$$\begin{aligned} \Gamma \;=\; & \arg\max_{\#M} EU \left\{ \left[\left(1 - \tau_s^{\#M} - \tau_g^{\#M}\right) k \right.\right. \\ & \left.\left. - \, 2\left[1 - F\left(S^h, G\right)\right] k \left(\frac{\#M - 1}{\#M}\right)\right] (\#M)^{-1} \sum_{i \in M} \lambda^i \right\}, \end{aligned} \tag{4.15}$$

where $\left\{\tau_s^{\#M}, \tau_g^{\#M}\right\}$ are the solutions to equations (4.10) and (4.11). We may also deduce that Γ, assuming $\Gamma \neq \{\emptyset\}$, can have at most two elements. This is a corollary of **Proposition 4.3**.

Proposition 4.3 *Expected utility is continuous and strictly increasing in the endowment; market size is weakly increasing in the endowment.*

Proof. Where $\#M = 1$ it is clear from equation (4.9) that expected utility is increasing in k. We can see from the equation for exchange costs that $\lim_{k\to 0} \alpha = k$. As such, $EU_{\#M=1} > EU_{\text{any } \#M>1}$ for some $k > 0$ must hold. Given that τ (and so also α) is continuous in k for a given market size, expected utility is also continuous in k for a given $\#M$. Let $\kappa = (k-y)/k = \tau_s + \tau_g + 2\left(\frac{\#M-1}{\#M}\right)(1-\pi)$ be the proportional cost of optimal amelioration and exchange, itself continuous in k. The parameter κ is, then, the extent of per capita *non*-production costs. The choice of optimal market size will be determined by κ and the risk aversion parameter; i.e., agents are keen to diversify but will only do so when the marginal cost of doing so meets the marginal benefit. We know that as $\kappa \to 1$, no diversification is optimal. If diversification to some market size $\#M' > 1$ takes place, it must be that $\frac{\partial\kappa}{\partial k} = \frac{\partial\tau_s}{\partial k} + \frac{\partial\tau_g}{\partial k} - 2\left(\frac{\#M'-1}{\#M'}\right)\frac{\partial\pi}{\partial k} < 0$ at some $k_{\#M'} > 0$ at which $EU_{\#M=1} = EU_{\#M=\#M'}$, i.e., that the share allocated to production, and so expected utility, is increasing in k before and beyond $k_{\#M'}$.

$$\begin{aligned}\frac{\partial\kappa}{\partial k} &= \frac{\partial\tau_s}{\partial k} + \frac{\partial\tau_g}{\partial k} - \theta\left[\frac{\partial F}{\partial S}\left(k\frac{\partial\tau_s}{\partial k} + \tau_s\right) + \frac{\partial F}{\partial G}\left(k\frac{\partial\tau_g}{\partial k} + \tau_g\right)\right]; \\ &= -\left(\theta k\frac{\partial F}{\partial S} - 1\right)\frac{\partial\tau_s}{\partial k} - \left(\theta k\frac{\partial F}{\partial G} - 1\right)\frac{\partial\tau_g}{\partial k} - \theta\left(\tau_s\frac{\partial F}{\partial S} + \tau_g\frac{\partial F}{\partial G}\right),\end{aligned}$$

which, using the optimality conditions, becomes,

$$\frac{\partial\kappa}{\partial k} = -(\tau_s + \tau_g)/k < 0. \tag{4.16}$$

As k increases the same argument follows for further diversification: For diversification to $\#M'' > \#M'$, it must be that expected utility is sufficiently increasing before and after some $k_{\#M''} > k_{\#M'}$ at which $EU_{\#M'} = EU_{\#M''}$. Globally, expected utility is the envelope of expected utility for each market size.

In other words, for some finite market size $\#M'$ to be an equilibrium, it must be that $\#M' + i$, for $i = 1, 2, ...$, exhibits a greater degree of non-production costs and that those larger production costs are not sufficiently compensated by additional risk insurance. As k increases κ falls across all $\#M$, so the addition to non-production costs from diversifying further becomes equal to the marginal utility gain from diversification.

Even in where exchange costs are rising in k, then, diversification is possible if the share of the endowment spent on amelioration and exchange is decreasing fast enough and if agents value diversification enough. When exchange costs fall as the level of the endowment increases further, for a given market size, optimal contributions rise but exchange costs fall. Under some conditions on F, expected utility will be strictly concave in k. ■

The above serves as a description of the mechanics of our model: As the endowment increases, contributions to amelioration rise while also increasing expected utility. As the proportional cost of diversifying further falls, so the optimal market size increases.

A further implication of **Proposition 4.3** is that Γ can be comprised of two elements. We make the simplifying, but not unrealistic, assertion that there exists some arbitrarily small exchange cost, ε, to coordinating general amelioration; we could think of the cost of collecting taxes, for example. As such, the institution would wish to minimise the measure of intermediaries from which it had to receive a contribution to general amelioration. If, in the case that $\varepsilon = 0$, optimal market size is not unique, i.e., that $\Gamma = \{\#M', \#M' + 1\}$, then a positive ε would cause the larger of the two optimal market sizes to be selected. In general we allow this arbitrarily small ε to approach zero. Let us suppose that an institution can enforce intermediary choice of the supremum. As such, we do not formally model it, but still select $\max\{\Gamma\}$ as the solution to equation (4.15). We are then in a position to define the unique first-best allocation. Given this concept, and following **Propositions 4.1–4.3** and **Lemma 4.1**, we have that the endowment and optimal contributions are bijective: For any k there is a unique pair of τ^*; for any τ^* there can be only one k.

The relative sizes of τ_s and τ, as well as the interaction between them, will depend on the form of F. We develop a tractable analysis for a CES amelioration technology in Section 4.6 below.

4.2.1 Core and equilibrium

Definition 4.2 *The **first-best allocation**, given technologies, preferences and endowments, is a unique composition for each market $M \in \mathcal{A}$, comprising a feasible choice of market size for each $h \in H$, $\#M^* = \max\{\Gamma\}$, and the unique optimal allocations $\{\tau^i\} = \{\tau^*\}$ for all $i \in I$ which are the solutions to equations (4.10) and (4.11) with $\#M^h = \#M^*$. It follows that there is an unique optimal level of general ameliorating capital, $G^* = \tau_g^* k$, and unique optimal specific ameliorating capital, $S^{M^*} = S^* = \tau_s^* k$, for each market $M \in \mathcal{A}$, and, as such, a unique optimal degree of ex ante arrangement effectiveness $\pi^* = F(S^*, G^*)$ and exchange cost $\alpha^* = (1 - \pi^*)k$.*

Clearly, this definition relies on us stretching in a somewhat unappealing way the notion of "cooperation"; it requires all other intermediaries to allocate resources to general amelioration. We return to this issue below.

Proposition 4.4 *The core of a cooperative economy can support the first-best allocation.*

Proof. Consider the proposed allocation: $x^h = \{c^*, y^*, \tau^*, \#M^*; i \in M^h\}$ for all $h \in H$. This n-tuple determines $F(S^*, G^*)$, π^* and hence α^*. Let us suppose a strict subset of agents in the market intermediated by h, $B \subset M^h$, can form a blocking coalition. We may view this set of agents as a market. Our notion of the cooperative equilibrium across intermediaries necessitates that the blocking coalition cannot deviate from contributing τ_g^* on average. (Otherwise, as demonstrated in **Proposition 4.5** an allocation with $\tau_g^i = 0$ for all $i \in I$ would block the first-best allocation. We reserve these questions for the purely non-cooperative case.)

So, the blocking intermediary is choosing τ_s^B given the market size $\#B < \#M^h$. These agents are better off with the following program which implies an alternative, unique allocation, x^B:

$$\begin{aligned} c^i &= \bar{c}, \qquad \tau_g^i = \tau_g^*, \qquad \forall i \in B; \\ 1 &= 2\left(\frac{\#B - 1}{\#B}\right) F_s'\left(S^B, G^*\right). \end{aligned}$$

The consumption profile follows from optimizing over CRRA utility functions and the second condition was derived above. By **Proposition 4.1** we note that taxes are strictly lower, arrangements strictly more ineffective and exchange costs strictly higher. There are fewer transactions in this proposed market but each is more costly. In addition, the investment portfolio is less diversified. By **Proposition 4.3**, and given that market size and contributions deviate from first-best, it must be that the higher exchange cost and less diversification is not compensated for by the fewer transactions, and must cause expected utility to be lower. Now consider the case $B \supset M$. The same logic follows: In this proposed market, exchange costs are strictly lower, taxes higher and arrangements more ineffective. The deviation from first-best means that expected utility must be lower than in market M. We conclude that the cooperative allocation is in the core. Further, if $\Gamma \neq \emptyset$ the core is non-empty and $\#M$ is the unique, optimal market size; hence core allocations are allocations of the cooperative economy. If $\Gamma = \emptyset$ core allocations do not exist. ■

4.3 Non-cooperative equilibrium

We need to consider whether the cooperative equilibria characterised in the previous Subsection can be supported as equilibrium outcomes in a non-cooperative economy. We suggested above that voluntary contributions to general amelioration will be zero in the absence of inter-coalition cooperation; here we see that outcome in a fully non-cooperative set-up. Given this result, we need to look at mechanisms that can support some level of general amelioration, and whether this level is optimal.

Proposition 4.5 *In the core, voluntary contributions to general amelioration are zero.*

Proof. Consider an equilibrium in which $\tau_g^i = 0 \; \forall \; i \in I$, so that $G = 0$. We can find a level of specific amelioration that is optimal, with $\tau_s^i = \tau_s^{**}$, $\forall \; i \in I$. Let expected utility for each agent $i \in I$ be $EU^i = EU_0$. Suppose there exists a blocking coalition B such that an agent $b \in B$ proposes $\tau_g^b > 0$ for $i \in B$. If $\#B < \aleph_0$, then $G = 0$ obtains. As such, it follows that $EU^i < EU_0$ for all $i \in B$. Suppose, however that $\#B = \aleph_0$ and that an agent $b \in B$ proposes $\tau_g^b > 0$ for $i \in B$. In this case $G = G^b > 0$. It follows from **Lemma 4.1** that some positive level of G is optimal, so we can have $EU^i > EU_0$ for each $i \in B$. But there exists a blocking coalition $B' \subset B$ in which some agent $b' \in B'$ proposes $\tau_g^{b'} = 0$ for each $i \in B'$. Since $\#B' < \aleph_0$ we still have $G = G^b$. For each $i \in B'$, EU^i must be greater than or equal to that for each $i \in B$. So while there can exist blocking coalitions which propose $\tau_g^b > 0$ for some $i \in B$, they themselves are not in the core. Core equilibria can only be strategies in which $\tau_g^i = 0$ for all $i \in I$. Voluntary contributions to general amelioration are zero. ■

As we suggested in the previous Section, **Proposition 4.5** can also be proven using a weakened cooperative concept of the core, i.e., one with intra- but not inter-coalitional cooperation. This result is invariant to the level of specific amelioration: *Any* intermediation strategy in which the allocation to general amelioration is non-zero will be blocked. We have seen that some cooperative equilibrium concepts can retrieve optimal outcomes in both specific and general amelioration. Given the failure of general amelioration under weaker definitions of cooperative equilibrium we must also consider a different equilibrium concept to resurrect the existence of the public good features of our setup. We must proceed to see whether the optimal non-cooperative strategies of both intermediating and non-intermediating agents can be supported in some forms of decentralised, non-cooperative strategies.

Noncooperative Strategies following Townsend (1978)

We now develop an understanding of the non-cooperative equilibria of our economy. We proceed, first with optimisation for a given market size, and, second, optimal choice of market size. The first step is substantially more involved than the second. Townsend (1978) has demonstrated that cooperative equilibria in a special case of the economy we have developed can be decentralised. In our set-up things are less straightforward because of both the public good aspect of general amelioration and the decision to ameliorate, but we follow his general procedure for decentralisation. Let any agent $h \in I$ propose strategy P^h for intermediating in a market. This strategy has eight components: M^h is the market proposed by agent h; P_1^h is the yield in terms of the consumption good for one share in the portfolio of agent h; P_2^h is the price in terms of the capital good at which agent h is willing to buy an unlimited number of shares in any project i of M^h; P_3^h is a fixed fee in terms of the capital good for the purchase of shares in the portfolio of agent h by $i \in M^h$; P_4^h is the price in terms of the capital good at which agent h is willing to sell an unlimited number of shares in her portfolio to agents i of M^h; τ_s^h is the proportion of the capital endowment that agent h proposes to use in improving the excludable transactions technology; τ_g^h is the proportion of the capital endowment that agent h proposes to use to contribute to the public good transactions technology. Finally, $\alpha^h(\pi)$ is the choice as to the overall effectiveness of transactions technology, our ex post exchange cost. Of course, it is not strictly necessary to include $\alpha^h(\pi)$ in the definition of the strategy space. However, it aids intuition, we think, to do so. For example, it is useful to think of an intermediary optimizing over $\alpha^h(\pi)$ and then optimizing over the pair $\{\tau_s^h, \tau_g^h\}$. In what follows Q_D^{ih} is the quantity of shares purchased by i in h's portfolio, whilst Q_S^{hi} is the quantity of shares sold by i to h. A^{ih} is a switching function, where $A^{ih} = 1$ if agent i buys shares in the portfolio of intermediary h, and $A^{ih} = 0$ otherwise.

Step 1: Optimisation for a given #M. Optimisation by a non-intermediary entails the maximisation of expected utility by choice of $\{Q_D^{ih}, Q_S^{hi}, y^i, A^{ih}\}$,

$$\sum_{h:i \in M^h} \left[A^{ih}\right] \left[EU\left(Q_D^{ih} P_1^h + y^i \lambda^i - Q_S^{hi} \lambda^i\right)\right], \tag{4.17}$$

subject to the following contraints regarding feasibility and participation,

$$\left[\sum_{h:i \in M^h} \left(A^{ih}\right) \left((1 - \tau_s^h - \tau_g^h) k^i + Q_S^{hi} P_2^h - Q_D^{ih} P_4^h - P_3^h - \alpha(\pi^h) - y^i\right)\right] \geq 0; \tag{4.18}$$

$$c^i(\omega) = \left[\sum_{h:i \in M^h} \left(A^{ih}\right) \left(Q_D^{ih} P_1^h(\omega) + y^i \lambda^i(\omega) - Q_S^{hi} \lambda^i(\omega)\right)\right] \geq 0, \forall \omega \in \Omega; \tag{4.19}$$

$$A^{ih} = 1 \Rightarrow A^{ij} = 0, \forall i \text{ such that } i \in M^j, \forall j \neq h. \tag{4.20}$$

An intermediary chooses strategies to maximise,

$$EU\left(y^h\lambda^h + \sum_{i\in M^h}\left[Q_S^{hi}\lambda^i - Q_D^{ih}P_1^h\right]\right), \tag{4.21}$$

subject to,

$$\left[(1-\tau_s^h-\tau_g^h)k^h - \alpha(\pi^h)\left(\#M-1\right) - y^h + \sum_{i\in M^h}\left[Q_D^{ih}P_4^h + P_3^h - Q_S^{hi}P_2^h\right]\right] \geq 0; \tag{4.22}$$

$$c^h(\omega) = \left[\sum_{i\in M^h}\left[Q_S^{hi}\lambda^i(\omega) - Q_D^{ih}P_1^h(\omega)\right] + y^h\lambda^h(\omega)\right] \geq 0, \forall\omega\in\Omega; \tag{4.23}$$

$$\begin{aligned}
G \leq \mathcal{G} &= \frac{1}{\#I}\sum_{i\in I}\tau_g^i k; &(4.24)\\
S^h \leq \mathcal{S}^h &= \frac{1}{\#M^h}\sum_{i\in M^h}\tau_s^i k; &(4.25)\\
\alpha^h &= (1-\pi^h)k; &(4.26)\\
\pi^h &= F(S^h, G); &(4.27)\\
\sum_{i\in M^h}P_3^h - \alpha\left(\#M-1\right) &= -2\alpha\frac{(\#M-1)}{\#M}; &(4.28)\\
P_3^h &= \frac{\alpha\left(\#M-2\right)}{\#M}. &(4.29)
\end{aligned}$$

An equilibrium of this non-cooperative game is a set of actions $\left\{Q_{S*}^{ij}, Q_{D*}^{ji}, A_*^{ij}\right\}$ and a strategy $P_*^i = \left\{M_*^i, P_{1*}^i, P_{2*}^i, P_{3*}^i, P_{4*}^i, \tau_{s*}^i, \tau_{g*}^i, \alpha_*^i(\pi)\right\}$ for each agent $i \in I$ (where for any variable x, we use the following convention: $x^{ii} = 0$), an allocation $\{c_*^i, y_*^i; i \in I\}$ and a set of markets which satisfy:

1. If agent i is not an intermediary ($Q_{D*}^{ji} = 0, \forall j \in I$) then $\left\{y_*^i, Q_{S*}^{ij}, Q_{D*}^{ji}, A_*^{ih}, \tau_{s*}^{hi}, \tau_{g*}^{hi}, \alpha_*^{hi}(\pi)\right\}$ maximises (4.17) subject to (4.18), (4.19) and (4.20) and $P^h = P_*^h$ $\forall h \neq i$. c_*^i is given by (4.19).

2. All agents participate in one market, M, the union of which covers the popu-

lation. In each market there is one intermediary such that $h \in M$, $M_*^h = M$, and $\left\{\tau_s^{ih}, \tau_g^{ih}, \alpha^{ih}(\pi)\right\} = \left\{\tau_{s*}^{ih}, \tau_{g*}^{ih}, \alpha_*^{ih}(\pi)\right\}$. For each $i \in M - h$, $A^{ih} = 1$. For every such h, P^h is feasible, with y^h chosen to maximise (4.21) subject to (4.22) and (4.23). c_*^h is given by (4.23).

3. There exist no blocking strategies for any agent of I.

For solutions in the core, we know that contributions to general amelioration are zero; we demonstrated that in **Proposition 4.5** and so by Property 3 above such contributions are not part of the equilibrium.

In the same way as with the cooperative case, we can characterise the optimal strategies for intermediaries and non-intermediaries for a given market size. As such, the following unconstrained optimization delivers the supporting price vector holding $\#M$ constant. In a sense, agents carry out the optimisation for a sufficiently large range of $\#M$ and each optimally selects that $\#M$ which maximises expected utility. In what follows we reduce notational clutter by noting that when we write x we really mean $x(\omega)$.

The Lagrangian for agents in an economy who are not an intermediary is given by,

$$\begin{aligned}\mathcal{L}^{NI} &= EU\left(Q_D^{ih}P_1^h + y^i\lambda^i - Q_S^{hi}\lambda^i\right) \\ &-\phi\left[(1-\tau_s^h-\tau_g^h)k^i + Q_S^{hi}P_2^h - Q_D^{ih}P_4^h - P_3^h - \alpha(\pi^h) - y^i\right] \\ &+\mu\left[Q_D^{ih}P_1^h(\omega) + y^i\lambda^i(\omega) - Q_S^{hi}\lambda^i(\omega) - c(\omega)\right].\end{aligned} \tag{4.30}$$

First-order necessary conditions for non-intermediaries include,

$$U'(\cdot)\lambda^i + \phi + \mu\lambda^i = 0; \tag{4.31}$$

$$-U'(\cdot)\lambda^i - P_2^h\phi - \mu\lambda^i = 0; \tag{4.32}$$

$$U'(\cdot)P_1^h + \phi P_4^h + \mu P_1^h = 0. \tag{4.33}$$

Equations (4.31) and (4.32) together imply that $P_2^h = 1$, and so it follows that $P_4^h = 1$.

We can write down an intermediary's problem in a similar way. The Lagrangian follows,

$$\begin{aligned}\mathcal{L}^I \;=\;& EU\left(y^h\lambda^h + \sum_{i\in M^h}\left[Q_S^{hi}\lambda^i - Q_D^{ih}P_1^h\right]\right) \qquad (4.34)\\ &-\phi^h\left[(1-\tau_s^h-\tau_g^h)k^h - \alpha(\pi^h)\,(\#M-1) - y^h + \sum_{i\in M^h}\left[Q_D^{ih}P_4^h + P_3^h - Q_S^{hi}P_2^h\right]\right]\\ &+\mu^h\left[\sum_{i\in M^h}\left[Q_S^{hi}\lambda^i(\omega) - Q_D^{ih}P_1^h(\omega)\right] + y^h\lambda^h(\omega) - c(\omega)\right]\\ &-\eta^h\left[\sum_{i\in M^h}P_3^h - (1-F(S^h,G))k\,(\#M-1) + 2(1-F(S^h,G))k\frac{(\#M-1)}{\#M}\right]\end{aligned}$$

By **Proposition 4.5** we know that the shadow price of general amelioration will be zero, so $G = 0$ holds. The first-order necessary conditions simplify, and include:

$$U'(\cdot)\sum_{i\in M^h}\lambda^i + \phi^h\,(\#M)\,P_2^h + \mu^h\sum_{i\in M^h}\lambda^i = 0; \qquad (4.35)$$

$$-U'(\cdot)\,(\#M)\,P_1^h - \phi^h\,(\#M)\,P_4^h - \mu^h\,(\#M)\,P_1^h = 0; \qquad (4.36)$$

$$U'(\cdot)\lambda^h + \phi^h + \mu^h\lambda^h = 0; \qquad (4.37)$$

$$\phi^h k - \phi^h F'(S^h,0)(\#M-1)k - \eta^h F'(S^h,0)(\#M-1)k + 2\eta^h F'(S^h,0)k\frac{(\#M-1)}{\#M} = 0. \qquad (4.38)$$

Using $P_2 = P_4 = 1$, it follows that,

$$P_1^h = -\frac{\phi^h}{U'(\cdot)+\mu^h}, \qquad (4.39)$$

and,

$$U'(\cdot) = -\frac{\phi^h\,(\#M)\,P_2^h}{\sum_{i\in M^h}\lambda^i} - \mu^h. \qquad (4.40)$$

Therefore,

$$P_1^h = \left(\frac{1}{\#M}\right) \sum_{i \in M^h} \lambda^i. \tag{4.41}$$

Finally, we have,

$$P_3^h = [1 - F(S^h, 0)k] \frac{(\#M - 2)}{\#M}. \tag{4.42}$$

Since they are the shadow prices of dual constraints we have $\phi^h = -\eta^h$, so then (4.38) implies,

$$1 = 2F'(S^h, 0) \frac{(\#M - 1)}{\#M}. \tag{4.43}$$

This expression, when evaluated, is the same as equation (4.10), in the absence of general amelioration. Equation (4.43) determines the unique optimal contribution to specific amelioration, $\tau_s^{\#M}$, for a given market size, $\#M$. This completes the characterisation of the optimal strategy for all agents in a market intermediated by agent h as,

$$P_*^i = \Bigg\{ M^h, \left(\frac{1}{\#M^h}\right) \sum_{i \in M^h} \lambda^i, 1, [1 - F(S^h, 0)k] \frac{(\#M - 2)}{\#M}, \\ 1, \tau_{s*}, 0, \big(1 - F(S^h, 0)k\big) \Big\}, \forall i \in M^h, \forall h \in H.$$

Proposition 4.6 *In the core, specific amelioration is optimal and intermediary rents are zero.*

Proof. Consider some level of endowments, preferences and technologies such that $\hat{S}^M$ is the optimal level of specific amelioration for each $M \in \mathcal{A}$. There is no general amelioration. The optimal average allocation for each market is, therefore, $\hat{\tau}_s^{**} = \hat{S}^M / \hat{k}$ and utility is denoted by EU^{**}. For some market $M \in \mathcal{A}$, let $B \subseteq M$ be a blocking strategy in which an agent $b \in B$ proposes allocations in which $\tau^{ib} \neq \hat{\tau}_s^{**}$ such that $S^B \neq \hat{S}^M$. By definition of the optimality of $\hat{S}^M$ it must be that for some $i \in B$, $EU^i < EU^{**}$. This is the relevant contradiction. ■

Step 2: Optimisation over $\#M$. An optimal choice of $\#M$ then maximises expected utility. By **Proposition 4.3** we know that this optimal $\#M^*$ will be unique and our second best equilibrium is well-defined by P_*^i where $\#M^h = \#M^*$ for all $h \in H$. We have thus managed to retrieve one of the main cooperative outcomes in a non-cooperative setting. Townsend (1978) demonstrates that non-cooperative equilibria of this sort are in the core, and that all core allocations can be supported by such non-cooperative equilibria.

4.4 Economy-wide institutions

We have seen that in the absence of cooperation between intermediaries, general amelioration cannot be in the core. What is required is some government or institution that has the power to tax and spend. If somehow the "government is an exogenous, benevolent economic agent"[28] then perhaps the first-best, analyzed in Section 4.2, is attainable. The exogeneity of benevolent governance is a useful sidestep. But how stable is such governance? How might an heterogeneous case be resolved? Can a government be formed to represent the median voter, or could it represent the interests with most to gain from a policy that suits a minority?

Questions of public goods, redistribution and governance in general equilibrium have been addressed in Aumann and Kurz (1977), among others.[29] The key motivating notion there is that a government must reflect in its policies the forces that keep it in power. This literature has shown that the distribution of voter power can matter if the task of the government is a redistribution of income. Agents in a sense can make side-payments to 'buy' the votes of other agents. Since in the homogeneous case there is one public good and one type of agent, we do not need to worry about these effects in this Section. However, we introduce here a form of

[28] Aumann and Kurz (1977), p. 1137.

[29] Also there is Becker (1983) and Lindbeck (1985).

'non-cooperative democracy', and outline its application to an heterogeneous case in Section 4.5.

In the process of electioneering,[30] a group of agents $\mathcal{V} \subseteq I$ offer manifestos to be voted upon. The manifesto $\mathcal{M}^g = \{\{G^{ig}\}_{i\in I}, G\}$ of each agent $g \in \mathcal{V}$ includes taxation levels G^{ig} for each agent $i \in I$ as well as promised general amelioration level $G \leq \frac{1}{\#M}\sum_{i\in I} G^{ig}k^i$. We assume that there is no deception here; agents who promise a manifesto follow through with it to the letter. Agent $i \in I$ votes for the manifesto of the agent that will deliver her the highest expected utility where $EU^i|_{g'}$ is the expected utility to agent i if agent g' forms the government. We let $V^{g'}$ denote the set of agents who vote for the manifesto of agent g'. So, if $\#V^{g'} > \#V^{g''}$ for every $g'' \in \mathcal{V}\backslash g'$, then agent g' forms the government, imposes taxation levels and delivers general amelioration. It is simple to show that core equilibria can be characterised by each agent $i \in I$ being taxed equally and where there is no rent from governing, i.e., $G = \frac{1}{\#I}\sum_{i\in I} \tau_g^i k$. Consider the alternative to this. If an agent $g' \in \mathcal{V}$ offers a manifesto $\mathcal{M}^{g'} = \{\{G^{ig'} = \tau_g'\}_{i\in I}, G'\}$ in which $\frac{1}{\#I}\sum_{i\in I}\tau_g' k > G'$ there is some other agent $g'' \in \mathcal{V}$ who offers a manifesto $\mathcal{M}^{g''} = \{\{G^{ig''} = \tau_g'\}_{i\in I}, G''\}$ in which $\frac{1}{\#I}\sum_{i\in I}\tau_g' k > G'' > G'$ which delivers $\#V^{g''} > \#V^{g'}$. Given perfect competition in the political process, the rent from governing is driven to zero.

Proposition 4.7 *General and specific amelioration are jointly-optimal and in the core of an economy with a non-cooperative democracy.*

Proof. This is a straightforward corollary of **Proposition 4.4**. ■

[30] In what we are describing here, there is a relation to the work of Besley and Coate (1997) but, for our analysis, the findings of Aumann, *et al.*, are more directly related. For Besley and Coate, a candidate possesses a preferred policy and seeks a mandate from citizens to carry through only that policy; in our economy, agents seek to form a government for purely selfish reasons. Our model, with a countable infinity of agents and a single (purely) public good, fits more easily into an economy in which agents bid strategies for governance; the prevailing policy need not be any individual agent's idiosyncratic preference. Moreover, a central assumption of Besley and Coate is that a 'representative democracy' exists; if no candidate prevails, the 'default policy' does. In our model, and in reality, the absence of governance is anarchy, i.e., no policy.

We have demonstrated that some forms of decentralised behaviour can capture the equilibria of the cooperative environment described in Section 4.2.

4.5 Heterogeneous Agents and Political Economy

The existence of authority and power in bilateral or multilateral contracting underlies a considerable amount of contract theory. In the context of general equilibrium the existence of a power imbalance will be reflected in a stable equilibrium in which one party extracts rents from another party or group of parties. The analysis above demonstrated that such rents cannot exist in large and homogeneous economies of general equilibrium: Specifically, there can be no equilibrium in which one agent extracts a rent from her market since there is free-entry to the establishment of markets. Exogenously distinguishing between types of agents along, for example, employer/employee lines does not easily fit into our concept of exchange, even though we are looking at similar problems to those involving such differences. An economy populated by agents differentiated by their ability to form arrangements might give us more interesting results, however, particularly given that agents of different type will share the public good. The outcomes of a democratic procedure in an heterogeneous economy might be expected to differ from the homogeneous case: In the words of Aumann and Kurz (1977, p. 1137),

> the actions of the government, and in particular its tax policies, can be understood only as an endogenous consequences of the political forces that enable it to maintain power.

Extensions of our basic model to an heterogenous economy follow, but we leave the majority of this extension to future work in Nolan and Trew (Forthcoming).

4.5.1 The Heterogeneous Set-up

Consider that, before agents make any decisions, there is an exogenous and random allocation of agents to 'Type'. Agents can be one of a finite number of Types, indexed on $q \in Q = \{1, ..., n\}$. We denote the Type of agent i by q^i. For any agent $i \in I$, the probability of being Type $q' \in Q$, $\Pr[q^i = q'] \in (0, 1)$, is the same for all $q' \in Q$ and for all $i \in I$. So each agent has the same probability of being any Type. This necessitates, of course, that $\Pr[q^i = q] = n^{-1}$ for all $i \in I$ and all $q \in Q$. Let $I^q \subset I$ be the set of all Type q agents, so $\cup_{q \in Q} I^q = I$. Since n is finite each set of Types comprises a countable infinity of agents. There is complete information following the realisation of Type; every agent $i \in I$ knows the Type of any agent $j \in I$.

Let us restrict heterogeneity to enter only with different specific ameliorating functions; all other aspects, the general ameliorating function, the degree of risk aversion and the endowment, are all homogeneous across agents. In other words, for reasons of education, status, geographic location, intrinsic ability, etc., some agents in the economy are better able[31] to form private and excludable arrangements that concern the conduct of exchange. An agent of Type 1 is better able to form arrangements than an agent of Type 2, who is in turn better able than a Type 3 agent, etc.

In Section 4.1 the cost of exchange is determined by the degree of effectiveness of the ex ante arrangement formed by agents, $\alpha^{ij} = (1 - \pi^{ij})\, k$, which in the homogeneous environment is independent of $\{i, j\}$. Furthermore, in an intermediated market, the exchange cost in a market intermediated by agent h is simply $\alpha^h = \left(1 - \pi^h\right) k$. Again, this concept is symmetric across all agents. If different agents have different production technologies over specific arrangements then π^{ij},

[31] We mean *better able* in the following sense: An agent a can form an arrangement of effectiveness $\hat{\pi}$ at a lower cost than can an agent b of lesser ability.

and so α^{ij}, is not independent of $\{i,j\}$. We can make a simple departure from the homogeneous case by viewing the agent Type as something akin to a production parameter. Suppose that ameliorating technologies are additively separable, such that the degree of effectiveness of an arrangement written between two agents, i and j, is, where $\tilde{q}^{ij}$ is some increasing function of q^i and q^j,

$$\pi^{ij} = F_S^{\tilde{q}^{ij}}\left(S^{ij}\right) + F_G\left(G\right).$$

Suppose also that $F_S^{\tilde{q}^{ij}}\left(S^{ij}\right)$ is strictly decreasing and strictly concave in $\tilde{q}^{ij} = \left(q^i + q^j\right)/2$. So, for a given $\bar{S}^{ij} = \bar{S}^{lm}$, we have that $F_S^{\tilde{q}^{ij}}\left(\bar{S}^{ij}\right) > F_S^{\tilde{q}^{lm}}\left(\bar{S}^{lm}\right)$ for all $\{i,j,l,m\} \in \times_4 I$ iff $\tilde{q}^{ij} < \tilde{q}^{lm}$. As such, the first derivative, $\partial F_S^{\tilde{q}^{ij}}/\partial\tau_s$ is also decreasing in $\tilde{q}^{ij}$. By the conditions for optimality, equations (4.10) and (4.11), we know also that as $\tilde{q}^{ij}$ rises, so τ_s^* falls for a given market size. As such, since $\pi^{ij} = F_S^{\tilde{q}^{ij}}\left(S^{ij}\right) + F_G\left(G\right)$ is decreasing, we know that $\alpha^{ij} = \left(1 - \pi^{ij}\right)k$ is increasing in $\tilde{q}^{ij}$: Ex ante arrangements become more effective as the average Type falls. So long as market size is constant, changing $\tilde{q}$ does not affect τ_g^*, the optimal contribution to general amelioration. Type 1 agents possess the most efficient specific amelioration technology; Type 2 agents possess the second most, etc. And a combination of Type 1 and Type 2 agents can be more efficient at forming arrangements than two Type 2 agents.

The extension to an intermediated environment follows: $\pi^h = F_S^{\tilde{q}^h}\left(S^h\right) + F_G\left(G\right)$, where the production technology F_S is strictly decreasing and strictly concave in $\tilde{q}^h = \frac{1}{\#M^h}\sum_{i \in M^h} q^i$. For a given market size, what happens to expected utility will depend on its effect on the extent of non-production costs, $\kappa = \left(k - y\right)/k = \tau_s + \tau_g + 2\left(\frac{\#M-1}{\#M}\right)\left(1 - \pi\right)$. Suppose that q can take on a continuum of values. Given that π^{ij} is additively separable, we can take the total derivative of π with respect to q and show, using also $\frac{\partial \tau_g}{\partial q}|_{\#M} = 0$, that,

$$\begin{aligned}\frac{d\kappa}{dq} &= \frac{d\tau_s}{dq} - 2\left(\frac{\#M-1}{\#M}\right)\left(\frac{\partial F_S}{\partial q}\frac{dq}{dq} + k\frac{\partial F_S}{\partial S}\frac{d\tau_s}{dq}\right); \\ &= -2\left(\frac{\#M-1}{\#M}\right)\frac{\partial F_S}{\partial q} - \left[2\left(\frac{\#M-1}{\#M}\right)k\frac{\partial F_S}{\partial S} - 1\right]\frac{d\tau_s}{dq},\end{aligned}$$

which, using equation (4.10), leaves,

$$\frac{d\kappa}{dq} = -2\left(\frac{\#M-1}{\#M}\right)\frac{\partial F_S}{\partial q}. \tag{4.44}$$

Since we have that $\frac{\partial F_S}{\partial q} < 0$, it follows that $\frac{d\kappa}{dq} > 0$ where $\#M > 1$; by $\frac{\partial^2 F_S}{\partial q^2} > 0$, we have $\frac{d^2\kappa}{dq^2} < 0$. For a given market size, arrangements formed by agents with a higher average Type spend more of their endowment on exchange and amelioration. So, for a given market size greater than one, arrangements formed by agents of lower Type have strictly greater expected utility. Of course, we must also consider the effect of $\tilde{q}$ on optimal market size. If changing $\tilde{q}$ induces a change in optimal $\#M$ then it must be in the direction of reducing optimal market size with increasing average Type. This not only further reduces expected utility but, by equation (4.44), also increases the the rate at which κ changes. **Proposition 4.8** develops a useful simplification that we can make.

Proposition 4.8 *Agents of different Type do not exchange and have different optimal strategies.*

Proof. Suppose that it is an equilibrium for agent i to write an ex ante arrangement with agent j, and for no further diversification to take place. Suppose also that $q^i < q^j$. If costs are equally shared, agent i could do strictly better by forming an arrangement with any other agent of type $q' < q^j$. But agent j has an incentive to make compensating side-payments to i in order to induce her into forming an ex ante arrangement. It can only be feasible to do so if the gain in utility for the agent j is greater than the loss in utility for agent i, and this is ruled-out by the strict concavity of the function F_S in $\tilde{q}$. Expected utility can be written as $EU^i = \sum_{\omega \in \Omega} p(\omega) u \left\{ [(1-\kappa) k] \bar{\lambda} \left(\omega, M^h \right) \right\}$. For a given market size, then, EU^i is also strictly decreasing and strictly concave in $\tilde{q}$. Consider $EU^i_{M^{ij}}$ to be the expected utility to agent i from diversifying into a market composed of agents i and j. Let us loosely denote then $EU^i_{M^{ii}}$ the expected utility to i from diversifying into a market with another agent of her own Type[32], and $EU^j_{M^{jj}}$ likewise. So it is only *feasible* for agent j to make any side-payments if $EU^j_{M^{ij}} - EU^j_{M^{jj}} > EU^i_{M^{ii}} - EU^i_{M^{ii}}$. By the strict concavity of F_S in $\tilde{q}$, we know that this cannot be the case since $2EU_{M^{ij}} < EU^i_{M^{ii}} + EU^j_{M^{ii}}$. This argument continues to hold good with more than two agents, and is made stronger if the inclusion of the less able agent causes a reduction in the optimal market size.

For the arrangement to be an equilibrium, then, it must be that $q^i \geq q^j$. Suppose then that $q^i > q^j$. This too is a contradiction since it cannot be feasible for agent i to make sufficient side-payments: It must be that, in equilibrium, all ex ante relationships are formed between agents of the same Type, i.e., $q^i = q^j$.

Moreover, each agent Type will always have different optimal specific contributions, and will have different optimal general contributions iff optimal market sizes differ. Specifically, the 'better' an agent's specific amelioration technology, the greater will be her optimal contributions to each form of amelioration. This is a straightforward corollary of the conditions for optimality in the homogeneous case, equations (4.10) and (4.11). ■

Agents do not mix Type in the formation of ex ante arrangements, and so do not mix Type when it comes to exchange. Different agent Types essentially operate in economies joined only by the public good.

[32]We must remember that we are in a world in which a countable infinity of agents of the same Type as agent i exist.

4.5.2 A Two-Type Case and Democratic Outcomes

Suppose there are two Types of agents, $Q = \{1, 2\}$. Each agent of Type q has a first-best strategy P_q^i. This is the strategy that would be optimal if the economy was comprised of only agents of type q. Let us suppose that the agent Types are such that $\#M_1^h > \#M_2^h > 1$ always holds, so we know that $\tau_x^1 > \tau_x^2 > 0$ for $x \in \{s, g\}$. In other words, agents of Type 1 tend to favour larger markets and higher taxes than those of Type 2. Which tax rate, or compromise, prevails?

Agents wish to decide the flat rate of tax, τ_g, by means of a majority vote. Could a coalition D of voters comprising slightly greater than a half of the population tax the complement of D their whole endowment and redistribute it? Following Aumann and Kurz (1977), we consider an environment in which any agent can, if she wishes, destroy her endowment in the face of what she perceives to be unfair taxation. So the majority coalition D cannot pillage those in $I \backslash D$ with abandon; it must consider the credible threat of agents in $I \backslash D$ to destroy their own endowments. The procedure is a cooperative one in the sense that threats are credible and that each agent agrees to abide by the democratic result. The resultant allocations are competitive, however, in that they can be supported by a vector of prices.

Can the resultant tax differ from one which a benevolent dictator would impose? Aggregate social welfare is, put simply,

$$W = \sum_{i \in I} \theta^i E\left[u^i\left(\tau_g\right)\right], \tag{4.45}$$

where each θ^i is a weight on the utility of each individual $i \in I$. The choice of weights is evidently critical to the tax chosen by the benevolent dictator. Consider the standard in which the utility of each agent is valued equally, $\theta^i = \hat{\theta}$, $\forall i \in I$. The solution to (4.45), $\hat{\tau}_g = \arg\max_{\tau_g} W$, is then the socially optimal level of tax in a purely egalitarian sense. But where agents can vote upon taxation regimes, the weights on utility themselves become endogenous; they reflect the power of each agent. Aumann and Kurz (1977) develop these ideas in the context of a value allocation following Shapley (1953). We very briefly introduce the concept of the Shapley value before going on to its implications for political economy.

The Shapley (1953) Value

The Shapley-value allocation in cooperative games reflects the marginal contribution of each agent to aggregate utility. The classic exposition in a general equilibrium context is Aumann (1975). Let N be a finite set of *players*, indexed by i; subsets of N are *coalitions*. A *game* is a function ν that associates with a coalition, S, its worth, $\nu(S)$. The *value* of a game ν is given by a payoff vector $\varphi\nu$ where φ is a function satisfying conditions of, among other things, efficiency, symmetry and additivity. The central result is as follows:

Proposition 4.9 *There is a unique value φ given by the following equation,*

$$(\varphi\nu)(\{i\}) = E(\nu(S_i \cup \{i\}) - \nu(S_i)), \tag{4.46}$$

where S_i is the set of all players preceding i ordered randomly and E is the expectations operator when all $|N|!$ such orders are assigned equal probability.

Proof. See Shapley (1953). ■

The intuition behind equation (4.46) is that the value of each agent is proportional to the expected marginal contribution that she makes to the coalition. Allocations work at the margin. Applications include that of Hart and Moore (1990) in the context of optimal firm structures. Aumann (1975) shows that a central result follows in a competitive (non-atomic) continuum market economy: Every value allocation is competitive; and, under some conditions on preferences, every competitive allocation is a value allocation. This is the Value Equivalence

Theorem.

For our purposes, the most important interpretation of the Shapley value is as an index for measuring the power of agents in a game.

Power and Political Outcomes

Aumann and Kurz (1977) develop the Shapley value allocations in the context of power and taxes in an income redistribution game. Political outcomes depend on both voting resources *and* economic resources.

This contrasts with the standard political economy setting in which each agent has one vote and maximises her own utility. In a sense, we have two concepts of 'optimality': An individually rational, *power*, approach that incorporates the power of each agents into weighting; and an *egalitarian* approach, in which the optimal tax is that which maximises the welfare of every agent given equal weighting. The former is that which results endogenously from competition among interest groups in an economy, in a sense 'might is right'; the latter is what we would normally consider to be optimal, an aggregate utility with weights equal regardless of the relative power of agents.

We can think of a political economy game as one of competing coalitions of agents, each of whom attempts to maximise their own surplus by winning in a majority-vote procedure. As noted above, the key threat that agents in the minority can make is to destroy their own endowment.

What is the aggregate utility of a political coalition and what power does that political coalition have relative to other political coalitions with whom it is competing to set taxes and reallocate resources. In the income-redistribution game of Aumann and Kurz (1977), if a coalition S is the whole economy $T = S$ then it maximises aggregate utility given a vector of weights, $\boldsymbol{\theta} = \{\theta^i\}$ on each agent i.[33]

[33] Aumann and Kurz (1977) employ a continuum economy, so $\{i\}$ here is not an individual agent but an infinitesimally small subset di of the ocean of all agents.

Given endowments $\boldsymbol{e}$, equilibrium allocations $\boldsymbol{x}$ are those that solve,

$$\nu_{\boldsymbol{\theta}}(S) = \max \left\{ \int \boldsymbol{\theta} u(\boldsymbol{x}) : \int \boldsymbol{x} = \int \boldsymbol{e} \right\}. \tag{4.47}$$

There may be, however, gains to be made for a coalition $S \subset T$ to allocate resources amongst itself, bargaining as a whole with agents in $T \backslash S$. There are two problems: First, the vector of weights for finding the worth of each coalition can be chosen arbitrarily; second, the worth of the coalition with respect to its complement plays a central role in its bargaining position. The insight of Aumann and Kurz (1977) is that, by using the Shapley value, these two problems solve each other.

The detailed procedure is given in that paper. Consider for now a simple optimisation problem terms of the following Lagrangian,

$$\mathcal{L} = \int \boldsymbol{\theta} u(\boldsymbol{x}) + p \left(\int \boldsymbol{e} - \int \boldsymbol{x} \right). \tag{4.48}$$

The first-partial with respect to $\boldsymbol{x}$ obtains $\boldsymbol{\theta} u'(\boldsymbol{x}) = p$. This relates to the Value Equivalence Theorem: $\boldsymbol{x}$ is the Walrasian allocation corresponding to the price system p. The problem in a world where coalitions form to decide on political outcomes is that the allocations determined by this partial condition are *rarely* feasible. Aumann and Kurz (1977) demonstrate that the unique feasible allocation corresponds to weights determined by the Shapley value.

So, in the final compromise of the income redistribution game, the weight upon an individual agent is $\theta^i = p/u'\left(x^i\right)$ where p is some constant. This results from the efficiency axiom: Marginal utility, $\theta^i u'\left(x^i\right)$ given an ex post allocation x^i is equal for all agents i with positive income.

Power and Taxes with Endogenous Exchange Costs

In the context of the model presented above, we need to view these resulting allocations in a slightly different way. For feasibility, it must be that the vector of weights $\boldsymbol{\theta}$ across the whole economy is feasible given the resultant socially optimal tax rate τ_g and vector of net incomes $\mathbf{x}$. Since our game is not one of direct income redistribution, but instead of redistribution via the choice of τ^*, we must find,

$$\tau_g^* = \arg\max_{\tau_g} W = \sum_{q\in Q}\sum_{i\in I^q} \theta_q^i E\left[u_q^i\left(\tau_g\right)\right], \tag{4.49}$$

such that,

$$\theta_1 = 1/u_1'\left(\tau_g^*\right); \tag{4.50}$$

$$\theta_2 = 1/u_2'\left(\tau_g^*\right). \tag{4.51}$$

The problem in the context of our exchange cost economy is simplified by the fact that the ‘government’ is not directly redistributing resources between agents. It is enough for the tax rate to satisfy equations (4.49)–(4.51); the resulting ‘allocation’ will always be feasible.

Given a strictly concave utility function, the *power* approach will, following Aumann and Kurz (1977) mean that a strictly higher weight is placed on those agents whose Type is lower; as such, the socially optimal tax in a *power* world would be strictly higher. Given that, in equilibrium, markets formed by Type 1 agents are larger than those formed by Type 2 agents, the ‘measure’[34] of Type 1 intermediaries is lower than the measure of Type 2 intermediaries. In a sense, the smaller group of relatively more adept intermediaries form a political elite in the economy and are able to shape the tax regime to suit themselves better. Such a result suggests deeper foundations for the work, such as that of Acemoglu and

[34] And, of course, we are being loose in our terminology here.

Johnson (2006), which appeals to the existence of elites to deliver institutional outcomes. We develop some preliminary numerical implications in Section 4.6, but further pursue most of these deeper issues in Nolan and Trew (Forthcoming).

4.6 A Numerical Application with a CES Production Technology

We began this Chapter with a quote from Coase (1992). It asserted that a large portion of what we consider to be economic activity is actually aimed at achieving what high transactions costs would otherwise prevent. We have developed a framework in which we can consider the validity of this assertion: Under a reasonable specification, can the extent of non-production costs in our economy be 'large'? Furthermore, we might ask whether the implied levels of taxation can be reconciled with what we know about the size of governments in modern developed and under-developed economies. Moreover, what might the interactions between wealth, taxation and market size look like in quantitative terms?

To examine the quantitative implications of the exchange cost model laid out above we first return to the case in which agents are homogeneous. We use MATLAB to compute equilibria; code is reproduced in Appendix Section A.3.

Consider a general CES technology for the production of ex ante arrangements,

$$F\left(S^h, G\right) = \left[\delta F_s \left(S^h\right)^\sigma + (1-\delta) F_g \left(G\right)^\sigma\right]^{1/\sigma}, \tag{4.52}$$

where $\delta \in (0,1)$ and $s = 1/(1-\sigma)$ is the constant elasticity of substitution. Functions F_x are for both $x = s$ and $x = g$ continuous, decreasing and strictly concave in its only argument; with $F_x' > 0$, $F_x'' < 0$, $F_x(0) = 0$, $F_x(y) \to 1$ as $y \to \infty$, $F_x'(y) \to 0$ as $y \to \infty$ and $F_x'(0) = \infty$ for $x \in \{s, g\}$. We need not require that F_g

and F_s are identical. Let us impose simply,

$$F_s = [1 - \exp(-\beta\tau_s k)]^{\gamma_s};$$
$$F_g = [1 - \exp(-\beta\tau_s k)]^{\gamma_g},$$

which, with $\beta, \gamma_s, \gamma_g \in (0, 1)$, satisfy our requirements. These functions have a number of useful properties that both satisfy our requirements on arrangement technologies and aid us in the numerical exercise. Taking the first partial of F_s with respect to τ_s we obtain,

$$F_s' = \beta k \gamma_s \exp(-\beta\tau_s k)[1 - \exp(-\beta\tau_s k)]^{\gamma_s - 1}$$

Our Inada-type requirements, that $F_s(0) = 0$, that $F_s(\tau_s k) \to 1$ as $\tau_s k \to \infty$, $F_s'(\tau_s k) \to \infty$ as $\tau_s k \to 0$ and $F_s'(\tau_s k) \to 0$ as $\tau_s k \to \infty$, thus hold. In addition, the parameters $\beta, \gamma_s, \gamma_g \in (0, 1)$ allow us to control numerical results with some ease. Calibrating the model *is* difficult: There are a large number of equilibria in which optimal market size is infinite, and a large number in which the optimal market size is unity. Further, computational limitations mean that even with a small number of potential technologies the combinatorial numbers reach the limit of floating-point accuracy quickly. Finding a calibration which results in optimally finite-sized markets that is also numerically computable using MATLAB v.7 requires us to choose parameter values with care. Roughly speaking, β impacts upon the level of the exchange cost; γ_s, γ_g impact upon the level of the proportional contributions.

We use the parameters γ_s, γ_g to make distinctions between the relative efficiency of specific and general amelioration. Of course, we could also have done this with separate β_s and β_g. Because $F_x \in [0, 1]$ for $x \in \{s, g\}$ we know that $\frac{\partial F_x}{\partial \gamma_x} \leq 0$. So we can take agent Types of 'better ability' to have a strictly lower γ_s.

Let us define a CRRA utility function as $U(x) = [x^{1-\gamma} - 1] / (1 - \gamma)$ with $\gamma > 0$

being the coefficient of relative risk aversion. We must also restrict the space of technologies. For numerical tractability, we reduce to a two-state case in which the set of technologies is restricted to $\Lambda = \{\lambda_1, \lambda_2\}$ where $\lambda_1 < \lambda_2$ and $p(\lambda_1) = \rho$ so $p(\lambda_2) = 1 - \rho$.

The expected average technology for a market of size $\#M$, denoted $\bar{\lambda}_{\#M} = \frac{\sum_{i \in M} \lambda^i}{\#M}$, is invariant to market size, namely $E[\bar{\lambda}_{\#M}] = \rho\lambda_1 + (1-\rho)\lambda_2, \forall \#M$. The benefit from increasing the market size is in increasing the distribution of expected average technologies. In the case where $\#M = 1$, the average technology for the market can be λ_1 or λ_2 with probabilities ρ and $(1-\rho)$. With $\#M = 2$, technologies can be $\{\lambda_1, \lambda_1\}$, $\{\lambda_1, \lambda_2\}$, $\{\lambda_2, \lambda_1\}$ or $\{\lambda_2, \lambda_2\}$. In this case the average technology can in addition be $(\lambda_1 + \lambda_2)/2$ with probability $2\rho(1-\rho)$. With a strictly concave utility function (and with an abuse of notation), $U[(\lambda_1 + \lambda_2)/2] > \{U[\lambda_1] + U[\lambda_2]\}/2$. The logic follows that as $\#M$ increases, so the range of expected average technologies fill out the range $[\lambda_1, \lambda_2]$, and so expected utility increases.

So while expected average technology will be constant, the range of average technology differs. The expected average technology for a market of size $\#M$ is given by the following,

$$E\left[\bar{\lambda}_{\#M}\right] = \sum_{i=0}^{\#M} \left\{ \binom{\#M}{i} \rho^{\#M-i}(1-\rho)^i \left[(\#M - i)\lambda_1 + i\lambda_2\right] \right\} / \#M, \tag{4.53}$$

which, as described above, is invariant to $\#M$. What is important, is the effect this has on expected utility. The optimisation problem becomes,

$$\Gamma = \arg\max_{\#M} \sum_{i=0}^{\#M} \binom{\#M}{i} \rho^{\#M-i}(1-\rho)^i U \left[\begin{array}{c} \left\{ \left(1 - \tau_s - \tau_g\right) k - \frac{(2\alpha)(\#M-1)}{\#M} \right\} \times \\ \times \left[(\#M - i)\lambda_1 + i\lambda_2\right] / \#M \end{array} \right]$$

subject to,

$$2\left(\frac{\#M-1}{\#M}\right)\left[\delta\left[1-\exp\left(-\beta\tau_s k\right)\right]^{\gamma_s\sigma}+(1-\delta)\left[1-\exp\left(-\beta\tau_g k\right)\right]^{\gamma_g\sigma}\right]^{\frac{1-\sigma}{\sigma}}\times \\ \times\delta\beta k\gamma_s\exp\left(-\beta\tau_s k\right)\left[1-\exp\left(-\beta\tau_s k\right)\right]^{\gamma_s\sigma-1}=1, \tag{4.54}$$

$$2\left(\frac{\#M-1}{\#M}\right)\left\{\delta\left[1-\exp\left(-\beta\tau_s k\right)\right]^{\gamma_s\sigma}+(1-\delta)\left[1-\exp\left(-\beta\tau_g k\right)\right]^{\gamma_g\sigma}\right\}^{\frac{1-\sigma}{\sigma}}\times \\ \times(1-\delta)\beta k\gamma_g\exp\left(-\beta\tau_g k\right)\left[1-\exp\left(-\beta\tau_g k\right)\right]^{\gamma_g\sigma-1}=1, \tag{4.55}$$

where $\alpha = \left\{1-\left[\delta\left[1-\exp\left(-\beta\tau_s k\right)\right]^{\gamma_s\sigma}+(1-\delta)\left[1-\exp\left(-\beta\tau_g k\right)\right]^{\gamma_g\sigma}\right]^{1/\sigma}\right\}k$ and where we select $\sup\{\Gamma\}$ for a unique solution.

The relative shares of each form of amelioration will be determined by,

$$\frac{F_g'\left(\tau_g k\right)}{F_s'\left(\tau_s k\right)}=\left[\frac{(1-\delta)}{\delta}\right]^s. \tag{4.56}$$

We can solve equation (4.54) for τ_g,

$$\tau_g=-\frac{1}{\beta k}\ln\left\{1-\left\{\left(\frac{1}{1-\delta}\right)\left[\begin{array}{c}\left[2\left(\frac{\#M-1}{\#M}\right)\gamma_s\delta\beta k\exp\left(-\beta\tau_s k\right)\left[1-\exp\left(-\beta\tau_s k\right)\right]^{\gamma_s\sigma-1}\right]^{\frac{\sigma}{\sigma-1}} \\ -\delta\left[1-\exp\left(-\beta\tau_s k\right)\right]^{\gamma_s\sigma}\end{array}\right]\right\}^{\frac{1}{\gamma_g\sigma}}\right\} \tag{4.57}$$

and then substitute for τ_g in equation (4.55) to find a closed-form solution for τ_s^*. This can then be used in equation (4.57) to find the unique optimal pair of contributions $\tau^*=\left\{\tau_s^*,\tau_g^*\right\}$. Equation (4.57) also makes plain that our requirement on F, for τ_s and τ_g are increasing in k, will hold.

4.6.1 Amelioration, Diversification and Utility with a Fixed Endowment

As an example, we can run a baseline calibration of the choices over differing levels of specific and general amelioration. For a given pair $\{\tau_s, \tau_g\}$ agents choose $\#M$ to maximise expected utility. Table 4.1 gives the parameter values in the baseline case.

Table 4.1: Baseline Calibration for Endogenous Exchange Costs

endowment	k	25
coefficient of relative risk aversion	γ	3
specific amelioration curvature	γ_s	0.065
general amelioration curvature	γ_g	0.085
ameliorating capital factor	β	0.03
weight on specific amelioration	δ	0.5
CES coefficient s, $1 - \frac{1}{s} =$	σ	0.5
low technology	λ_1	1
high technology	λ_2	5
probability of low technology	ρ	0.5

Our baseline case supposes a difference in the technology of specific and general amelioration: Since $\gamma_g > \gamma_s$ we have imposed that, relative to specific, general amelioration is a less efficient ameliorator of exchange costs: For a given level of ameliorating capital, general amelioration obtains a lesser degree of ex post arrangement effectiveness than does specific amelioration.

First we consider the effect on market choices and expected utility from different levels of ameliorating allocations: $\tau \in \{\tau_s, \tau_g | \tau_s \in [0,1], \tau_g \in [0,1], \tau_s + \tau_g < 1\}$. This is something of a brute-force method of finding optimal combinations of $\{\tau_s, \tau_g\}$, but it helps us to see the workings of our model more clearly. We consider in this first simulation *allocations* of amelioration costs, so even if agents do not diversify, the cost of the allocated contribution is still deducted from their endowment. The problem of agents is simply to form optimally-sized markets given the combination of $\{\tau_s, \tau_g\}$. In other words, if an agent is not going to diversify at some pair of contributions, then she is better off not making any contributions. Figure 4.1 reflects this, and gives the expected utility of decisions over optimal market size for a grid of pairs within this range of τ.

The figure demonstrates optimal responses in terms of choices over diversification; for each allocated $\{\tau_s, \tau_g\}$ pair, agents choose how far (if at all) to diversify given their residual endowment and given exchange costs resulting from the amelioration determined by that $\{\tau_s, \tau_g\}$ pair. We draw the picture twice: The upper panel depicts in three dimensions the interaction between expected utility and both amelioration allocations but does not clearly show a peak in utility for some non-zero pair of $\{\tau_s, \tau_g\}$; the lower panel shows in two dimensions that this pair exists. Figure 4.2 demonstrates the choice over diversification more explicitly by giving the underlying choice of market size against allocations of $\{\tau_s, \tau_g\}$ pairs.

We can see from Figures 4.1 and 4.2 that over some convex combinations of the $\{\tau_s, \tau_g\}$ pairs, expected utility from diversification is higher than from not diversifying.

There is a peak in expected utility at some unique combination of τ_s and τ_g. Using a grid search for maximum expected utility with steps every 0.002 of τ_s and τ_g, the numerical maximum expected utility is obtained with diversification to a market size of 13 with optimal amelioration expenditure of $\hat{\tau}_s^* k = 1.15$ and $\hat{\tau}_g^* k = 1.5$. This can be seen in Figure 4.1 as an optimal pair $\hat{\tau}^* = \{0.046, 0.06\}$. The optimal arrangements have a degree of effectiveness of $\hat{\pi}^* = 0.7846$, with an exchange cost of $\hat{\alpha}^* = 5.3854$. We begin to interpret the relatively low optimal tax rate in the next Subsection, once we have seen that it holds over a range of endowments.

Interestingly, market size does not peak at the peak of expected utility: Sub-optimal allocations can induce *larger* markets as a second-best. If, for example,

Figure 4.1: Expected Utility and Amelioration Allocations

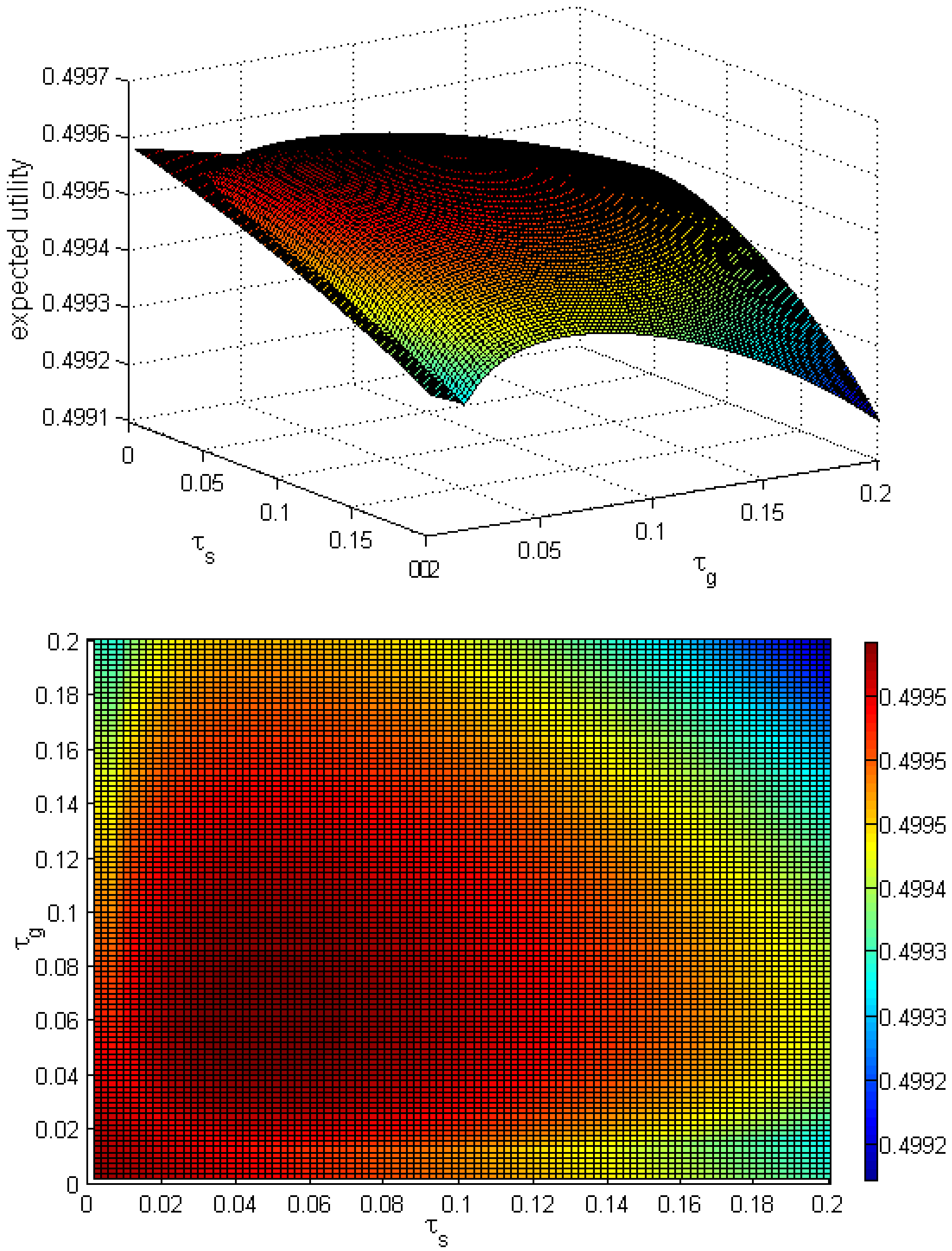
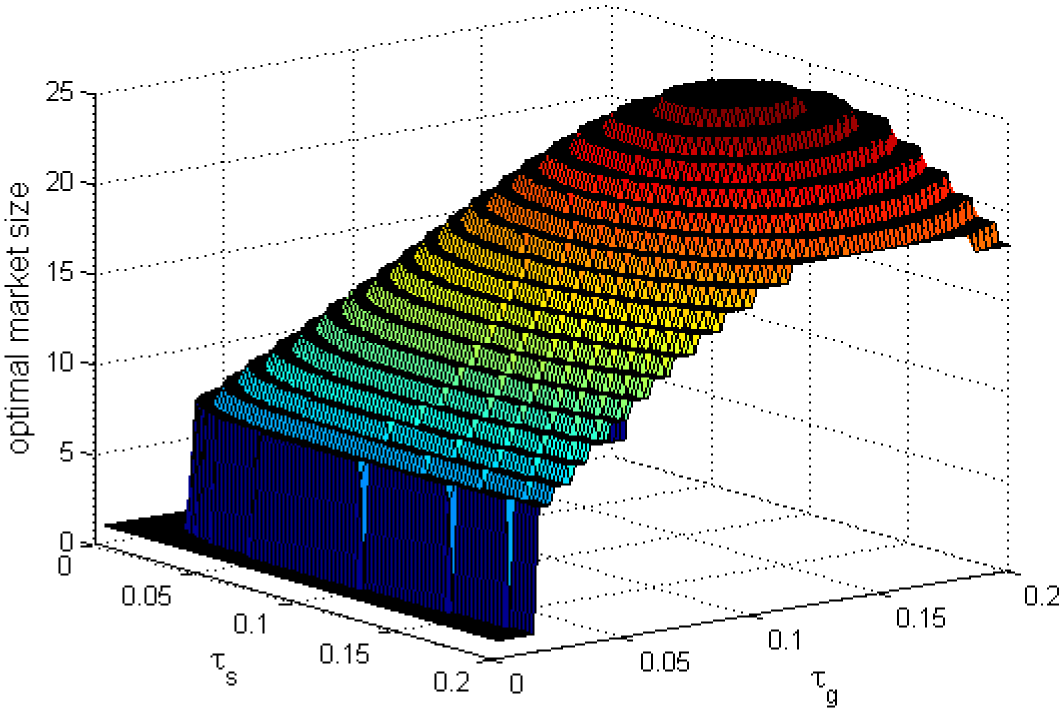
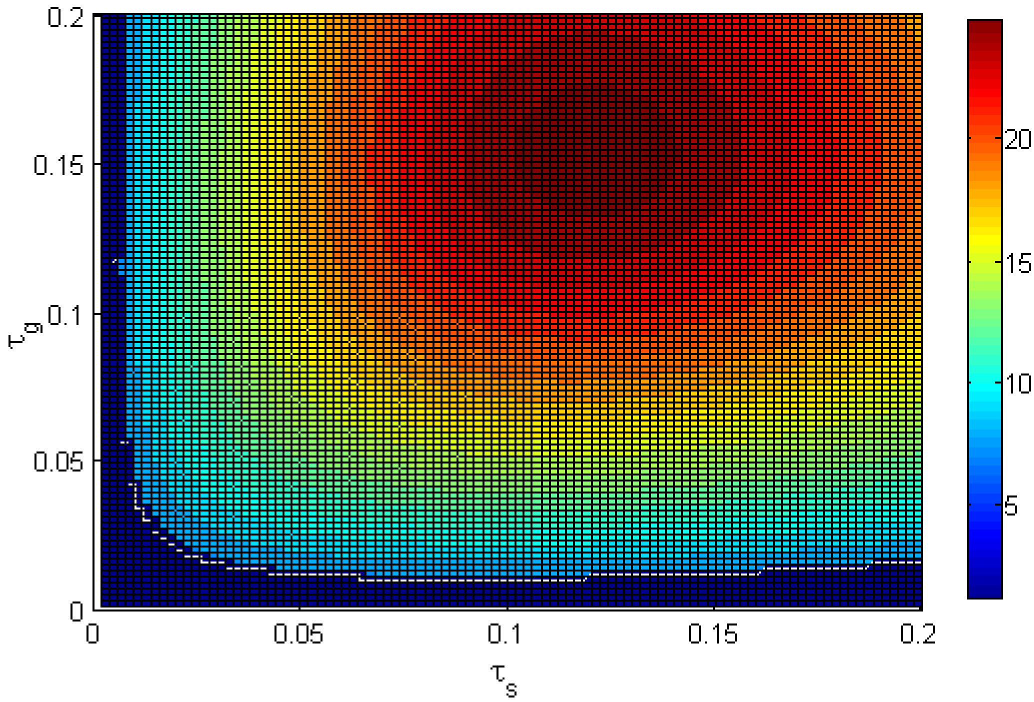


Figure 4.2: Optimal Market Size and Amelioration Allocations





the level of τ_g is, for whatever reason, forced higher than an agent would choose optimally, then we can see by the distribution of colour in the lower panel of Figure 4.1 that a constrained optimal response (i.e., one that maximises expected utility holding τ_g fixed) can be to choose a higher τ_s. So agents are, in second best, spending much larger amounts on non-production costs. Exchange costs are consequently much lower, and so agents diversify more. Further, there is a limit to this positive relationship: The market size does peak in Figure 4.2, it begins to fall after $\tau_g \approx 0.16$. There is, then, a hump-shaped response in both market size and τ_s to increases in τ_g. If, conversely, τ_g is forced lower than its optimal level, agents choose a lower τ_s, exchange costs are higher and diversification is always (weakly) diminished. This will become an important observation later, when we come to look at economies with heterogeneous agents.

An alternative solution method can evaluate the marginal conditions using equation (4.57) find $\Gamma = \{\#M^*\}$. Naturally, this yields similar, but improved, results. It also serves to check our brute-force method. A market size of 13 and optimal amelioration expenditure of $\tau_s^* k = 1.1703$ and $\tau_g^* k = 1.4874$ ($\tau = \{0.046811, 0.059497\}$).

Of course, in solving the marginal conditions numerically the method using marginal conditions also invokes something like a grid search to find optimal pairs of contributions. The brute-force method, however, searches over the entire range of $\tau \in \{\tau_s, \tau_g | \tau_s \in [0,1], \tau_g \in [0,1], \tau_s + \tau_g < 1\}$ for each market size under consideration, while the second only searches over solutions to equation (4.55) with equation (4.57) substituted into it.

The simulation code is such that for each τ on the search grid we consider a range of market sizes, $\{1, 2, ..., \#\bar{M}\}$, to which agents could diversify. The optimal market size for that τ is the one in the range $\{1, 2, ..., \#\bar{M} - 1\}$ which maximises expected utility. If expected utility is maximised at $\#\bar{M}$, we must re-run the experiment with a higher $\#\bar{M}$. So, consider the calibration in Table 1 with a triangular grid search with steps every 0.005 and maximum market size of 25; we have to carry out $\bar{M}$ calculations for $\sum_{i=1}^{n}(n-i)$ pairs of τ which for $n = 1 + (1/0.005) = 201$ and $\#\bar{M} = 25$, is 502,500 calculations in all. By contrast, in the numerical solution to the analytical method, we only need to carry out the search over only one pair of τ for each market size, i.e., only 25 calculations. It is therefore vastly more efficient and, generally, more accurate. We provide MATLAB code in Appendix Section A.3.

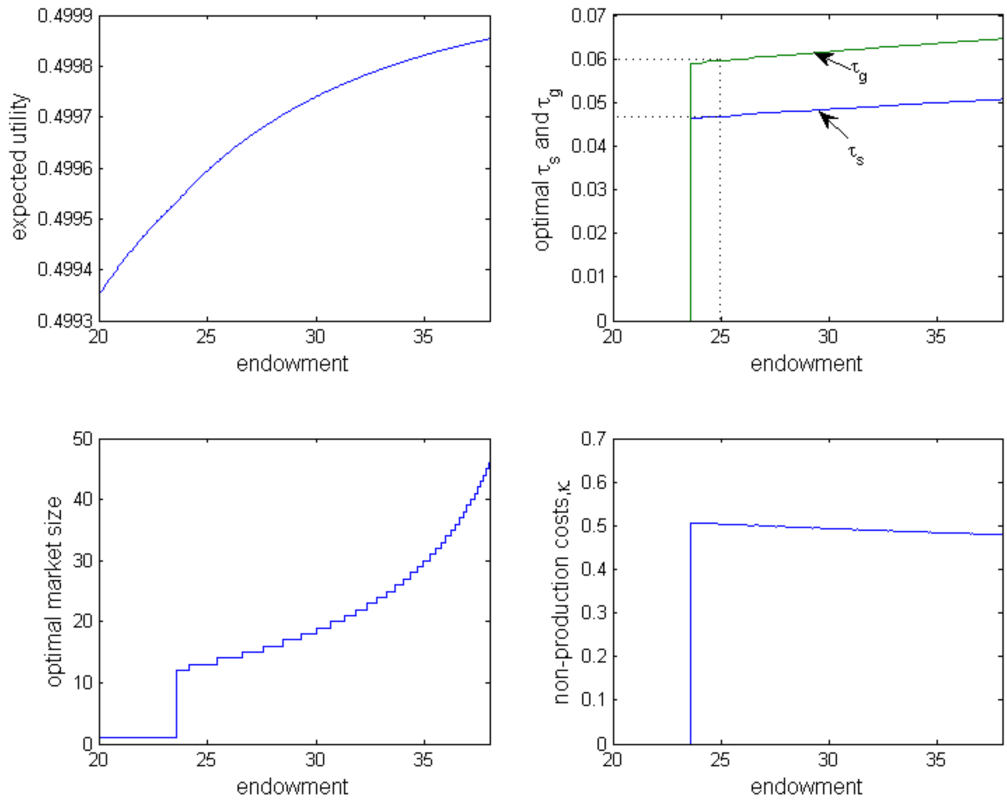
4.6.2 Amelioration, Diversification and Utility over a Range of Endowments

We can consider the nature of optimal amelioration in economies with different attributes. Our analytical results above, specifically **Propositions 4.1–4.3**, have formalised a number of interactions in the model. But we can get a handle of the magnitude of these interactions by simulating the economy over a range of endowment levels.

Using the parameter values given in Table 1, but allowing k to vary over the range $[20, 38]$, we solve the marginal conditions for the optimal amelioration allocations for a range of market sizes and select the diversification level that maximises expected utility, following the procedure laid-out in Section 4.2. Figures 4.3 gives results for these simulations.

The previous numerical simulation is of course nested here: Where $k = 25$, $\tau = \{0.046811, 0.059497\}$ is shown by the dotted lines. **Propositions 4.1–4.3** are confirmed numerically. For a given endowment, optimal contributions rise, ex ante arrangements become more effective and exchange costs fall as the market size increases. For a given market size, optimal contributions rise and ex ante arrangements become more effective as the endowment increases. Moreover, expected utility is continuous, strictly increasing and strictly concave in the endowment.

Figure 4.3: Amelioration, Diversification, Utility and the Endowment



Interpreted literally, these numerical results suggest that the optimal tax rate over a range of endowments is quite low. A modern developed economy might expect the level of taxation as a percentage of GDP to be around 30–45%; OECD (2005) have, for 2002, tax receipts as a percentage of GDP ranging from 18.1% in Mexico to 50.2% in Sweden. The OECD average was 36.3%. In a sense, our numbers are then too small. But our set-up and its calibration are restricted by two things: The desire to maintain a realistic degree of relative risk aversion; and, computational limits. We can see from Figure 4.3 that the level of the optimal tax is not rising sharply in the level of the endowment. Tax rates will not quickly reach those observed in developed economies, so we need not be too concerned that computational limitations are driving these results. Further, deviating from a reasonable degree of risk aversion, while it will have an impact (as we will see below), would only allow us to match realistic tax levels at the expense of unrealistically risk averse agents. But perhaps these numbers *are* realistic: The governments of modern economies have (at their minimum) both efficiency and equity aims. The numbers coming out of this exchange cost model reflect only taxes which fund institutions that improve the efficiency of the market. It may be that, of the tax burden imposed by modern governments, only between 5 and 10 percentage points, around a quarter of the take, are 'efficiency taxes'.

A few additional remarks can be made. Market size increases slowly at low levels of the endowment, and more quickly as the endowment becomes larger: There is a take-off effect in regard to the non-linearity between wealth and market size. The extent of non-production costs also follows the pattern we expected, falling in k for a given $\#M$ and increasing in $\#M$ for a given k. Here the former effect outweighs the latter: As the endowment becomes larger so the share of non-production costs trends downwards. What is interesting, though, is the *extent* of non-production costs that induce finite-sized coalitions of agents to form. Here, agents can prefer to forsake half of their endowment in order to diversify, most of which comes in the form of the ex post exchange cost. As the economy becomes richer, so agents shift their non-production spending toward higher ameliorating contributions and greater ex ante arrangement effectiveness. We cannot yet say whether these sorts of numbers are realistic, but they are the product of reasonable parameterisation and are, as we shall see below, not worryingly sensitive to alternative parameter values. As Coase noted, a lot of what we might think of as market activity is designed to achieve what high transactions costs would otherwise prevent. Here we see that behaviour played out in a general equilibrium framework.

4.6.3 Robustness to Risk Aversion and Ameliorating Technologies

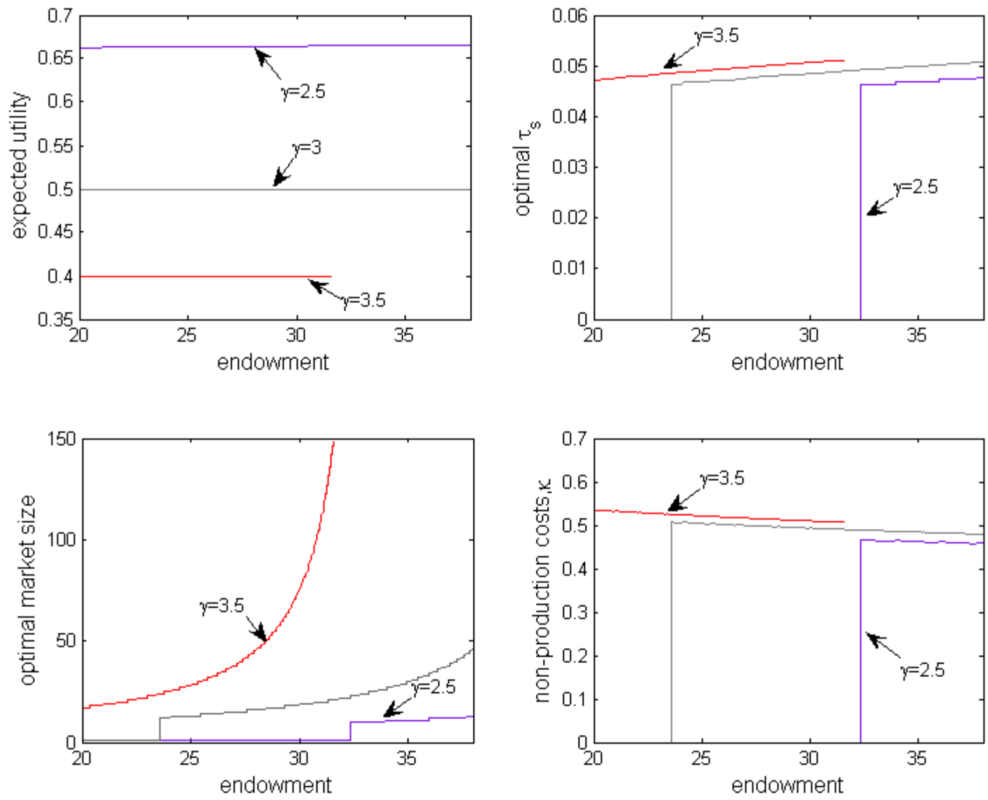
Each of the numerical findings discussed above arise out of reasonable calibrations of quite general functional forms. But we need to say something about the sensitivity of these results to the choice of different parameter values. We have already noted that some sensitivity exists, but a lot of this can be put down to computational limitations. MATLAB v.7, for example, will only calculate up to 170!. As can be seen in the bottom-left panel of Figure 4.3, this will quickly become a limiting problem.

Regardless, we can see the effect of relatively small changes in both the risk aversion parameter and the technological parameters. Figures 4.4 and 4.5 give results for different parameter values. In each, the grey-coloured, central case is that resulting from the parameterisation in Table 4.1. Variations on the coefficient of relative risk aversion, γ, and on the ameliorating technology, β, are Table 4.1 values multiplied by $1 \pm \frac{1}{6}$.

Figure 4.4 shows some variation in optimal behaviour in regard to risk aversion. Higher risk aversion means that agents are more keen to diversify. The level of the endowment at which any diversification becomes optimal is decreasing in the level of risk aversion. For a given technology, this means that as risk aversion increases, optimal market size for a given endowment level weakly increases. Further, agents are willing to spend more on diversification, as can be seen by the extent of non-production costs in each case.

Figure 4.5 demonstrates the effect of varying the coefficient on ameliorating capital in the exchange technology. Making exchange arrangements less costly, by reducing β, means that agents can diversify with much lower endowments. The extent of non-production costs is lower for a given endowment, but unchanged for a given market size. The effect of changing the exchange technology is primarily

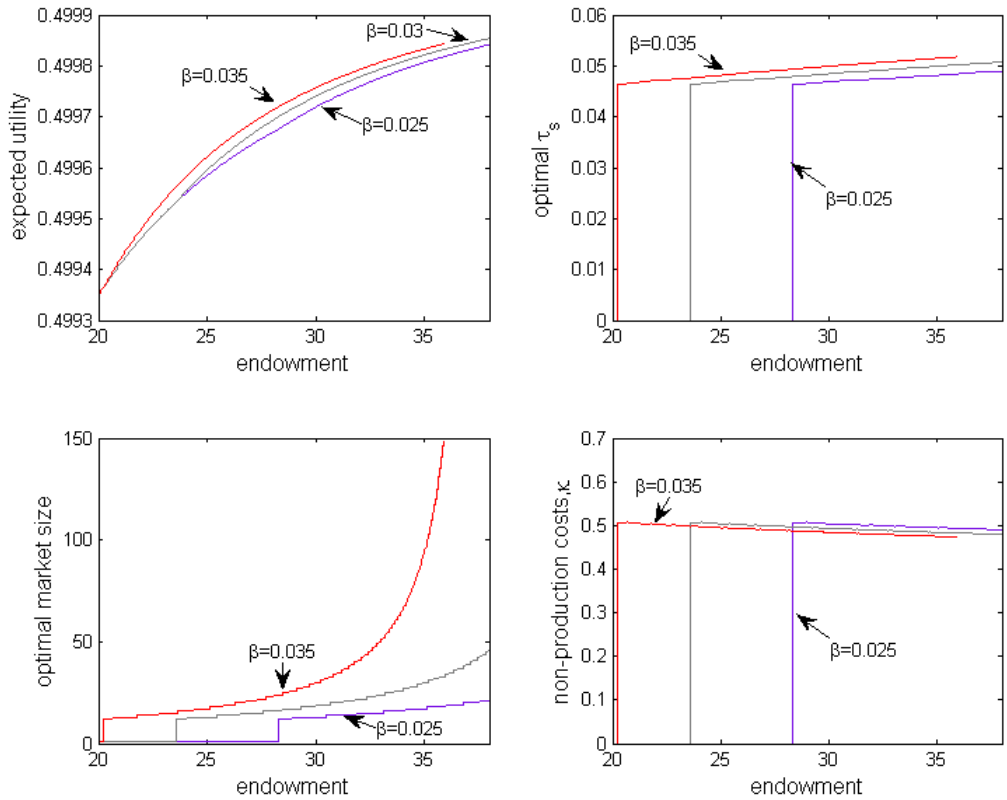
Figure 4.4: Risk Aversion and Endogenous Exchange



to make it feasible for agents to diversify with a lower endowment: The nature of that diversification is little affected.

So there is some sensitivity to parameter values but, as discussed above, this causes difficulties largely because of the computational limitations on market size. It is not the case that the model parameterisations are what would normally be termed "knife-edge" – the model behaves in the ways we can expect over a reasonably large range of parameterisations.

Figure 4.5: Ameliorating Technology and Endogenous Exchange



4.6.4 Heterogeneity and Taxes

We can consider the numerical implications of Shapley-allocations in determining the equilibrium tax rates in an heterogeneous environment. Consider the set-up in Subsection 4.5.2, in which two Types of agent exist, so $Q = \{1, 2\}$. Agents are differentiated by their ability to ameliorate exchange costs using specific amelioration; suppose that Type 1 agents have a more efficient technology of specific amelioration. By **Proposition 4.8**, we know that agents of different Type will not form into the same coalition. Further, if agents are sufficiently different we found that they will wish to diversify into coalitions of different size; agents of Type 1 will seek to form larger markets. This will mean that optimal general amelioration will be greater for the Type 1 agents.

In terms of the functional forms introduced above, equation (4.52), this can be captured by letting $\gamma_s^1 < \gamma_s^2$. The effect of changing γ_s is not trivial, as we shall see in the calibration below. In general, we can say that agent Types with better specific amelioration will always weakly prefer to diversify into larger markets. By the homogeneity of the general amelioration technology, their optimal τ_g will be weakly greater. The impact on optimal τ_s is ambiguous, however: Greater diversification induces a higher τ_s^* while the better technology allows a lower τ_s^*.

We wish to solve for the equilibrium tax rate following Aumann and Kurz (1977). The tax rate must satisfy equations (4.49)–(4.51). First, the weight on each agent Type yields a welfare-maximising tax rate; second, each Type's inverse marginal utility resulting from that tax rate must be proportional to those weights. We can think of a general iterative algorithm for the two-Type case as follows:

1. Solve for equilibrium behaviour and expected utility over the range of $\tau_g \in [\tau_g^2, \tau_g^1]$;
2. For each $\tau_g^\diamond \in [\tau_g^2, \tau_g^1]$ find the implied power of each agent Type, $1/u_1'\left(\tau_g^\diamond\right)$ and $1/u_2'\left(\tau_g^\diamond\right)$;
3. Solve for τ_g^* to maximise equation (4.49) given weights $\theta^1 = 1/u_1'\left(\tau_g^\diamond\right)$ and $\theta^2 = 1/u_2'\left(\tau_g^\diamond\right)$;
4. If the solution τ_g^* from 3. coincides with the $\tau_g^\diamond$ that resulted in the ex post marginal utilities then we have satisfied equations (4.49)–(4.51).

Our procedure for finding the marginal utilities is by using two-sided differences: Calculate utility with $(1 \pm h)k$ (where h is the step size), holding $\#M$ constant but allowing τ_s to change, and then use the formula for the first derivative; i.e., $f'(x) \doteq [f(x+h) - f(x-h)]/(2h)$ which, for sufficiently small h, will be accurate enough for our purposes (see Judd, 1998 §7.7, for a full description). MATLAB code is reproduced in Appendix Section A.3. In general, we have found that an exact and unique solution will exist.

Table 4.2 gives our baseline calibration for the heterogeneous case. Table 4.3 gives preliminary results for four cases: Full weighting given to each Type; an egalitarian case in which each Type is weighted equally; and, a case in which the tax is set according to Aumann-Kurz power allocations using the algorithm above.

Table 4.2: Calibration for Endogenous Exchange Costs with Heterogeneity

endowment	k	25
coefficient of relative risk aversion	γ	3
specific amelioration curvature for Type 1 agents	γ_s^1	0.065
specific amelioration curvature for Type 2 agents	γ_s^2	0.100
general amelioration curvature	γ_g	0.085
ameliorating capital factor	β	0.03
weight on specific amelioration	δ	0.5
CES coefficient s, $1 - \frac{1}{s} =$	σ	0.5
low technology	λ_1	1
high technology	λ_2	5
probability of low technology	ρ	0.5

Table 4.3: Results for Endogenous Exchange Costs with Heterogeneity

	$\boldsymbol{\theta_1}$	$\boldsymbol{\theta_2}$	$\#\boldsymbol{M^1}$	$\#\boldsymbol{M^2}$	$\boldsymbol{\tau_s^{1*}}$	$\boldsymbol{\tau_s^{2*}}$	$\boldsymbol{\tau_g^*}$	$\boldsymbol{\kappa^1}$	$\boldsymbol{\kappa^2}$
1. Type 1	1	0	13	17	0.0470	0.0708	0.0595	0.5037	0.5916
2. Type 2	0	1	6	7	0.0667	0.0625	0.0538	0.4696	0.5496
3. Egalitarian	1	1	7	8	0.0432	0.0638	0.0550	0.4750	0.5587
4. Power	590.9	354.8	9	11	0.0467	0.0622	0.0572	0.4889	0.5758

By favouring only Type 1 agents in welfare maximisation, Case 1 demonstrates that our heterogeneous economy nests the homogeneous economy from Subsection 4.6.1. The optimal market size of Type 1 agents, and their optimal $\{\tau_s, \tau_g\}$ pair, is that from the homogeneous economy (there is a slight rounding error apparent in

τ_s because τ_g is not solved using the marginal conditions. This is the unconstrained optimal behaviour of Type 1 agents, with an optimal tax level of 6.0%. Type 2 agents respond by diversifying more, and by allocating more resources to specific amelioration. We developed an understanding of this phenomenon in Subsection by inspecting Figure 4.2. Case 2 depicts the unconstrained optimal behaviour of Type 2 agents: Smaller markets and a lower tax of 5.4%. The optimal response of Type 1 agents is to reduce their optimal market size but *increase* the level of specific amelioration. As can be seen by the extent of non-production costs, despite this increase in τ_s^1, κ^1 falls markedly. Agents with better ability to form specific arrangements, when allocated a lower than optimal τ_g, respond by allocating more resources to the amelioration with which they have a comparative advantage.

Most general equilibrium welfare analysis takes the weight on all agents to be equal. An egalitarian welfare maximisation obtains Case 3. We see that the equilibrium is somewhat closer to the Type 2 optimum, with a tax of 5.5%, and market sizes only slightly greater than those in Case 2. This follows from the strict concavity utility function: The marginal gain to agents of Type 1 of some increase in τ_g is less than the marginal loss for agents of Type 2, so an equal weighting will tend towards reproducing Case 2.

Case 4 incorporates power allocations, following the Aumann-Kurz developments of the Shapley value. Type 1 agents, as a result of their greater expected contribution to aggregate utility, hold a greater weight in welfare maximisation, equivalent to around $\theta_1 = \frac{5}{8}$ and $\theta_2 = \frac{3}{8}$. The optimal tax level is, given this weighting, 5.7%. This is still sub-optimal for agents of each Type but, relative to the egalitarian arrangement, Type 1 agents gain in expected utility terms, and Type 2 agents lose out. So, in an economy where powerful agents can have greater say over political economy outcomes, we can deviate from the efficient outcomes of a traditional welfare-maximising equilibrium. Private activity is, in some senses, less efficient in an economy in which institutions are dominated by a powerful elite.

The results from the numerical analysis of the heterogeneous economy are intriguing. They also shed light on a number of unexplored theoretical phenomena. We observe that agents of the Type with less power can actually form into *larger* markets than those agents of the Type with greater power. Agents who have less ability to form effective arrangements are, relative to the egalitarian case normally addressed in welfare analysis, worse-off. There is a great deal of interesting work to be done in understanding these interactions. How does the power allocation relate to the formation of markets? Can these power allocations be supported in a fully non-cooperative environment? What is the relation between the level of wealth and the power of the better able agents? Can we say something more general about the relationship between market efficiency and the power of elites in the formation of institutions? We leave a fuller analysis to Nolan and Trew (Forthcoming).

4.7 Discussion and Concluding Remarks

Of course, given the highly abstract nature of our set-up, it is difficult to pin any specific interpretation on our results. We suggested in the introduction, however, that our framework could be interpreted in the light of the recent literature on incomplete and partially complete contracts.[35] In many ways, the microeconomic frictions invoked by a large range of the macroeconomic literatures can be seen in this light.

4.7.1 Incomplete Contracting and Exchange

Maskin and Tirole (1999) argue that incomplete contracts are poorly motivated in models where agents are otherwise fully rational. On the other hand Hart and

[35] See the very recent Hart and Moore (2004, 2006) work on *partial* contracts.

Moore (1999) have argued that writing comprehensively complete contracts is "prohibitively expensive". There seems little disagreement that bounded rationality of some sort may well provide a motivation for incomplete contracts. However, there is no widely agreed upon way of modelling this bounded rationality. We could interpret our framework by distinguishing between the costs of forming contracts of some degree of completeness, and the costs of that degree of contract completeness in terms of a *residual* costly exchange. So, although it may be *possible* to form comprehensively complete contracts, doing so may involve some exorbitant cost.

The standard view of contractual incompleteness is one in which a contract is incomplete if there are certain choice variables that cannot be contracted on. This is termed a "limitation in contractual language" by Bolton and Dewatripont (2005, pp.489–90). We believe that the concept of contractual incompleteness in terms of limited contractual language is overly restrictive. What we wish to consider, however, is some choice regarding the *detail* of a contract. In the Hart and Moore (2006) set-up, the "broad outlines of ex post trade are contractible, but finer points are not." A contract that specifies actions for all states is not unique. Take the state of nature space Ω in the framework above. It is fully observed ex post and so can be contracted over ex ante. Suppose two contracts: The first lists a single action to be carried out regardless of the realised state $\omega \in \Omega$; the second lists a specific action for each specific $\omega \in \Omega$. Both contracts are, in a strict sense, contingent on all choice variables. But it is our view that we can call the former contract more incomplete than the latter. Suppose now that contracting costs bias the contracts that agents write away from the second form of contract. This sort of relation between complexity costs and incomplete contracts was identified in Anderlini and Felli (1999).

Ex ante contracting and ex post exchange

The general equilibrium framework laid-out above can be interpreted in these incomplete contracting terms. An ex ante contract which defines broad outlines of trade is one in which *ranges of actions* are defined over *ranges of possible outcomes.* Ex post, there are costs to fill in the detail of the finer points. Or the ex post exchange cost reflects some expected loss from having not written a complete contract. For Hart and Moore (2006) the costs from partial incompleteness can take the form of a deadweight loss.

So in a world in which the formation of contracts is a costly process, there can be both ex ante contracts to commit to exchange and ex post contracts (or renegotiations) to specify the details of exchange. Ex ante, there is perfect competition in the provision of, to some degree incomplete, contracts concerning future exchange; ex post there are further costs within a monopolistic bilateral exchange. In a sense, our exchange cost is then also a pure deadweight loss.

A definition of a complete contract follows:

Definition 4.3 *A* ***complete contract*** *is an ex ante arrangement that yields no ex post exchange cost.*

This is our working definition. We provide below a number of real-world examples that can fit into this more general concept of contractual incompleteness.

Some Examples

Here we provide some simple examples of the contracting process. Agents in our model always wish to commit to diversifying ex ante by writing an initial contract; the further question is whether agents choose to formalise all potential ex post interactions in the ex ante contract or whether the ex ante contract is optimally incomplete with positive exchange costs left to be incurred ex post. These examples go through situations in which these types of arrangements do exist in reality, though not all are examples of diversification and exchange *per se*.

i) Bookkeeping. Suppose that the nature of any realised investment output is sufficiently complex such that communication between agents in the conduct of exchange is not easy. In such circumstances, agents can carry out bookkeeping since this enables information to flow more smoothly. The bookkeeping requires the expenditure of resources, let's say in fees to an accountant. For each bilateral exchange, accounts must be drawn up between the exchanging parties. Given the large number of possible contingencies, it is too expensive to write an ex ante contract that specifies accounts for transactions between each agent for every state of nature. As such, the ex ante contract commits each agent to diversifying but is not comprehensively complete. In defining the market, however, resources can also be allocated to constructing a market-specific standard for the reporting of accounts; this is essentially a code of conduct for the market. The code of conduct reduces the cost of writing accounts for each ex post exchange. Additional resources can be allocated to a public institution that upholds, for example, corporate governance standards or minimum requirements on accounting standards. After the realisation of outputs, exchange takes place with additional costs that reflect the completeness of the ex ante contract.

ii) Housing Market. A number of specialised private agencies, in addition to public institutions, are involved in housing transactions. There are legislated requirements covering the process of house-buying, and in addition local groups of agencies can come together to form specific standards to suit local requirements. Each of these arrangements require resources to construct ex ante and enforce ex post a framework for the purchase of housing. Consequently, for any given pair of agents, there is then a ready-made template through which the transaction can take place. Even using this template requires the additional expenditure of resources; but this additional expenditure, the exchange cost, is diminished by the specific and general arrangements established ex ante. In Scotland, for example, the highest bidder in a one-shot game obtains the property under sale. There is an informal agreement that submitting a bid is a commitment to purchase in the event of having entered the winning bid. Of course, this is open to abuse from individuals that, after submitting a winning bid, find a better deal and renege. A local solution emerged in Edinburgh: A group of property lawyers formed a specific contract that committed each of their clients to not reneging. Here, a partially incomplete ex ante contract was made more complete by private and excludable action.

iii) Sporting Rules and The Law. A Premiership football player was recently alleged to have bitten ("just a nibble") one of his opponents. The Football Association was not sure what to do, because there was no specific law about biting. It was reported that the behaviour came under violent misconduct, for which a range of penalties could be imposed, and that the FA would clarify the law for the future. (In the end they did nothing.) In a sense, the FA laws were, ex ante, partly incomplete here. Here, an unwritten contingency is considered ex post in relation to the broad outlines of the ex ante laws. This example is not uncommon, and nor even is it specific to sport: Law often defines only the broad outlines within which specific behaviours are included. When some specific and unprecedented something happens, it goes to court, judgements are handed down, the law evolves. In the terms of our setup, an ex ante contract does not necessarily define specific contingencies and specific actions but ranges of contingencies and ranges of actions. Our range of contingency would be "violent misconduct", the unspecified detail includes such things as "just a nibble". So the decision is over how detailed your ex ante contract is; how many broad categories there are in your ex ante framework of law and hence how much work you have to do ex post to fill in the additional detail.

iv) Infrastructure. We do not wish to confuse the exchange costs that cause coalitions to form with exchange (i.e., *freight*) costs ameliorated by better infrastructure. Rather, the latter are, largely, internal to the firm and can be thought of as reflected in the prices of its goods. The former can be the product of contracting costs. Infrastructures are a good example of where the contracting conditions come into sharper relief, and this is why we spent time on them in Chapter 3. We saw that local, excludable arrangements can emerge to finance (and benefit from the return on) investment in infrastructure. Public works can plan, build and maintain an infrastructure for the facilitation of private exchange but local arrangements such as turnpike trusts, canal and railway companies can further employ local resources to improve the standard of infrastructure. The local financial coalitions act to form the behavioural rules of the coalition, and we argued that this was, relative to full public planning, an efficient outcome. Moreover, the ease with which local coalitions can form is related to the provision of pure public goods such as the quality of the legislative environment. In other words, a combination of local and public arrangements facilitate the contracting of local agents into a financial coalition.

Relation to Other Work

Our notion of incompleteness is close in spirit to Hart and Moore (1999, p. 116) who suggest that "the degree of partial incompleteness depends on the parties' ability to describe the nature of trade." However, we also consider, in addition to an agent's *ability* to describe the nature of trade, her *willingness* to do so, given the costs of contracting. The greater the complexity of the private contract, the more costly it is to write; agents *choose* the degree of complexity given technologies and resources. From the perspective of our framework, the contractual incompleteness is manifested in contracts that favour broad outlines over finer detail, rather than one which leaves out specified actions for behaviour under every contingency. The general-specific distinction then follows in the context of ex ante contracts: We have both private and social contracts.

By looking at the macroeconomic implications of incomplete contracting, and since this perspective allows us an understanding of the endogenous existence of institutions, this framework might also be seen as an attempt to bridge the gap between contract theory and institutional economics. Boyd and Prescott (1985, fn.2) also reflect this: "Our intermediary-coalitions could also be viewed as a nexus of contracts... or as an arrangement to economise on transactions costs..." In looking at the implications of this framework from a macroeconomic perspective, it can also have implications for the recent work on the underpinnings of political economy as in Acemoglu and Johnson (2006). We come onto some of these issues in the next Chapter.

4.7.2 Further Work

Given the large range of literatures whose theoretical results are based on the existence of frictions to exchange, we considered a stylised method of introducing the concept of an endogenous exchange cost. We have made an initial foray into establishing a framework within which we can model exchange costs in a more acceptable manner. A great deal is left to future work, however: A full exposition of political economy issues in the heterogeneous environment; an investigation of optimal taxation and regulation; the relation to a theory of institutions; and, not least, a consideration of the dynamic elements in the context of, for example, a theory of finance and growth. Indeed, we noted in Chapter 1 the finding of Demirgüç-Kunt and Levine (2001) that institutions that affect contracts and property rights were found to be central in an empirical sense. We turn to consider in more detail some of these extensions in Chapter 5.

Chapter 5

Towards the Microfoundations of Finance and Growth

We developed in Chapter 4 a framework model that can account for the existence, extent and consequences of costly exchange in general equilibrium. The framework touched upon a wide range of subjects: The size of markets, financial intermediation, the partial incompleteness of contracts and political economy outcomes. We have made each of these aspects endogenous to the distribution of resources, technologies and preferences among rational economic agents.

But what relation does this framework and its implications have to our central thesis on finance and growth? How does it relate to the critique of Chapters 1 and 2? Further, in which ways does it lead towards a model capable of accounting for the rich story of finance and growth described in Chapter 3? In other words, how might it be extended to develop more fully the finance and growth implications?

In this short Chapter we intend to tie-together some of our findings. We can then think about directions for future work, i.e., what we need to do before we can speak of *the* microfoundations of finance and growth with any confidence. Section 5.1 goes over the central messages coming from Chapters 1–3. Section 5.2 then folds the framework of Chapter 4 into the wider context of our perspective on finance and growth, as well as sketching directions for future research.

5.1 Finance and Growth Orthodoxy

The orthodox treatment of the connection between the extent of financial development and the level of economic growth has neglected a number of important aspects. We placed stress in the critique of Chapter 1 on the disconnection between the empirical and theoretical literatures. Chapter 2 then firmed-up some of the more intuitive criticisms by developing a representative, parsimonious finance and growth theory that could be calibrated to some historical data. Chapter 3 took a step away from the standard approach to finance and growth and instead took the historical literature as a guide to the development of a more realistic theory of finance and growth.

In the context of the early stages of industrial development, we observed strongly spatially concentrated elements in the finance of investments where fixed costs were high. We needed to consider why these coalitions emerged in this way, and argued that it had to do with conditions of information asymmetry and, by consequence, the costly formation of financial contracts. We also observed a dynamic aspect. These spatially concentrated coalitions changed over the course of the industrial revolution. In the early stages, small coalitions were highly spatially concentrated, and formed the basis of almost disjointed regional components of a larger national economy. As the economy took hold, coalitions became larger, more centralised, and more specialised.

Of course, this story is one stylised by a macroeconomist seeking to fit a nice model to intuitive historical trends. But the historical evidence consistently supports the general story. There were disaggregated patterns to financial coalitions and their nature *did* change over the course of industrial development.

This means that there is something missing both from empirical analyses of the aggregate and from theoretical conceptions of static efficiency. We sought to reconcile our more qualified interpretation of the macroeconomics of finance and growth with the microeconomic phenomena that might underly them by considering the role that costly exchange can play. Townsend (1978) established that financial coalitions of finite-size could be part of an equilibrium in economies with costly bilateral exchange. These exchange costs are, in that work, taken to be fixed and exogenous. The omission, that has carried through to recent work, is some foundation for these exchange costs.

That omission is, from the perspective of this thesis, a stark one. Given that exchange costs underlie financial structures, and given also that we observe dramatically changing financial structures over the course of an industrial revolution, it is surely not sufficient for a finance and growth story to hold exchange costs to be exogenous and fixed. Our view must be, then, that exchange costs are *not* a fundamental.

The purpose of Chapter 4 was to develop a general equilibrium theory that could capture a number of these stylised elements in the process of making exchange costs endogenous. The next Section goes through some of the implications of this model in the context of our study of finance and growth.

5.2 The Microfoundations of Finance and Growth

What steps in research will be required before we can speak of the microfoundations of finance and growth with confidence? First, a dynamic version of the model in Chapter 4 will look at the implications of endogenous exchange costs in the context of economic growth: Can we capture the dynamic and disaggregated elements of the finance and growth story? Second, we ought to specify more fully the political economy implications, specifically in the context of incomplete contracts, institutions and political outcomes. Each of these research avenues will be major undertakings, but we sketch their structure and potential implications below.

5.2.1 Dynamic Spatial Coalitions and Growth

The model of Chapter 4 can, in some regards, be seen in the light of the story of finance and growth developed through Chapters 1–3.

Consider a model of endogenous exchange costs like that in Chapter 4 but in which agents are spatially separated. We could think, for example, of a countable infinity of agents equidistantly ordered on the real line. Suppose, further, that the costs of forming an arrangement is increasing in the distance between agents. This reflects what we observed in Chapter 3: The costs of excessive informational asymmetry can be mitigated by trading only with those within a limited proximity. In some ways, an agent 'knows' and 'trusts' her neighbour better than someone from far away. This has obvious interpretations in terms of contracts. So, the costs of forming optimal arrangements is increasing in the sum of the distances involved in the intermediated exchange among agents of a coalition. This happens for two reasons: The cost of arrangements of a given degree arrangement effectiveness is higher; and, the arrangement will, optimally, be less effective. These are implications of the heterogeneous extension to the exchange cost model. In this setting, then, we would observe finite-sized coalitions that are *also* spatially concentrated. Further, economies with a higher endowment will observe larger coalitions, both in terms of the number of agents and in terms of their greater spatial dispersion.

So risk averse agents do still have the incentive to contract into a diversification strategy, but now they compose strategies that include the spatial decision in addition to everything else. We can interpret the model in Chapter 3 in some limited directions. Let us think of a simple exogenous growth story in which there are no dynamic interactions, i.e, suppose that our cross-sectional results from Chapter 4 can broadly reflect a dynamic story. Townsend (1983a), Boyd and Prescott (1985) and Prescott and Boyd (1987) show ways of introducing some forms of exogenous growth into simple general equilibrium models related to that presented here.

We have seen from both the analytical and numerical results coming out of our exchange cost model that exchange costs, market size and optimal institutions vary as the endowment increases. In the early stages of growth, exchange costs are high, and financial coalitions are small and spatially concentrated. Here we capture, in some senses, the high levels of trust in the 'village economies' intimated by Kiyotaki and Moore (2006). As the economy develops, the level of wealth grows and so the extent of non-production costs falls. As this happens, agents can, at intervals, diversify into coalitions of both greater number and wider geography. As they do so, the level of financial depth in the economy increases. Within our framework model of endogenous exchange costs, extended to capture spatial effects, we are thus able to capture a number of the observed finance and growth phenomena.

But further to demonstrating that a static equilibrium with spatial elements exists, formalising this extension in a dynamic setting is not necessarily straightforward. There are interesting games being played in such a dynamic setup. To note a few: The ordering of development and finance; the choice along the pecking order of finance; the linking together of small coalitions/infrastructures into larger ones; and, the movement towards centralised (and specialised) financial markets. We can look toward a large literature on sequential games, the optimal contract literature engendered by Townsend (1979), as well as work such as Prescott and Vischer (1977) on the location of firms in general equilibrium, for clues as to how to proceed in this regard.

This is still a story of exogenous growth. We might support growth within an endogenous growth framework in the manner of the new growth literature following Lucas (1988), Romer (1990) and Rebelo (1991). But we should not view an exogenous growth model as inferior by definition. Parente (2001) provides a powerful critique of the applied success of the endogenous growth literature. The important work of Parente and Prescott (2000) extends this critique by forming an alternative paradigm for our understanding of growth and convergence. Here, barriers to growth are caused by firms with monopoly power who have an interest in maintaining inferior technologies. The microeconomic foundations of the existence and persistence of these barriers are lacking, however. Parente and Prescott largely put them down to insider power, vested interests and a lack of accountability in political institutions. In other words, inappropriate institutions supported by an inequitable distribution of power among agents mean that technological diffusion does not occur as it would under perfect competition. The coalitions that form in our exchange cost economy are, in some regards, reminiscent of imperfectly competitive markets. Further, our analysis of political outcomes that reflect underlying power structures might also be interpreted in these terms. A major objective of the work coming out of this thesis is to map the aggregate output in a dynamic version of our exchange cost economy into a neoclassical production function, and link this with a theory of TFP in the manner of Parente and Prescott (2000).

5.2.2 Contracts, Institutions and Growth

If we believe that exchange costs can be the product of the partial incompleteness of contracts, as argued in Chapter 4, then institutions, contracts and economic growth can be related via the costliness of exchange. This connection has long been hinted-at by, among others, Boyd and Prescott (1985), Dixit (1996) and Williamson (2000). The incompleteness of contracts is said to be mitigated by the existence of institutions. Further, institutions are often supported by the invocation of transaction costs. If we can rigorously define a connection between incomplete contracts and institutions via the costliness of exchange then we will have moved a long way towards articulating, in an economically meaningful way, the interaction between private contracts and institutions (or social contracts). With an extension to a growth model, this would also provide a framework for the study of appropriate institutions in the context of economic development. As Bolton and Dewatripont (2005, p.489) point out,

> The analysis of institutions is clearly of fundamental importance, and an incomplete-contracting approach offers a vehicle to explore these issues systematically. The fourth stage of research in contract theory is thus a natural development, which should eventually produce an economic theory of institutions as rich as the theory of incentives developed in the last three decades.

Of course, there is a great deal to be done in making these interactions rigorous, and even more to make them convincing. Primarily, the link between incomplete contracts and exchange costs needs to be formalised at the microeconomic level. The recent work on partial contracts of Hart and Moore (2004, 2006) suggests that this might be possible. But how precisely does incomplete contracting relate to optimal institutions; what do we mean by institutions; what are the optimal arrangements at a micro level; what policies can be implemented to most efficiently combat contract incompleteness; how does the interaction between contracts and institutions change as the economy moves through a period of industrialisation?

It has been suggested that we might incorporate Shapley allocations into the individual payoffs to individuals *within* a given market, something which can lead to a form of contractual incompleteness. The consequent losses could potentially be mapped into the exchange costs introduced in Chapter 4. Doing so, we would be able to think of contractual incompleteness on a more fundamental level and at the same time retain the larger implications of our general equilibrium framework.

Further to these more fundamental questions, there are interesting technical problems to resolve. In invoking the Shapley value to support power-based political outcomes on top of the core-equilibrium concept that captures market equilibria we are mixing solution concepts in a single model. If we could simultaneously model market and political economy decisions with a single solution concept then the implications of the model could be trusted more. This may require us to develop a continuum-economy version of our exchange cost economy.

We leave all of these extensions to future work.

Conclusions

Chapter 5 has covered much of what would normally comprise a section of concluding remarks so I will make this brief.

This thesis has offered a critique of the standard approach to the connection between financial development and economic growth in the context of economies going through industrial revolution. We found approaches to this classic question to be evidence of the dangers of neglecting the *interface* between theory and evidence. By looking at the question from an historical perspective we were able quickly to recognise the inadequacies in current thinking. In doing so, we also managed to draw-out clear implications for the actual disaggregated and dynamic elements of a more appropriate story of finance and growth. By matching these implications to a general equilibrium theory of exchange costs, we took a few steps towards a proper understanding of the microfoundations of finance and growth.

In this sense, Landes (1969) was right: History can indeed be the laboratory of the social sciences. It is hoped that with the conclusion of Shea's *Handbook of Corporate Finance*, we can further improve our understanding of a wide range of contemporary questions in economics. We will have hard data that is disaggregated to the point of biography on the existence, constitution and development of actual financial coalitions right across the most important period of the economic development of the UK. The potential for developments coming out of this dataset is very great.

We have been through a number of potential extensions of this work. It seems that a great deal of exciting research remains to be done. The challenge of understanding financial development and economic takeoff at a microeconomic level, via the formation of contracts and the setting of the institutional environment, is a potentially very fruitful one.

Bibliography

Acemoglu, D. and Johnson, S. (2006). *Economic Origins of Democracy and Dictatorship*. Cambridge University Press.

Acemoglu, D. and Zillibotti, F. (1997). 'Was Prometheus Unbound By Chance? Risk, Diversification and Growth'. *Journal of Political Economy*, 105(5):709–51.

Aghion, P. and Bolton, P. (1997). 'A Theory of Trickle-Down Growth and Development'. *Review of Economic Studies*, 64(1):151–72.

Aghion, P. and Howitt, P. (1992). 'A Model of Growth Through Creative Destruction'. *Econometrica*, 60(2):323–51.

Aghion, P. and Howitt, P. (1998). *Endogenous Growth Theory*. The MIT Press.

Aghion, P., Howitt, P., and Mayer-Foulkes, D. (2005). 'The Effect of Financial Development on Convergence: Theory and Evidence'. *The Quarterly Journal of Economics*, 120(1). 173-222.

Alesina, A., Spolaore, E., and Wacziarg, R. (2000). 'Economic Integration and Political Disintegration'. *American Economic Review*, 90(5):1276–96.

Anderlini, L. and Felli, L. (1999). 'Incomplete Contracts and Complexity Costs'. *Theory and Decision*, 46(1):23–50.

Auerbach, P. and Siddiki, J. U. (2004). 'Financial Liberalisation and Economic Development: An Assessment'. *Journal of Economic Surveys*, 18(3):231–65.

Aumann, R. J. (1975). 'Values of Markets with a Continuum of Traders'. *Econometrica*, 43(4):611–46.

Aumann, R. J. and Kurz, M. (1977). 'Power and Taxes'. *Econometrica*, 45(5):1137–61.

Bagehot, W. (1873). *Lombard Street: A Description of the Money Market*. Richard D. Irwin. Available at www.econlib.org.

Bairoch, P. (1982). 'International industrialization levels from 1750 to 1980'. *Journal of European Economic History*, 11(1):269–334.

Barro, R. J. and Sala-i-Martin, X. (2004). *Economic Growth*. MIT Press, 2nd edition.

Beck, T. and de la Torre, A. (2006). 'The Basic Analytics of Access to Financial Services'. Policy Research Working Paper Number 4026. World Bank.

Beck, T., Demirgüç-Kunt, A., Laeven, L., and Levine, R. (2005). 'Finance, Firm Size, and Growth'. Policy Research Working Paper Series. The World Bank. No. 3485.

Beck, T., Demirgüç-Kunt, A., and Levine, R. (1999). 'A New Database on Financial Development and Structure'. Policy Research Working Paper Number 2146. World Bank. The dataset is updated at the World Bank Finance Research website, http://econ.worldbank.org/.

Beck, T., Demirgüç-Kunt, A., and Levine, R. (2004). 'Finance, Inequality, and Poverty: Cross-Country Evidence'. NBER Working Paper Number 10979.

Beck, T. and Levine, R. (2005). 'Legal Institutions and Financial Development', in Ménard, C. and Shirley, M. M. (eds.). *Handbook of New Institutional Economics*. Springer.

Beck, T., Levine, R., and Loayza, N. (2000). 'Finance and the Sources of Growth'. *Journal of Financial Economics*, 58(1–2):261–300.

Becker, G. S. (1983). 'A Theory of Competition Amond Pressure Groups for Political Influence'. *Quarterly Journal of Economics*, 98(3):372–400.

Bencivenga, V. R. and Smith, B. D. (1991). 'Financial Intermediation and Endogenous Growth'. *The Review of Economic Studies*, 58(2):195–209.

Besley, T. and Coate, S. (1997). 'An Economic Model of Representative Democracy'. *Quarterly Journal of Economics*, 112(1):85–114.

Blackburn, K., Bose, N., and Capasso, S. (2005). 'Financial Development, Financing Choice and Economic Growth'. *Review of Development Economics*, 9(2):135–49.

Blackburn, K. and Hung, V. T. Y. (1998). 'A Theory of Growth, Financial Development and Trade'. *Economica*, 65(1):107–24.

Bogart, D. (2005). 'Turnpike Trusts and the Transportation Revolution in Eighteenth-Century England'. *Explorations in Economic History*, 42:479–508.

Bogart, D. (2006a). 'Neighbors, Networks, and the Development of Transport Systems: Explaining the Diffusion of Turnpike Trusts in Eighteenth-Century England'. Forthcoming in *Journal of Urban Economics*.

Bogart, D. (2006b). 'Private Ownership and the Devlopment of Transport Systems: Cross-Country Evidence from the Diffusion of Railroads in the 19th Century'. University of California-Irvine, mimeo.

Bogart, D. and Richardson, G. (2006). 'Parliament, Property Rights, and Public Goods in England: 1600 to 1815'. University of California-Irvine, mimeo.

Bolton, P. and Dewatripont, M. (2005). *Contract Theory*. MIT Press.

Bordo, M. S. and Rousseau, P. L. (2006). 'Legal-Political Factors and the Historical Evolution of the Finance-Growth Link'. NBER Working Paper No. 12035.

Bose, N. and Cothren, R. (1996). 'Equilibrium Loan Contracts and Endogenous Growth in the Presence of Asymmetric Information'. *Journal of Monetary Economics*, 38(3):363–76.

Boyd, J. H. and Prescott, E. C. (1985). 'Financial Intermediary-Coalitions'. *Journal of Economic Theory*, 38(2):211–32.

Broadbridge, S. (1955). 'The Early Capital Market: The Lancashire and Yorkshire Railway'. *The Economic History Review*, 8(2):200–12.

Broecker, T. (1990). 'Credit-Worthiness Tests and Interbank Competition'. *Econometrica*, 58(2):429–52.

Buchanan, B. J. (1986). 'The Evolution of the English Turnpike Trusts: Lessons from a Case Study'. *The Economic History Review*, 39(2):223–43.

Calderón, C. and Servén, L. (2004). 'The Effects of Infrastructure Development on Growth and Income Distribution'. Policy Research Working Paper Number 3400. World Bank.

Cameron, R., editor (1967). *Banking in the Early Stages of Industrialization*. Oxford University Press.

Cameron, R., editor (1972). *Banking and Economic Development*. Oxford University Press.

Capasso, S. (2004). 'Financial Markets, Development and Growth: Tales of Information Asymmetry'. *Journal of Economic Surveys*, 18(3):267–92.

Coase, R. H. (1992). 'The Institutional Structure of Production'. *American Economic Review*, 82(4):713–9.

Colombo, E. and Driffill, J., editors (2003). *The Role of Financial Markets in the Transition Process*. Physica-Verlag.

Cottrell, P. L. (1980). *Industrial Finance 1830-1914: The finance and organization of English manufacturing industry*. Methuen.

Crafts, N. F. R. (1995a). 'Macroinventions, Economic Growth, and 'Industrial Revolution' in Britain and France'. *The Economic History Review*, 48(3):591–8.

Crafts, N. F. R. (1995b). 'Exogenous or Endogenous Growth? The Industrial Revolution Reconsidered'. *The Journal of Economic History*, 55(4):745–72.

Crafts, N. F. R. and Harley, C. K. (1992). 'Output Growth and the British Industrial Revolution: A Restatement of the Crafts-Harley View'. *The Economic History Review*, 45(4):703–30.

Demirgüç-Kunt, A. (2006). 'Finance and Economic Development: Policy Choices for Developing Countries'. Policy Research Working Paper Number 3955. World Bank.

Demirgüç-Kunt, A. and Levine, R., editors (2001). *Financial Structure and Economic Growth: A Cross-Country Comparison of Banks, Markets and Development*. MIT Press.

Dixit, A. (1996). *The Making of Economic Policy: A Transaction-Cost Perspective*. MIT Press.

Driffill, J. (2003). 'Growth and Finance'. *The Manchester School*, 71(4):363–80.

Fuente, de la, A. and Marín, J. M. (1996). 'Innovation, Bank Monitoring, and Endogenous Financial Development'. *Journal of Monetary Economics*, 38(1):269–301.

Gehrig, T. and Stenbacka, R. (2003). 'Venture Cycles: Theory and Evidence', in Kanniaianen, V. and Keuschnigg, C. (eds.). *Venture Capital, Entrepreneurship, and Public Policy*. MIT Press.

Goldsmith, R. (1969). *Financial Structure and Development*. Yale University Press.

Greenwood, J. and Jovanovic, B. (1990). 'Financial Development, Growth and the Distribution of Income'. *Journal of Political Economy*, 98(5):1076–1107.

Greenwood, J. and Smith, B. D. (1997). 'Financial Markets in Development, and the Development of Financial Markets'. *Journal of Economic Dynamics and Control*, 21(1):145–81.

Gregorio, de, J. and Kim, S. J. (1998). 'Credit Markets with Differences in Abilities: education, distribution and growth'. *International Economic Review*, 41(3):579–607.

Grossman, G. and Helpman, E. (1991). 'Quality Ladders in the Theory of Economic Growth'. *Review of Economic Studies*, 58(1):43–61.

Grossman, S. J. and Stiglitz, J. E. (1980). 'On the Impossibility of Informationally Efficient Markets'. *American Economic Review*, 70(3):393–408.

Harris, R. (2000). *Industrializing English Law: Entrepreneurship and Business Organization, 1720-1844*. Cambridge University Press.

Hart, O. and Moore, J. H. (1990). 'Property Rights and the Nature of the Firm'. *Journal of Political Economy*, 98(6):1119–58.

Hart, O. and Moore, J. H. (1999). 'Foundations of Incomplete Contracts'. *Review of Economic Studies*, 66(1):115–38.

Hart, O. and Moore, J. H. (2004). 'Agreeing Now to Agree Later: Contracts that Rule Out but do not Rule In'. Harvard University, mimeo.

Hart, O. and Moore, J. H. (2006). 'Contracts as Reference Points'. Harvard University, mimeo.

Heston, A., Summers, R., and Aten, B. (2002). Penn World Table Version 6.1. Center for International Comparisons at the University of Pennsylvania (CICUP).

Howitt, P. and Aghion, P. (1998). 'Capital Accumulation and Innovation as Complementary Factors in Long-Run Growth'. *Journal of Economic Growth*, 3(2):111–30.

Hudson, P. (2002). *The Genesis of Industrial Capital: A Study of the West Riding Wool Textile Industry*. Cambridge University Press.

Hulten, C. R. (1996). 'Infrastructure Capital and Economic Growth: How Well You Use It May Well Be More Important Than How Much You Have'. NBER Working Paper No. 5847.

Hunt, B. C. (1935a). 'The Joint-Stock Company in England, 1800-1825'. *The Journal of Political Economy*, 43(1):1–33.

Hunt, B. C. (1935b). 'The Joint-Stock Company in England, 1830-1844'. *The Journal of Political Economy*, 43(3):331–64.

Jayaratne, J. and Strahan, P. E. (1996). 'The Finance-Growth Nexus: Evidence from Bank Branch Deregulation'. *The Quarterly Journal of Economics*, 111(3):639–70.

Judd, K. L. (1998). *Numerical Methods in Economics*. MIT Press.

Kenwood, A. G. (1965). 'Railway Investment in Britain, 1825–1875'. *Economica*, 32(127):313–22.

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

King, R. G. and Levine, R. (1993b). 'Finance, Entrepreneurship, and Growth: theory and evidence'. *Journal of Monetary Economics*, 32(3):513–42.

Kiyotaki, N. and Moore, J. H. (2006). 'Financial Deepening'. *American Economic Review: Papers and Proceedings.*

Krugman, P. R. and Venables, A. J. (1995). 'Globalization and the Inequality of Nations'. *The Quarterly Journal of Economics*, 110(4):857–80.

Kuznets, S. (1955). 'Economic Growth and Income Inequality'. *American Economic Review*, 45(1):1–28.

Laffont, J. J. and Martimort, D. (2002). *The Theory of Incentives.* Princeton University Press.

Landes, D. S. (1969). *The Unbound Prometheus: Technical Change and Industrial Development in Western Europe from 1750 to Present.* Cambridge University Press.

Landes, D. S. (1994). 'What Room for Accident in History? Explaining Big Changes by Small Events'. *The Economic History Review*, 47(4):637–56.

Landes, D. S. (1995). 'Some Further Thoughts on Accident in History: A Reply to Professor Crafts'. *The Economic History Review*, 48(3):599–601.

Landes, D. S. (1998). *The Wealth and Poverty of Nations.* Abacus.

Levine, R. (2004). 'The Corporate Governance of Banks: A Concise Discussion of Concepts and Evidence'. World Bank Policy Research Working Paper Number 3404.

Levine, R. (2005). 'Finance and Growth: Theory, Mechanisms and Evidence', in Aghion, P. and Durlauf, S.N. (eds.). *Handbook of Economic Growth.* Elsevier.

Levine, R., Loayza, N., and Beck, T. (2000). 'Financial Intermediation and Growth: Causality and Causes'. *Journal of Monetary Economics*, 46(1):31–77.

Levine, R. and Zervos, S. (1998). 'Stock Markets, Banks, and Economic Growth'. *American Economic Review*, 88(3):537–88.

Lévy-Leboyer, M. (1978). 'Capital Investment and Economic Growth in France, 1820–1930', in Mathias, P. and Postan, M. M. (eds.). *The Cambridge History of Europe. Volume VII(1).* Cambridge University Press.

Lindbeck, A. (1985). 'Redistribution Policy and the Expansionof the Public Sector'. *Journal of Public Economics*, 28(1):309–28.

Lucas, Jr., R. E. (1988). 'On the Mechanics of Economic Development'. *Journal of Monetary Economics*, 22(1):3–42.

Manning, M. J. (2003). 'Finance Causes Growth: Can We Be So Sure?'. *Contributions to Macroeconomics*, 3(1):Art.12.

Mathias, P. and Postan, M. M., editors (1978). *The Industrial Economies: Capital, Labour, and Enterprise*, volume VII. Cambridge University Press.

McDermott, J. (2002). 'Development Dynamics: Economic Integration and the Demographic Transition'. *Journal of Economic Growth*, 7:371–409.

McKinnon, R. I. (1973). *Money and Capital in Economic Development.* The Brookings Institution.

Milward, A. S. and Saul, S. B. (1973). *The Economic Development of Europe, 1780-1870.* George Allen & Unwin Ltd.

Mokyr, J. (1990). *The Lever of Riches: Technological Creativity and Technological Progress.* Oxford University Press.

Morales, M. F. (2003). 'Financial Intermediation in a Model of Growth Through Creative Destruction'. *Macroeconomic Dynamics*, 7(3):363–93.

Murphy, K., Shleifer, A., and Vishny (1989). 'Industrialization and the Big Push'. *Journal of Political Economy*, 97(5):1003–26.

Myers, S. C. (1984). 'The Capital Structure Puzzle'. *Journal of Finance*, 39(3):575–92.

Nolan, C. and Trew, A. W. (Forthcoming). 'Endogenous Exchange Costs in General Equilibrium with Heterogeneous Agents'. Centre for Dynamic Macroeconomic Analysis Working Paper.

North, D. C. (1990). *Institutions, Institutional Change and Economic Performance.* Cambridge University Press.

O'Brien, P. (2006). 'Provincializing the First Industrial Revolution'. Working Papers of the Global Economic History Network (GEHN) No. 17/06.

Obstfeld, M. and Rogoff, K. (2000). *NBER Macroeconomics Annual 2000*, chapter 'The Six Major Puzzles in International Macroeconomics: Is there a Common Cause?', pages 339–90. MIT Press.

OECD (2005). *OECD in Figures.* Organisation for Economic Co-operation and Development. Available at www.oecd.org.

Pagano, M. (1993). 'Financial Markets and Growth: An Overview'. *European Economic Review*, 37(2–3):613–22.

Papoulis, A. (1984). *Probability, Random Variables, and Stochastic Processes, 2nd ed.* McGraw-Hill.

Parente, S. L. (2001). 'The Failure of Endogenous Growth'. *Knowledge, Technology and Policy*, 13:49–58.

Parente, S. L. and Prescott, E. C. (2000). *Barriers to Riches*. MIT Press.

Parente, S. L. and Prescott, E. C. (2005). 'A Unified Theory of the Evolution of International Income Levels', in Aghion, P. and Durlauf, S.N. (eds.). *Handbook of Economic Growth*. Elsevier.

Pollins, H. (1954). 'The Marketing of Railway Shares in the First Half of the Nineteenth Century'. *The Economic History Review*, 7(2):230–9.

Prescott, E. C. and Boyd, J. H. (1987). 'Dynamic Coalitions: Engines of Growth'. *American Economic Review: Papers and Proceedings*, 77(2):63–7.

Prescott, E. C. and Vischer, M. (1977). 'Sequential Location among Firms with Foresight'. *Bell Journal of Economics*, 8(2):378–93.

Rajan, R. and Zingales, L. (1998). 'Financial Dependence and Growth'. *American Economic Review*, 88(3):559–86.

Rajan, R. and Zingales, L. (2004). *Saving Capitalism from the Capitalists*. Princeton University Press.

Rebelo, S. T. (1991). 'Long-Run Policy Analysis and Long-Run Growth'. *Journal of Political Economy*, 99(3):500–21.

Rivera-Batiz, L. A. and Romer, P. (1991). 'Economic Integration and Endogenous Growth'. *Quarterly Journal of Economics*, 106(2):531–55.

Robinson, J. (1952). *The Rate of Interest and Other Essays*, chapter 'The Generalization of The General Theory'. MacMillan.

Rodrik, D. (1996). 'Understanding Economic Policy Reform'. *Journal of Economic Literature*, 34(1):9–41.

Romer, P. (1986). 'Increasing Returns and Long-Run Growth'. *Journal of Political Economy*, 94(5):1002–37.

Romer, P. (1990). 'Endogenous Technological Change'. *Journal of Political Economy*, 98(5.2):S71–S102.

Rousseau, P. L. and Sylla, R. (2005). 'Emerging Financial Markets and Early US Growth'. *Explorations in Economic History*, 42(1):1–26.

Rousseau, P. L. and Wachtel, P. (1998). 'Financial Intermediation and Economic Performance: Historical Evidence from Five Industrialized Economies'. *Journal of Money, Credit and Banking*, 30(4):657–78.

Saint-Paul, G. (1992). 'Technological Choice, Financial Markets and Economic Development'. *European Economic Review*, 36(4):763–81.

Schumpeter, J. A. (1934). *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle*. Harvard University Press.

Shapley, L. S. (1953). 'A Value for n-Person Games', in Kuhn, H. W. and Tucker, A. W. (eds.). *Contributions to the Theory of Games: Vol. 2*. Princeton University Press.

Shea, G. S. (2007). 'The Handbook of 18th and Early 19th Century British Corporate Finance'. University of St. Andrews, mimeo.

Smith, Jr., C. O. S. (1990). 'The Longest Run: Public Engineers and Planning in France'. *The American Historical Review*, 95(3):657–92.

Sylla, R., Wilson, J. W., and Wright, R. E. (2006). 'Integration of Trans-Atlantic Capital Markets, 1790-1845'. *Review of Finance*, 42(1):1–26.

Townsend, R. M. (1978). 'Intermediation with Costly Bilateral Exchange'. *The Review of Economic Studies*, 55(3):417–25.

Townsend, R. M. (1979). 'Optimal Contracts and Competitive Markets with Costly State Verification'. *Journal of Economic Theory*, 21(2):265–93.

Townsend, R. M. (1983a). 'Financial Structure and Economic Activity'. *American Economic Review*, 73(5):895–911.

Townsend, R. M. (1983b). 'Theories of Intermediated Structures'. *Carnegie Rochester Conference Series on Public Policy*, 18(Spring):221–72.

Townsend, R. M. and Ueda, K. (2006). 'Financial Deepening, Inequality, and Growth: A Model-Based Quantitative Evaluation'. *Review of Economic Studies*, 73(1):251–293.

Trew, A. W. (2004). 'Endogenous Growth and Welfare with Asymmetric Information in Financial Intermediation'. Unpublished Masters thesis, University of Warwick.

Trew, A. W. (2006a). 'Finance and Growth: A Critical Survey'. Centre for Dynamic Macroeconomic Analysis, Working Paper No. 0507. Available at www.st-and.ac.uk/cdma/papers.html.

Trew, A. W. (2006b). 'Finance and Growth: A Critical Survey'. *Economic Record*, 82(259):481–90.

Trew, A. W. (2007). 'Endogenous Financial Development and Industrial Takeoff'. Centre for Dynamic Macroeconomic Analysis, Working Paper No. 0702. Available at www.st-and.ac.uk/cdma/papers.html.

Tsiddon, D. (1992). 'A Moral Hazard Trap to Growth'. *International Economic Review*, 33(2):299–321.

Turnbull, G. (1987). ‘Canals, Coal and Regional Growth during the Industrial Revolution’. *The Economic History Review*, 40(4):537–60.

Ward, J. R. (1974). *The Finance of Canal Building in Eighteenth Century England*. Oxford University Press.

Williamson, O. (2000). ‘The New Institutional Economics: Taking Stock, Looking Ahead’. *Journal of Economic Literature*, 38(September):595–613.

Wright, R. E. (2002). *The Wealth of Nations Rediscovered: Integration and Expansion in American Financial Markets*. Cambridge University Press.

Wright, R. E. (2004). ‘Testing the Finance-Led Growth Hypothesis: Early 19th Century America, Britain, and Canada’. Terry School of Business Workshop, University of Georgia.

Young, A. (1995). ‘The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience’. *Quarterly Journal of Economics*, 110(3):641–80.

Appendix A

MATLAB Code

Below is enough code to enable the reader to replicate the simulations carried out in the thesis. This code has been successfully compiled in MATLAB v.7(R14). Earlier versions may not accept the 'anonymous functions' of the form `U=@(W)W^gamma`. Otherwise, the code should be compatible with most earlier versions. Unintended line breaks are denoted "...". Hard copies of the code files, and of simulation output used to draw Figures, are available from the author by email at alex.trew@dunelm.org.uk.

A.1 Code for Chapter 2

The following is the file `morhazgrowth.m`. This calls the growth estimates in `gamma(:,1)` to find implied financial efficiency under a range of parameters for moral hazard. The data for these estimates are given in Appendix Table B.2.

■morhazgrowth.m

```
global kappa
global etaf
global etax
global alpha
global rho
global theta
global A
global h
global s
global varphi1
global varphi2

kappa=1
```

```
etaf=.1
etax=.01
s=10
varphi1=.4
varphi2=.2
alpha=.67
rho=.02
theta=5
A=30

setaf=zeros
betaf=zeros
for i=1:213
    gamma(i,2)=(1-(((theta*gamma(i,1)+rho)/A)^(1/(1-alpha)))*...
       (1/(1-alpha)));   %first find the associated t*
    t=gamma(i,2)
    j=0;
    i
    for s=0.5:0.5:8
       j=j+1;
       y = @(etaf)((((2/3)*(kappa*(1-t))+((((kappa*(1-t))^2)/9)+(2/3)*s*t)...
         ^(1/2))*(2*s*t-(((2/3)*(kappa*(1-t))+((((kappa*(1-t))^2)/9)+(2/3)...
         *s*t)^(1/2))-(kappa*(1-t)))^2))-2*s*(etax+(varphi2/(1-varphi1))...
         *etaf));
       if sign(y(-5))~=sign(y(5));
           etaf=fzero(y, [-50 50]);
           setaf(i,j,1)=i;
           setaf(i,j,2)=s;
           setaf(i,j,3)=(etaf);
           setaf(i,j,4)=1/etaf;
           betanought=kappa*(1-t);
           betastar=(2/3)*betanought+(((betanought^2)/9)+(2/3)*s*t)^(1/2);
           betaf(i,j,1)=i;
           betaf(i,j,2)=t;
           betaf(i,j,3)=s;
           betaf(i,j,4)=(etaf);
           betaf(i,j,5)=betanought;    %min effort level
           betaf(i,j,6)=betastar;      %optimal effort forced
           betaf(i,j,7)=((betastar-betanought)^2)/(2*s); %total monitoring spend
       end
    end
end

figure; surf (setaf(:,:,1), setaf(:,:,2), 1-setaf(:,:,3)); figure(gcf)
```

A.2 Code for Chapter 3

This code is relatively self-explanatory. The file `findev.m` solves for both regional and national growth rates using the function `infsolve.m`, then chooses the optimal growth rate and tracks prevailing levels of capital and infrastructure.

■findev.m

```
global theta
global A
global Ahat
global alpha
global psi
global Psi
global phi
global Phi
global rho
global nu
global omega

global i
global infdev

theta=5;
A=.2;
Ahat=.35;
alpha=2/3;
psi=.4;
Psi=.6;
phi=.4;
Phi=.6;
rho=.02;
nu=5;
omega=5;

infdev=zeros;

I0=20;  %initial condition for infrastructure
K0=I0/alpha;

reggrowthlim=(1/theta)*(A*((((1-alpha)*(1-psi))/(1+phi))^(1-alpha))-rho);
natgrowthlim=(1/theta)*(Ahat*((((1-alpha)*(1-Psi))/(1+Phi))^(1-alpha))-rho);

infdev(1,1)=I0;
infdev(1,6)=K0;
```

```
for i=2:200;
    Iguess=infdev(i-1,1)*1.05;
    if isreal(infsolve(Iguess))
        infdev(i,1)=fzero(@infsolve,Iguess);   %make a starting guess for fzero
        infdev(i,2)=(infdev(i,1)-infdev(i-1,1))/(infdev(i-1,1));
        infdev(i,3)=natgrowthlim;
        infdev(i,4)=reggrowthlim;
        infdev(i,5)=max(infdev(i,4),infdev(i,2));  %prevailing growth rate
        infdev(i,6)=infdev(i-1,6)*(1+infdev(i,5));
        infdev(i,1)=infdev(i-1,1)*(1+infdev(i,5));
    else
        infdev(i,2)=NaN;
        infdev(i,3)=natgrowthlim;
        infdev(i,4)=reggrowthlim;
        infdev(i,5)=infdev(i,4);
        infdev(i,6)=infdev(i-1,6)*(1+infdev(i,5));
        infdev(i,1)=infdev(i-1,1)*(1+infdev(i,5));
    end
end

infdev(1,:)=NaN;

figure;plot(infdev(:,2:5));
```

■**infsolve.m**

```
function y=infsolve(I)

global theta
global A
global Ahat
global alpha
global psi
global Psi
global phi
global Phi
global rho
global nu
global omega

global i
global infdev
```

```
Ih=infdev(i-1,1);

y=(1+Phi+((2*nu)/I))*(((Ahat^(-1))*(rho+theta*((I-Ih)/(Ih))))^...
  (1/(1-alpha)))-(1-alpha)*(1-Psi-((2*omega)/I));
```

A.3 Code for Chapter 4

Printed below is code for three simulations described in Chapter 4. Three `.m`-files contain the core models. The first, `taurange.m`, solves for optimal market size and expected utility for a triangular grid of $\{\tau_s, \tau_g\}$ pairs. The second, `opttau.m`, solves for optimal $\{\tau_s, \tau_g\}$ using the marginal conditions over a range of parameterisations (in this case, over a range of k). The third, `taxes.m`, develops the numerical implications of power and taxes in our model. Two functions, `alpha.m` and `tausolve.m` solve respectively for equilibrium α and, using the marginal conditions, optimal $\{\tau_s, \tau_g\}$.

■taurange.m

```
global beta     %weight on ameliorating capital
global delta    %weight on specific contracting
global sigma    %elasticity of substitution
global gamma    %exponent on utility function
global rho      %probability of low tech
global lambda1  %low tech
global lambda2  %high tech
global curves   %gamma_s
global curveg   %gamma_g
global taus     %specific tau
global taug     %general tau
global k        %endowment level
global noM      %market size
global b        %agent Type
global incomp   %"incompleteness"
global alph     %exchange cost

k=25;
beta=.03;
curves=.065;
curveg=.085;
delta=.5;
sigma=0.5;
rho=.5;
lambda1=1;
lambda2=5;
```

```
ro=3;
U=@(W)(((W)^(1-ro)-1)/(1-ro));
b=1;
top=30;      %set the top of the range of coalition sizes through which we
               index for each alpha
optimalsize=zeros;
h=0
for taus=0:.005:1    %cycle through taus
    h=h+1
    j=0;
    for taug=0:.005:1   %cycle through taug
        j=j+1;
        if taus+taug>=1
            optimalsize(h,j,1)=NaN;
            optimalsize(h,j,2)=NaN;
            optimalsize(h,j,3)=NaN;
            optimalsize(h,j,4)=NaN;
            optimalsize(h,j,5)=NaN;
            optimalsize(h,j,6)=NaN;
        else
            maxutility=0;
            for noM=1:1:top
                utilityi=0;
                utilityizero=0;
                for i=0:1:noM
                    utilityi=utilityi+(factorial(noM)/(factorial(i)*...
                      factorial(noM-i)))*(rho^(noM-i))*((1-rho)^i)*U((((...
                      1-taus-taug)*k)-(2*alpha((taus*k),(taug*k))*(((noM-1)...
                      /noM))))*(((noM-i)*lambda1+i*lambda2)/noM));
                end
                if isreal(utilityi)      %need to take out complex numbers...
                                            for the graphing
                    if utilityi>maxutility   %keeps tabs on the optimal...
                                                market size
                        if noM>=top   %this bit is just so that if it goes...
                                        to the limit of our search we...
                                        don't have a flat line at 'top'
                            optimalsize(h,j,1)=taus;
                            optimalsize(h,j,2)=taug;
                            optimalsize(h,j,3)=NaN;
                            optimalsize(h,j,4)=NaN;
                            optimalsize(h,j,5)=NaN;
                            optimalsize(h,j,6)=NaN;
                        else
                                maxutility=utilityi;
```

```
                                optimalsize(h,j,1)=taus;
                                optimalsize(h,j,2)=taug;
                                optimalsize(h,j,3)=noM;
                                optimalsize(h,j,4)=utilityi;
                                optimalsize(h,j,5)=incomp;
                                optimalsize(h,j,6)=alph;
                        end
                    end
                else
                    optimalsize(h,j,1)=taus;
                    optimalsize(h,j,2)=taug;
                    optimalsize(h,j,3)=NaN;
                    optimalsize(h,j,4)=NaN;
                    optimalsize(h,j,5)=NaN;
                    optimalsize(h,j,6)=NaN;
                end
            end
        end
    end
end
figure; surf(optimalsize(:,:,1),optimalsize(:,:,2),optimalsize(:,:,4))
figure; surf(optimalsize(:,:,1),optimalsize(:,:,2),optimalsize(:,:,3))
```

■opttau.m

```
global beta     %weight on ameliorating capital
global delta    %weight on specific contracting
global sigma    %elasticity of substitution
global gamma    %exponent on utility function
global rho      %probability of low tech
global lambda1  %low tech
global lambda2  %high tech
global curves   %gamma_s
global curveg   %gamma_g
global taus     %specific tau
global taug     %general tau
global k        %endowment level
global noM      %market size
global b        %agent Type
global incomp   %"incompleteness"
global alph     %exchange cost

%k=25;
beta=.03;
```

```
curves=.065;
curveg=.085;
delta=.5;
sigma=0.5;
rho=.5;
lambda1=1;
lambda2=5;
ro=3;
U=@(W)(((W)^(1-ro)-1)/(1-ro));
b=1;

optimaltau=zeros;

top=50;     %set the top of the range of coalition sizes through
               which we index for each alpha
h=0
for k=20:.05:38
    h=h+1
    maxutility=0;
    fguesses=zeros;
    for noM=1:1:top
        if noM<2
            taus=0;
            taug=0;
        else
            if noM==2  %finding the fguess here is difficile, this is a...
                          rough automation, but not always does it work
                fguess=.025;     %first stab is important
            else
                fguess=fguesses(noM-1);
            end
            while isreal(tausolve(fguess))==0   %cycle until tausolve is real
                fguess=fguess+.0001;
            end
            options = optimset('Display','off','TolFun',1e-8);
            [x,y,z]=fzero(@tausolve,fguess,options);
            while z~=1  %cycle until fsolve works
                fguess=fguess+.0001;
                [x,y,z]=fzero(@tausolve,fguess,options);
                if fguess>1
                    warning('fguess>1; reset noM=2 fguess')
                    break
                end
            end
            taus=fzero(@tausolve, fguess,options);
```

```
        end
        fguesses(noM)=fguess;
        utilityi=0;
        for i=0:1:noM
            utilityi=utilityi+(factorial(noM)/(factorial(i)*...
              factorial(noM-i)))*(rho^(noM-i))*((1-rho)^i)*...
              U((((1-taus-taug)*k)-(2*alpha((taus*k),(taug*k))...
              *((noM-1)/noM)))*(((noM-i)*lambda1+i*lambda2)/noM));
        end
        if isreal(utilityi)      %need to take out complex numbers for...
                                    the graphing: isreal(A) returns 1 if all...
                                    elements of A are real, 0 otherwise
            if utilityi>maxutility   %keeps tabs on the optimal market size
                if noM>=top   %this bit is just so that if it goes to the...
                                 limit of our search we don't have a flat...
                                 line at 'top'
                    optimaltau(h,1)=k;
                    optimaltau(h,2)=NaN;
                    optimaltau(h,3)=NaN;
                    optimaltau(h,4)=NaN;
                    optimaltau(h,5)=NaN;
                    optimaltau(h,6)=NaN;
                    optimaltau(h,7)=NaN;
                    optimaltau(h,8)=NaN;
                    optimaltau(h,9)=NaN;
                    optimaltau(h,10)=NaN;
                else
                    maxutility=utilityi;
                    optimaltau(h,1)=k;
                    optimaltau(h,2)=taus;
                    optimaltau(h,3)=taug;
                    optimaltau(h,4)=noM;
                    optimaltau(h,5)=utilityi;
                    optimaltau(h,6)=incomp;
                    if noM>1
                        optimaltau(h,7)=alph;
                        optimaltau(h,8)=((noM-1)/noM)*alph;
                        optimaltau(h,9)=(taus+taug)*k+2*((noM-1)/noM)*alph;...
                          %per capita spend on tau and exchange
                        optimaltau(h,10)=optimaltau(h,9)/k;  %kappa
                    else
                        optimaltau(h,7)=0;
                        optimaltau(h,8)=0;
                        optimaltau(h,9)=0;
                        optimaltau(h,10)=0;
```

```
                    end
                end
            end
        end
    end
end
figure; stairs(optimaltau(:,1),optimaltau(:,2:3))
figure; stairs(optimaltau(:,1),optimaltau(:,8))
figure; stairs(optimaltau(:,1),optimaltau(:,10))
```

■**taxes.m**

```
function alph = alpha(S,G)

global beta     %weight on ameliorating capital
global delta    %weight on specific contracting
global sigma    %elasticity of substitution
global gamma    %exponent on utility function
global rho      %probability of low tech
global lambda1  %low tech
global lambda2  %high tech
global curves   %gamma_s
global curveg   %gamma_g
global taus     %specific tau
global taug     %general tau
global k        %endowment level
global noM      %market size
global b        %agent Type
global incomp   %"incompleteness"
global alph     %exchange cost

optimaltau=zeros;

beta=.03;
curves=[.065 .1];
curveg=.085;
delta=.5;
sigma=0.5;
rho=.5;
lambda1=1;
lambda2=5;
ro=3;
U=@(W)(((W)^(1-ro)-1)/(1-ro));    %CRRA utility
```

```
top=50;
k=25
taus=zeros;
h=0;
for taug=0.03:.0001:.08
    maxutility=[0 0];
    h=h+1
    for b=1:2;   %cycle through agent Types
        for noM=1:1:top
            taus(b)=(1/(beta*k))*(-log(1-((1/(1-delta))*(((2*((noM-1)/noM)...
              *beta*k*exp(-beta*taug*k)*curveg*delta*((1-exp(-beta*taug*k))...
              ^(curveg*sigma-1)))^(sigma/(sigma-1)))-delta*(1-exp(-beta*taug...
              *k))^(curveg*sigma)))^(1/(curves(b)*sigma))));
            if isreal(taus(b))
                if taus(b)~=0
                    if taus(b)+taug<1
                        utilityi(b)=0;
                        for i=0:1:noM
                            utilityi(b)=utilityi(b)+(factorial(noM)/(factorial...
                              (i)*factorial(noM-i)))*(rho^(noM-i))*((1-rho)...
                              ^i)*U((((1-taus(b)-taug)*k)-(2*alpha((taus(b)*k),...
                              (taug*k))*((noM-1)/noM)))*(((noM-i)*lambda1+...
                              i*lambda2)/noM));
                        end
                        if isreal(utilityi(b))
                            if utilityi(b)>maxutility(b)
                                if noM>=top
                                    optimaltau(h,b,1)=k;
                                    optimaltau(h,b,2)=NaN;
                                    optimaltau(h,b,3)=NaN;
                                    optimaltau(h,b,4)=NaN;
                                    optimaltau(h,b,5)=NaN;
                                    optimaltau(h,b,6)=NaN;
                                    optimaltau(h,b,7)=NaN;
                                    optimaltau(h,b,8)=NaN;
                                    optimaltau(h,b,9)=NaN;
                                    optimaltau(h,b,10)=NaN;
                                    optimaltau(h,b,11)=curves(b);
                                    optimaltau(h,b,12)=NaN;
                                    optimaltau(h,b,13)=NaN;
                                else
                                    hstep=.01;   %two-sided differences
                                    for j=0:1
                                        k2=k*(1+(-1+2*j)*hstep);
                                        taus2(b)=(1/(beta*k2))*(-log(1-((1/(1-...
```

```
          delta))*(((2*((noM-1)/noM)*beta*k2*...
          exp(-beta*taug*k2)*curveg*delta*((1...
          -exp(-beta*taug*k2))^(curveg*sigma-...
          1)))^(sigma/(sigma-1)))-delta*(1-...
          exp(-beta*taug*k2))^(curveg*sigma)))...
          ^(1/(curves(b)*sigma))));
        utilityi2(b)=0;
        for i=0:1:noM
            utilityi2(b)=utilityi2(b)+(...
              factorial(noM)/(factorial(i)*...
              factorial(noM-i))*(rho^(noM...
              -i))*((1-rho)^i)*U((((1-taus2(b)...
              -taug)*k2)-(2*alpha((taus2(b)*k2)...
              ,(taug*k2))*((noM-1).../noM)))*...
              (((noM-i)*lambda1+i*lambda2)/noM));
        end
        marginals(j+1)=utilityi2(b);
    end
    marginal=(marginals(2)-marginals(1))/(2*hstep);
    maxutility=utilityi;
    optimaltau(h,b,1)=k;
    optimaltau(h,b,2)=taus(b);
    optimaltau(h,b,3)=taug;
    optimaltau(h,b,4)=noM;
    optimaltau(h,b,5)=utilityi(b);
    optimaltau(h,b,6)=incomp;
    if noM>1
        optimaltau(h,b,7)=alph;
        optimaltau(h,b,8)=((noM-1)/noM)*alph;
        optimaltau(h,b,9)=(taus(b)+taug)*k+2*...
          ((noM-1)/noM)*alph;
        optimaltau(h,b,10)=optimaltau(h,b,9)/k;
        optimaltau(h,b,11)=curves(b);
        optimaltau(h,b,12)=marginal;
        optimaltau(h,b,13)=1/marginal;
    else
        optimaltau(h,b,7)=0;
        optimaltau(h,b,8)=0;
        optimaltau(h,b,9)=0;
        optimaltau(h,b,10)=0;
        optimaltau(h,b,11)=0;
        optimaltau(h,b,12)=0;
        optimaltau(h,b,13)=0;
    end
end
```

```
                        end
                    end
                end
            else
                optimaltau(h,b,1)=k;
                optimaltau(h,b,2)=NaN;
                optimaltau(h,b,3)=taug;
                optimaltau(h,b,4)=NaN;
                optimaltau(h,b,5)=NaN;
                optimaltau(h,b,6)=NaN;
                optimaltau(h,b,7)=NaN;
                optimaltau(h,b,8)=NaN;
                optimaltau(h,b,9)=NaN;
                optimaltau(h,b,10)=NaN;
                optimaltau(h,b,11)=NaN;
                optimaltau(h,b,12)=NaN;
                optimaltau(h,b,13)=NaN;
            end
        else
            optimaltau(h,b,1)=k;
            optimaltau(h,b,2)=NaN;
            optimaltau(h,b,3)=taug;
            optimaltau(h,b,4)=NaN;
            optimaltau(h,b,5)=NaN;
            optimaltau(h,b,6)=NaN;
            optimaltau(h,b,7)=NaN;
            optimaltau(h,b,8)=NaN;
            optimaltau(h,b,9)=NaN;
            optimaltau(h,b,10)=NaN;
            optimaltau(h,b,11)=NaN;
            optimaltau(h,b,12)=NaN;
            optimaltau(h,b,13)=NaN;
        end
    end
end
end


max=zeros;   %finds optimal taug etc. for each Type
for b=1:2
    maxu=0;
    [m n]=size(optimaltau);
    for i=1:1:m
        if optimaltau(i,b,5)>maxu
            maxu=optimaltau(i,b,5);
```

```
            max(1,b)=optimaltau(i,b,1);
            max(2,b)=optimaltau(i,b,2);
            max(3,b)=optimaltau(i,b,3);
            max(4,b)=optimaltau(i,b,4);
            max(5,b)=optimaltau(i,b,5);
            max(6,b)=optimaltau(i,b,6);
            max(7,b)=optimaltau(i,b,10);
            max(8,b)=optimaltau(i,b,11);
            max(9,b)=i;
            max(10,b)=optimaltau(i,b,12);
        end
    end
end
max

theta1=1;         %finding the egalitarian allocation
theta2=1;
maxmax=zeros;
maxweight=0;
[m n]=size(optimaltau);
for i=1:1:m
    if theta1*optimaltau(i,1,5)+theta2*optimaltau(i,2,5)>maxweight
        maxweight=theta1*optimaltau(i,1,5)+theta2*optimaltau(i,2,5);
        for b=1:2
            maxmax(1,b)=optimaltau(i,b,1);
            maxmax(2,b)=optimaltau(i,b,2);
            maxmax(3,b)=optimaltau(i,b,3);
            maxmax(4,b)=optimaltau(i,b,4);
            maxmax(5,b)=optimaltau(i,b,5);
            maxmax(6,b)=optimaltau(i,b,6);
            maxmax(7,b)=optimaltau(i,b,10);
            maxmax(8,b)=optimaltau(i,b,11);
            maxmax(9,b)=i;
            maxmax(10,b)=maxweight;
            maxmax(11,b)=optimaltau(i,b,12);
        end
    end
end
maxmax

maxpower=zeros;       %index over taug for finding power allocations
maxweight=zeros;
i=0;
for l=max(9,2):max(9,1)   %cycle from tau^2*_g to tau^1*_g
    i=i+1;
```

```
    theta1=optimaltau(l,1,13)/(optimaltau(l,1,13)+optimaltau(l,2,13));
    theta2=optimaltau(l,2,13)/(optimaltau(l,1,13)+optimaltau(l,2,13));
    j=0;
    maxweight(i)=0;
    for m=max(9,2):max(9,1)
        j=j+1;
        if theta1*optimaltau(m,1,5)+theta2*optimaltau(m,2,5)>maxweight(i)
            maxweight(i)=theta1*optimaltau(m,1,5)+theta2*optimaltau(m,2,5);
            for b=1:2
                maxpower(i,1,b)=optimaltau(l,b,3);
                maxpower(i,2,b)=optimaltau(m,b,3);
                maxpower(i,3,b)=l;
                maxpower(i,4,b)=m;
                maxpower(i,5,b)=theta1;    %these are the weights implied...
                                                 by the taug
                maxpower(i,6,b)=theta2;
                maxpower(i,7,b)=optimaltau(m,1,13)/(optimaltau(m,1,13)+
                  optimaltau(m,2,13));     %these are the weights implied...
                                                 by the ex post marginal utilities
                maxpower(i,8,b)=optimaltau(m,2,13)/(optimaltau(m,1,13)+
                  optimaltau(m,2,13));
            end
        end
    end
end

maxedpower=zeros;
[p q]=size(maxpower);    %pick out power equilibria from maxpower
for s=1:p
    t=0;
    if maxpower(s,3,b)==maxpower(s,4,b)
        t=t+1;
        for b=1:2
            maxedpower(t,1,b)=optimaltau(maxpower(s,3,b),b,1);
            maxedpower(t,2,b)=optimaltau(maxpower(s,3,b),b,2);
            maxedpower(t,3,b)=optimaltau(maxpower(s,3,b),b,3);
            maxedpower(t,4,b)=optimaltau(maxpower(s,3,b),b,4);
            maxedpower(t,5,b)=optimaltau(maxpower(s,3,b),b,5);
            maxedpower(t,6,b)=optimaltau(maxpower(s,3,b),b,6);
            maxedpower(t,7,b)=optimaltau(maxpower(s,3,b),b,10);
            maxedpower(t,8,b)=optimaltau(maxpower(s,3,b),b,11);
            maxedpower(t,9,b)=maxpower(s,3,b);
            maxedpower(t,10,b)=optimaltau(maxpower(s,3,b),b,12);
            maxedpower(t,11,b)=optimaltau(maxpower(s,3,b),b,13);
        end
```

```
    end
end
```

■alpha.m

```
function alph = alpha(S,G)

global beta     %weight on ameliorating capital
global delta    %weight on specific contracting
global sigma    %elasticity of substitution
global gamma    %exponent on utility function
global rho      %probability of low tech
global lambda1  %low tech
global lambda2  %high tech
global curves   %gamma_s
global curveg   %gamma_g
global taus     %specific tau
global taug     %general tau
global k        %endowment level
global noM      %market size
global b        %agent Type

alph=(1-(delta*(1-(exp(-beta*taus(b)*k)))^(curves(b)*sigma)+(1-delta)*(1...
  -(exp(-beta*taug*k)))^(curveg*sigma))^(1/sigma))*k;
incomp=(delta*(1-(exp(-beta*taus(b)*k)))^(curves(b)*sigma)+(1-delta)*(1...
  -(exp(-beta*taug*k)))^(curveg*sigma))^(1/sigma);
```

■tausolve.m

```
function y=tausolve(taus)

global beta     %weight on ameliorating capital
global delta    %weight on specific contracting
global sigma    %elasticity of substitution
global gamma    %exponent on utility function
global rho      %probability of low tech
global lambda1  %low tech
global lambda2  %high tech
global curves   %gamma_s
global curveg   %gamma_g
global taus     %specific tau
global taug     %general tau
global k        %endowment level
```

```
global noM      %market size

y=2*((noM-1)/noM)*(((delta*(1-exp(-beta*taus*k))^(curves*sigma))+...
  ((1-delta)*(1-exp(-beta*(1/(beta*k))*(-log(1-((1/(1-delta))*(((2*...
  ((noM-1)/noM)*beta*k*exp(-beta*taus*k)*curves*delta*((1-exp(-beta*...
  taus*k))^(curves*sigma-1)))^(sigma/(sigma-1)))-delta*(1-exp(-beta*...
  taus*k))^(curves*sigma)))^(1/(curveg*sigma))))*k))^(curveg*sigma)))...
  ^((1-sigma)/sigma))*beta*k*exp(-beta*(1/(beta*k))*(-log(1-((1/(1-...
  delta))*(((2*((noM-1)/noM)*beta*k*exp(-beta*taus*k)*curves*delta*...
  ((1-exp(-beta*taus*k))^(curves*sigma-1)))^(sigma/(sigma-1)))-delta*...
  (1-exp(-beta*taus*k))^(curves*sigma)))^(1/(curveg*sigma))))*k)*curveg*...
  (1-delta)*(1-exp(-beta*(1/(beta*k))*(-log(1-((1/(1-delta))*(((2*((noM-1)...
  /noM)*beta*k*exp(-beta*taus*k)*curves*delta*((1-exp(-beta*taus*k))^...
  (curves*sigma-1)))^(sigma/(sigma-1)))-delta*(1-exp(-beta*taus*k))^...
  (curves*sigma)))^(1/(curveg*sigma))))*k))^(curveg*sigma-1)-1;

taug=(1/(beta*k))*(-log(1-((1/(1-delta))*(((2*((noM-1)/noM)*beta*k*...
  exp(-beta*taus*k)*curves*delta*((1-exp(-beta*taus*k))^(curves*sigma-...
  1)))^(sigma/(sigma-1)))-delta*(1-exp(-beta*taus*k))^(curves*sigma)))...
  ^(1/(curveg*sigma))));
```

Appendix B

Data for Chapters 1, 2 and 3

All of the following data are also available in electronic form from the author by email at alex.trew@dunelm.org.uk.

Data for Figure 1.1 is derived from the full database in Beck et al. (1999), which is available from http://econ.worldbank.org/programs/finance/. We restrict the sample to the “pure cross-sectional dataset” used by Levine et al. (2000), i.e., for the period 1960–95, for 71 of the 192 countries in the complete dataset as listed in their Table 9. Table B.1 below gives for each year the number of countries with missing observations for three series. The codes are used in later revisions of the database: `pcrdbofgdp` is ‘Private credit by deposit money banks and other financial institutions to GDP’; `llgdp` is ‘Liquid liabilities to GDP’; and, `dbagdp` is ‘Deposit Money Bank Assets to GDP’. These are the three significant variables in Aghion et al. (2005).

Figure 2.1 is generated using the ‘Revised Best Guess’ index of industrial production in Crafts and Harley (1992), in their Table A3.I. The index is detrended using a Hodrick-Prescott filter with different smoothing parameters before taking growth rates. Table B.2 gives these data.

Table B.3 gives the series of per capita industrial production used for Figure 3.1, as taken from Bairoch (1982) Tables 9 and 12.

Table B.1: Data for Figure 1.1. Finance to GDP Ratios. Source: Levine et al. (2000) and own calculations.

year	`pcrdbofgdp`	`llgdp`	`dbagdp`
1960	25	29	25
1961	24	28	23
1962	24	27	23
1963	21	24	20
1964	16	20	15
1965	16	21	16
1966	13	18	13
1967	14	17	13
1968	12	16	11
1969	9	14	9
1970	8	13	8
1971	7	11	6
1972	7	11	6
1973	6	11	6
1974	5	10	5
1975	5	9	5
1976	5	9	5
1977	5	9	5
1978	4	9	5
1979	4	9	5
1980	3	7	3
1981	3	7	3
1982	5	9	5
1983	5	10	5
1984	4	9	4
1985	4	9	4
1986	5	10	5
1987	5	10	5
1988	3	8	3
1989	4	8	4
1990	4	10	5
1991	5	10	5
1992	6	11	6
1993	5	11	5
1994	2	9	3
1995	3	9	3

Table B.2: Data for Figure 2.1: Index of Industrial Production, Smoothing and Growth Rates. Source: Crafts and Harley (1992) and own calculations.

year	**Revised Best Guess**	**growth (non-HP)**	**HP λ=10,000**	**growth λ=10,000**	**HP λ=100**	**growth λ=100**
1700	1.92		2.068112		2.09101	
1701	2.11	9.895833	2.085856	0.858006	2.112083	1.007796
1702	1.85	-12.3223	2.103586	0.849997	2.131446	0.916773
1703	2.6	40.54054	2.121289	0.841539	2.147368	0.747004
1704	2.4	-7.69231	2.138926	0.831454	2.155304	0.369549
1705	2.35	-2.08333	2.156509	0.822034	2.155234	-0.00324
1706	2.02	-14.0426	2.174073	0.814468	2.149586	-0.26204
1707	2.08	2.970297	2.191674	0.809587	2.142737	-0.31864
1708	2.19	5.288462	2.209352	0.806604	2.137765	-0.23203
1709	2.15	-1.82648	2.227136	0.804953	2.137123	-0.03002
1710	1.77	-17.6744	2.245054	0.804504	2.143786	0.311769
1711	2.1	18.64407	2.263124	0.804877	2.160857	0.796304
1712	2.12	0.952381	2.281318	0.803946	2.187702	1.242322
1713	2.26	6.603774	2.299592	0.801037	2.223077	1.617021
1714	2.18	-3.53982	2.317886	0.79551	2.265064	1.888661
1715	2.22	1.834862	2.336134	0.787279	2.31211	2.077047
1716	2.37	6.756757	2.354258	0.775838	2.361815	2.149771
1717	2.58	8.860759	2.372169	0.760793	2.410856	2.07642
1718	2.52	-2.32558	2.389778	0.742319	2.455993	1.872227
1719	2.62	3.968254	2.407018	0.721385	2.495676	1.615761
1720	2.52	-3.81679	2.423833	0.698597	2.528996	1.335114
1721	2.44	-3.1746	2.440191	0.674878	2.556288	1.079137
1722	2.68	9.836066	2.456068	0.650638	2.577794	0.841325
1723	2.69	0.373134	2.47144	0.625879	2.592598	0.574263
1724	2.58	-4.08922	2.486306	0.601509	2.600801	0.31641
1725	2.63	1.937984	2.500686	0.578385	2.603481	0.103064
1726	2.7	2.661597	2.514611	0.55684	2.601508	-0.07578
1727	2.66	-1.48148	2.528123	0.537343	2.596017	-0.2111
1728	2.51	-5.6391	2.541284	0.520575	2.589125	-0.26545
1729	2.41	-3.98406	2.554168	0.506989	2.583594	-0.21364
1730	2.54	5.394191	2.566847	0.496395	2.58139	-0.08529
1731	2.5	-1.5748	2.579377	0.488168	2.582747	0.052541
1732	2.53	1.2	2.591814	0.482158	2.587481	0.183308
1733	2.65	4.743083	2.604203	0.478014	2.594584	0.27452
1734	2.64	-0.37736	2.616585	0.475463	2.602472	0.304012
1735	2.65	0.378788	2.629005	0.47465	2.610115	0.293667
1736	2.69	1.509434	2.641509	0.475633	2.616857	0.258329
1737	2.63	-2.23048	2.654148	0.478455	2.622444	0.213498
1738	2.67	1.520913	2.666974	0.483266	2.627351	0.18711
1739	2.7	1.123596	2.680041	0.489935	2.632129	0.181853
1740	2.57	-4.81481	2.693399	0.498435	2.637756	0.213769

cont.

year	Revised Best Guess	growth (non-HP)	HP λ=10,000	growth λ=10,000	HP λ=100	growth λ=100
1741	2.64	2.723735	2.707103	0.508803	2.645887	0.308285
1742	2.58	-2.27273	2.721195	0.520541	2.657503	0.439012
1743	2.63	1.937984	2.735709	0.53337	2.673523	0.602809
1744	2.69	2.281369	2.750666	0.546742	2.694091	0.769341
1745	2.64	-1.85874	2.766077	0.560257	2.718918	0.921536
1746	2.69	1.893939	2.781945	0.573682	2.747673	1.057564
1747	2.76	2.60223	2.798263	0.586559	2.779234	1.148652
1748	2.82	2.173913	2.815013	0.598565	2.811905	1.175538
1749	2.87	1.77305	2.832172	0.609576	2.843796	1.134151
1750	2.97	3.484321	2.849721	0.619627	2.873099	1.030419
1751	2.95	-0.6734	2.867642	0.62887	2.898267	0.875999
1752	3.03	2.711864	2.88593	0.637737	2.918724	0.705806
1753	3.02	-0.33003	2.904588	0.646519	2.934407	0.537352
1754	2.94	-2.64901	2.923634	0.655711	2.946371	0.407718
1755	2.99	1.70068	2.943096	0.66569	2.956524	0.344592
1756	2.83	-5.35117	2.963006	0.676487	2.966711	0.344547
1757	2.98	5.300353	2.983398	0.688233	2.979111	0.417968
1758	2.94	-1.34228	3.004296	0.700449	2.994537	0.517793
1759	2.99	1.70068	3.025719	0.713103	3.013809	0.643597
1760	2.99	0	3.047685	0.725961	3.037205	0.776288
1761	3.09	3.344482	3.070204	0.738893	3.064762	0.90731
1762	3.05	-1.2945	3.093283	0.751698	3.096046	1.020756
1763	3.03	-0.65574	3.116929	0.764436	3.130875	1.124946
1764	3.22	6.270627	3.141146	0.776959	3.168606	1.20515
1765	3.13	-2.79503	3.165929	0.788989	3.20759	1.230309
1766	3.25	3.833866	3.191281	0.800781	3.246689	1.218938
1767	3.42	5.230769	3.217202	0.81222	3.283989	1.148878
1768	3.29	-3.80117	3.243695	0.823491	3.317612	1.023835
1769	3.43	4.255319	3.270787	0.835217	3.347037	0.886935
1770	3.39	-1.16618	3.298507	0.847516	3.371468	0.729949
1771	3.49	2.949853	3.326902	0.860841	3.390941	0.577555
1772	3.57	2.292264	3.356027	0.875421	3.405672	0.434441
1773	3.5	-1.96078	3.385952	0.891689	3.416873	0.328889
1774	3.23	-7.71429	3.416771	0.910202	3.427396	0.307975
1775	3.32	2.786378	3.448588	0.931197	3.440926	0.39475
1776	3.46	4.216867	3.481488	0.954021	3.459172	0.530277
1777	3.51	1.445087	3.515544	0.97821	3.482637	0.678322
1778	3.62	3.133903	3.550827	1.003622	3.511828	0.838201
1779	3.38	-6.62983	3.587407	1.030162	3.54753	1.016611
1780	3.49	3.254438	3.625359	1.057939	3.591606	1.242455
1781	3.61	3.438395	3.664741	1.086293	3.644247	1.465657
1782	3.65	1.108033	3.705595	1.114787	3.704625	1.656807
1783	3.73	2.191781	3.747958	1.143223	3.771571	1.807106
1784	3.95	5.898123	3.791862	1.171409	3.843371	1.903701

cont.

year	**Revised Best Guess**	**growth (non-HP)**	**HP λ=10,000**	**growth λ=10,000**	**HP λ=100**	**growth λ=100**
1785	3.96	0.253165	3.837337	1.199262	3.917892	1.938957
1786	3.93	-0.75758	3.884427	1.227163	3.99407	1.94437
1787	4.13	5.089059	3.933191	1.255378	4.071262	1.932655
1788	4.16	0.726392	3.983692	1.283957	4.148182	1.88935
1789	4.2	0.961538	4.036011	1.313325	4.224135	1.830974
1790	4.3	2.380952	4.090248	1.343823	4.29854	1.761426
1791	4.42	2.790698	4.146519	1.375735	4.370577	1.675863
1792	4.66	5.429864	4.204961	1.409428	4.439442	1.575641
1793	4.49	-3.64807	4.265739	1.445384	4.504822	1.472711
1794	4.47	-0.44543	4.329062	1.484465	4.568612	1.416036
1795	4.64	3.803132	4.395163	1.52692	4.632558	1.399672
1796	4.78	3.017241	4.464289	1.572768	4.697419	1.400114
1797	4.56	-4.60251	4.53671	1.62223	4.764029	1.418031
1798	4.71	3.289474	4.612729	1.675638	4.83405	1.469779
1799	5.13	8.917197	4.69265	1.732629	4.9071	1.511161
1800	5.07	-1.16959	4.776789	1.792989	4.98156	1.517376
1801	4.87	-3.94477	4.865503	1.857194	5.058036	1.535188
1802	5.14	5.544148	4.959181	1.925337	5.138022	1.581367
1803	5.19	0.972763	5.058209	1.996873	5.22113	1.617515
1804	5.36	3.27553	5.162995	2.071601	5.306993	1.644519
1805	5.48	2.238806	5.273958	2.14919	5.39493	1.657011
1806	5.52	0.729927	5.391536	2.229413	5.484794	1.6657
1807	5.74	3.985507	5.51619	2.312024	5.577284	1.686313
1808	5.46	-4.87805	5.648391	2.396608	5.673456	1.724348
1809	5.6	2.564103	5.788635	2.482904	5.77599	1.807252
1810	6.09	8.75	5.937398	2.569912	5.885432	1.894769
1811	6.4	5.090312	6.095137	2.656697	6.000568	1.956289
1812	6.02	-5.9375	6.262324	2.742957	6.12223	2.027515
1813	6.07	0.830565	6.439462	2.828633	6.255245	2.172658
1814	6.17	1.647446	6.62703	2.91279	6.403417	2.368761
1815	6.71	8.752026	6.825469	2.994397	6.568697	2.581124
1816	6.55	-2.3845	7.035176	3.072421	6.750703	2.770805
1817	7.16	9.312977	7.256536	3.146463	6.950465	2.959128
1818	7.46	4.189944	7.489883	3.215684	7.167006	3.115497
1819	7.21	-3.35121	7.735545	3.279913	7.401446	3.271099
1820	7.4	2.635229	7.993844	3.339122	7.657833	3.46401
1821	7.71	4.189189	8.265052	3.392707	7.938301	3.6625
1822	8.13	5.447471	8.549379	3.440118	8.242406	3.830861
1823	8.56	5.289053	8.846982	3.480992	8.567422	3.94321
1824	9.02	5.373832	9.157975	3.515241	8.909496	3.992733
1825	9.85	9.201774	9.482443	3.543006	9.264704	3.986845
1826	8.99	-8.73096	9.820457	3.564629	9.630225	3.945307
1827	10.2	13.4594	10.17212	3.580978	10.00909	3.934143
1828	10.9	6.862745	10.53747	3.591659	10.39793	3.884898

cont.

year	Revised Best Guess	growth (non-HP)	HP λ=10,000	growth λ=10,000	HP λ=100	growth λ=100
1829	10.5	-3.66972	10.91653	3.597215	10.79529	3.821515
1830	11.5	9.52381	11.30936	3.598464	11.20473	3.792712
1831	11.6	0.869565	11.71598	3.595445	11.62685	3.767314
1832	11.6	0	12.13643	3.588753	12.06521	3.770241
1833	12.2	5.172414	12.57076	3.578674	12.5231	3.795164
1834	12.9	5.737705	13.01892	3.565144	12.99917	3.801519
1835	13.4	3.875969	13.48087	3.548275	13.48882	3.76678
1836	14.7	9.701493	13.95653	3.528378	13.98647	3.689343
1837	14	-4.7619	14.44581	3.505774	14.48564	3.568993
1838	15.3	9.285714	14.94872	3.481329	14.98701	3.461132
1839	16.6	8.496732	15.4652	3.454993	15.48638	3.331992
1840	16.2	-2.40964	15.99523	3.427259	15.98268	3.204765
1841	16.5	1.851852	16.53891	3.399046	16.48599	3.149107
1842	15.7	-4.84848	17.09637	3.370573	17.00856	3.169771
1843	16.6	5.732484	17.66771	3.341888	17.56277	3.258428
1844	18.7	12.6506	18.25291	3.312265	18.14792	3.331801
1845	19.7	5.347594	18.85184	3.28128	18.7537	3.337985
1846	19.7	0	19.46441	3.249388	19.37529	3.314483
1847	19.1	-3.04569	20.09062	3.21719	20.01735	3.313802
1848	20.9	9.424084	20.73048	3.184898	20.68778	3.349264
1849	21.2	1.435407	21.38393	3.152103	21.38532	3.371749
1850	21.2	0	22.05089	3.119004	22.11082	3.392512
1851	22.1	4.245283	22.7313	3.085613	22.86328	3.403126
1852	23.9	8.144796	23.42498	3.051653	23.63259	3.364831
1853	25.6	7.112971	24.13171	3.016984	24.40101	3.251533
1854	25.6	0	24.8513	2.981947	25.15348	3.083746
1855	25.5	-0.39063	25.58373	2.947244	25.88691	2.915829
1856	27.3	7.058824	26.32903	2.913193	26.6027	2.765056
1857	28.4	4.029304	27.08725	2.879755	27.29836	2.615007
1858	27.1	-4.57746	27.8585	2.84728	27.97839	2.491099
1859	28.8	6.273063	28.64305	2.816195	28.65829	2.430108
1860	30.4	5.555556	29.44108	2.786138	29.3448	2.395486
1861	30	-1.31579	30.2528	2.757104	30.04605	2.389684
1862	30.2	0.666667	31.07851	2.72935	30.78073	2.445195
1863	30	-0.66225	31.91847	2.702705	31.56708	2.554696
1864	31.9	6.333333	32.77287	2.67682	32.41753	2.694105
1865	33.9	6.269592	33.6417	2.65107	33.32884	2.811138
1866	34.9	2.949853	34.52487	2.625219	34.29257	2.89159
1867	34.4	-1.43266	35.4223	2.599389	35.30603	2.955323
1868	35.5	3.197674	36.33397	2.573721	36.37257	3.020859
1869	35.3	-0.56338	37.25975	2.547953	37.48651	3.062583
1870	38.4	8.78187	38.1994	2.521909	38.63343	3.05954
1871	41.3	7.552083	39.15253	2.495137	39.77703	2.960142
1872	42.5	2.905569	40.11873	2.467797	40.8787	2.769625

cont.

year	**Revised Best Guess**	**growth (non-HP)**	**HP λ=10,000**	**growth λ=10,000**	**HP λ=100**	**growth λ=100**
1873	43.7	2.823529	41.09784	2.440516	41.91506	2.535195
1874	44.2	1.144165	42.0899	2.413904	42.87892	2.299552
1875	44.2	0	43.09524	2.388556	43.78095	2.10368
1876	44.7	1.131222	44.11439	2.364878	44.64505	1.973678
1877	45.4	1.565996	45.14799	2.343	45.49928	1.913379
1878	44.6	-1.76211	46.19674	2.322917	46.37226	1.918681
1879	42.8	-4.03587	47.26137	2.304547	47.29164	1.982595
1880	48.2	12.61682	48.34244	2.287421	48.26731	2.063093
1881	50.9	5.60166	49.44006	2.270523	49.26427	2.065494
1882	53.5	5.108055	50.55435	2.253822	50.24683	1.994477
1883	54.5	1.869159	51.68556	2.237605	51.19568	1.888372
1884	52.2	-4.22018	52.83423	2.222415	52.12401	1.813312
1885	50.4	-3.44828	54.00118	2.208712	53.07809	1.8304
1886	49.9	-0.99206	55.18719	2.196263	54.10492	1.934564
1887	53.7	7.61523	56.39265	2.184317	55.22473	2.069703
1888	57.2	6.517691	57.61745	2.171902	56.41571	2.156595
1889	60.9	6.468531	58.86118	2.158597	57.64078	2.171504
1890	61.8	1.477833	60.1234	2.144415	58.87072	2.133803
1891	62.3	0.809061	61.4039	2.129777	60.1089	2.10322
1892	59.1	-5.13644	62.70259	2.115006	61.38798	2.127942
1893	58	-1.86125	64.01952	2.100269	62.76254	2.239129
1894	61.8	6.551724	65.35433	2.085012	64.26426	2.392704
1895	64.2	3.883495	66.7061	2.068369	65.87721	2.509879
1896	68.8	7.165109	68.07353	2.049932	67.56083	2.555683
1897	69.6	1.162791	69.45508	2.029493	69.25775	2.511697
1898	72.8	4.597701	70.84927	2.007333	70.92303	2.404463
1899	76.3	4.807692	72.25466	1.983628	72.51512	2.244826
1900	76.4	0.131062	73.66997	1.958788	74.01128	2.063232
1901	75.8	-0.78534	75.09436	1.933472	75.42658	1.912273
1902	76.6	1.055409	76.52724	1.908103	76.79999	1.82086
1903	75.5	-1.43603	77.9681	1.882799	78.17422	1.789367
1904	76.2	0.927152	79.41642	1.857587	79.58999	1.811034
1905	81.7	7.217848	80.87146	1.832169	81.06124	1.848542
1906	85.9	5.140759	82.33215	1.806183	82.56805	1.858855
1907	88.7	3.259604	83.79749	1.779789	84.09687	1.851588
1908	82.6	-6.87711	85.26684	1.753461	85.66747	1.867609
1909	83.6	1.210654	86.74007	1.727787	87.34566	1.958951
1910	84.9	1.555024	88.21677	1.702433	89.16655	2.084701
1911	90.8	6.949352	89.6962	1.677039	91.12783	2.199568
1912	93.3	2.753304	91.17731	1.651251	93.1845	2.256905
1913	100	7.181136	92.65915	1.625233	95.28828	2.257656

Table B.3: Data for Figure 3.1. Volume of Per Capita Industrial Production (UK 1900=100). Source: Bairoch (1982)

year	**UK**	**US**	**Japan**	**France**	**Germany**	**Russia**
1750	10	4	7	9	8	6
1800	16	9	7	9	8	6
1830	25	14	7	12	9	7
1860	64	21	7	20	15	8
1880	87	38	9	28	25	10
1900	100	69	12	39	52	15
1913	115	126	20	59	85	20
1928	122	182	30	78	101	20
1953	210	354	40	90	138	73
1980	325	629	353	265	393	252