Growth with Cross-Sectional Heterogeneity

Thesis submitted for the degree of
Doctor of Philosophy (PhD) by Claudio Michelacci,
registered at the London School of Economics

May 1998

Abstract

Different agents experience different histories and pursue different economic functions. This implies that once a picture of the economic system is taken, a lot of cross-sectional heterogeneity appears. This thesis consists of four essays each one of them makes a case where the intrinsic heterogeneity of the economic system is crucial for understanding macroeconomic performance.

Firstly, it shows when and how an increase in the level of business cycle volatility harms the growth process in a Keynesian world where the decision to free resources and to take advantage of them lies on different agents.

Secondly, it analyses the effects of an increase in research effort in a Schumpeterian world where innovation requires an entrepreneur to implement a valuable invention. In this context the observed decreasing returns in *R&D* might be the outcome of lack of entrepreneurial skills rather than any vanishing of investment opportunities.

Thirdly, it extends the Solow-Swan growth model allowing for cross-sectional heterogeneity. In doing so it reconciles apparently conflicting results on cross-sectional convergence and stochastic output dynamics.

Finally, it argues that cross-sectional heterogeneity is an important transmission mechanism. In the context of a stylized vintage model it is shown how the mechanism generating heterogeneity in the real world also generates persistence in the aggregate fluctuations. Moreover, as aggregate shocks create very high degree of persistence without affecting either the number of firms in the market or technological progress, this degree of persistence is simply attributed to cross-sectional heterogeneity.

Contents

List of Figures

List of Tables

0.1 Acknowledgements

Writing a PhD dissertation might be a very tough experience. You leave your native country, you lose contact with some very good friends, you miss a lot of potentially enriching experiences. During the last three years, however, I have been so lucky that I will never regret what I missed. I would like to thank all of them that contributed to this. I am deeply indebted to a great person and excellent supervisor, Chris Pissarides, that taught me both how to write (what I hope to be) good papers in economics and mostly important that nothing is worth the loss of a smile. The Centre for Economic Performance and the London School of Economics have been an incredibly lively place to work in.

Finally a special thanks to my parents, who never doubted I would make it, to the friends I lost, who will always remain deeply in my heart and to the friends I have because they are my greatest wealth.

0.2 Introduction

Different agents experience different histories and pursue different economic functions. This implies that once a picture of the economic system is taken, a lot of cross-sectional heterogeneity appears. This observation is neither new nor particularly deep unless it is shown that the intrinsic heterogeneity of the economic system is crucial for understanding macroeconomic performance. This thesis consists of four essays each one of them makes a case where cross-sectional heterogeneity might be relevant for macroeconomics.

The first chapter, *Long-Run Growth and Business Cycle Volatility*, analyses when and how an increase in the level of business cycle volatility harms the growth process. In a Keynesian world (Keynes, 1936) where the decision to free resources and to take advantage of them lies on different agents, the business cycle affects growth through two margins: the incentive to do research and implement ideas and the amount of resources available in the economy. In this chapter I show how the business cycle creates disequilibrium and imbalances in the economy. In fact, it destroys the possible good balance between amount of resources available and incentive to use them, raising the former and reducing the latter in recession and vice-versa during boom. As both margins are strictly required to innovate, an economy with high level of business cycle volatility will not be able to innovate much.

The second chapter, *Lack of Entrepreneurial Skills and Decreasing Returns in R&D*, analyses the effects of an increase in research effort in a Schumpeterian world (Schumpeter, 1939) where innovation requires an entrepreneur to implement a valuable invention. In standard endogenous growth models, the higher the research effort the higher is the innovation rate of the economy, but if research and entrepreneurial skills compete in the allocation of aggregate resources, the relation between growth and $R\&D$ is hump-shaped. This chapter proposes a general equilibrium model of endogenous growth in which the observed decreasing returns in $R\&D$ might be the outcome of lack of entrepreneurial skills rather any of vanishing of investment opportunities. If this is so, the amount of resources devoted to research has been excessive and there may be a case for policy intervention. Indirect inference on the model and a temptative empirical investigation seem to support the relevance of this hypothesis.

The third chapter, *(Fractional) Beta Convergence*, jointly written with Paolo Zaffaroni, notes how unit roots in output (Nelson and Plosser, 1982) , an exponential 2% rate of convergence (Barro, 1991) and no change in the underlying dynamics of output (Jones, 1995a) seem to be three stylized facts that can not go together. This chapter extends the Solow-Swan growth model allowing for cross-sectional heterogeneity. In this framework, aggregate shocks might vanish at a hyperbolic rather than at an exponential rate. This implies that the level of output can exhibit long memory and that standard tests fail to reject the null of a unit root despite mean reversion. Exploiting secular time series properties of GDP, we conclude that traditional approaches to test for uniform (conditional and unconditional) convergence (see i.e Barro 1991, Barro and Sala-I-Martin 1991) suit first step approximation. It shows both theoretically and empirically how the uniform 2% rate of convergence repeatedly found in the empirical literature is the outcome of an underlying parameter of fractional integration strictly between 0.5 and 1.

This is consistent with both time series and cross-sectional evidence recently produced.

The fourth chapter, *The Macroeconometrics of Cross-Sectional Heterogeneity*, starts from the observation that detrended aggregate time series display a very high degree of persistence. In frequency domain this shows up in a typical spectral shape (Granger, 1966). The chapter uses this well known empirical regularity as testing bench for a simple transmission mechanism: cross-sectional heterogeneity. To highlight its relevance the chapter analyses a vintage model (see i.e. Solow, 1960) where all aggregate uncertainty takes the form of a reallocative shock that can be driven by either nominal or real factors.

The chapter argues that this transmission mechanism is an important one as it is powerful, realistic, robust and general. It is *powerful* because a sufficient amount of cross-sectional heterogeneity is able to generate typical spectral shapes without necessarily relying on real shocks. It is *realistic* because it is shown how two of the most striking empirical regularities in economics, the typical spectral shape and the Gibrat's law, might be just two faces of the same coin. It is *robust* because it is able to generate almost any kind of spectral shapes provided that reallocations cleanse the economy reallocating between very low and very high technological states. It is *general* because the chapter gives micro-foundation for fractional cointegration of aggregate economic variables high-lightening why a typical spectral shape might be such a generalized phenomenon.

Chapter 1

Long-Run Growth and Business Cycle Volatility

Abstract

Before implementation, a new idea is a private good as it is both rivalrous and excludable. Its widespread economic consequences arise only when a researcher finds in the market resources that suit its economic applicability. In this context we analyse how an increase in the size of business fluctuations (business cycle volatility) affects the growth process. The business cycle affects growth through two margins: the incentive to do research and implement ideas and the amount of resources available in the economy. In this chapter we show how the business cycle creates disequilibrium and imbalances in the economy. In fact, it destroys the possible good balance between amount of resources available and incentive to use them, raising the former and reducing the latter in recession and vice-versa during boom. As both margins are strictly required to innovate, an economy with high level of business cycle volatility will not be able to innovate much. This and other implications of the model seem to broadly match recent both theoretical and

empirical evidence.

1.1 Introduction

The close interrelationship between productivity growth and business cycles has long been recognized (see for example Harrod, 1939 and Schumpeter, 1939). Armed with the newly developed tools of endogenous growth theory (pioneered among others by Romer 1986, Lucas 1988, Stokey, 1988, Romer, 1990, Grossman and Helpman, 1991 and Aghion and Howitt, 1992) we come back to the origins of economic thought to analyse how the amplitude of business fluctuations affect the growth process.

The argument of this chapter is based on three premises.

The first is that widespread economic consequences of a new idea arise with implementation. As pointed out by Romer (1990) economists studying public finance have identified two fundamental attributes of any economic good: the degree to which it is rivalrous and the degree to which it is excludable. Before implementation, a new idea is a private good as it is both rivalrous and excludable. It is rivalrous because the extra-profits associated with its implementation accrues to who is first able to exploit its economic applicability. It is excludable because in principle a researcher is able to keep its content secret until it is first implemented. After implementation it becomes, at least partially both non rivalrous and non excludable as the society learns from the experience of the inventor at almost no cost. That the economic consequences of an invention arise with implementation is at the core of the Schumpeterian theory. In fact, Schumpeter (1934) notes how "an idea or scientific principle is not, by itself, of any importance for economic practice: the fact that Greek science had probably produced all that is necessary in order to construct a steam engine did not help the Greeks or Romans to build a steam engine; the fact that Leibnitz suggested the idea of the Suez canal exerted no influence what ever on economic history for two hundreds years". In Romer (1990), for example, the implementation of a new idea breaks the binding constraint of decreasing marginal productivity simply because final good producers are able to diversify across a greater number of intermediates, while Helpman and Trajtemberg (1994) and Aghion and Howitt (1996) note that firms learn to use a new technology, not discovering everything on their own but using and or learning from the experience of other firms in similar situations. Technological spill-overs spread across society through a process of social learning.

The second premise is that the implementation of a new idea is not a mechanicistic process. The researcher must not only search in the market for resources that suit his idea, but must also compete in doing so with other researchers with the same intent. As a result, the amount of resources available in the market are a scarce good that different researchers with competing ideas try to take advantage of.

The third is that the resources required to implement new ideas must be available and free to be utilised in a new economic process (hereafter we refer to them as *slacked*). The basic intuition here, is a Schumpeterian (1934) one, that is "firms must be driven into the bankruptcy court and people thrown out of employment, before the ground is clear and the way paved for new achievement of the kind which has created modern civilizations and made the greatness of this country". As a result, the market is populated by resources freed by old investors that potential new investors try to take advantage of. The fact that the supply of resources and the desire to take advantage of them lie in different subjects is a old Keynesian (1936) idea. It implies that an increase in the amount of resources that are kept slacked does not necessarily increase the innovation rate of the economy unless it increases the prospective yield of an innovation[1]. Consequently, a lack of coordination between the "demand" and the "supply" of slacked resources may arise.

We then go on to show how cross-sectional heterogeneity together with the central role of the implementation of a new idea (hereafter simply *innovation*) sheds new light on the relation between the size of the business fluctuations and the growth process. We will see how an increase in the size of business fluctuations (hereafter *business cycle volatility*) raises the average total amount of slackness in the market, the average profitability of an innovation and therefore stimulates the total amount of research pursued in the economy. Moreover, the higher the level of business cycle volatility the higher the degree of renewal in the economy as the "troubles" associated with recessions "are the means to reconstruct each time the economic system on a more efficient plan" (Schumpeter, 1934). As a result, the higher the level of business cycle volatility the closer the economy is to its technology frontier. Notwithstanding this, the level of business cycle volatility in the economy can be excessive and harm the growth process. In order to inno-

[1]Keynes (1936) clearly explains this point with regard to saving and investment decisions: "If, therefore an act of saving does nothing to improve prospective yield, it does nothing to improve investment".

vate an economy, requires both slacked resources (*slackness*) and researchers trying to implement new ideas (*implementation effort)*. The business cycle induces a wedge between the decision to free resources and the decision to take advantages of them, so that at high levels of business cycle volatility the economy fails to coordinate properly. In fact, the business cycle breaks the required to grow balance between slackness and implementation effort. On the one hand, during a boom slackness is binding and it is more so the bigger the size of the expansion, on the other hand during a recession the willingness to take advantage of slackness is binding and it is more so the deeper the size of the slump. Despite the increase in the average amount of slackness and implementation effort, resources never match because once one is high the other is low. As a result an economy with high level of business cycle volatility will not be able to innovate much.

The main contribution of the chapter is to show that in a world in which recessions cleanse the economy and permanently increase the long-run level of output, an increase in both the average amount of slackness and implementation effort, is not enough to stimulate growth. For an economy to grow, it requires a good balance between slackness and implementation effort. The business cycle breaks this balance and creates disequilibrium and imbalances. In section 6 we will relate these results to both theoretical and empirical strands of research in the literature. We then note how most of the predictions of the model seems to broadly match empirical evidence. It is only cross-sectional heterogeneity that might be able to explain why business fluctuations harm the growth process, in a world in which averages seem to stimulate growth.

Section 2 highlights further the general set-up of the model. Section 3 presents the model in steady states. The analysis confirms standard results in endogenous growth theory. It also shows how the correlation between growth rates and excess capacity in the market is spurious, as first pointed out by Bean and Pissarides (1993). This generalises recent theoretical and empirical results. Section 4 introduces business fluctuations in the basic set up and shows the basic intuition behind the fact that business fluctuations can be excessive. Section 5 analyses more formally how an increase in business cycle volatility affects long- run growth. Section 6 analyses the patterns of technological adoption of the economy at higher levels of business cycle volatility. Section 7 relates the chapter to both theoretical and empirical strands of research in the literature. Section 8 concludes. The appendixes contain the derivation of some results used in the text.

1.2 The General Set-Up

The three premises highlighted in the introduction imply that an innovation requires the simultaneous occurrence of three elements. First of all it requires the existence of an idea suitable for economic applicability, second the availability of resources that suit the economic exploitation of the idea and last, but not least, that the profitability of the innovation be higher than the cost of recovering the required resources in the market.

To evaluate these claims, it might help to have a specific case in mind. Suppose that a researcher gets the (very strange) idea of producing quasi-disposable plastic watches. In this phasis the researcher ensures that the content of his idea is kept secret in order to avoid other competitors exploiting its economic potential. The researcher must then decide if it is worth either entering the market and starting a costly search process in order to find engineers specialised in plastic materials, machines to melt plastics and deal with aluminium and possibly employ a top manager with a good knowledge in engineering, marketing and accounting or waiting for more suitable time. If the expectation of future profits are higher than the cost of recovering the required resources, the researcher can decide that is worth trying to implement his idea. If eventually the process ends successfully, the idea becomes a public good from which other competitors and in general the whole society (like car or mobile phones manufacturers) can learn from. Widespread economic consequences of an idea arise at this point therefore through an innovation. This basic structure of the process of technical change is described in Figure 1.1.

In order to model these features of technical change, we assume that vertical innovations arise when a researcher with an idea matches with a free entrepreneur. As in Schumpeter (1949) "when we speak of the entrepreneur we do not mean so much a physical person as we do a function". Following Schumpeter (1947) we define the word entrepreneurial function, as the complex number of activities idea specific that are required in order to "get things done"[2].

To capture the fact that the required resources are idea specific and so

[2]The interaction between research and entrepreneurial activities can also occur across independent departments inside the same firm. In any case it seems unlikely that even very large firms can find all the required resources for innovating without entering the market.

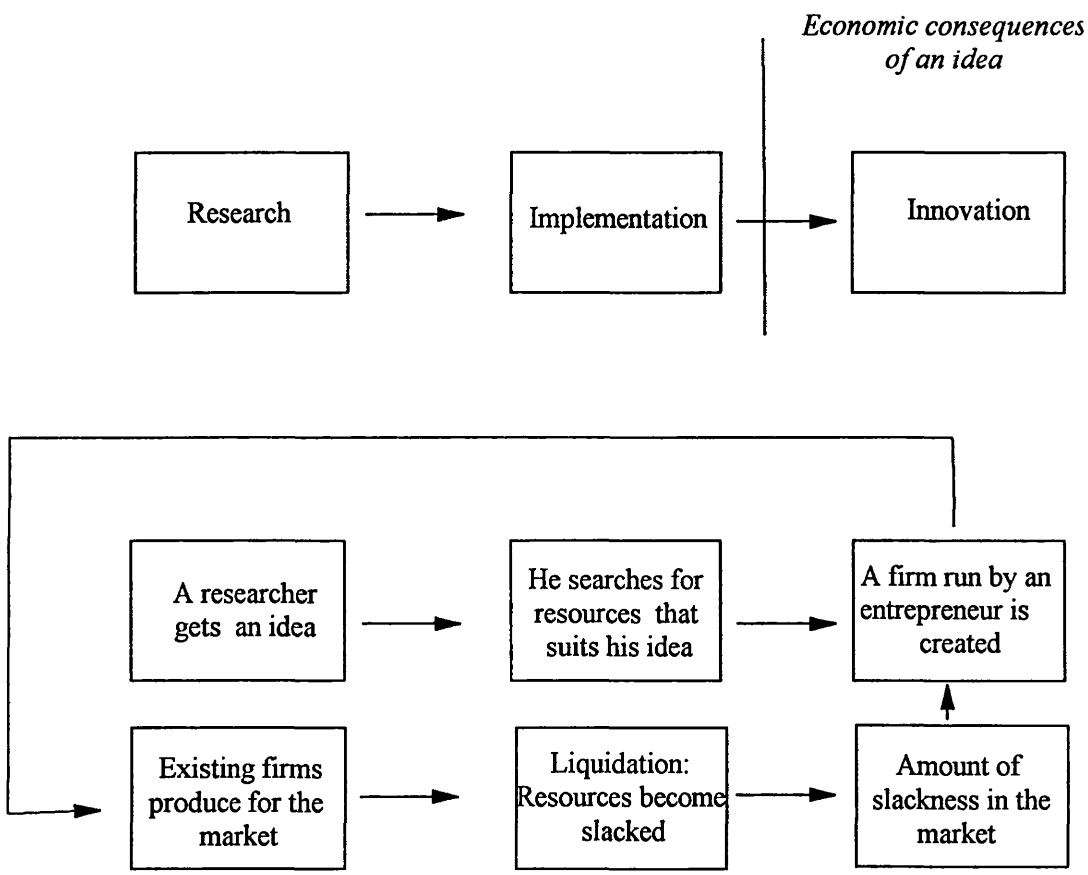

Figure 1.1: The Structure of the Innovation Process

that finding them is time consuming, we rely on an homogeneous-of-degree one matching function (see for example Pissarides, 1990). The matching function gives the total number of matches between researchers with an idea and free entrepreneurs as an increasing but concave function of the total number of free entrepreneurs and researchers with an idea to implement. The matching function allows to represent in a very parsimonious way two key characteristics of the implementation of an idea: the fact that resources are heterogeneous so that search is time consuming and that the amount of resources available in the market are a scarce good different researchers are competing for.

The basic structure of the model is described in the second row of Figure 1.1. When a researcher gets an idea suitable for economic applicability. He can then either look for a free entrepreneur or wait for better time (this is the Leibnitz case with the Suez canal). If he opts for the first alternative and is able to find a free entrepreneur, a new firm run by the entrepreneur is created and it starts producing for the market. The probability of finding a free entrepreneur depends on how many other researchers are competing for the same resources and on the number of entrepreneurs available. If we assume that an entrepreneur already running a firm is not therefore available to implement a new idea (the third premise highlighted in the introduction), then resources must become slacked before being employed again[3].

[3]Schumpeter (1949) observes how "the man whose mind is entirely absorbed by a struggle for entrepreneurial success has as a rule very little energy left for serious activity in any other direction". The basic intuition is either that an entrepreneur involved in a profitable activity is not interested in searching for something else or more generally that searching for a new economic idea to implement is a costly activity incompatible with running a firm.

We now proceed to describe the formal structure of the model.

Within our economy there exists a continuum of researchers and a populations of entrepreneurs of measure C. Both entrepreneurs and researchers are infinitely lived, risk-neutral and maximise expected returns in output units discounted at rate $r > 0$.

A researcher gets ideas suitable for economic application according to a Poisson process with intensity κ. He must then decide if it is worth either trying to implement the idea now or waiting. If he opts for the first alternative, he starts searching for a free entrepreneur. The flow cost of searching is given by $\omega x(t)$ where $x(t)$ indexes the leading technology in the economy at time t. As in the meantime the technological environment changes, the researcher must also update his discovery during the search process. We assume that old ideas depreciate at the same rate as the rate of growth of the leading technology, $g_t = \frac{\dot{x}(t)}{x(t)}$, so that at each point in time a fraction g_t of the value of the idea is lost in keeping it up-to-date.

An entrepreneur can be either running an enterprise or producing at home. If he runs a firm, the entrepreneur gets a real flow of profit equal to $\pi(\epsilon, \tau, t)$ where ϵ is a firm specific component of productivity, τ is the date of creation of the firm and t is the current date. Profits of the entrepreneur are chosen so as to share with the researcher the gains from the innovation at each point in time. The entrepreneur's share is β. If he works at home, the entrepreneur is able to produce a flow of goods equal to $hx(\tau)$ where $x(\tau)$ is the technology that the entrepreneur last used in a firm. $hx(\tau)$ is a measure of the level of human capital of the entrepreneur at time $t > \tau$. We assume that the entrepreneur can only upgrade his level of human capital when he meets a researcher with a new idea. This implies that τ is the date of creation of the enterprise that the entrepreneur last ran. Upgrading human capital, however, is costly. In fact, the entrepreneur must allocate to the scope, a fraction $\Gamma(\tau,t) = [x(t) - x(\tau)]/x(t)$ of his wealth. This cost is sunk and is paid before the meeting takes place, therefore it is not shared with the researcher.

This way of modelling the process of accumulating human capital implies that the upgrading cost depends on how far left behind the level of human capital of the entrepreneur was left, on the leading technology currently used, and that it has the nature of an interaction cost that must be paid, for the meeting to take place and be profitable. For instance, in the quasi-disposable plastic watches example, an engineer that had previously worked in the food industry must first update his human capital before knowing whether he suits the specific requirement of the new idea. Thus the timing of the bargaining is as follows: first the entrepreneur updates his human capital and then the researcher and the entrepreneur bargain to share the potential profits of the firm exploiting the new idea.

An innovation requires a researcher with an idea and a free entrepreneur. Once they meet, a new firm run by the entrepreneur is created. Following the third premise in the introduction we assume that an entrepreneur running a firm is not available for implementing new ideas.

An innovation that occurs at time τ will open access of that enterprise to the leading technology $x(\tau)$ as of that date. Each firm created at time τ is characterised by a fixed irreversible technology and produces at each point in time a flow of goods equal to $x(\tau)(P + \sigma\epsilon)$. P and σ are common to all enterprises in the market, whereas ϵ is firm specific. P is an aggregate component of productivity. In this section we assume that all aggregate uncertainty is completely solved so that P is a constant. The parameter σ reflects idiosyncratic volatility, an increase in σ representing a symmetric mean preserving spread in the dispersion of the firm specific component $\sigma\epsilon$.

As in Mortensen and Pissarides (1994) the process that changes the idiosyncratic component ϵ is Poisson with arrival rate λ. When there is a change, the new value of ϵ is a drawing from the fixed distribution $F(y)$, which has bounded support $[\epsilon^l, \epsilon^u]$ and no mass points. Firms are always created with an idiosyncratic component equal to the upper support of the distribution ϵ^u. Once the firm is created, however, the entrepreneur has no choice in the firm's productivity. Thus, the productivity of the firm is a stochastic process with initial condition the upper support of the distribution F and terminal state the level of idiosyncratic productivity at the firm closing date. The firm is shut down because either the idiosyncratic productivity falls below some critical value $\epsilon^d < \epsilon^u$ or because of a catastrophic exogenous event that arrives with frequency δ. This implies that at each point in time existing firms are closed at rate $\lambda F(\epsilon^d) + \delta$.

The rate at which researchers with ideas and free entrepreneurs meet is determined by the homogeneous-of-degree one matching function $m(i_t; s_t)$. Where i_t is the number of researchers trying to implement an idea and s_t is the number of entrepreneurs working at home. i_t is a measure of the aggregate level of *implementation effort* in the economy at time t while s_t is a measure of the aggregate amount of resources suitable for economic exploitation (*slackness*). We assume that the matching function is increasing and concave in each of its arguments. Since prospective new firms offer the highest human capital opportunities in the market and some profits, no free entrepreneur turns down a firm-creation opportunity. The probability that a researcher meets an entrepreneur is given by $q(\theta_t) = m(i_t; s_t)/i_t$. Where θ_t is defined as i_t/s_t and measures the *willingness* to take advantage of a given amount of slackness at time t. By analogous considerations it follows that $\theta_t q(\theta_t)$ is the instantaneous probability that a free entrepreneur finds a researcher with an idea. So that $z = \theta_t q(\theta_t)$ measures, at the aggregate level, the *actual use* that is made at time t of a given amount of slackness.

We assume that vertical innovations are the unique source of growth. We follow Aghion and Howitt (1992, 1994) in assuming that the rate of growth of the technological parameter $x(t)$ is given by the product of the size of the innovation (γ) and the frequency of innovation. Our assumptions imply that $\theta_t q(\theta_t) s_t$ is the number of innovations created by the economy at time t, so that the growth rate g_t of the technological parameter $x(t)$ is given by $\theta_t q(\theta_t) s_t ln\gamma$.

The unknowns of the model are the level of willingness (θ_t), the growth rate of the technological parameter (g_t) and the level of the critical idiosyncratic productivity (ϵ^d) that determines the amount of firms that are closed at each point in time.

1.2.1 The R&D Sector

The assumption that a researcher gets an idea with intensity κ implies that

$$R(t) = E_{\tau \geq 0}\{[sup(D(t+\tau); 0)]e^{-r\tau}\}, \tag{1.1}$$

where $R(t)$, and $D(t+\tau)$ are respectively the value at time t of doing research, and the value of trying to implement the idea at time $t+\tau$. The above equation embodies the option for the researcher to wait for better time, so that $D(t+\tau)$ is always greater than 0.

If the researcher decides that it is worth trying to implement his idea he enters the market and searches for a free entrepreneur. The assumption that searching costs $x(t)\omega$ and that ideas depreciate at the rate of growth of the technological parameter $x(t)$, g_t, implies that the value of the attempt of implementation, $D(t)$, follows the Bellman equation

$$(r+g_t)D(t) = -\omega x(t) + q(\theta)[I(\epsilon^u;t,t) - D(t)] + \dot{D}, \tag{1.2}$$

where $I(\epsilon^u, t, t)$, and $\dot{D}$ are respectively the present discounted value of the rents of an innovation introduced at time t with idiosyncratic component ϵ^u and the time derivative of $D(t)$.

Free entry until the exhaustion of all rents from research implies that $R(t) = 0$. Using equation (4.10) and (1.2) researchers will try to implement ideas up to the point that

$$I(\epsilon^u;t,t) = \omega\frac{x(t)}{q(\theta)}, \tag{1.3}$$

so that the value of an innovation, $I(\epsilon^u;t,t)$, equates the expected cost of implementation $(\omega\frac{x(t)}{q(\theta)})$.

1.2.2 The Value of an Innovation, Human Capital and Profits

Since researchers and entrepreneurs have the option to close non-profitable enterprises at no cost, a firm will be shut down only if either its value is below zero or the exogenous event with frequency δ arrives. For any idiosyncratic component ϵ, the asset value at time t of the rents from an innovation introduced at time τ, $I(\epsilon, \tau, t)$, solves

$$\begin{aligned} rI(\epsilon,\tau,t) &= Px(\tau)+\sigma\epsilon x(\tau)-\pi(\epsilon,\tau,t)+\lambda[\int_{\epsilon^l}^{\epsilon^u} max(I(s,\tau,t);0)dF(s)+ \\ &- I(\epsilon,\tau,t)]+\delta[0-I(\epsilon,\tau,t)]+\dot{I}. \end{aligned} \tag{1.4}$$

$(P+\sigma\epsilon)x(\tau)$, $\pi(\epsilon,\tau,t)$ and $\dot{I}$ are respectively the flow of goods that a firm created at time τ is able to produce, the profits that the entrepreneur gets for running the firm and the time derivative of the rent from innovation. The assumption that profits are the outcome of a bilateral bargain between the researcher with an idea and the entrepreneur implies

$$\beta[I(\epsilon,\tau,t)-R(t)] = (1-\beta)[E(\epsilon,\tau,t)-H(\tau,t)], \tag{1.5}$$

where $E(\epsilon,\tau,t)$ measures the value at time t of an enterprise created at time τ with idiosyncratic component ϵ and $H(\tau,t)$ measures the asset value of an entrepreneur working at home whose human capital is given by $hx(\tau)$.

The assumption that the entrepreneur can update his human capital only when he meets a researcher with an idea and that he has to pay as upgrading cost, a fraction $\Gamma(\tau,t)$ of his wealth implies that

$$rH(\tau,t) = hx(\tau)+\theta q(\theta)\{[(1-\Gamma(\tau,t)]E(\epsilon^u,t,t)-H(\tau,t)\}+\dot{H},$$

where $\theta q(\theta)$ is the probability of meeting a researcher with a new idea and $\dot{H}$ represents the time derivative of H.

Similarly the value of an enterprise, $E(\epsilon, \tau, t)$, follows the Bellman equation

$$\begin{aligned} rE(\epsilon,\tau,t) &= \pi(\epsilon,\tau,t) + \lambda[\int_{\epsilon^l}^{\epsilon^u} max(E(s,\tau,t), H(\tau,t))dF(s) + \\ &- E(\epsilon,\tau,t)] + \delta[H(\tau,t) - E(\epsilon,\tau,t)] + \dot{E}, \qquad (1.6) \end{aligned}$$

where $\dot{E}$ indicates the time derivative of $E(\epsilon, \tau, t)$. Using equations (1.4), (1.6) and substituting in (1.5) we find that profits are equal to

$$\begin{aligned} \pi(\epsilon,\tau,t) &= \beta(P+\sigma\epsilon)x(\tau) + (1-\beta)\theta q(\theta)hx(\tau) + \\ &+ (1-\beta)\theta q(\theta)\{[1-\Gamma(\tau,t)]E(\epsilon^u,t,t) - H(\tau,t)\}. \qquad (1.7) \end{aligned}$$

Substituting this expression into equation (1.4) and after integration by parts we find that

$$\begin{aligned} (r+\delta+\lambda)I(\epsilon,\tau,t) &= (1-\beta)x(\tau)(P+\sigma\epsilon-h) + \\ &- (1-\beta)\theta q(\theta)\{[1-\Gamma(\tau,t)]E(\epsilon^u,t,t) - H(\tau,t)\} + \\ &+ x(\tau)\frac{\sigma\lambda(1-\beta)}{r+\delta+\lambda}\int_{\epsilon^d}^{\epsilon^u}[1-F(s)]ds + \dot{I}. \qquad (1.8) \end{aligned}$$

Basic dynamic programming arguments guarantee that the value of an innovation strictly increases in the idiosyncratic component ϵ, so that a firm will decide to shut down for all idiosyncratic components lower than a given reservation productivity ϵ^d. We conjecture that

$$\begin{aligned} I(\epsilon;\tau,t) &= I(\epsilon;\tau,\tau) = I(\epsilon)x(\tau), \qquad \forall\epsilon,\ \forall\tau,\ \forall t, \\ E(\epsilon;\tau,t) &= E(\epsilon;\tau,\tau) = E(\epsilon)x(\tau), \qquad \forall\epsilon,\ \forall\tau,\ \forall t. \end{aligned}$$

As our conjecture turns out to be satisfied in equilibrium, using equation (1.8) and the previous considerations, we find that the reservation productivity (ϵ^d) solves

$$P - h + \sigma\epsilon^d + \frac{\sigma\lambda}{r+\delta+\lambda}\int_{\epsilon^d}^{\epsilon^u}[1-F(s)]ds - \frac{\beta}{1-\beta}\theta q(\theta)I(\epsilon^u) = 0. \qquad (1.9)$$

Moreover given our conjecture on the functional form $I(\epsilon;\tau,t)$, the free entry condition (1.3) can be written as

$$I(\epsilon^u) = \omega\frac{1}{q(\theta)}, \qquad \forall t. \qquad (1.10)$$

Equation (1.10) defines the willingness to take advantage of a given amount of slackness in the market, θ, as a monotone increasing function of the value of an innovation ($\dot{q}(\theta) < 0$). The higher the value of an innovation, the bigger is the incentive to enter the market and try to implement an idea.

1.2.3 Steady State Analysis

The assumption that vertical innovations are the unique source of growth and that γ is the size of technological improvement implies that the growth rate of the technological parameter $x(t)$ is given by

$$g_t = \frac{\dot{x}(t)}{x(t)} = \theta_t q(\theta_t)\, s_t\, ln\gamma, \qquad (1.11)$$

where $\theta_t q(\theta_t) s_t$ is the number of innovations created by the economy at time t. As the population of entrepreneurs has fixed measure C, and at each point in time firms are shut down at rate $\lambda F(\epsilon^d)+\delta$, slackness in the market evolves according to the differential equation

$$\dot{s}_t = [\delta + \lambda F(\epsilon^d)](C - s_t) - \theta q(\theta) s_t.$$

This implies that in steady state the amount of slackness in the economy is equal to

$$s = C\,\frac{\delta + \lambda F(\epsilon^d)}{\delta + \lambda F(\epsilon^d) + \theta q(\theta)}. \tag{1.12}$$

As a result, the steady state rate of growth, g, is equal to

$$g = z\,C\,\frac{\delta + \lambda F(\epsilon^d)}{\delta + \lambda F(\epsilon^d) + z}\,ln\gamma, \tag{1.13}$$

where we have indicated with $z = \theta q(\theta)$ the actual use that is made of the steady state amount of slackness. This equation defines the growth rate of the leading technology $x(t)$, g, as the product of the steady state value of slackness, s, given by equation (1.12) times the actual use that is made of it, z. As the probability that a firm will last forever is equal to 0, g is also the expected long-run growth rate of the economy. Equation (1.13) together with equations (1.9) and (1.10) and the auxilium of (1.5) and (1.8) solve the model for the key parameters g, ϵ^d and θ.

This simple simultaneous equation system in three equations with three unknowns has several implications. C is the size of the population of entrepreneurs and measures the total amount of resources available in the economy. Equation (1.13) shows how our economy exhibits scale effect. The bigger the size of the economy, the easier it is to implement new ideas. The higher the amount of implementation effort sustained by the economy, the higher is the innovation rate in equilibrium. This is a very well known result in endogenous growth theory (see, for example, Barro and Sala-I-Martin 1995)[4].

[4]The existence of scale effect in standard endogenous growth models has recently been questioned by Jones (1995a, 1995b). In a related chapter, I show why the empirical observation of decreasing returns in *R&D* does not exclude the possible relevance of scale

This formula is also consistent with results found by Romer (1990a, 1990b). In order to sustain growth we must assign some degree of market power to researchers. If for example β is equal to 1, all the gains of an innovation would accrue to the entrepreneurs and the value of an innovation would be equal to 0 (see equation 1.8)[5]. In this case the willingness to take advantage of slackness would tend to be equal to 0 and no idea would be implemented in equilibrium. In order to sustain growth we need a value of β strictly less than 1 so that a positive value of willingness can be sustained in equilibrium.

The above formula can be read also in another way. It shows how the correlation between the long-run rate of growth and the amount of slackness is spurious. If we think of the unemployment rate in an economy as positively correlated to the amount of slackness, this very simple formula generalises the results by Pissarides (1990) (Chapter 2), Aghion and Howitt (1994) and Mortensen and Pissarides (1995). For example, an increase in the value of an innovation, drives through equations (1.9) and (1.10) an increase in willingness θ and as a consequence an increase in the actual use made of slackness z. This increases the growth rate and reduces the amount of slackness in the market given by equations (1.13) and (1.12) respectively. On the other hand, an increase in the liquidation rate $\lambda F(\epsilon^d) + \delta$ that does not alter the value

effects. The main intuition is that if research crowds out more socially useful entrepreneurial skills the growth rate of the economy can fall despite a rise in the amount of research effort. In this case, the observed decreasing returns in $R\&D$ would be the outcome of the lack of entrepreneurial skills (see chapter 2.

[5]The world analysed by Romer (1990), Romer (1990b), Aghion and Howitt (1992) and more generally by Barro and Sala-I- Martin (1995), roughly corresponds to the case $\beta = 0$. In fact, in these models, the researchers accrue all the rents from innovating. Here, on the contrary, the rentiers are two.

of an innovation, does not alter willingness θ and so the actual use made of slackness z remains also unchanged. In this case, the growth rate tends to increase as well as the amount of slackness. As most changes tend to affect both willingness and the liquidation rates, we have at hand a very simple explanation for the lack of empirical correlation between growth rates and the amount of slackness in the market, see for example Bean and Pissarides (1993) and Caballero (1993).

As the next section deals with the effects of changes in business cycle volatility on long-run growth, it is interesting to investigate the effects of a change in idiosyncratic volatility. An increase in idiosyncratic volatility is equivalent to an increase in σ. This is easy to assess, differentiating equation (1.8),(1.9) and (1.10), as it increases the value of an innovation, of willingness, θ and thus of the actual use made of slackness, z. Moreover, also the reservation productivity ϵ^d increases, driving a rise in the liquidation rate of the economy. This implies that the growth rate will tend unequivocally to increase even if the effects on the average amount of slackness are ambiguous. It might be useful to understand why this happens. The reason is that an increase in σ drives at the same time an increase in the value of an innovation and in the amount of resources that are freed and so available in the economy. The business cycle, on the contrary, induces a wedge between the decision to free resources and to take advantage of them. It raises the amount of liquidation in the economy but reduces the willingness to use slackness in the recession and vice-versa in the boom. Because of this reason, an increase in business cycle volatility can result in very different effects from those caused by idiosyncratic volatility, as it breaks the required to grow balance between implementation effort and slackness in the aggregate economy.

1.3 Cyclical Fluctuations

In this section we extend the model to the case where one of the variable changes probabilistically. We model in detail the case where the common productivity component P takes two values, a high value P_B and a low value P_R, according to a Poisson process with rate μ. The Poisson process captures the important feature of cyclical shocks, a positive probability less than one that boom or recession will end within a finite period of time. In this context we will analyse the effects of mean preserving increases in business cycle volatility. We will show how the business cycle drives a wedge between the decision to free resources and to take advantage of them. Because of this, an increase in business cycle volatility breaks the balance between slackness and implementation effort in the economy and eventually harms the growth process.

1.3.1 The R&D Sector

As people discount the change in aggregate regime, the value of doing research at time t in aggregate state i, $R_i(t)$, is equal to

$$\begin{aligned} rR_i(t) &= -\kappa[sup(D_i(t);0) - R_i(t)] + \mu[sup(R_j(t);0) - R_i(t)] + \dot{R}_i(t), \\ &\qquad i = B, R, \qquad j \neq i, \end{aligned} \tag{1.14}$$

where μ is the instantaneous probability that an aggregate switch occurs. The structure of this Bellman equation is exactly as in equation (4.10), where we just added a term to take into account the possible gains and losses associated with a change in aggregate regime. The above equation embodies the option for the researcher to wait for better time, so that when the aggregate environment changes, the researcher can always decide to stop researching.

Reasoning analogously, we find that the value of the implementation attempt at time t, $D_i(t)$, follows the Bellman equation

$$\begin{aligned}(r+g_t)D_i(t) &= -\omega x(t) + q(\theta_t)[I_i(\epsilon^u;t,t) - D_i(t)] + \mu[sup(D_j(t);0) - D_i(t)] + \\ &+ \dot{D}_i, \qquad i = B, R, \qquad j \neq i, \end{aligned} \tag{1.15}$$

if the economy is in state i. θ_t, $q(\theta_t)$, $I_i(\epsilon^u;t,t)$ are respectively the level of willingness in the market for resources at time t, the probability of finding a free entrepreneur and the value of the rents from innovating. The above equation embodies the option for the researcher to wait for better times, so that $D_i(t)$ is always greater than 0. The exhaustion of the rents condition implies that

$$R_i(t) = R_j(t) = 0, \;\; \forall t, \qquad i = B, R, \qquad j \neq i,$$

and using equation (1.14) and (1.15) we find that researchers will try to implement ideas up to the point that

$$I_i(\epsilon^u, t, t) = \omega \frac{x(t)}{q(\theta_t)}, \; \forall t, \qquad i = B, R. \tag{1.16}$$

This is a marginal condition that states that in each aggregate state researchers will find it profitable to try to implement an idea up to the point that the expected implementation cost is equal to the value of an innovation.

1.3.2 Value of an Innovation, Human Capital and Profits

Reasoning in the same way as in the previous section, we find that the value of an innovation in state i, $I_i(\epsilon, \tau, t)$, is equal to

$$\begin{aligned} rI_i(\epsilon,\tau,t) &= x(\tau)(P_i+\sigma\epsilon) - \pi_i(\epsilon,\tau,t) + \lambda[\int_{\epsilon^l}^{\epsilon^u} max(I_i(s,\tau,t);0)dF(s) + \\ -I_i(\epsilon,\tau)] &+ \delta[0 - I_i(\epsilon,\tau,t)] + \mu[max(I_j(s,\tau,t);0) - I_i(\epsilon,\tau,t)] + \dot{I}_i, \\ & \qquad i = B, R, \qquad j \neq i, \end{aligned} \tag{1.17}$$

where t, τ and ϵ are respectively the current date, the implementation date and the idiosyncratic component of productivity. As in the previous section, the equation embodies the option to close non-profitable firms so that the value of an innovation is always greater than 0. $(P_i + \sigma\epsilon)x(\tau)$, π_i are respectively the flow of goods that the firm created at time τ is able to produce and the profit that the entrepreneur gets for running the enterprise in aggregate state i. The assumption that profits and rents are split according to the Nash sharing rule and that the entrepreneur can update his level of human capital only when he meets a researcher with a new idea paying an upgrading cost equal to fraction $\Gamma(\tau, t)$ of his wealth imply that

$$\begin{aligned} \pi_i(\epsilon,\tau,t) &= \beta(P_i+\sigma\epsilon)x(\tau) + (1-\beta)\theta_t q(\theta_t) h x(\tau) + \\ &+ (1-\beta)\theta_t q(\theta_t)\{[1-\Gamma(\tau,t)][E_i(\epsilon^u,t,t) - H_i(\tau,t)]\}, \\ i &= B, R. \end{aligned} \tag{1.18}$$

If we conjecture that

$$I_i(\epsilon;\tau,t) = I_i(\epsilon;\tau,\tau) = I_i(\epsilon)x(\tau), \qquad \forall\epsilon,\ \forall\tau,\ \forall t,$$

$$\begin{aligned} E_i(\epsilon;\tau,t) &= E_i(\epsilon;\tau,\tau) = E_i(\epsilon)x(\tau), \qquad \forall\epsilon,\ \forall\tau,\ \forall t, \\ i &= B, R, \end{aligned}$$

and we then substitute equation (1.18) into (1.17) we find that equation (1.16) can be written as

$$I_i(\epsilon^u) = \omega\frac{1}{q(\theta_i)}, \qquad \forall t, \qquad i = B, R. \tag{1.19}$$

This equation tells us that willingness to use a given amount of slackness can take just two values, equal to θ_R in recession and equal to θ_B in a boom.

We now want to solve for the reservation productivity in boom and recession, equal to ϵ_B^d and ϵ_R^d, respectively. Substituting the equation for profits (1.18) into equation (1.17), we get an expression for the value of an innovation. Basic dynamic programming arguments and the fact that the value of an innovation strictly increases in the idiosyncratic component ϵ, imply that a firm will decide to shut down for all idiosyncratic component lower than a given reservation productivity. If we describe ϵ_R^d as the reservation productivity in a recession and ϵ_B^d as the reservation productivity in a boom, we can then prove (see Appendix A) that

$$\epsilon_R^d \geq \epsilon_B^d.$$

The liquidation rates reflect the value of the option to keep the firm operating. Clearly the higher the value of this option, the lower is the incentive to liquidate and thus the lower is the reservation productivity. During a boom the aggregate value of productivity is higher and higher also is the value of the option to keep the firm operating and this in turn lowers the value of the critical reservation productivity ϵ_B^d . Using equations (1.17), (1.18), the previous considerations, the Nash sharing rule and after integration by parts we find that

$$\begin{aligned} \nu I_i(\epsilon) &= (1-\beta)(P_i - h + \sigma\epsilon) + (1-\beta)\frac{\sigma\lambda}{\nu}\int_{\epsilon_i^d}^{\epsilon^u}[1-F(s)]ds + \\ &- \beta\theta_i q(\theta_i) I_i(\epsilon^u) + \mu I_i(\epsilon), \qquad if\ \epsilon \geq \epsilon_R^d, \qquad i = B, R, \qquad j \neq i. \end{aligned} \tag{1.20}$$

On the other hand, firms with idiosyncratic productivity ϵ greater than ϵ_B^d but lower than ϵ_R^d will operate only in a boom, so that

$$\begin{aligned} \nu I_B(\epsilon) &= (1-\beta)(P_B - h + \sigma\epsilon) + (1-\beta)\frac{\sigma\lambda}{\nu-\mu}\int_{\epsilon_B^d}^{\epsilon_R^d}[1-F(s)]ds + \\ &+ (1-\beta)\frac{\sigma\lambda}{\nu}\int_{\epsilon_R^d}^{\epsilon^u}[1-F(s)]ds - \beta\theta_B q(\theta_B) I_i(\epsilon^u), \qquad if\ \epsilon_B^d \leq \epsilon \leq \epsilon_R^d, \end{aligned} \tag{1.21}$$

while

$$I_R(\epsilon) = 0, \qquad if\ \epsilon_B^d \leq \epsilon \leq \epsilon_R^d,$$

because of the option to close non-profitable firms. The previous considerations and equations (1.20), (1.21) imply that the two reservation productivities ϵ_R^d and ϵ_B^d are such that

$$P_R - h + \sigma\epsilon_R^d = -\frac{\sigma\lambda}{\nu}\int_{\epsilon_R^d}^{\epsilon^u}[1-F(s)]ds + \frac{\beta}{1-\beta}\theta_R q(\theta_R) I(\epsilon^u) - \mu I_B(\epsilon_R^d), \tag{1.22}$$

$$\begin{aligned} P_B - h + \sigma\epsilon_B^d &= -\frac{\sigma\lambda}{\nu-\mu}\int_{\epsilon_B^d}^{\epsilon_R^d}[1-F(s)]ds - \frac{\sigma\lambda}{\nu}\int_{\epsilon_R^d}^{\epsilon^u}[1-F(s)]ds + \\ &+ \frac{\beta}{1-\beta}\theta_B q(\theta_B) I_B(\epsilon^u). \end{aligned} \tag{1.23}$$

Equations (1.22) and (1.23) together with equations (1.19) and (1.20) define a simultaneous equation system with 4 endogenous variables ϵ_R^d, ϵ_B^d, θ_R and θ_B.

1.3.3 Creating an Innovation

At each point in time the rate of growth of the technology parameter $x(t)$ is given by

$$g_t = \frac{\dot{x}(t)}{x(t)} = m(i_t; s_t)\, ln\gamma = \theta_t q(\theta_t) s_t\, ln\gamma,$$

where $\theta_t q(\theta_t) s_t = m(i_t; s_t)$ is the number of innovations created by the economy at time t. As the probability that a firm will last forever is equal to 0, the expected long-run rate of growth is given by the unconditional expected value of $\dot{x}(t)/x(t)$. Condition (1.19) implies that θ_t can take just two values, equal to θ_R in recession and to θ_B in boom. $\theta_t q(\theta_t)$ measures the actual use that is made of slackness at time t. We indicate with z_B and z_R its value in boom and recession respectively. All these considerations and the fact that in the long run the economy will spend half of its time in a recession and half in a boom, implies that the long-run growth rate of the economy is given by

$$\begin{aligned} g = E[\frac{\dot{x}(t)}{x(t)}] &= \frac{1}{2} ln\gamma\, [E(m(i_t; s_t)|Boom) + E(m(i_t; s_t)|Recession)] \quad (1.24) \\ &= \frac{1}{2} ln\gamma\, [z_B s_B + z_R s_R], \;\; where \quad (1.25) \\ s_B &= E(s_t|Boom), \\ s_R &= E(s_t|Recession). \end{aligned}$$

$E(\cdot|i)$ indicates the expected value of the random variable "$\cdot$" conditioned to being in aggregate state i. The amount of slackness in aggregate state i, s_t^i, evolves according to the differential equation

$$\dot{s}_t^i = [\delta + \lambda F(\epsilon_i^d)](c - s_t^i) - \theta_i q(\theta_i) s_t^i, \qquad i = B, R,$$

where a change in the aggregate state implies a change in the way in which the system evolves over time. The stochastic dynamic system is further characterized by the initial condition of slackness at the time of an aggregate switch. In fact, when the economy switches from a boom to a recession, it will exist a positive mass of marginal enterprises with idiosyncratic productivity $\epsilon_B^d \leq \epsilon \leq \epsilon_R^d$, that will decide to shut down. Appendix B derives the long-run expected percentual mass of firms that at the end of the boom, have idiosyncratic productivity between ϵ_B^d and ϵ_R^d and shows that it is equal to

$$\Delta = \frac{\lambda[F(\epsilon_R^d) - F(\epsilon_B^d)]}{\lambda F(\epsilon_R^d) + \delta + \mu}. \tag{1.26}$$

Δ measures the instantaneous cleansing effect that takes place when the economy shifts from a boom to a recession. Appendix B derives also the expected value of the amount of slackness in boom and in recession. It shows that

$$\begin{aligned} s_B &= c[\frac{\mu^2 \Delta}{H} + \frac{J + \alpha_B \phi_B^2 (\mu + \phi_R)}{(\mu + \phi_B) H}], \\ s_R &= c[\frac{\mu(\mu + \phi_B(1 - \alpha_B))\Delta}{H} + \frac{J + \alpha_R \phi_R^2 (\mu + \phi_B)}{(\mu + \phi_R) H}], \end{aligned}$$

where

$$\begin{aligned} \alpha_i &= \frac{\delta + \lambda F(\epsilon_i^d)}{\phi_i}, \qquad i = B, R, \\ \phi_i &= z_i + \delta + \lambda F(\epsilon_i^d), \qquad i = B, R. \end{aligned}$$

α_i is the long-run steady state value of the slackness rate if the economy forever remained in aggregate state i while ϕ_i is the speed of adjustment of slackness in aggregate state i. J and H are appropriately defined constant[6]. Appendix B shows how s_R is always greater than s_B so that the amount of slackness in the economy is counter-cyclical.

1.4 Growth and Business Cycle Volatility

In this section we analyse the effects on long-run growth of a mean preserving increase in business cycle volatility. In other words we consider how the growth rate changes when

$$dP_B + dP_R = 0, \;\; dP_B > 0. \tag{1.27}$$

We will show how the business cycle creates disequilibrium as it breaks the balance between slackness and implementation effort in the aggregate economy. We will first highlight the main intuition in a stylised framework, we will then analyse the results of the chapter fully exploiting the dynamic nature of the model.

1.4.1 Growth and Business Cycle Volatility: the Basic Intuition

A new idea needs to be implemented to exert widespread economic consequences. As a result innovating requires a certain amount of both slackness (s_t) and implementation effort (i_t), so that $g_t = m(i_t; s_t)$. This implies that

[6]Appendix B shows how $J = \mu^2(\alpha_B\phi_B + \alpha_R\phi_R) + \mu(\alpha_R\phi_R\phi_B + \alpha_B\phi_B\phi_R)$, while $H = \mu(\phi_R + \phi_B) + \phi_R\phi_B + \mu^2\Delta$.

slackness and implementation effort are two strictly complementary inputs of a hypothetical aggregate production function, $m(i_t; s_t)$ that gives the total number of ideas that a society transforms into a public good (innovation).

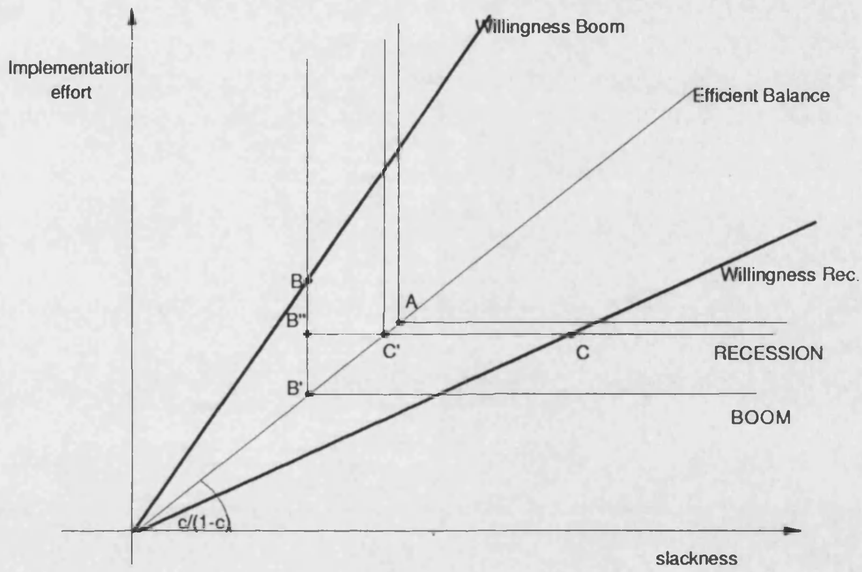

Figure 1.2: BCV harms Growth

Suppose that in the absence of business fluctuations the economy would settle into a point- like A (Figure 1.2) with a balanced level of slackness s_t and implementation effort i_t. The two orthogonal lines crossing A represent the isoquant of the aggregate production function $m(i_t; s_t)$ with an amount of innovation equal to that in A. In this context, an increase in business cycle volatility breaks the perfect balance between slackness and implementation effort. During a boom the economy settles into a point-like $B$with an enhanced level of willingness $\theta_B = i_t/s_t$ driven by an increase in the value of an innovation (see equation 1.19) but with a reduced level of slackness as now the value of the option of keeping old firms operating has increased (see equation 1.22). An analogous reasoning suggests that during a recession the economy settles in a point like C (Figure 1.2), with an enhanced level of slackness (because of the fall in the value of keeping old firms operating, see equation 1.22) but with a reduced level of the willingness $\theta_R = i_t/s_t$ to use it (see equation 1.19). As implementation effort is now binding, the amount of innovation in the economy falls to a level equal to that associated with point C' (Figure 1.2). Despite the increase in the average amount of slackness and implementation effort measured respectively by the midpoint between $B''C$ and BB', the innovation rate of the economy unequivocally falls. The business cycle creates disequilibrium and the economy fails to coordinate properly. There are two Keynesian (1936) features here. The first is the one highlighted in the introduction: in a world with cross-sectional heterogeneity where the decision to free resources and to take advantage of them lie in different subjects, an increase in the amount of resources that are kept slacked raises the innovation rate of the economy if and only if it raises the prospective yield of an innovation. The second is more subtle and shows the source of the coordination failure in the economy. "Though an individual can safely neglect the fact that demand is not a one-sided transaction, it makes nonsense to neglect it when we come to the aggregate" in general "there cannot be a buyer without a seller or a seller without a buyer" (see Keynes, 1936). Researchers look at the amount of slackness in the economy and at the value of a successful innovation and decide whether or enter the market. At the aggregate level, however, this effort can be too much (in a boom) or too little (in a recession) as a lot of buyers (researchers) enter a market with too few sellers (free entrepreneurs), a similar situation as in the game of the musical chairs[7].

In what follows to capture the appropriate degree of complementarity between slackness and implementation effort we rely on a CES aggregate matching function with elasticity of substitution strictly less than 1, that is

$$m(i_t; s_t) = [(c\, i_t)^\rho + ((1-c)s_t)^\rho]^{\frac{1}{\rho}}, \qquad \rho < 0.$$

This function implies that one-sided constant percentual increases in either i_t or s_t have decreasing effects, that is the elasticity is a decreasing function in the relevant variable. Figure 1.3 plots the long-run growth rate of the economy at different levels of business cycle volatility measured as $\frac{P_B - P_R}{2}$ (a). We have chosen a level of complementarity between slackness and implementation effort such that $\rho = -100$[8]. Despite the increase in both the average amount of implementation effort (panel d) and slackness (panel b)

[7]The "coordination failure" analysed here is of a nature strictly different from that at the core of the earlier neo-Keynesian literature (i.e. Cooper and John, 1988), here no multiple equilibra arise. The point that "idiosyncrasies" can arise in a laissez faire economy was first made by Caballero and Hammour (1994a) in a similar environment but in a different context.

[8]The other parameters are chosen as follows. $c = 0.5$ so that slackness and implementation effort are equally important in the matching process, $\mu = 0.333$ to capture a business cycle of duration of about 7 years. The distribution function F is chosen without loss of generality to be uniform with upper support equal to $2\sqrt{3}$ and zero expected value. $\sigma = 3$, $r = 0.1$, $\delta = 0$ and $h = 7$. The value of λ is chosen to be equal to 0.2 implying that change in the idiosyncratic productivity occurs every 5 years. The average level of productivity $\frac{P_R + P_B}{2}$ is chosen to be equal to 6. The amount of resources in the economy C and the size of technological progress $\ln \gamma$ are chosen to be equal respectively to 100 and 0.01 to capture the fact that we are modelling an aggregate economy with very small innovations. Following standard practice in search theory the bargaining parameter, β, is set to be equal to 0.5. The cost of searching, ω, is set to be equal to 7.6 in order to start with a balanced level of slackness and implementation effort given the other parameter values, so that $\theta_R = \theta_B = 1$ when $P_B = P_R$.

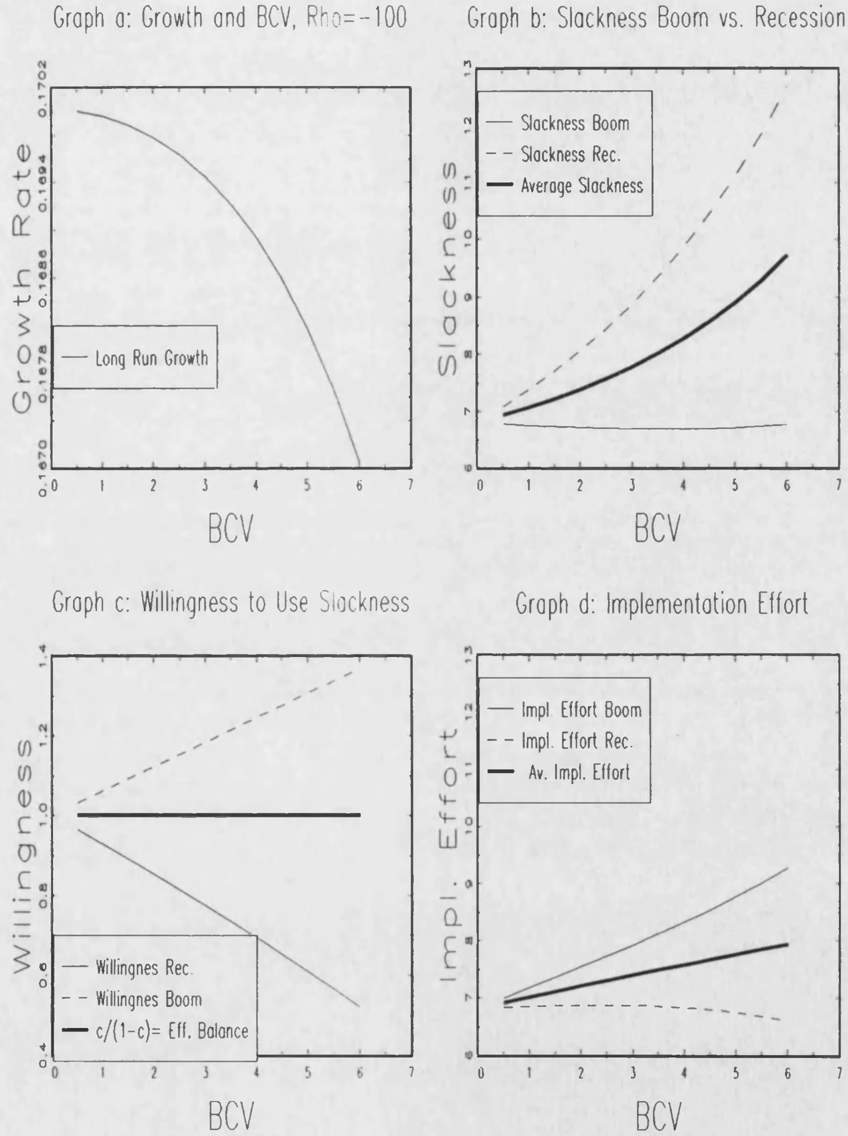

Figure 1.3: BCV Harms Growth (Simulations)

the long-run growth rate of the economy falls as the balance between the two quantities i_t and s_t breaks down (panel c).

In the previous paragraph we analysed the effects of an increase in the level of business cycle volatility under two extreme hypotheses: the level of complementarity was set at a sufficiently high level ($\rho = -100$) and we assumed that in the absence of business fluctuations the balance between slackness and implementation effort was perfect. If we relax either of these two assumptions, an increase in business cycle volatility can stimulate growth. Suppose, for example that the economy in the absence of business fluctuations would settle into a point-like A (Figure 1.4) that is with a binding value of slackness[9]. Now if the level of business cycle volatility increases, the economy would settle in a point- like B in the boom and a point-like C in the recession. We will see in the next section that the business cycle helps to free resources increasing the amount of liquidation in the economy. This implies that the average amount of slackness given by the midpoint between CC' tends to be higher than the one associated with A. As slackness is binding and as an increase in business cycle volatility helps in liquidating resources it also stimulates the growth process. Nonetheless, further increases in business cycle volatility will eventually harm the growth process and the same case will arise as that analysed in Figure 1.2.

Figure 1.5 shows the relation between long-run growth and business cycle volatility once we start from a binding initial value of slackness. The graph is generated with a value of $\rho = -10^{10}$. In the next section we will see how

[9]Similar considerations would hold if implementation effort were binding.

[10].Figure 1.5 is generated with the same parameters as Figure 1.3 except for the search

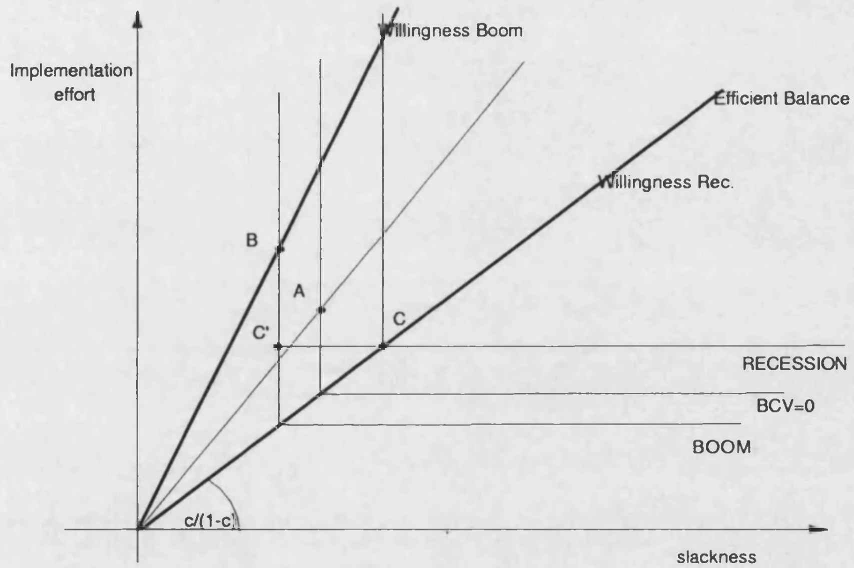

Figure 1.4: Small Amount of BCV Stimulates Growth

this is the typical shape of the relation between growth rate and business cycle volatility provided that $\rho < 0$. Two elements are worth noting in Figure 1.5. The first is that the long-run growth rate of the economy initially increases as business cycle volatility increases (panel a). The second is that the threshold beyond which further increases in business cycle volatility harm the long-run growth of the economy tends to be higher the lower the degree of complementarity between slackness and implementation effort (high value of ρ)[11].

Despite their usefulness these figures completely neglect the dynamic nature of the model. In fact, slackness is an endogenous variable with a spurious correlation with the growth rate. Both Figure 1.2 and 1.4 seem to suggest that the innovation rate of the economy can not rise unless the level of slackness also raises. We have already noted in Section 2 how this proposition is false, as for example, an increase in the actual use made of slackness z raises the growth rate but reduces the amount of slackness in the market[12]. In the next subsection we will exploit the dynamic nature of the model to analyse more carefully the mechanism through which the business cycle creates dis-

cost ω which is set at a smaller value equal to 5 in order to start from a situation where the aggregate amount of slackness is binding. That is, $\theta_R = \theta_B > 1$, when the level of business cycle volatility is equal to 0.

[11] In a previous version of the chapter we compared the behaviour of the economy for positive value of ρ against negative one. When $\rho > 0$ an increase in business cycle volatility always increases the growth rate of the economy. The intuition is simple. A positive value of ρ implies that slackness and implementation effort are substitutes so that the innovation rate of the economy depends on the average amount of resources in the market rather than their composition. In this version we ruled out this case as highly implausible. Because of their intrinsic nature, slackness and implementation effort seem to be strictly complements rather than substitutes and a temptative guess should assign a very negative value to ρ.

[12] For example a more careful analysis of both Figure 1.3 and 1.5 show how the level of slackness in boom tends to increase at high levels of business cycle volatility.

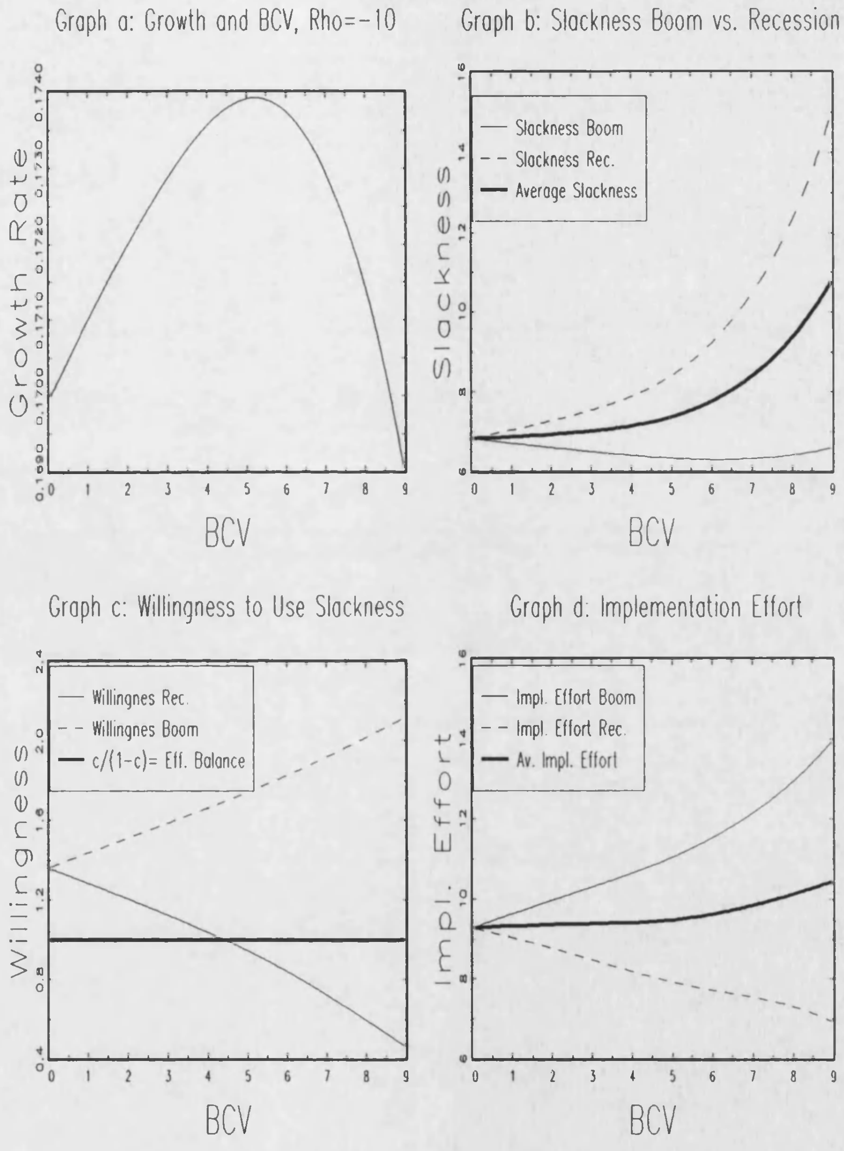

Figure 1.5: Small Level of BCV Stimulates Growth (Simulations)

equilibrium in the economy.

1.4.2 Growth and Business Cycle Volatility: a Dynamic Approach

The rate of growth of the technological parameter given by equation (1.25) can be written as

$$g = \frac{1}{4} ln\gamma [(z_B + z_R)(s_B + s_R) - (z_B - z_R)(s_R - s_B)].$$

An increase in either the average amount of slackness in the economy (s_B+s_R) or in its average actual use $(z_B + z_R)$ would necessarily increase the long run growth rate of the economy if they were uniformly spread out over both the aggregate states. The business cycle, however, tends to move the amount of slackness and the willingness θ to use it in opposite directions. The terms $(z_B - z_R)$ and $(s_R - s_B)$ represent a first measure of the waste of resources produced by the business cycle. It is not however the only one. We can decompose the effects of an increase in BCV on the long-run growth rate of the economy into three components:

$$\begin{aligned} dg &= \frac{1}{4} ln\gamma (A + B + C), && (1.28) \\ A &= (z_R + z_B)(ds_R + ds_B), \\ B &= (dz_B + dz_R)(s_B + s_R), \\ C &= -(s_R - s_B)(dz_B - dz_R) - (ds_R - ds_B)(z_B - z_R). \end{aligned}$$

All three terms are potentially able to capture different aspects of the destructive effects of an increase in the level of business cycle volatility on growth. The terms A, B and C respectively measure the impact of an increase in business cycle volatility on the average amount of slackness in the economy, the average use that is made of it and the waste of resources.

Average amount of slackness in the market

The term A in (1.28) measures the impact of an increase in business cycle volatility on the average amount of slackness. In Section 2 we established that there are two independent sources affecting the amount of slackness in the economy; the amount of liquidation in the economy and the actual use that is made of slacked resources. An increase in the level of business cycle volatility tends to increase the level of the former and reduce the level of the latter so that the average amount of slackness unequivocally increases. In fact, every time the economy switches from a boom to a recession, the economy is "cleansed" of all marginal enterprises with idiosyncratic productivity is between ϵ_B^d and ϵ_R^d. The percent of firms that are shut down at the time of the aggregate switch is given by the "cleansing" parameter Δ in equation (1.26). As the level of business cycle volatility increases, the amount of cleansing in the economy represented by the parameter Δ also increases, as the difference between the reservation productivities in recession and in boom given respectively by equations (1.22) and (1.23) widen (see Appendix A). In this sense the business cycle frees resources making them available for the implementation of new ideas. At high levels of business cycle volatility old firms are more likely to be destroyed, therefore the business cycle helps in liquidating old investments and business, making resources available for new investment and technological opportunities. The equilibrium average amount of slackness is also a function of the actual use that is made of the slacked resources currently available. We turn now to the analysis of the term B in equation (1.28).

The average actual use made of slackness

The average actual use of a given amount of slackness z depends of both the profitability of an innovation $I_i(\epsilon^u, \tau, t)$ and the technology with which the economy transforms ideas and slackness in innovations represented by $m(i_t; s_t)$. The higher the profitability of an innovation, the higher is the willingness to use slackness θ_i and also the higher the actual use that is made of it. However, an increase in willingness only has a relevant effect on the actual use of slackness, if the market is asking for it. For example, in the case of Figure 1.2, a further increase in the profitability of an innovation in a boom has no effect at all on the actual use because what is binding is the amount of slackness and not the willingness to use it. If the market power of researchers (β close to 0) is sufficiently high the average profitability of an innovation increases but despite this the average actual use of slackness eventually falls as business cycle volatility increases.

In fact, implicit in a firm is an option to stop losses that increases the average value of the innovation as the variance increases. This implies that the increase in the value of I_B tends to be higher in absolute terms than the decrease in I_R. The mathematical reason behind this fact is that firms have a margin that they can adjust when things go wrong. If we differentiate the value of an innovation given by equation (1.20) and (1.21) with respect to a change in business cycle volatility, we note how the impact of this change does not depend on the particular value of the idiosyncratic component ϵ. Moreover, we note that the differential of the value of the innovation in a boom depends on the interval in which the idiosyncratic component lies. If we indicate with dI_B^1 the differential of the value of an innovation in a boom when $\epsilon_R^d \leq \epsilon \leq \epsilon^u$, and with dI_B^2 the differential of the value of an innovation in a boom when $\epsilon_B^d \leq \epsilon \leq \epsilon_R^d$, we find that the impact of a change in business cycle volatility on the average value of an innovation solves

$$\{\nu-\lambda[1-F(\epsilon_R^d)]-\mu\}(dI_B^1+dI_R) = \lambda F(\epsilon_R^d)I_B(\epsilon_R^d)\,d\epsilon_R^d+\lambda[F(\epsilon_R^d)-F(\epsilon_B^d)]\,dI_B^2,$$

where we have assumed that $\beta = 0$. This result shows that the average value of an innovation increases as the level of business cycle volatility increases if both the two differential on the right are positive. Appendix A shows that this is actually the case. The option to stop mechanism implies that as business cycle volatility increases, also the size of the cake that the researcher and the entrepreneur have to share increases. These additional gains accrue to researchers only when they have a sufficient amount of market power. On the other hand, if β is closer to one, most of the additional gains accrue to entrepreneurs. In this context, therefore, an increase in business cycle volatility can make the average value of an innovation fall as the increase in the average profitability of innovating positively affect the outside option of entrepreneur.

Despite the increase in the average profitability of an innovation, the average actual use of slackness eventually falls. By using equation (1.19) and (1.20) the value of rents from innovating at the upper support of the idiosyncratic distribution F, can be expressed in the general form

$$x(t)I_i(\epsilon_u) = x(t)\frac{K_i}{\nu + \beta z_i}, \qquad i = B, R, \tag{1.29}$$

where K_i is a complex function of the parameter P_R, P_B and the reservation productivities in boom and recession. An increase in the level of business cycle volatility increases (decreases) the value of K_B (K_R). Using equation (1.19) and equation (1.20), it can be shown that the impact of a change in K_i on the actual use made of slackness z_i is given by

$$\begin{aligned} dz_i &= h(\theta)\, dK_i, \qquad i = B, R, \\ h(\theta) &= \frac{1}{\omega\nu\frac{1}{-q(\theta)^2[\theta+\frac{q(\theta)}{\dot{q}(\theta)}]} + \frac{\omega\beta}{q(\theta)+\dot{q}(\theta)\theta}}. \end{aligned} \tag{1.30}$$

Appendix C shows how the function $h(\theta)$ in equation (1.30)is decreasing in θ, and satisfies the following boundary conditions:

$$\begin{aligned} h(\theta) &\to \xi \;\; (0), \qquad if\; \theta \downarrow \; 0, \\ h(\theta) &\to 0 \;\; (\xi), \qquad if\; \theta \uparrow \; \infty, \end{aligned}$$

where $\xi = \frac{c^{\rho(1-\rho)}}{\omega\beta}$. At high levels of business cycle volatility the net impact of a further increase (decrease) in the value of an innovation on the actual use made of slackness tends to vanish (be magnified). Eventually an increase in business cycle volatility will necessarily reduce the average actual use made of slackness, as Figure 1.2 suggests.

Waste of resources associated with business fluctuations

The term C in equation (1.28) represents a further measure of the destructive effects of an increase in business cycle volatility increases. Both the two terms in C are negative and their absolute value tends to increase as business cycle volatility. The business cycle tends to move slackness and the willingness to use it in opposite directions. In particular, the amount of slackness tends to increase when research is less profitable (the term $(z_R - z_B)(ds_R - ds_B)$ in C) and the willingness to use a given amount of slackness tends to increase when less resources are available to implement new ideas (the term $(s_R - s_B)(dz_R - dz_B)$ in C).

Figure 1.6 summarises the results of this section[13]. Initially the business cycle frees resources as it helps to liquidate bad investment (panel b) therefore increases the growth rate (panel a). Eventually, however, despite the increase in the average value of an innovation (panel f), the average actual use made of slackness falls (panel e) and the waste of resources captured by the term C in equation (1.28) (panel c and d) climbs. At this point, a further increase in business cycle volatility harms the growth process (panel a) and slackness increases too (panel b) as resources become slack when no one is willing to use them.

1.5 Unit Roots, Cleansing and Technological Adoption

According to the liquidationist view "depressions are not simply evils" but "are the means to reconstruct each time the economic system on a more efficient plan" (Shumpeter, 1937). Recent both theoretical[14] and empirical

[13]The parameter values are the same as those used for Figure 1.5.

[14]See for example Bradford De Long (1990), Davis and Haltinwanger (1991), Mortensen and Pissarides (1994) and Caballero and Hammour (1994b).

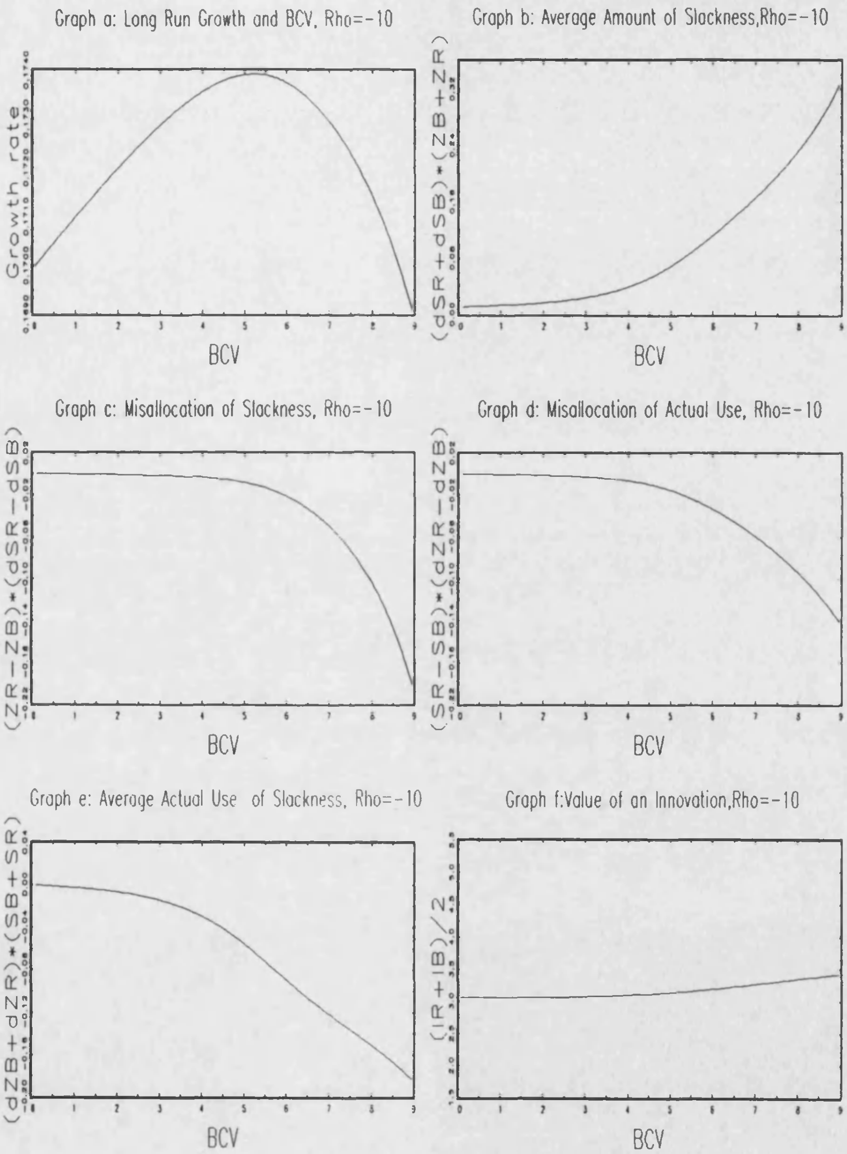

Figure 1.6: Growth and BCV: a Dynamic Explanation

evidence[15] has shown how this view deserves credit: recessions tend to cleanse the economy and raise the long-run level of output. Therefore, if a theory explains why business cycle volatility harms growth, it must do so in a world in which the "cleansing" property of recessions is satisfied. Figure 1.7 shows, in an intuitive way, how our model satisfies this property. Panel *a* in Figure 1.7 shows the simulated time series for the level of total output[16] when the value of business cycle volatility is equal to 4.5[17].

Panel *b* shows how a recession cleanses the economy. In fact, it liquidates businesses adopting old technologies, and makes these resources available for the creation of firms adopting the current leading technology. As a result, when aggregate conditions turn good the economy is able to produce a greater output (Figure 1.7, panel *b*). The cleansing effect of recessions can be measured by the sum of the Wold coefficients of the moving average representation of the growth rate of total output. This is carried out for different levels of business cycle volatility in Figure 1.8 panel *c* [18]. This sum is always negative and decreasing, therefore the degree of cleansing of recessions tends

[15]See for example Bean (1990), Gali and Hammour (1991), Saint-Paul (1993) and Nickell et al. (1992).

[16]We have defined the total output (GDP) of our economy as the sum of output produced by firms operating in the economy plus the amount of output produced by entrepreneurs working at home.

[17]The other parameter values are the same as those for Figure 1.5. The time series in this section is generated as discrete limits of the continuous model described in the previous sections. Each interval lasts one month and the simulations are run for 1200 intervals (100 years). The programs used for running the simulations are available on request.

[18]The parameter values are the same as the one used for Figure 1.5. In order to calculate the moving average representation of the growth rate of total output we looked for the best ARMA representation of the series according to the Akaike criterium. For a discussion on how these time series have been generated see the previous note.

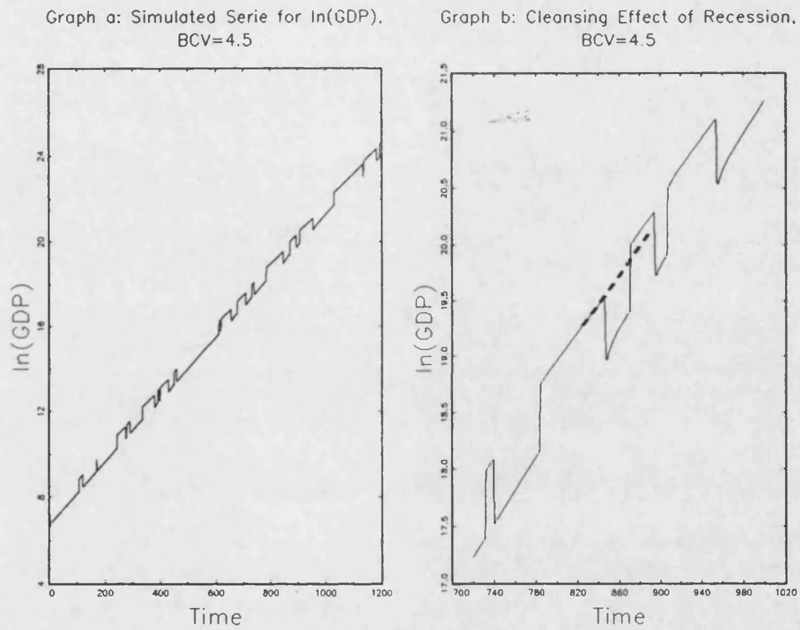

Figure 1.7: The Cleansing Effect of Recessions

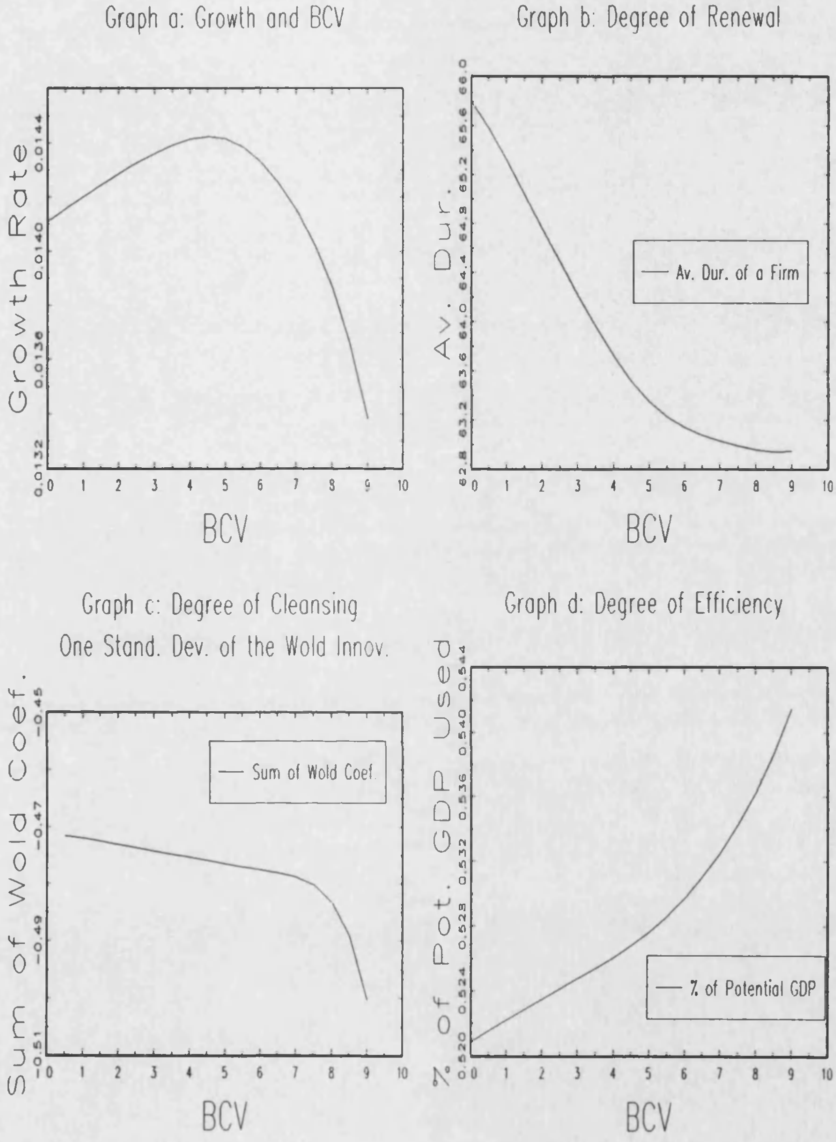

Figure 1.8: BCV and Technological Adoption

to increase as business cycle volatility increases. In fact, when business cycle volatility increases, the average duration of firms falls as the probability of being one of the marginal firms with idiosyncratic productivity between ϵ_B^d and ϵ_R^d at the time when aggregate conditions switch from a boom to a recession is greater (panel *b*). If we define the level of potential output in the economy as the level of output that would be produced if all economic units were using the current leading technology in the economy, we cab then note that the higher the level of business cycle volatility, the higher the average ratio between actual output and potential output (see Figure 1.8, panel *d*). As a result, at high levels of business cycle volatility the economy tends to stay closer to its technology frontier independently of what is the impact of business cycle volatility on the long run growth rate of the economy (panel *a*).

1.6 Some Empirical Implications

The model described in the previous sections has some further empirical implications that might be worth exploring. This section tries to throw some light on them, leaving a more exhaustive investigation to further research.

One of the main implications of the model is the existence of possible non-linearities in the relationship between business cycle volatility and long-run growth. If the balance between slackness and implementation effort is not perfect, small cyclical fluctuations help in liquidating old businesses, increase the profitability of doing research and stimulate growth. As the level of business cycle volatility increases, however, the business cycle creates disequilibrium and imbalances destroying the required to grow balance between the amount of resources available in the market and the willingness to use them. Figure 1.9[19] shows how the relation between growth and business cycle volatility might exhibit a threshold effect using cross-sectional evidence for OECD countries in the period 1970-1990. Ramey and Ramey (1995) find in their sample however, a significant negative relationship between growth and business cycle volatility once conditioned by a set variables identified by Levine and Renelt (1992). According to our model this implies that world economies have settled on the negative sloped-arm of their growth business cycle volatility technological frontier. If we consider the simulations used for building Figure 1.8, we assume that world economies settle on the negative sloped arm of their growth business cycle volatility technological frontier, we are then able to regress the growth rate of the technological parameter $x(t)$, g_t, on its standard deviation, sdg_t, we find that

$$\begin{aligned} g_t &= \underset{(187.93)}{0.0153} - \underset{(-13.35)}{0.354}\, sdg_t \end{aligned}$$

(t statistics in parenthesis). This coefficient of -0.354 resembles very much the coefficient equal to -0.385 found by Ramey and Ramey (1995), Table 1, for the set of OECD-countries in the period 1952-1988.

The size of the population of entrepreneurs, C, measures the total amount of resources available in the economy. Our model implies that an economy with higher C will have a magnified effect in the relation between long-run growth and business cycle volatility. As empirically the relation is found to

[19]Data are taken from the OECD-CEP data set, see Bagliano et al. (1991).

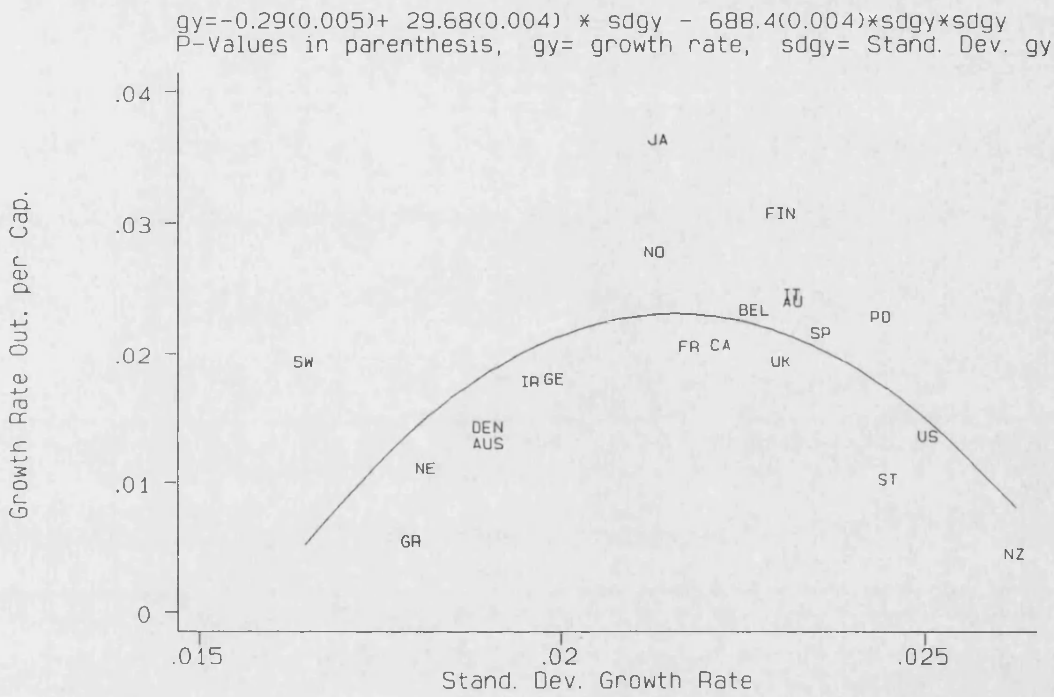

Figure 1.9: OECD (1970-1990), Growth Rates and BCV

be negative, the model implies that richer countries should have a stronger negative correlation. This is what in Ramey and Ramey found in Table 1. The regression for the set of OECD countries exhibits a coefficient in front of the standard deviation of growth rates that is more negative than that for the world economy, that is -0.385 against -0.211.

1.7 Conclusions

We have presented in this chapter a model of endogenous growth in which the widespread economic consequences of a new idea arise only when a researcher recovers in the market resources that suit its economic applicability. In this context we have analysed how the level of business cycle volatility affects the long- run growth rate of the economy. We have disentangled the question at three levels. Can an increase in the level of business cycle volatility stimulates growth? This chapter suggests a positive answer if the amount of resources suitable of economic exploitation represents in the absence of business fluctuations a binding constraint for the economy. Can the level of business cycle volatility be excessive and harm the growth process? This chapter suggests a positive answer again, as the business cycle eventually creates disequilibrium and imbalances in the economy destroying the balance between the amount of resources available in the market and the willingness to use them. So why have people sometimes argued that even deep recession are beneficial? We have seen how at high level of business cycle volatility an economy tends unequivocally to be more efficient in the sense that it stays closer to its technology frontier. Notwithstanding this, at high levels of business cycle volatility an economy will not be able to innovate much, as it fails in transforming new economic ideas from private to public goods the whole society can learn from .

Schumpeter (1934) observes that "all those features of depressions, which spell widespread suffering and needless waste, can yet be taken care of". In this chapter we claim that the level of business cycle volatility experienced by most countries seems to have represented a case of "needless waste". On the other hand, if a country has been detrimentally affected by a lack of business cycle volatility, it remains a call for further research.

1.8 Appendixes

1.8.1 Appendix A: Properties of Reservation Productivities

In this appendix we will prove 4 results:

- $\epsilon_R^d \geq \epsilon_B^d$,
- the difference between ϵ_R^d and ϵ_B^d tends to increase as business cycle volatility increases,
- an increase in business cycle volatility raises the value of the rents from innovation during a boom if $\epsilon_B^d \leq \epsilon \leq \epsilon_R^d$,
- an increase in business cycle volatility raises the value of the rents from innovation during a boom if $\epsilon_R^d \leq \epsilon$. The converse holds for the value of rents from innovation in recession.

First of all, we note that it is enough to prove these results for any fixed level of market tightness θ. As the free entry condition given by equation (1.19) is increasing in θ, while the value of an innovation is decreasing in θ by equation (1.23), these partial equilibrium results extend easily in the general setup. In general equilibrium the size of these effects will be dampened but their signs will not be changed.

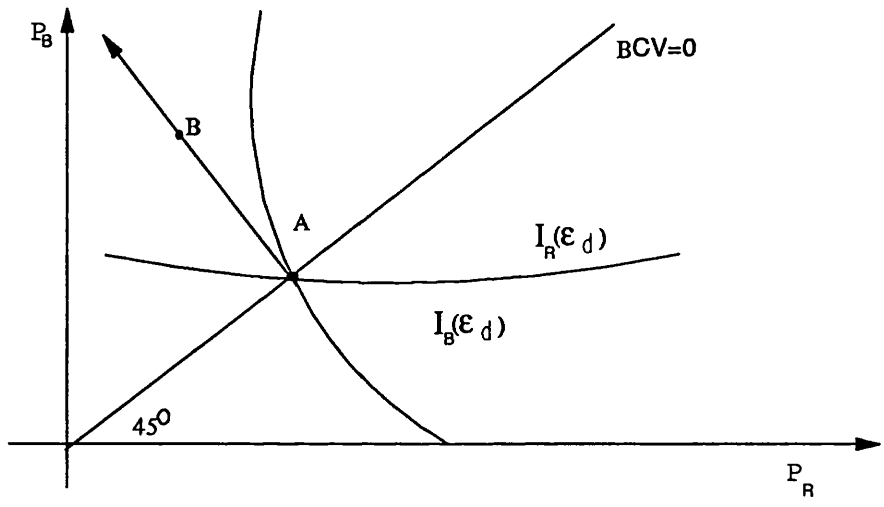

Figure 1.10: Structure of Reservation Productivities

For any given level of market profitability, basic dynamic programming arguments imply that the value of an innovation is an increasing function of both the level of aggregate productivity in a boom (P_B) and a recession (P_R). Moreover, discounting and persistence of the business fluctuations implies that, for any given value of ϵ, the rate of substitution between P_B and P_R is higher for the value of an innovation in recession ($I_R(\epsilon)$) than for that in a boom ($I_B(\epsilon)$). Symmetry implies that the first will be higher than one, while the second will be lower than one. Consider now, for each given idiosyncratic productivity, the set of indifference curves in the state space $P_B - P_R$ for both I_R and I_B. Now suppose that the reservation productivity in boom and recession be equal to ϵ^d. Let us now consider the indifference curve of I_R evaluated at the reservation productivity ϵ^d. The value of this indifference curve is clearly zero. Consider now the indifference curve of $I_B(\epsilon^d)$ delivering the same value as the one for recession. This indifference curve would cross the indifference curve for $I_R(\epsilon^d)$ in point A (see Figure 1.10), where the level of BCV is equal to 0. This implies that the indifference curve in boom through the actual value of BCV is associated with a positive value. Monotonicity with respect to the idiosyncratic component proves the first point. An increase in the level of BCV is equivalent to a movement from A to B (Figure 1.10). The structure of the indifference curves clearly implies that ϵ_R^d will tend to rise while ϵ_B^d will tend to fall. This proves the second point. To prove the third point just note that the indifference curve associated to point B is always on an higher level than that trough A, so that the value of rents in a boom increases as we move from A to B. This proves the third point. Point four follows easily from the same lines of reasoning.

1.8.2 Appendix B: Expected Slackness in Boom and Recession

The amount of slackness in the market in aggregate state i evolves according to the differential equation

$$\dot{s}_t^i \;=\; \delta_i(c - s_t^i) - \theta_i q(\theta_i) s_t^i, \qquad i = B, R,$$

$$\delta_i \;=\; \delta + \lambda F(\epsilon_i^d), \qquad if \quad i = B, R.$$

We indicate with

$$\alpha_i = \frac{\delta_i}{\delta_i + \theta_i q(\theta_i)}, \qquad i = B, R,$$

the long-run steady state level of the slackness rate if the economy stayed forever in state i, and we indicate with

$$\phi_i = \theta_i q(\theta_i) + \delta_i, \qquad i = B, R,$$

the speed of adjustment of the amount of slackness. Then the value of slackness after t_i periods in state i, is given by

$$s_t^i = e^{-\phi_i t_i} s_j + c\alpha_i (1 - e^{-\phi_i t_i}), \qquad i = B, R, \qquad j = 0, 1.$$

where s_j represents the value of slackness at the beginning of state i. So that

- s_0=value of slackness at the beginning of the boom,
- s_{0-}=value of slackness at the end of recession,
- s_1=value of slackness at the beginning of recession,
- s_{1-} =value of slackness at the end of boom.

Our model implies that

$$\begin{aligned} s_1 &= \Delta_{tB} + s_{1-}, \\ s_0 &= s_{0-}, \end{aligned}$$

where Δ_{tB} is the measure of enterprises that at the end of a boom have an idiosyncratic productivity between ϵ_B^d and ϵ_R^d.

Given that a Poisson process is a renewal process with inter-occurrence probability given by a negative exponential, we find that the expected value of slackness in state i is

$$s_i = E(s_t^i) = \frac{\mu}{\mu+\phi_i}E(s_j) + \frac{C\alpha_i\phi_i}{\mu+\phi_i}, \qquad i = B, R, \qquad j = 0, 1, \qquad (1.31)$$

where E(s_j) represents the best forecast of the value of slackness at the beginning of state i.

To have a close form solution for the value of slackness we have to find an expression for E(s_j), $j = 0, 1$. In equilibrium s_0 solves

$$s_0 = e^{-\phi_R t_R}[e^{-\phi_B t_B}s_0 + c\alpha_B(1 - e^{-\phi_B t_B}) + \Delta_{tB}] + C\alpha_R(1 - e^{-\phi_R t_R}).$$

$t_R(t_B)$ is a negative exponential random variable representing the duration of the recession (boom).

As the inter-occurrence random variables are independent, the expected value of s_0 is equal to

$$E(s_0) = \frac{\mu}{\mu+\phi_R}[\frac{\mu}{\mu+\phi_B}E(s_0) + \frac{\alpha_B\phi_B}{\mu+\phi_B} + E(\Delta_{tB})] + \frac{\alpha_R\phi_R}{\mu+\phi_R}, \qquad (1.32)$$

where $E(\Delta_{tB})$ is the expected measure of enterprises that at the end of a boom have an idiosyncratic productivity between ϵ_B^d and ϵ_R^d. To solve for $E(s_0)$ we need an expression for $E(\Delta_{tB})$. Δ_{tB} solves the simple Cauchy problem given by

$$\dot{\Delta}_t = -\delta_B\Delta_t + \lambda[F(\epsilon_R^d) - F(\epsilon_B^d)](C - s_t^B - \Delta_t), \qquad \Delta_0 = 0.$$

Together with the differential equation for slackness given by equation (??) we have to solve a recursive linear dynamic system. This implies that

$$\begin{aligned}\Delta_{tB} &= \lambda[F(\epsilon_R^d) - F(\epsilon_B^d)]\{\frac{C(1-\alpha_B)}{\delta_R}[1-e^{-\delta_R t_B}] \\ &- \frac{s_0 - C\alpha_B}{\delta_R - \phi_B}[e^{-\phi_B t_B} - e^{-\delta_R t_B}]\}.\end{aligned}$$

As a result

$$\begin{aligned}E(\Delta_{tB}) &= \lambda[F(\epsilon_R^d) - F(\epsilon_B^d)]\{\frac{C(1-\alpha_B)}{\delta_R}[1-\frac{\mu}{\delta_R+\mu}] \\ &- \frac{E(s_0) - C\alpha_B}{\delta_R - \phi_B}[\frac{\mu}{\mu+\phi_B} - \frac{\mu}{\delta_R+\mu}]\}. \qquad (1.33)\end{aligned}$$

Substituting this expression in that for $E(s_0)$ as given by equation (2.7) we find

$$E(s_0) = c\,\frac{\mu(\alpha_B\phi_B + \alpha_R\phi_R) + \alpha_R\phi_R\phi_B + \Delta[\mu^2 + \mu\phi_B(1-\alpha_B)]}{H}, \qquad (1.34)$$

and

$$E(\Delta_{tB}) = \Delta C(1 - \frac{\alpha_B\phi_B}{\mu\phi_B}) - \Delta\frac{\mu}{\mu+\phi_B}E(s_0),$$

where

$$\begin{aligned}H &= \mu(\phi_B+\phi_R) + \phi_R\phi_B + \mu^2\Delta, \\ \Delta &= \frac{\lambda[F(\epsilon_R^d) - F(\epsilon_B^d)]}{\lambda F(\epsilon_R^d) + \delta + \mu}.\end{aligned}$$

Analogously it can be shown that the expected value at the beginning of recession is equal to

$$\begin{aligned}E(s_1) &= C\{\frac{\mu(\alpha_B\phi_B + \alpha_R\phi_R) + \alpha_B\phi_B\phi_R + \Delta[(\mu+\phi_B)(\mu+\phi_R)}{H} + \\ &- \frac{\mu(\alpha_B\phi_B + \alpha_R\phi_R) - \alpha_B\phi_B\phi_R]}{H}\}.\end{aligned}$$

To obtain the average value of slackness in boom we substitute the expression for the expected value of s_0 in equation (1.31). We then have

$$s_B = E(s_t^B) = C\,[\frac{\mu^2\Delta}{H} + \frac{J + \alpha_B\phi_B^2(\mu+\phi_R)}{(\mu+\phi_B)H}],$$

where

$$J = \mu^2(\alpha_B\phi_B + \alpha_R\phi_R) + \mu(\alpha_R\phi_R\phi_B + \alpha_B\phi_B\phi_R).$$

Reasoning analogously, we find that the expected value of slackness in a recession is equal to

$$s_R = E(s_t^R) = C\,\{\frac{\mu[\mu+\phi_B(1-\alpha_B)]\Delta}{H} + \frac{J + \alpha_R\phi_R^2(\mu+\phi_B)}{(\mu+\phi_R)H}\}.$$

It is easy to show the following two propositions.

Proposition 1:

The average amount of slackness during a recession is higher than that during a boom if $\alpha_B < \alpha_R$. This means

$$s_B < s_R.$$

Proof

The difference between the level of slackness in the two states is equal to

$$s_R - s_B = C[\frac{\Delta\phi_B(1-\alpha_B)}{H} + \frac{J(\phi_B-\phi_R) + \alpha_R\phi_R^2(\mu+\phi_B)^2 - \alpha_B\phi_B^2(\mu+\phi_R)^2}{(\mu+\phi_B)(\mu+\phi_R)H}].$$

The first term is positive as the steady state value of the slackness rate is a number between 0 and 1. To prove the positiveness of the second term we just have to check that the numerator is positive. Substituting back the value of J into the above equation, we find that the numerator has the same sign as

$$\alpha_R\phi_R^2\phi_B(\mu+\phi_B)-(\mu+\phi_R)\alpha_B\phi_B^2\phi_R+\mu\alpha_R\phi_R\phi_B(\mu+\phi_B)-\mu\alpha_B\phi_B\phi_R(\mu+\phi_R).$$

This is equal to

$$\begin{aligned}\alpha_R &\quad \phi_R\phi_B(\mu+\phi_B)(\mu+\phi_R)-\alpha_B\phi_B\phi_R(\mu+\phi_B)(\mu+\phi_R)=\\ = &\quad (\alpha_R-\alpha_B)\phi_R\phi_B(\mu+\phi_B)(\mu+\phi_R)>0.\\ CVD &\quad .\end{aligned}$$

Corollary 1:

The amount of slackness in the economy is counter-cyclical. That is

$$s_B < s_R.$$

Proof

This follows easily from proposition one, the fact that $\epsilon_R^d \geq \epsilon_B^d$ (appendix A) and last from the procyclicality of willingness θ_i.

Proposition 2:

Both the average slackness rates in recession and boom are increasing function of the parameter Δ. Moreover, as Δ increases, the difference between the two average amounts of slackness in a recession and boom also increases.

Proof:

A fraction

$$\frac{a+cx}{b+dx},$$

is an increasing function in x iff

$$cb > ad.$$

If we apply this simple result to the case of the average amount of slackness in the case of a boom, we find that it is an increasing function with respect to Δ, iff

$$\begin{aligned}
[\quad & \mu\alpha_R\phi_R(\mu+\phi_B) + \mu\alpha_B\phi_B(\mu+\phi_R)](\phi_B - \phi_R) + \\
+ \quad & \alpha_R\phi_R^2(\mu+\phi_B)^2 - \alpha_B\phi_B^2(\mu+\phi_R)^2 > 0,
\end{aligned}$$

this is equivalent to

$$\begin{aligned}
(\quad & \mu+\phi_B)[\mu(\phi_R+\phi_B) + \phi_R\phi_B] \geq \mu^2(\alpha_B\phi_B + \alpha_R\phi_R) + \\
+ \quad & \mu(\alpha_R\phi_R\phi_B + \alpha_B\phi_B\phi_R) + \alpha_B\phi_B^2(\mu+\phi_R).
\end{aligned}$$

This inequality is automatically satisfied, as the average slackness rate in a boom is a quantity between 0 and 1.

Reasoning analogously, we can then prove the same result for the average slackness rate in recession.

To prove the second part of the proposition, we note that the difference of the slackness rate is equal to

$$\begin{aligned}
s_R - s_B &= C\,[\frac{\Delta\phi_B(1-\alpha_B)}{H} + \\
&+ \frac{J(\phi_B-\phi_R) + \alpha_R\phi_R^2(\mu+\phi_B)^2 - \alpha_B\phi_B^2(\mu+\phi_R)^2}{(\mu+\phi_B)(\mu+\phi_R)H}].
\end{aligned}$$

Reasoning in the same way as before and using the definition of H, it follows that this quantity is increasing in Δ. CVD.

1.8.3 Appendix C: Average Actual Use of Slackness in the Case of a CES Matching Function

This appendix shows that at high levels of BCV, the impact of an increase in the value of an innovation on the actual use made of slackness tends to vanish.

The time independent component of the rents from an innovation at the upper support of the idiosyncratic distribution F can be expressed in the general form

$$I_i(\epsilon_u) = \frac{K_i}{\nu + \beta z_i}, \qquad i = B, R,$$

where K_i is a complex function of the parameter P_B, P_R and the reservation productivities in boom and recession. Using the free entry condition (1.19) we find that the impact of a change of K_i on the actual use made of slackness in boom and recession, z_i is given by

$$\begin{aligned} dz_i &= \frac{1}{\omega\nu\frac{1}{f(\theta)} + \frac{\omega\beta}{g(\theta)}} dK_i, \qquad i = B, R, \\ f(\theta) &= -q(\theta)^2[\theta + \frac{q(\theta)}{\dot{q}(\theta)}], \\ g(\theta) &= q(\theta) + \dot{q}(\theta)\theta. \end{aligned}$$

Assuming that the matching function is CES,

$$m(i_t; s_t) = [(ci_t)^\rho + ((1-c)s_t)^\rho]^{\frac{1}{\rho}}, \qquad \rho < 0,$$

we find that

$$
\begin{aligned}
g(\theta) &= [c^{\rho} + (1-c)^{\rho}\,\theta^{-\rho}]^{\frac{1}{\rho}-1}, \\
f(\theta) &= \theta^{\rho+1}\,(\frac{c}{1-c})^{\rho}\,[c^{\rho} + (1-c)^{\rho}\,\theta^{-\rho}]^{\frac{2}{\rho}}.
\end{aligned}
$$

It is easy to check that both the function $f(\theta)$ and $g(\theta)$ are decreasing in θ. This also implies that the function $h(\theta)$ that measures the impact of an increase of the component K_i on the actual use of slackness, z_i, is also a decreasing function in θ.

Moreover it is easy to prove that $f(\theta)$ and $g(\theta)$ satisfy the following properties at the boundaries

$$
\begin{aligned}
g(\theta) &\to 0, && if\ \theta \uparrow \infty, \\
f(\theta) &\to 0, && if\ \theta \uparrow \infty, \\
g(\theta) &\to c^{\rho(1-\rho)}, && if\ \theta \downarrow 0, \\
f(\theta) &\to \infty, && if\ \theta \downarrow 0.
\end{aligned}
$$

This implies that the function $h(\theta)$ has the following properties

$$
\begin{aligned}
h(\theta) \to \xi, &\qquad if\ \theta \downarrow\ 0, \\
h(\theta) \to 0, &\qquad if\ \theta \uparrow\ \infty,
\end{aligned}
$$

where $\xi = \frac{c^{\rho(1-\rho)}}{\omega\beta}$.

This proves the result stated in the text.

Chapter 2

Lack of Entrepreneurial Skills and Decreasing Returns in R&D

Abstract

In standard endogenous growth models, the higher the research effort the higher is the innovation rate of the economy. Innovating, however, is a complex process that requires an entrepreneur to implement a valuable invention. If research and entrepreneurial skills compete in the allocation of aggregate resources, the relation between growth and $R\&D$ is hump-shaped. This chapter proposes a general equilibrium model of endogenous growth in which the observed decreasing returns in $R\&D$ might be the outcome of a lack of entrepreneurial skills rather than any vanishing of investment opportunities. If this is so, the amount of resources devoted to research has been excessive and there may be a case for policy intervention. Indirect inference on the model, observed changes in fiscal structure and more formal empirical investigation seem all to support the relevance of this hypothesis.

2.1 Introduction

Since the Malthusian (1798) and Ricardian (1817) prophecies of the eventual coming of a stationary state, the spectre of diminishing returns has hovered over economics. Recent empirical evidence seems to support some form of diminishing returns in research over time. For example, Griliches (1990) and Kortum (1993) note how the ratio of the number of patent applications to the scientists and engineers involved in *R&D* has fallen over time in the post-war period (Figure 1), while Jones (1995a, 1995b) points out that the increase in the amount of *R&D* effort (Figure 2) has been translated into stagnant or declining growth rates. Moreover, Griliches (1979) observes that "the exhaustion of inventive and technological opportunities remains a major suspect for the productivity slow-down in the 70's". According to Schumpeter (1943, 1946) the vanishing of investment opportunities is, however, just one of two competing explanations of the "state of decay" of the "capitalist society". A "more plausible" one states that "the individual leadership of the entrepreneur tends to lose importance".

According to Schumpeter, the amount of entrepreneurial as well as research ability is important in determining the growth rate of an economy. Moreover, Schumpeter (1947) notes how the inventor and the entrepreneur are distinct entities: "It is particularly important to distinguish the entrepreneur from the inventor....The inventor produces ideas, the entrepreneur "gets things done"....an idea or scientific principle is not, by itself, of any importance for economic practice". This chapter then proposes a general equilibrium model of endogenous growth in which a vertical innovation arises only when an entrepreneur matches with a valuable invention. In other words, in order to grow, an economy requires both researchers producing inventions and entrepreneurs implementing them. If these two inputs compete in the allocation of aggregate resources, the relation between growth and *R&D* effort is hump-shaped.

The trade off between allocating resources to either research or entrepreneurial activities arises in many situations. For example, the economist must allocate his time between reading old papers in order to write new ones and teaching students, talking with colleagues, presenting work at the NBER meetings, etc. Per se, scientific knowledge has no economic impact unless some effort is made to spread it. As a result, a no growth equilibrium can be the outcome of absence of either research or entrepreneurial skills. In a world where the allocation between research and entrepreneurial activities takes place through the market, where the rents from innovating are ex-post shared by researchers and entrepreneurs and where there are externalities associated with innovating, there is no compelling reason why private incentives should coincide with social ones. If the allocation of resources takes place through the market, agents choose how to allocate their skills as a function of their private incentives. If there are then both rents to be shared among different rentiers and externalities associated with innovating, it will be mere chance that private and social interests coincide. This implies that any point on the growth *R&D* technological frontier can be sustained as a competitive equilibrium. In this framework, the decreasing returns in research observed in some industrialised countries (see Jones, 1995b) could be the result of an inefficient equilibrium shift that has increased the amount of research effort to the detriment of more socially useful entrepreneurial skills.

We try to test for the plausibility of this hypothesis. We initially show how in the US the increase in the amount of resources allocated to research, and more generally to education, has gone together with a dramatic fall in the amount of entrepreneurship once this is proxied by the population of self-employed (see e.g. Evans and Jovanovic, 1989 and Evans and Leighton, 1989). We then try to explain why entrepreneurship might have become progressively less profitable from a private point of view. According to the model, the observed tendency for the real interest rate to fall as an economy develops (see Barro and Sala-I-Martin, 1995, p.6), the fall in the obsolescence rate of an innovation (Caballero and Jaffe, 1993) and the increase in the degree of appropriability of the rents of an invention (consistent with the empirical evidence contained in Schankerman and Pakes 1986), might all be able to explain the alleged equilibrium shift. Changes in the fiscal structure are, however, the most obvious candidates. As they merely redistribute resources across occupations, fiscal changes alter the private profitability of an occupation without altering its social value. We show that basic and applied research funded by nonprofit institutions, government expenditure in education and the amount of subsidies to firms have all substantially increased in the postwar period in the US. All these changes tend unequivocally to make research relatively more profitable than entrepreneurship without changing its underlying social value and they might thus explain why the US economy ended up by suffering from a lack of entrepreneurial skills. Finally, we check for some more direct empirical evidence. Looking at both time series (relative to the US) and cross sectional (OECD countries) evidence, we show how the claim of a non-monotonic relation between growth and relative $R\&D$ effort is consistent with the data. Moreover, the hypothesis that US growth rates have stagnated as a direct result of a lack of entrepreneurial skills can not be rejected at standard level of significance.

Relation to the Literature. This chapter relates to several strands of research in the literature. It shares with Helpman and Trajtenberg (1994) and chapter one the premise that the widespread economic consequences of an invention arise only with its implementation. The statement that research effort can be excessive is also a feature that arises in Tirole (1988, p.399) and in Aghion and Howitt (1992), while the claim that the allocation of talent matters for economic performance is contained also in Baumol (1990) and Murphy, Shleifer and Vishny (1991). It is the joint interaction between these strands of the literature that is new here. In a world where the inter-temporal spill-overs at the heart of endogenous growth models (see i.e. Romer, 1990) arise only with the implementation of endogenously produced inventions, it is the allocation between research and entrepreneurial activities that drives growth performance. The main reference is, however, the recent literature on growth and scale effects. Jones (1995b), Kortum (1994) and Young (1995) propose a theoretical solution to the empirically observed lack of any scale effect in the relation between $R\&D$ effort and growth. These theoretical solutions share some important features. In particular, all models assume that at higher levels of development the output cost of an innovation progressively increases. In other words, the decreasing returns in $R\&D$ derives from some form (at least in relative terms) of vanishing of investment opportunities. The side effect of this assumption is that the solutions of all these models satisfy some optimality property. For example, both in Jones (1995b) and Kortum (1994) the economy growth rate is the social optimal one, while in Young (1995) the amount of $R\&D$ effort is optimal even if its allocation among different dimensions of innovation might not be. In any case, all these models explain the empirically observed increase in the amount of $R\&D$ effort as the welfare improving response of a competitive economy to an increase in the scale of the economy. The contribution of this chapter is to propose a different reading of the same empirical facts. We show both theoretically and empirically how the alleged decreasing returns in $R\&D$ might be the outcome of an inefficient equilibrium shift that has increased the amount of research effort to the detriment of more socially useful entrepreneurial skills. This would be consistent with the Schumpeterian hypothesis that the "state of decay of the capitalist society" will ultimately be driven by the lack of entrepreneurial skills.

Section 2 expounds the general set-up of the model. A more formal presentation is contained in section 3. Section 4 explores how the theoretical framework can account for the alleged decreasing returns in $R\&D$. Section 5 checks for some direct empirical evidence in favor of the "lack of entrepreneurial" skill hypothesis.

2.2 The General Set-up

Our economy is populated by a continuum of agents of size C. Each agent is infinitely lived, risk-neutral and maximizes expected returns in output units discounted at rate $r > 0$. At each point in time agents can choose either to become researchers in order to produce inventions or entrepreneurs. We indicate with f_t the relative (with respect to C) size of the population of researchers, at time t.

A researcher discovers inventions according to a Poisson process with intensity λ. When he makes a discovery, he starts searching for entrepreneurs able to implement it. The flow cost of searching is given by $\chi x(t)$, $\chi \geq 0$, where $x(t)$ represents the leading technology in the economy at time t . We assume that the innovation opportunities associated with the invention vanish according to a Poisson process with rate of arrival ν. This implies that at each point in time the stock of scientific knowledge, κ_t, measured by the number of inventions suitable for economic exploitation, tends to increase in response to the discovery of new inventions, while it tends to fall as the old ones become obsolete. This allows one to capture some key characteristics of the process of accumulating scientific knowledge (see Adams, 1990). Firstly, the stock of scientific knowledge is *fundamental* as it is the outcome of research. Secondly, it recognises the *heterogeneity* of information, as the implementation of an invention requires a costly search process in order to find suitable entrepreneurs. Thirdly, the use of the stock of knowledge is *repetitive,* as an invention can give rise to a "cluster" of innovations concentrated "in certain sectors and their surroundings" (Schumpeter, 1939, pp.100- 101). Finally it recognises the *time specificity* of information, as scientific knowledge becomes obsolete as time goes by[1].

[1] It might be argued that the rate at which inventions become obsolete, depends on the rate of growth of the leading technology of the economy (*technological obsolescence*). We

An innovation requires an invention discovered by a researcher and an entrepreneur. Once they match, a new firm run by the entrepreneur is created. An innovation that occurs at time t will open access of that enterprise to the leading technology $x(t)$ as of that date. Each firm can produce, at each point in time t, a flow of goods equal to $Px(t)$ and is shut down according to a Poisson process with rate of arrival δ.

The entrepreneur can be either running an enterprise or producing at home (see Benhabib et al., 1991). If he runs a firm, the entrepreneur obtains a real flow of profit equal to π_t. Profits of the entrepreneur are chosen so as to share with the researcher the gains from innovating at each point in time. The entrepreneur's share is β. If he works at home, the entrepreneur is able to produce a flow of goods equal to $hx(t)$, a measure of the level of human capital of the entrepreneur at time t. In equilibrium an entrepreneur running a firm has no incentive to search. This implies that only entrepreneurs working at home are "available" for innovating.

The rate at which free entrepreneurs find suitable inventions is determined by the homogeneous-degree one matching function $m(\kappa_t; s_t)$ (see Pissarides, 1990), where κ_t represents the stock of scientific knowledge suitable for economic exploitation at time t and s_t is the number of entrepreneurs working at home[2]. s_t is a measure of the amount of *entrepreneurial slackness* avail-

consider this the kind of simplifying assumption that does not affect the main results of the chapter. Moreover, it might be argued that the economic applicability of an invention might be time dependent (*clock time obsolescence*) if some uncertainty is associated with it, if the inventor is finitely lived or if there is some loss of memory in the accumulation of intangible knowledge. In this chapter we follow the empirical work by Adams (1990) in assuming a constant rate of obsolescence of the stock of scientific knowledge κ_t.

[2]Schumpeter (1949) was very worried that his model looked like a model of "exogenous" technological growth: "a stumbling block" of the theory "may be expressed by saying that

able in the market. We assume that the matching function is increasing and concave in each of its complement arguments. The matching function allows one to represent in a parsimonious fashion two key characteristics of the implementation of inventions: the fact that both entrepreneurial skills and inventions are heterogenous, so that search is time consuming, and the fact that entrepreneurial skill is a scarce good for which different researchers are competing.

Since an invention offers the highest human capital opportunities in the market plus some profits, no free entrepreneur turns down an investment opportunity. Thus, the probability that an invention matches with an entrepreneur is given by $q(\theta_t) = m(\kappa_t; s_t)/\kappa_t$ where θ_t defines tightness in the market for entrepreneurial skills and is equal to κ_t/s_t. By analogous considerations it follows that $p(\theta_t) = \theta_t q(\theta_t)$ is the instantaneous probability that a free entrepreneur finds a valuable invention. We also assume that

$$p(0) = q(\infty) = m(0; s_t) = m(\kappa_t; 0) = 0, \qquad p(\infty) = q(0) = \infty. \qquad \text{(A1)}$$

Vertical innovations are the unique source of growth. We follow Aghion and Howitt (1992, 1994) in assuming that the rate of growth of the technological parameter $x(t)$ is given by the product of the size of the innovation, γ, and the frequency of innovation. Our assumptions imply that $\theta_t q(\theta_t) s_t$ is the number of innovations introduced in the economy at time t, so that the

the entrepreneur simply does nothing but take advantage of technological progress, which therefore appears, implicitly or explicitly as something that goes along entirely independently of entrepreneurial activity...It is perhaps not difficult to understand that technological progress, so obvious in some societies and so nearly absent in others, is a phenomenon that needs to be explained". This set-up allows us to model the entrepreneurial function in an endogenous growth framework.

growth rate g_t of the technological parameter $x(t)$ is given by $\theta_t q(\theta_t) s_t ln\gamma$. Finite present values and positive economic growth require

$$r > g, \tag{A2}$$

$$\lambda(1-\beta)(P-h) > 0, \tag{A3}$$

respectively. Condition (*A2*) is standard, while condition (*A3*) is the same as that in Romer (1990). In order to sustain growth we must assign a strictly positive degree of market power to researchers, $\beta < 1$. Research is an investment that requires the costs sustained today to be compensated by strictly positive economic rents in the future.

The unknown in the model are the level of tightness θ, the relative size of the population of researchers f_t and the steady state growth rate of the economy g.

2.3 The Model

In this section we first introduce the formal structure of the economy previously described. We then analyse the steady state equilibrium and its welfare properties.

2.3.1 Research, Innovation and Profits

The assumption that a researcher produces a valuable invention with rate of arrival λ and the option he has of becoming an entrepreneur imply that

$$R_t = R\,x(t) = E_{\tau \geq 0}\{[(L_{t+\tau} + sup(R_{t+\tau}; H_{t+\tau})]e^{-r\tau}\}, \tag{2.1}$$

where R_t, L_t and H_t are respectively the value at time t of doing research, of producing an invention and of being a self-employed entrepreneur (working at home). $t + \tau$ is the arrival date of the first invention produced by the researcher.

As a researcher is unable to implement the invention himself, when he makes a discovery he starts searching for entrepreneurs able to implement it. The assumption that the flow cost of searching is $\chi x(t), \chi \geq 0$, and that inventions vanish with arrival rate ν imply that the value of an invention follows the following Bellman equation

$$(r + \nu)L_t = -\chi x(t) + q(\theta)I_t + \dot{L}_t, \tag{2.2}$$

where I_t measures the present discounted value of an innovation introduced at time t and $\dot{L}_t$ is the time derivative of the value of an invention. As, at each point in time, a firm produces a flow of goods equal to $Px(t)$ and entrepreneurs get a real flow of profits equal to π_t , the asset value of an innovation to the researcher is equal to

$$(r + \delta)I_t = Px(t) - \pi_t + \dot{I}_t, \tag{2.3}$$

where $\dot{I}_t$ represents the time derivative of I_t while δ is the rate at which enterprises are shut down and thus measures the obsolescence rate of innovations.

A self-employed entrepreneur working at home at time t gets, in present value terms, a real income equal to H_t. H_t solves the Bellman equation

$$rH_t = hx(t) + \theta q(\theta)(E_t - H_t) + \dot{H}_t, \tag{2.4}$$

where $hx(t)$, E_t and $\dot{H}_t$ are respectively the real flow of goods produced by an entrepreneur working at home, the value of an enterprise run by the entrepreneur and the time derivative of H_t. $\theta q(\theta)$ is the instantaneous probability that a free entrepreneur finds a valuable invention. Analogously, E_t, the present value of an enterprise solves

$$rE_t = \pi_t + \delta[sup(R_t; H_t) - E_t] + \dot{E}_t, \tag{2.5}$$

where the above equation embodies the option entrepreneurs have of becoming researchers. The assumption that profits are the outcome of bilateral bargaining between the researcher with the invention and the entrepreneur implies that profits maximize the weighted product of the entrepreneur's and researcher's net return from the creation of a new firm:

$$\pi_t = \arg\max_{\pi_t}(I_t + L_t - L_t)^{1-\beta}(E_t - H_t)^{\beta}, \;\; 0 < \beta < 1,$$

where β measures the bargaining power of entrepreneurs. As a result, profits are such that

$$I_t = (1-\beta)(E_t - H_t + I_t) = (1-\beta)S_t, \tag{2.6}$$

where $S_t = E_t - H_t + I_t$ is the private net surplus associated with the creation of a new firm. Of this surplus, researchers and entrepreneurs appropriate fractions $1-\beta$ and β respectively. If we now impose a the free entry condition, it follows that in equilibrium an entrepreneur working at home must be indifferent between searching for an investment opportunity and becoming a researcher, i.e. $H_t = R_t$. Substituting in (2.3), (2.4), (2.5) and (2.6), we then find that profits are equal to

$$\pi_t = \beta P x(t) + (1-\beta) h x(t) + \beta\theta q(\theta) I_t. \tag{2.7}$$

2.3.2 Steady State Equilibrium

In steady state each variable is growing at the same rate g as the economy so that

$$R_t = Rx(t), \quad I_t = Ix(t), \quad L_t = Lx(t), \quad H_t = Hx(t), \quad E_t = Ex(t). \tag{2.8}$$

We can then make use of (2.8) to rewrite equations (4.11), (2.3), (2.4) and (2.5) as

$$\begin{aligned}
L &= \frac{1}{r+\nu-g}\left[-\chi + \frac{q(\theta)(1-\beta)(P-h)}{r+\delta+\beta\theta q(\theta)-g}\right], && (4.11') \\
I &= \frac{(1-\beta)(P-h)}{r+\delta+\beta\theta q(\theta)-g}, && (2.3') \\
H &= \left[h + \frac{\beta\theta q(\theta)(P-h)}{r+\beta\theta q(\theta)+\delta-g}\right]\frac{1}{r-g}, && (2.4') \\
E &= \beta I + (1-\beta)H. && (2.5')
\end{aligned}$$

Moreover, using (2.8) and the free entry condition ($R_t = H_t$), we find that

$$\begin{aligned}
R &= (L+R)E_{\tau\geq 0}(e^{(r-g)\tau}) \\
&= (L+R)\int_0^\infty e^{-(r-g)\tau}\lambda e^{-\lambda\tau}d\tau \\
&= \frac{\lambda L}{r-g}.
\end{aligned}$$

Together with equations (4.11') and (2.4') this implies that the free entry condition ($R_t = H_t$) can be written as

$$\beta\theta q(\theta)S + h = \lambda\frac{(1-\beta)q(\theta)S - \chi}{r + \nu - g}, \tag{FE}$$

where $S = I + E - H$ measures the private net surplus associated with the creation of a new firm and is equal to

$$S = \frac{P - h}{r + \delta + \beta\theta q(\theta) - g}.$$

The left-hand side of equation (*FE*) represents the relative profitability of being an entrepreneur as an increasing function of the amount of tightness θ in the market for entrepreneurial skills. The higher the amount of scientific knowledge κ_t available per entrepreneur, the more profitable is entrepreneurship. Analogously, an increase in θ reduces the profitability of being a researcher as the implementation of inventions becomes progressively more difficult. As a result the right hand side of equation (*FE*) defines a decreasing function in θ that measures the relative profitability of being a researcher. At the point at which the right-hand side and left hand side cross individuals are indifferent between becoming researchers or entrepreneurs.

The assumption that enterprises are closed according to a Poisson process with rate of arrival δ implies that slackness evolves according to the differential equation

$$\dot{s}_t = \delta[C(1-f) - s_t] - \theta q(\theta)s_t.$$

so that in steady state it is equal to

$$s_t = s = C\frac{\delta(1-f)}{\theta q(\theta)+\delta}. \tag{2.9}$$

Using the assumption that the frequency of innovation is equal to the number of successful matches between the steady state level of the stock of knowledge $\kappa_t = \frac{\lambda}{\nu} f$, and the amount of entrepreneurial slackness given by equation (2.9), we find that the growth rate g of the technological parameter $x(t)$ is equal to

$$g = ln\gamma\, C\, m\left(\frac{\lambda}{\nu} f; \frac{\delta(1-f)}{\theta q(\theta)+\delta}\right), \tag{DE}$$

where θ, or tightness in the market for entrepreneurial skills, solves the simple non-linear equation

$$\theta = \frac{\kappa_t}{s_t} = \frac{\lambda}{\nu}\frac{[\theta q(\theta)+\delta]f}{\delta(1-f)}. \tag{THE}$$

Equations (*DE*) and (*FE*), given the constraint imposed by (*THE*), completely solve the model in the growth rate g, relative research effort f space (Figure 3). Equation (*THE*) merely expresses tightness as the ratio of the steady state amount of scientific knowledge to the amount of slackness given by equation (2.9). Equation (*DE*) defines the frontier of technological possibilities of the economy. It defines a strictly concave relation between growth and relative research effort (Figure 3), satisfying the property that no growth can be the outcome of either too much or too little research effort (see appendix 7 for a formal derivation). As both research and entrepreneurial skills are required to sustain growth, an over-allocation to either factor harms the growth process.

Condition (*FE*) defines an equilibrium locus and tells us, for any given level of the growth rate, the relative amount of resources that the economy will end up devoting to research. This condition defines a strictly positive relation between growth rate and research effort, mapping the zero one interval over the whole real line (see appendix 7). The positive slope of the relation (Figure 3) is a consequence of the nature of research as an investment. A researcher incurs an immediate cost in the expectation of future rewards that will arrive only after the implementation of valuable inventions. The higher the growth rate, the lower will be the effective discount rate of the researcher and the higher the gains associated with a successful innovation. As a change in the discount factor always tends to have a greater effect on the investment with the longer time horizon, an increase in the level of the growth rate will always make research relatively more profitable than entrepreneurship. The steady-state equilibrium is defined by the point at which the (*FE*) and (*DE*) condition cross (point A in Figure 3). At that point no researcher has any incentive to become an entrepreneur (or vice-versa) and the economy will grow at the constant steady state growth rate defined by the technological frontier (*DE*). In appendix 7 we show how any point on the technological frontier (*DE*) can be sustained as equilibrium of our economy. The basic intuition is the same as the one in Hosios (1990) and Caballero and Hammour (1996). An economy characterised by ex-ante competitive relationships but ex- post bilateral monopolies has no compelling tendency to coordinate itself towards the social optimum, as there is no reason why the outcome of the bilateral bargain should reflect the actual social value of the function pursued by each agent. If the allocation of resources takes place through the market, agents choose how to allocate their skills according to their private incentives. Moreover, if there are both rents to be shared among different rentiers and externalities associated with innovating, it will be mere chance that private and social interests coincide. In the next subsection we investigate the structure of the constrained social optimum.

2.3.3 Welfare Implications

In this section we analyse the social planner's problem and show how the constrained social optimum always lies on the positively sloped arm of the technological frontier defined by condition (*DE*) plus (*THE*). In the steady state, the social planner maximizes the aggregate flow of utility in output units. If we indicate with $n(f)$, g, θ and $x(0)$ respectively the number of enterprises operating in the economy, the steady state growth rate, the level of tightness in the market for entrepreneurial skills, and the starting value of the level of technology, we find that, in the steady state, the social planner solves

$$\max_f \int_0^\infty \left\{ Pn(f) - \frac{\lambda}{\nu}\chi f + h[1 - f - n(f)] \right\} x(s) e^{-rs} ds =$$
$$\max_f \left[(P-h)n(f) - (\frac{\lambda}{\nu}\chi + h)f + h \right] \frac{x(0)}{r - g(f)}, \tag{2.10}$$

where

$$n(f) = (1-f)\frac{\theta q(\theta)}{\theta q(\theta) + \delta}, \tag{FI}$$

$$g(f) = ln\gamma\, C\,(1-f)\frac{\delta\theta q(\theta)}{\theta q(\theta) + \delta}, \tag{DE'}$$

$$\theta = \frac{\lambda}{\nu}\frac{[\theta q(\theta)+\delta]f}{\delta(1-f)}. \qquad \text{(THE')}$$

It is clear from equation (2.10) that the problem of the social planner consists of maximising both the growth rate and the number of firms operating in the economy, while keeping the number of researchers as low as possible. The economic intuition is pretty simple. Given the objective of maximising the present discounted value of output, the central planner will maximise the growth rate keeping as high as possible the level of utilisation of the resources. This involves keeping as high as possible the number of firms operating in the economy and as low as possible the costs of getting a given outcome, namely the number of researchers. From equation (*FI'*) and (*DE'*) we note how the value of f that maximise the number of operating firms, $n(f)$, maximises also the growth rate of the economy g. As condition (*FI'*) and condition together (*DE'*) define a strictly concave hump-shaped relation, while the cost component implied by f is linear, the social optimum will always lie on the strictly positive arm of the technological frontier. More formally, we find that the derivative of the objective function of the social planner with respect to f is equal to

$$\begin{aligned}(\quad & P-h)\frac{n'}{r-g(f)}+\frac{1}{[r-g(f)]^2}[(P-h)n(f)-\frac{\chi}{\nu}+h]g'+\\ - \quad & (\frac{\chi}{\nu}\chi+h)\frac{1}{r-g(f)},\end{aligned}$$

where g' and n' indicate respectively the derivatives of $g(f)$ and $n(f)$ with respect to f. This derivative is clearly negative at the point in which $n' = g' = 0$, implying that the social optimum will always lie to the left of the point that maximises the growth rate of the economy[3].

2.4 Lack of Entrepreneurial Skills

The model presented in the previous section shows that an increase in research effort can crowd out more socially useful entrepreneurial skills and ultimately harms the growth process. As the amount of relative research effort increases, the stock of scientific knowledge increase, but it also becomes progressively more difficult to implement inventions both because more researchers are competing for the same resources (congestion externalities) and because an increase in research effort crowds out useful entrepreneurial skills (thin market externalities). If the stock of scientific knowledge is already big while the amount of entrepreneurial skills is low, an increase in research effort reduces the growth rate of the economy, because it misallocates socially useful resources. If this is the case, the growth rate stagnates despite the increase in research effort because of the *lack of entrepreneurial skills.* More formally, an increase in research effort can be translated into stagnant or declining growth rates, if for example the equilibrium of the economy shifts from point A to point B (Figure 3). The considerations in the previous section show how this equilibrium shift is inefficient but nonetheless feasible. If so, research becomes more profitable relative to entrepreneurship despite there having been no increase in its social productivity.

Despite an upward trend in the relative amount of research effort (Figure 2) and high education expenditures (Figure 4), the US growth rate has

[3]The social optimum growth rate will also be different from zero as the derivative of the technological frontier at $f = 0$ is infinity.

stagnated or even declined, so that the productivity of research effort has been decreasing over time (Figure 1). At the same time, the amount of entrepreneurial skills in the US economy, as measured by the population of self-employed[4] (see i.e. Evans and Jovanovic, 1989 and Evans and Leighton, 1989) has dramatically fallen. If we examine Figure 1, 2 and 4 together, it seems that the increasing amount of education and research effort financed by the US society has been to the detriment of more socially useful entrepreneurial skills.

In what follows we try to explain theoretically why the private returns from entrepreneurship have progressively fallen in the US over the post-war period. Firstly, we look for shifts in the underlying structural parameters of the model. Secondly, we consider the effects of changes in the fiscal structure. In particular we focus on the effects of subsidies to research and to firms.

2.4.1 Changes in Structural Parameters

It is possible to posit changes in the underlying parameters of the model able to explain the increase in the relative profitability of research. We focus on three possible candidates: changes in the discount factor, changes in the private value of an innovation and changes in the obsolescence rate of an

[4]It is reasonable to ask how good this proxy might be. There are grounds to believe that it satisfies some basic Schumpeterian criteria (see Schumpeter, 1949). According to Schumpeter (1949) the entrepreneurial function can be identified only a posteriori: "our definitions of entrepreneur, entrepreneurial function and so on can only grow out of it *a posteriori*" and again "when we speak of the entrepreneur we do not mean so much a physical person as we do a function, but even if we look at individuals who at least at some juncture in their lives fill the entrepreneurial function it should be added that these individuals do not form a social class. They hail from all the corners of the social universe". In fact, Evans and Jovanovic (1989) note how, in their sample, 20 percent of the individuals who switched into self-employment formed incorporated businesses.

innovation.

A fall in the discount factor, r, makes research relatively more profitable (equation *FE*). This arises from the nature of research as an investment. A researcher incurs an immediate cost in the expectation of future rewards that only arrives after the implementation of valuable inventions. As a change in the discount factor always tends to affect the investment with the longer time horizon more severely, a fall in the discount factor will always make research more profitable relative to entrepreneurship. Barro and Sala-I-Martin (1995, p.6) cite empirical evidence that seems to support this, as they claim that real returns exhibit a tendency to fall over some ranges as the economy develops. This would be consistent with a decreasing value of relative risk aversion at higher level of wealth.

Innovating creates rents that must be shared. These rents are the rewards that society leaves to both researchers and entrepreneurs in order to remunerate them for the inter-temporal positive knowledge spillovers embodied in equation (*DE*) and to sustain the growth process. These rents are shared according to an ex-post bilateral bargaining whose outcome does not necessarily reflect the social contribution of each agent to the introduction of new technology in the economy. In our framework the rent-sharing is modelled as the Nash bargaining solution to a bilateral monopoly. Here a fall in the bargaining power of entrepreneurs β, makes research relatively more profitable (see equation *FE*) in the absence of changes in the fundamentals of the economy. Schankerman and Pakes (1986, Table 5) find that the average value of a patent in France, UK and Germanyin 1975 is on average 70% higher than it was at the beginning of 1955. If we assume that the value of a patent measures the rents from innovating accruing to the researcher, this would imply an upward trend in the value of I_t in our model[5]. Equation (2.8) implies that as time goes by the value of an innovation increases at the same rate as the growth rate of technology. If we assume that the trend observed by Shankerman and Pakes (1986) has been repeated in the US, with similar magnitude, and that the average growth rate of total factor productivity per year has been equal to about 0.4%, as found in Young (1995a), this would mean that about 62% of the increase in $I_t = Ix(t)$ is left unexplained and must therefore be attributed to a change in its steady state value I. Even if we take the growth rate of GDP per worker as the right proxy for the growth rate of the economy we still find 33% of this increase to be unexplained. In order to explain such an increase, one could perhaps invoke a rise in the level of the bargaining power of researchers, $1-\beta$. Schumpeter (1943, 1946) attributes the fall of the bargaining power of entrepreneurs to cultural and sociological shifts that tend to destroy the "protective strata" able to sustain the entrepreneurial function. More specifically, β measures the outcome of a bilateral bargaining problem here specified parsimoniously. For example, we could assume the presence of, but not explicitly modelled, asymmetric information, the size of which depends on the technological content of the

[5]Alternatively, we might think that the value of a patent measures the total private surplus associated with innovating, S_t. If this is the case, the calculations below would apply to S_t rather than I_t. In the model, any increase in the total private surplus from innovating increases the relative profitability of research (see equation *FE)*. This comes out because of both the nature of research as an investment and the fact that researchers and entrepreneurs appropriate fractions $1-\beta$ and β, respectively, of the total private surplus from innovating (see equation 2.6). As a change in the level of future rewards always tends to affect more the investment with the longer time horizon because of leverage, an increase in the total private surplus from innovating, S_t, will always make research relatively more profitable than entrepreneurship.

innovation, if so, there would be a bias in favour of the researcher. At higher level of development the ability of the researcher to appropriate rents would be increased.

Particularly rich are the dynamics associated with changes in the obsolescence rate of an innovation, δ. A fall in δ increases the relative profitability of research for any given level of tightness, θ, in the market for entrepreneurial skills (see condition *FE*). In fact, for given probability of matching with a free entrepreneur (i.e. given θ), the gains associated with research increase thanks to higher gains associated with a successful innovation (lower effective discount rate of an innovation). However, for a given value of f, the value of θ tends also to increase, as a fall in δ reduces the number of entrepreneurs who are currently searching for some invention to implement. The ultimate impact on the (*FE*) curve, defined as it is in the growth-relative research effort space, is thus ambiguous. What is not ambiguous is the impact on the technological frontier defined by (*DE*). A fall in the obsolescence rate of an innovation δ implies that the economy tends to renovate less, so that for a given allocation of effort the amount of entrepreneurial slackness is lower. This implies a downward shift of the technological frontier *DE*, with the hump moving leftward as the constraint imposed by the lack of entrepreneurial skills becomes more binding for each given level of relative research effort, f[6]. In other words, an economy that liquidates more often (higher

[6]To prove the downward movement of the (DE) condition we merely note that for given f a fall in δ reduces the steady state level of slackness given by (2.9) both directly through δ and indirectly through the increase in θ. As slackness falls for any given level of f, the constraint imposed by the thin market externality becomes more binding as f increases. This explains the leftward movement of the hump. A formal proof of this last proposition is contained in the appendix.

δ) has more frequent opportunity to rebuild its productive stock at a higher technological level and grows faster thanks to the inter-temporal spill-over typical of standard endogenous growth models (see i.e. Romer, 1990). Caballero and Jaffe (1993, Figure 9) show how the degree of obsolescence of innovations has exhibited a strong downward trend over the 1964-1990 period although they do question the reliability of their estimate at the end of their sample[7]. This support the notion that the constraint imposed by the amount of entrepreneurial slackness has become progressively more binding over time.

2.4.2 Changes in Fiscal Structure

As they merely redistribute resources across occupations, fiscal changes alter the private profitability of an occupation without affecting its social value. As a result, changes in the fiscal structure are the most obvious candidates to explain why the US economy ended up by suffering from a lack of entrepreneurial skills.

In what follows we focus on two forms of fiscal intervention: subsidies to research, and more generally to education, and subsidies to firms. In order to isolate their relative impact we assume that any form of fiscal intervention is financed with lump-sum transfers to the whole population, independent of the occupation currently chosen. We model the effects of a subsidy to research, b_r, by assuming that the actual flow cost of searching, χ, is thereby reduced

[7]If we had endogenised the obsolescence rates following along the lines of Mortensen and Pissarides (1994), δ would reflect the value of the firm and thus the option of keeping it operating. At high levels of development the technological content of a firm increases, raising its value and thus reducing the willingness to liquidate.

to $\chi - b_r$. From the researcher's point of view, a subsidy to research reduces the actual cost sustained today in the expectation of the future rewards that he gets after the implementation of inventions. We model the effects of a subsidy to firms, b_f, by assuming that the actual flow of goods produced by the firm is equal to $P + b_f$. A subsidy to a firm increases the return that agents get from running it. Fiscal changes do not change the technological frontier of the economy while they do modify the allocation of skills in the economy. Condition (*FE*) can then be rewritten as:

$$\beta\theta q(\theta)S + h = \lambda\frac{(1-\beta)q(\theta)S - \chi + b_r}{r + \nu - g}, \tag{FE}$$

where $S = I + E - H$ measures the private net surplus associated with the creation of a new firm and is equal to

$$S = \frac{P + b_f - h}{r + \delta + \beta\theta q(\theta) - g}.$$

Subsidies to research and subsidies to firms work in very different ways. An increase in the level of subsidies to research, b_r, reduces the cost of doing research and increases the relative profitability of being a researcher as measured by the right hand side of equation (*FE*). This drives up the relative amount of resources, f, that society ends up allocating to research. An increase in the level of subsidies to firms, b_f, increases the incentive to create firms. These rewards are appropriated only after the match has taken place and thus they are split between the researcher and the entrepreneur with fractions $1 - \beta$ and β respectively. In the model, any increase in the total private surplus from innovating, S, increases the relative profitability of research[8] (see equation *FE)*. As a change in the level of future rewards always tends to more severely affect the investment with the longer time horizon, an increase in the total private surplus from innovating , S, will always make research more profitable relative to entrepreneurship.

Figures 5 and 6 show how both subsidies to firms and subsidies to research and education, as proportion of GDP, have significantly increased over the post-war period in the US. According to our previous considerations, this should have made research more profitable relative to entrepreneurship despite there having been no change in its underlying social value.

2.5 Early Attempts to Test the Lack of Entrepreneurial Skills Hypothesis

In this section we use both time series and cross sectional variation in order to test implications of the lack of entrepreneurial skills hypothesis.

If we assume that the matching function is Cobb Douglas, that is

$$m(i_t; s_t) = A(i_t)^{\alpha}(s_t)^{\beta}, \qquad 0 < \alpha, \beta < 1, \tag{2.11}$$

equation (DE) then suggests we run the following regression

$$ln(g_t) = \alpha \ \ln(f_t C_t) + \beta \ \ln(s_t) + \alpha \ \ln(\frac{\lambda}{\nu}) + \ln(\ln \gamma) + \ln(A). \tag{2.12}$$

We implement equation (2.12) for the US over the period 1950-1990. We follow Griliches (1979, 1990) in taking the total number of patent applica-

[8]Provided that the nature of research as an investment is preserved, that is $-\chi + b_r$ remains lower than h.

tions as an index of innovative activity in order to proxy for g_t[9], and Jones (1995b) in considering the ratio of scientists and engineers involved in $R\&D$ over the population if working age as a measure of relative research effort f_t. As a proxy for the scale of the economy C_t in equation (2.12) we take the population if working age, while we follow Evans and Jovanovic (1989) and Evans and Leighton (1989) in using as a measure of the amount of entrepreneurship in the US economy the population of self-employed[10]. In order to control for cyclical disturbances we also consider a specification with the unemployment rate as an independent variable[11]. Table 1 shows the estimates obtained by running the regression (2.12), allowing for different lags in the

[9] Whether a patent corresponds more to the notion of invention or innovation, is a topic beyond the scope of this chapter. The Patents Laws-United State Code, U.S.C. 101, states that " whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent". The UK Patents Act 1977 states that the invention to be patented must be "capable of industrial application". More generally Cornish (1989) states that "a patent cannot be granted for a thing or process which however interesting or suggestive it might be to scientists, has no known practical application at the priority date". The requisite of "industrial applicability" is not contained in the US law. It seems however, that the requisite of "usefulness" roughly coincides with that of "industrial applicability", so that even if a patent still represents something between an invention and an innovation as defined in this chapter, it would seem to be closer to the latter.

[10] See footnote 4 for a discussion.

[11] A more formal and alternative justification for introducing the unemployment rate in running equation (2.12) derives from the direct substitution of the steady state value of the amount of entrepreneurial slackness, s, given by equation (2.9). After doing so equation (2.12) looks like

$$ln(g_t) = \alpha \ \ln(f_t C_t) + \beta \ln(C_t - f_t C_t) + \beta \ln\left(\frac{\delta_t}{z_t + \delta_t}\right) + \alpha \ \ln(\frac{\lambda}{\nu}) + \ln(\ln\gamma) + \ln(A),$$

where $\ln\left(\frac{\delta_t}{z_t+\delta_t}\right)$ measures the percentage of entrepreneurs currently searching and thus "unemployed". This last interpretation, however, does not seem all that compelling as the coefficient in front of the unemployment rate turns out to have the wrong sign (negative rather than positive).

Indep. Variable	1 lags	2 lags	3lags	4lags	1 lags	2 lags	3 lags
Const.	1.21 (.31)	0.98 (.31)	0.76 (.33)	0.39 (.36)	0.56 (2.21)	0.19 (.41)	0.05 (.49)
$\ln(f_{t-i}\,C_{t-i})$	0.47 (.03)	0.49 (.03)	0.50 (.04)	0.52 (.04)	.50 (.04)	.53 (.04)	.54 (.04)
$\ln(s_{t-i})$	0.24 (.10)	0.29 (.10)	0.37 (.10)	0.48 (.10)	.30 (.10)	.37 (.09)	.46 (.10)
Unempl. Rate	-	-	-	-	-.11 (.05)	-.12 (.036)	-.10 (.04)
R^2	0.84	0.86	0.86	0.86	0.87	0.88	0.88
Wald Test (p-value) $\alpha+\beta=1$	0.005	0.03	0.24	0.94	0.06	0.37	0.96

Table 2.1: **Results from running regression (2.12).** The regression was ran using OLS, the dependent variable being the number of patent applications. Similar results emerge when allowing for first order auto-correlation in the residuals and a non-scalar auto-covariance matrix. These last results are available upon request.

relation between research effort, entrepreneurship and innovation.

The model presented in the previous section has two important testable implications. The first is that the parameters α and β must both be positive numbers strictly between zero and one. Table 1 shows this is actually the case. In fact both α and β are positive and significantly different from zero with α ranging between 0.47 and 0.54 while β ranges from a value of 0.24 to 0.46. This implies that the relation between the number of innovations and research effort $f_t\,C_t$ turns out to be concave and hump-shaped. The second testable implication is stronger and allows one to discriminate between the two competiting theories purporting to explain the alleged decreasing returns in R&D: the vanishing of investment opportunities and the lack of entrepreneurial skill hypothesis. The fact that the parameters α and β are positive and significantly different from zero implies that this last one deserves some credit. It is however a more "plausible" one only if the matching function (2.11) exhibits constant returns of scale, that is $\alpha + \beta = 1$. In fact, if so, the economy exhibits scale effects and well-balanced increases in the amount of research effort and entrepreneurial skills have constant effects at the margin on the number of innovations introduced in the economy. If we consider reasonable a lag of over two years before research effort shows up in a patent application, Table 1 shows how the hypothesis that the matching function (2.11) exhibits constant returns to scale cannot be rejected even at a 10% level of significance. Moreover, in the specification with the unemployment rate as independent variable, this hypothesis cannot be rejected whatever the number of lags considered.

The main claim of this chapter is that in a world where the allocation between research and entrepreneurial activities takes place through the market, the rents from innovating are ex-post shared by researchers and entrepreneurs and there are externalities associated with innovation, there is no compelling reason why private incentives should coincide with social ones. This implies that the economy might end up by devoting too many resources to research while neglecting neglecting other important functions for sustaining growth and ultimately reaching a Pareto dominated equilibrium. If this is the case, the relation between relative research effort and growth is hump shaped. Figure 4 plots the average growth rate of the index of total factor productivity for the set of OECD countries considered by Coe and Helpman (1993) together with the ratio of the *R&D* expenditure over GDP. The hypothesis of non monotonicity in the relation between growth and *R&D* cannot be rejected according to cross-sectional evidence, with the US, the UK and Switzerland settled on the negatively sloped arm of the growth-R&D technological frontier.

2.6 Conclusions

This chapter has shown how an increase in the amount of resources devoted to research does not necessarily increase the growth rate of the economy. In a world with rent sharing and inter-temporal spill-overs an increase in research effort can crowd out more socially useful entrepreneurial skills, reduce the growth rate and ultimately be Pareto worsening.

If this is the case, the observed increases in research effort have been excessive, raising a call for policy intervention. Indirect inference on the model, observed changes in the fiscal structure and more formal empirical investigation all seem to support the relevance of this hypothesis for the US. The Schumpeterian warning, that the "state of decay of the capitalist society" will ultimately be driven by the lack of entrepreneurial skills, might have some theoretical and empirical foundation. If and how relevant this notion might be in practice remains a question for further research and one that only more careful micro-based empirical investigation can solve.

2.7 Appendixes

2.7.1 Appendix A: Technical Appendix

In this Appendix we derive some technical results discussed in the text.

Properties of the THE Condition

Proposition 1: The (*THE*) condition defines tightness in the market (θ) as a monotone increasing function of f, call it $\theta(f)$, with the property that $\theta(0) = 0$ and $\theta(1) = \infty$.

Proof: To prove monotonicity, we first note that the right hand side of condition (*THE*) is an increasing function with respect to θ, with elasticity less than 1. We then we note that the right hand side is also a monotone strictly increasing function with respect to f. The fact that the left hand side is a linear function proves monotonicity. The properties of the function $\theta(f)$ at the boundary follow from a direct analysis of condition (*THE*) together with the fact that the elasticity of the right hand side with respect to θ is less than one.

Properties of the DE Condition

Proposition 2: The (*DE*) condition defines the growth rate of the economy g, as a concave function of f, call it $d(f)$, with the property that $d(0) = d(1) = 0$.

Proof: We first note that (2.9) together with proposition 1 define slackness in the market for entrepreneurial skills as a decreasing function of f, call it $s(f)$. This implies that the function $d(f)$ defined by (*DE*) is equal to

$$g = d(f) = ln\gamma m\left(C\frac{\lambda}{\nu}f; s(f)\right).$$

It follows that the second derivative of d with respect to f is equal to

$$\frac{d^2g}{df^2} = ln\gamma[(C\lambda)^2 m_{11} + 2C\lambda m_{12}s'(f)]\frac{1}{1+\frac{m_2}{ln\gamma\delta}} < 0,$$

where m_{ij} indicates the second derivative of $m(\cdot,\cdot)$ with respect to the ijth argument. The negativeness of the derivatives follow from the concavity of the matching function and the complementarity of each of its arguments, with $m_{12} > 0$. The properties of function $d(f)$ at the boundaries follow from condition (*A1*) and proposition 1.

Properties of the FE Condition

Proposition 3: The (*FE*) condition defines the growth rate of the economy g as a monotone increasing function of f, call it $e(f)$, with the properties that $e(0) = -\infty$ and $e(1) = \infty$. Moreover, any point on the technological frontier defined by (*DE*) can be sustained as a competitive equilibrium of the economy.

Proof: The left hand side of condition (*FE*) defines a function strictly increasing in θ and equal to $h(r+\delta-g)$ when $\theta = 0$. The right hand side of condition (*FE*) defines a function strictly decreasing in θ which, if condition (*A3*) holds, approaches infinity when θ goes to zero. This implies that a solution always exists. An increase in g causes a downward (upward) shift in the left (right) hand side so that the equilibrium level of tightness unequivocally increases. Together with proposition 1, this proves that $e(f)$ is strictly increasing in f. The properties of the function $d(f)$ at the boundary follow from condition (*A1*), proposition 1 and the fact that the right hand side goes to plus (minus) infinity when θ goes to zero (infinity). To prove the last part of the proposition, we note how changes in P, h, χ and β change, for a given level of g, the value of θ implied by condition (*FE*) without altering condition (*THE*). This implies that, for given g, there does exist a combination of parameters such that any value of θ is sustained as an equilibrium. Proposition 1 and the fact that these changes do not affect either (*THE*) or (*DE*) imply that any combination of f and g on the technological frontier defined by (*DE*) can be sustained as an equilibrium.

A fall in the Obsolescence Rate, δ, moves the Hump Leftward

The derivative of the growth rate defined by equation (*DE*) with respect to f is equal to

$$\frac{d\,g(f)}{d\,f} = ln\gamma\,\frac{m_1\frac{\lambda}{\nu} - m_2 C}{1 + m_2\delta C \ln\gamma}$$

where $s(f)$ defines the steady state value of slackness as defined by equation (2.9), while m_i defines the derivative of the matching function with respect to the ith argument. Noting that for given f a fall in δ reduces the steady state value of slackness $s(f)$, that the matching function is concave, that m_{12} is positive and finally that at the peak of the hump $m_1\frac{\lambda}{\nu} - m_2 C = 0$, we see that a fall in δ reduces the slope of the technological frontier at the value of f corresponding to the previous peak. This proves the leftward movement of the hump.

2.7.2 Appendix B: Data Appendix

Total Factor Productivity Growth data for the period 1970-1990 are taken from Coe and Helpman (1993).

The data for the real value of the expenditure in $R\&D$ for the OECD countries for the period 1970-1990 are calculated using information in Coe and Helpman (1993).

Data on Patent Applications, self-employment, expenditures in education, number of Scientists and Engineers involved in $R\&D$, subsidies to firms, and on the labor force are taken from various issues of The Statistical Abstract of the United States and from Historical Statistics of the United States: Colonial Times to 1970.

Data on the sources of funds of basic and applied research expenditure are taken from National Science Foundation/SRS "National Patterns of $R\&D$ Resources, 1994".

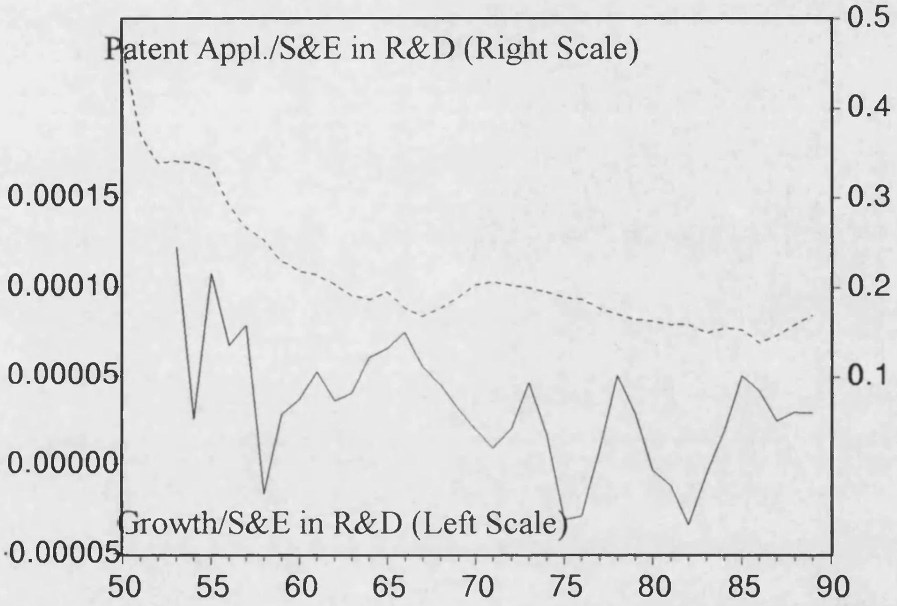

Figure 2.1: Decreasing Returns in R&D: US (1950-1990). Source: see data appendix. Regressions: Growth over S&E in R&D =5.43E-05(5.469)-1.34E-06(-2.86)tt. Patent Application over S&E in R&D= 0.26577(28.07)-0.003728(-8.3532)tt. Notes: t-statistics in parentheses, Growth= 3 year moving average of the growth rate of real GDP per Worker, S&E in R&D= Scientists and Engineers involved in R&D, tt=time trend.

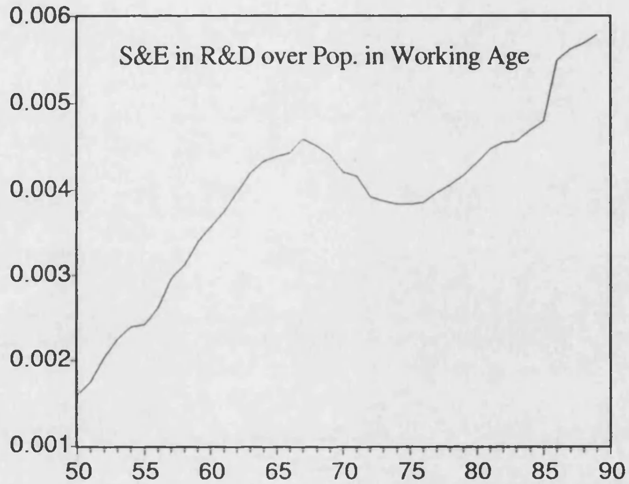

Figure 2.2: Relative allocation of resources in the R&D sector: US (1950-1990). Source: see data appendix.

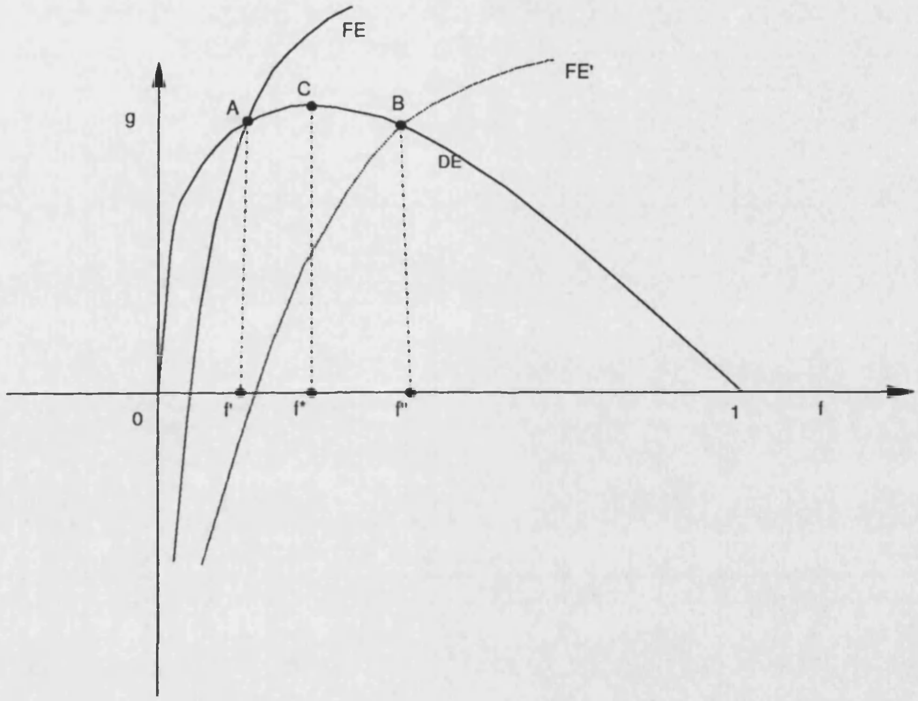

Figure 2.3: Steady state equilibrium

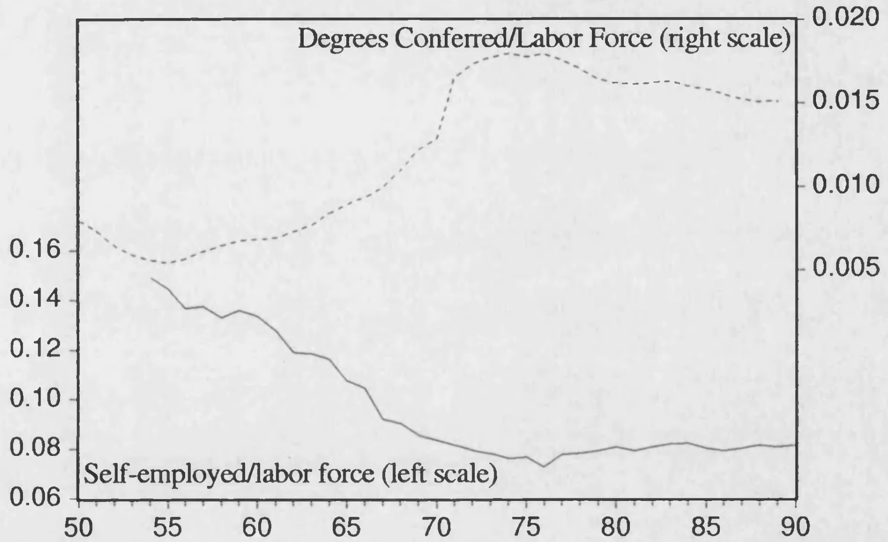

Figure 2.4: Lack of Entrepreneurial Skills (US 1950-1990). The dotted line shows the dynamics of the number of degrees conferred by institutions of higher education (Bachelor's, Master's, Doctor's) as a ratio of the labor force. The bold line shows the dynamics of the number of self-employed over the labor force. Source: see data appendix.

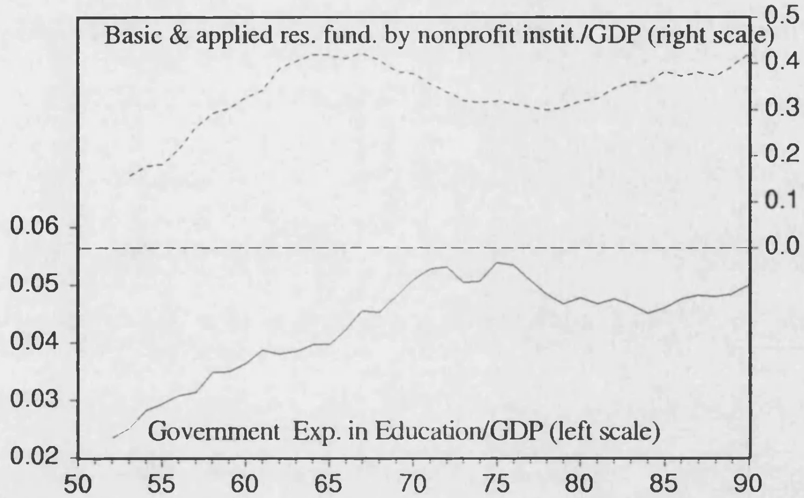

Figure 2.5: Subsidies to research (US 1953-1990). The dotted line shows the dynamics of the amount of basic and applied research expenditure funded by non profit institutions (Federal Government, Universities & Colleges, other nonprofits institutions) over GDP. The bold line shows the dynamics of the total amount of state and local government expenditures in education as a ratio of GDP. Source: see data appendix.

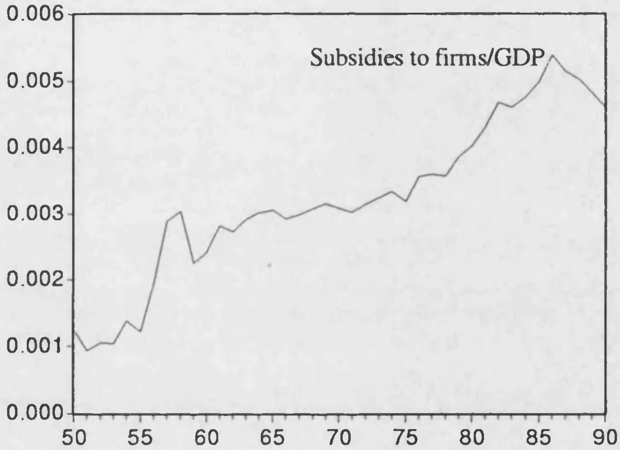

Figure 2.6: Subsidies to Firms (US 1950-1990). Source: see data appendix.

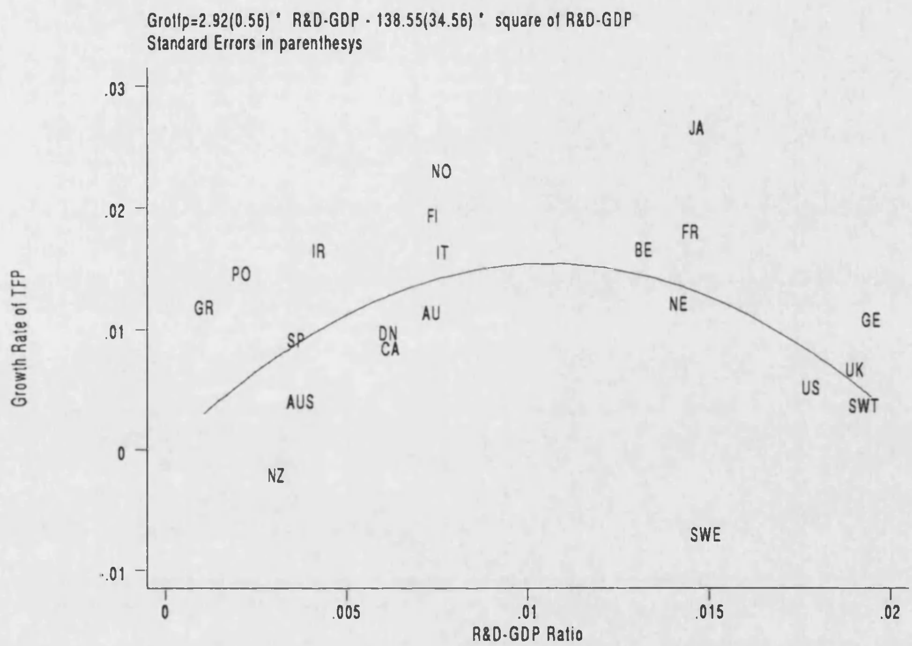

Figure 2.7: R&D and Growth: OECD Countries (1970-1990). Source: see data appendix.

Chapter 3

(Fractional) Beta Convergence

Abstract

Unit roots in output, an exponential 2% rate of convergence and no change in the underlying dynamics of output seem to be three stylized facts tha can not go together. This paper chapter the Solow-Swan growth model allowing for cross-sectional heterogeneity. In this framework, aggregate shocks might vanish at an hyperbolic rather than at an exponential rate. This implies that the level of output can exhibit long memory and that standard tests fail to reject the null of a unit root despite mean reversion.Exploiting secular time series properties of GDP, we conclude that traditional approaches to test for uniform (conditional and unconditional)convergence suit first step approximation. We show both theoretically and empirically how the uniform 2 % rate of convergence repeatedly found in the empirical literature is the outcome of an underlying parameter of fractional integration strictly between 0.5 and 1. This is consistent with both time series and cross-sectional evidence recently produced.

3.1 Introduction

The debate on unit roots and stochastic trends has dominated macroeconometrics over the eighties. Since the seminal work of Nelson and Plosser (1982), this literature has noted how standard unit roots tests have failed to reject the null of a unit root in output per capita. The nineties has signed the revival of the empirics on growth and convergence. Conditional uniform convergence, namely Beta convergence, means that aggregate shocks are absorbed at an uniform exponential rate. Most of empirical studies conclude that outputs per capita of very different economies converge to their long run steady state values at a uniform exponential rate of 2% for year,(see for example Barro , 1991, Barro and Sala-I-Martin, 1991, 1995, Mankiw, Romer and Weil 1992). These seem to be two of the most striking empirical regularities in modern empirical macroeconomics. More recently, Jones (1995b), has observed that in line with the standard exogenous growth Solow model, the trend of output per capita for OECD economies is pretty smooth over time and does not exhibit any persistent changes in the post World War era.

These three stylized facts seem to be inconsistent. On the one hand a unit root in output implies that shocks are permanent so that output does not exhibit mean reversion. On the other hand Beta convergence, henceforth β-convergence implies that output converges to its steady state level at a rate that even if very low it is positive and uniform across economies. The Jones invariance property implies that steady state output could well be represented by a smooth time dependent linear trend. If this is true, unit roots tests and β-convergence are testing for the same hypothesis.

This chapter starts from the observation that the "size of the unit root" component in GDP (the long-run effect of a unit shock) is usually found to be very low, (see Cochrane 1988, Cambell and Mankiw 1987, and Lippi and Reichlin 1991) and follows Quah (1995) in noting that cross-sectional and time series analysis can not get different conclusions. In agreement with Diebold and Rudebusch (1989), and Rudebusch (1993) we propose a different explanation. Perhaps the speed with which aggregate shocks are absorbed is so low that standard unit roots tests fail to reveal it[1]. This could actually be the case if GDP per capita exhibits long memory (Diebold and Rudebusch 1991). If we consider the standard Solow-Swan model and we allow for cross sectional heterogeneity in the speed with which different units in the same countries adjust, we show that the dynamics of output can exhibit long memory. We can then test for both uniform conditional and unconditional convergence allowing for rate of convergence different from the exponential one. In this framework, we show how a 2% percent rate of convergence superimposed as exponential and estimated over a time span that ranges from a minimum of 20 years to a maximum of 100 years correspond to a parameter of fractional integration that ranges from 0.51 to 0.99. This process is not covariance stationary but still mean reverting, so that standard unit roots test are likely not to reject the null of non stationarity despite the fact that convergence takes place. Using GDP per capita data for OECD countries for the period 1885- 1994, we test for this hypothesis. We conclude, that it can not be rejected, so that convergence takes place at an hyperbolic very slow

[1]Diebold and Senhadji (1996) that Rudebusch (1993) approach produces evidence that distinctly favors trend-stationarity using long spans of annual data.

rate.

The contribution of this chapter is to put together two different strands of research. On the one hand, time series analysis has concluded that shocks tend to have permanent effect on the level of output. On the other hand the literature on growth and convergence has concluded that countries converge to their long run steady state value at an exponential rate that is very low and uniform across countries. In this chapter we note that the two literatures are inconsistent once we allow for the invariance property by Jones and we follow the standard exogenous growth Solow model in approximating the dynamics of long run GDP per capita with a linear trend. In line with Diebold and Senhadji (1996) we propose a theoretical solution and we test for it. We conclude that standard tests for convergence suit first step approximation despite the mispecification of the empirical model. In doing so we show that the parameters of fractional integration of different OECD countries, though of similar magnitude and smaller than one, are significantly different one from the other. This delivers a possible explanation of why time series tests of convergence based on cointegration reject the null of convergence even among OECD countries (see for example Quah 1992, Bernard and Durlauf 1993 and Bernard and Durlauf 1996). As they are these tests are mispecified as different variables can be cointegrated only if they exhibit the same order of integration.

Section 1 reviews the Solow-Swan model. In this context we highlight further why the three stylized facts can not go together. Section2 briefly reviews the theory of long memory processes and shows how in a extension of the theoretical model, the path of adjustment of output can exhibit long memory. In this context we show why standard unit roots can not reject the null of a unit root while a uniform 2% rate of convergence can be found to be statistically significant. In this framework we check for uniform (conditional and unconditional) convergence. This is done in section 3. Section 4 concludes.

3.2 Empirics of the Solow Growth Model and Unit Roots

We begin by briefly reviewing the Solow growth model. We then focus on the time series properties of the reduced form of the model.

Solow Growth Model

Solow model takes the rates of saving and technological progress as exogenous. There are two inputs, capital and labor. We assume a Cobb-Douglas production function, so production at time t is given by

$$Y(t) = K(t)^{\alpha}(A(t)L(t))^{1-\alpha}, \qquad 0 < \alpha < 1.$$

The notation is standard:Y is output, K capital, L labor and A the level of technology. A is assumed to grow exogenously at rate g.

The model assumes that a constant fraction of output s is invested. Defining $\hat{k}$ and $\hat{y}$ as respectively the stock of capital and output per effective unit of labor, $\hat{k} = K/AL$ and $\hat{y} = Y/AL$, the evolution of $\hat{k}$ is governed by

$$\frac{d\,\hat{k}_t}{dt} = s\hat{k}_t^{\alpha} - (g+\delta)\hat{k}_t, \tag{3.1}$$

where δ is the depreciation rate. Equation (3.1) implies that $\hat{k}_t$ converges towards a steady state level $\hat{k}^*$ defined by

$$\hat{k}^* = (\frac{s}{g+\delta})^{\frac{1}{1-\alpha}}. \tag{3.2}$$

We can then consider a log-linear approximation of equation (3.1) around the steady state so that

$$\begin{aligned} \frac{d[ln(\hat{y}_t)]}{dt} &= -\beta[ln(\hat{y}_t) - ln(\hat{y}_t^*)], \\ &\text{with} \\ \beta &= (1-\alpha)(g+\delta), \end{aligned} \tag{3.3}$$

where $\hat{y}^* = (\hat{k}^*)^{\alpha}$. Discretizing equation (3.3) and indicating with y_t the log of output per capita, viz. $y = ln(Y/L)$ and by y_t^* the log of the level of output per capita in steady states we get

$$y_t - y_{t-1} = g + \beta y_{t-1}^* - \beta y_{t-1}, \qquad 0 < \beta < 1, \tag{3.4}$$

or equivalently

$$y_t - y_t^* = (1-\beta)[y_{t-1} - y_{t-1}^*]. \tag{3.5}$$

We now analyze the time series properties of both equations (3.4) and (3.5).

Equation (3.4) is the basic equation used to test for β-convergence (see for example Barro 1991, Barro and Sala-I-Martin 1991 and Mankiw, Romer and Weil 1992). β-convergence applies if a poor economy tends to grow faster than a rich one. This arises if the coefficient β in equation (3.4) is found to be positive and significantly different from zero. If this is the case, aggregate shocks that have pushed the current level of output away from the steady state level will be absorbed at the exponential rate β so that the dynamics of output will exhibit mean reversion. The standard approach to test for this property consists of approximating $g + \beta y^*_{t-1}$ with some control or environmental variables like the investment rate, population growth, government expenditure and so on, then estimating the regression (3.4) and eventually testing for the significance of the coefficient β. In practice, empirical studies repeatedly find a 2% coefficient, uniform across countries and significantly different from zero (Quah 1993) .

A test of unit root, like for example the Dickey Fuller's test (Dickey 1979), still uses an equation like (3.4) and tests for the coefficient β being significantly different from zero, where the term $g + \beta y^*_{t-1}$ is substituted by a smooth time dependent function. A value for the coefficient β not significantly different from zero is interpreted as an hint of the presence of a unit root in the underlying data generating process. If this is the case a temporary shock has permanent effects on the level of output and the dynamics of output does not exhibit mean reversion towards the smooth trend. Since the seminal work of Nelson and Plosser (1982) these tests have not been able to reject the null of a unit root in GDP per capita, even if their low power is well recognized (see for example Diebold and Rudebusch 1991, Rudebusch 1993 and Diebold and Senhadji 1996).

In general the existence of a unit root in output is not in contradiction with β-convergence if we allow for the steady state level of output to be cointegrated with the current level of output. In this case aggregate shocks are still absorbed at an exponential rate despite the fact that output is integrated, as implied by equation (3.5).

Jones (1995a, 1995b) has observed that the dynamics of aggregate output has moved smoothly and independently of most of the controlled variable used for testing β-convergence. This is in line with the standard exogenous growth Solow model where the level of long run GDP per capita, y_t^*, is represented by the linear trend, gt. If we take the data from Maddison (1995) for 16 OECD countries over the period 1885-1994 and we plot the dynamics of per capita GDP versus a common linear trend among all the countries in the sample, we note that this simple common trend fits long run per capita GDP extremely well. This is shown in Fig.1 where we plotted each series together with a country specific linear trend and a common linear trend obtained pooling together the series of all 16 OECD countries in our sample. The former has been estimated with OLS, the latter with GLS. Particular informative is the GLS estimate of the common trend. GLS estimating procedure implies that the better the fit of the specific trend the greater is the weight of this country in the determination of the common trend. In this case the US case outperforms by far all the other countries. This shows up in the final outcome, in fact the common GLS trend and the US OLS specific trend are almost undistinguishable (see Figure 3.1). Thus we can think of the US performance as representing the long run benchmark of all the other countries' performances. Nelson and kang (1984) argue, however, that regressions of driftless integrated series against a time trend can result in the inappropriate inference that the trend is significant and that it is a good description of the data, as Durlauf and Phillips (1988) show. Instead Jones (1995a) notes how a time trend calculated using data only from 1880 to 1929 forecasts extremely well the current level of GDP of the US economy. Following Diebold and Senhadji (1996) this is clearly incompatible with difference stationarity in aggregate output, as new information seems to be irrelevant for forecasting on very long horizons.

This suggests that, in accordance with the standard exogenous growth Solow model where the level of long run GDP per capita, y_t^*, is represented by the linear trend, $g\,t$, the dynamics of steady states output mimics a simple trend. As a deterministic function can not be cointegrated with a variable exhibiting stochastic trends, it turns out that β-convergence and unit root tests are both checking for mean reversion towards a smooth time dependent trend. In a time series formulation we can say that β-convergence is testing for trend stationarity in output where the stationary disturbance is superimposed as an autoregressive process of order one [2]. These simple con-

[2]The nature of the problem is just further complicated by the fact that growth theorists use panel data instead of just time series. Recent results (e.g. Levin and Lin 1992) show, however, that panel data just make dramatically increase the power of a unit root test as the cross sectional dimension increases.

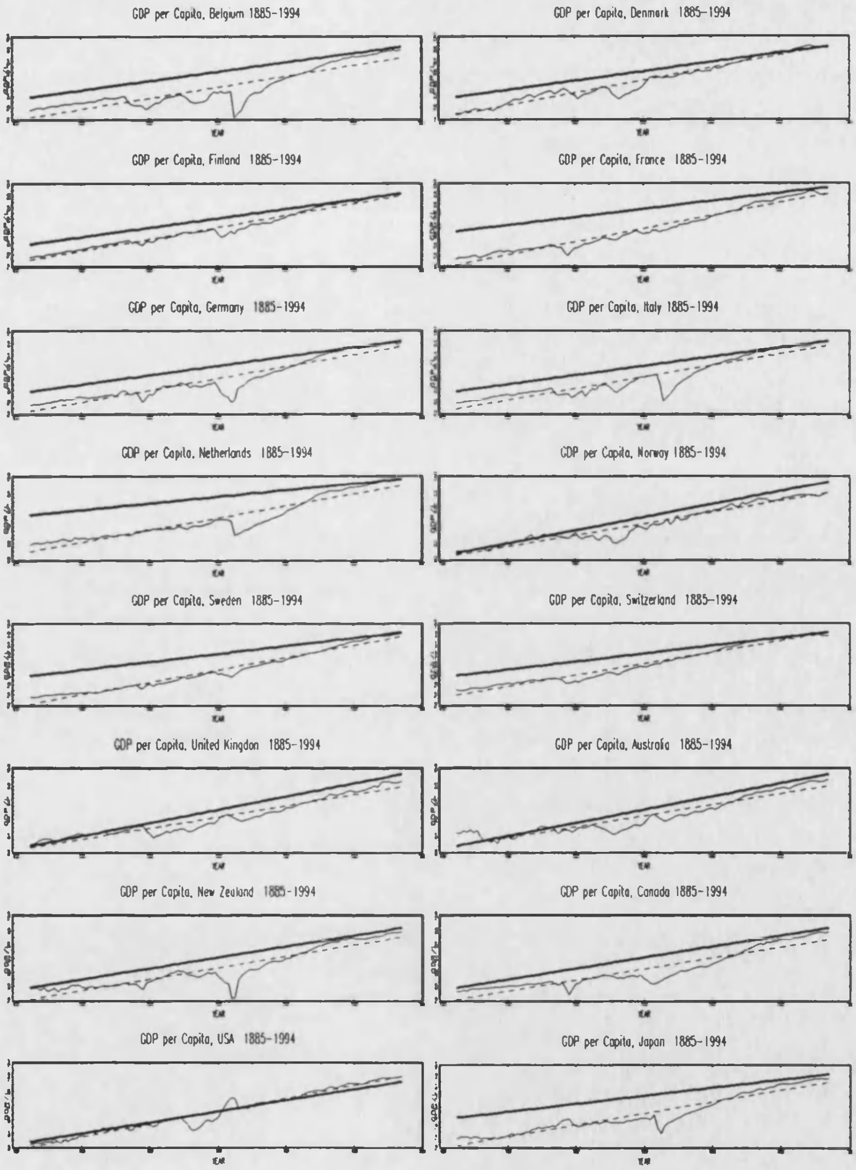

Figure 3.1: The dashed and bold lines represent the country-specific (OLS) and common (GLS) trend, respectively. The solid line represents logged GDP.

siderations imply that testing for β-convergence is meaningless if we assume the Jones invariance property together with the existence of a unit root in output[3]. As they stand, these three stylized facts can not go together. Our claim is that the (two equivalent) tests are both checking for a superimposed rate of "exponential" mean reversion.

If the rate of mean-reversion in (logged) GDP per capita or equivalently the rate of absorption of the shocks is hyperbolic (in a sense to be defined precisely below) instead of exponential, β-convergence would apply in the sense that poorer economy would grow faster and would converge towards their long run steady state and standard unit-root tests would fail to reject a unit-root albeit not present (see for example Diebold and Rudebusch 1991).

3.3 Theory of Long Memory and the Barro Regression

In this section we briefly review the theory of *long memory* processes which allows the possibility of "hyperbolic" mean reversion together with non-stationarity. We will then analyze why the Barro regression might be robust to rate of convergence different from the exponential one delivering the right answer to the problem of convergence.

[3]For example Den Haan (1995) notes that the slow speed of convergence observed in the data can be reconciled quantitatively with the neoclassical growth model assuming either a capital share equal to around 0.8 or a sufficient amount of persistence in the stochastic process driving technological progress. In either cases, the 2% rate of convergence is incompatible with aggregate output exhibiting a unit root (see his equation 3.4).

3.3.1 Theory of Long Memory Processes

Unit roots describe only a small set of nonstationary processes. A class that embeds either (covariance) stationary processes and unit roots is given by *strongly dependent* processes also known as *long memory* or *long range dependent* processes (see Robinson 1994 for a survey on the topic). Usually only the second moments properties are considered in order to characterize such a behaviour in terms of either the behaviour of the autocorrelation function at the long lags or the power spectrum at the zero frequency.

We shall assume that K denotes any positive constant (not necessarily the same) and $\sim$ asymptotic equivalence.

Definition 1

A real valued scalar discrete time process X_t is said to exhibit long memory in terms of the power spectrum with parameter $d > 0$ if

$$f(\lambda) \sim K\,\lambda^{-2d}, \qquad as \qquad \lambda \to 0^+.$$

In the nonstationary case ($d \geq 1/2$, see below) $f(\lambda)$ is not integrable and thus it is defined as a pseudo-spectrum.

The importance of this class of processes derives from smoothly bridging the gap between standard stationary processes and unit roots in an environment that maintains a greater degree of continuity (Robinson 1994). For the purpose, let us consider a parametric example.

Let $\{y_t\}$ be a discrete time scalar time series , $t = 1, 2, \ldots$, suppose v_t is an unobservable covariance stationary sequence with spectral density that is bounded and bounded away from zero at the origin, such that

$$(1-L)^d y_t = v_t, \qquad t = 1, 2, \ldots \tag{3.6}$$

where L is the lag operator. If $d = 0$, then y_t is a standard or better *weak memory* (covariance) stationary process with spectral density bounded away from zero (i.e. an ARMA process), whereas y_t is a random walk if $d = 1$. The parameter d however does not need to be an integer.

In what follows, we focus on the case in which y_t is a long memory process with parameter d positive, real with $0 < d < 1$. In this case, when v_t is assumed to be a white noise process, the process y_t defined in (3.6) is called an ARFIMA(0,d,0) process and more in general when v_t is an (inverted) ARMA(p,q) we obtain an ARFIMA(p,d,q) process.

The power spectrum of the y_t process is given by

$$f_y(\lambda) = \mid 1 - e^{i\lambda} \mid^{-2d} f_v(\lambda) = (2sin(\lambda/2))^{-2d} f_v(\lambda)\,, -\pi \leq \lambda \leq +\pi\,,$$

where $f_v(.)$ denotes the power spectrum of the v_t process. Thus from $sin(\omega)/\omega \sim 1\,, \omega \to 0$, when $d > 0$ as $\lambda \to 0^+$ we get

$$f_y(\lambda) \sim 4^{-d} f_v(0) \lambda^{-2d}\,.$$

Whenever $d > 0$ the power spectrum is unbounded at the zero frequency, implying that the series y_t exhibits long memory. This class of processes have many important properties. When $0 < d < 1/2$, y_t has both finite variance and exhibits mean reversion. When $1/2 < d < 1$ the process has infinite variance but it still exhibits mean reversion. This process is not (covariance) stationary but less "non stationary" than an unit root process so that standard unit root tests exhibits low power with respect to this alternative despite the presence of mean reversion (Diebold and Rudebusch 1991). When $d \geq 1$ the process has infinite variance and stops exhibiting mean reversion. In particular a unit root process is obtained when $d = 1$. This represent a particular case of a long memory process: a process with an infinite memory.

If $-1/2 < d$, (3.6) can be inverted so that

$$y_t = \sum_{i=0}^{\infty} \gamma_i v_{t-i}, \;\; \gamma_i = \prod_{k=1}^{i} \tfrac{k-1+d}{k}, \;\; i \geq 1, \;\; \gamma_0 = 1. \tag{3.7}$$

By use of Stirling's approximation it follows that as $i \to \infty$

$$\gamma_i \sim K i^{d-1}. \tag{3.8}$$

This can be interpreted such as the effect of a shock v_{t-i} , i periods ahead, vanishes at an hyperbolic rather than exponential rate exhibiting an high level of persistence, higher the bigger the parameter d. When $d = 1$, the unit root case arises where a shock arbitrarily far away in time exhibits permanent effects on the current level of y_t.

This persistence property reflects the characterization already given in the frequency domain. We have seen that a long memory process for $0 < d$ is defined by an unbounded spectrum at the origin. It is well accepted that the degree of persistence of a shock can be expressed by the "level" of the spectral density at zero frequency (Cochrane 1988) . The definition of long memory and the previous considerations suggest to take as an exact measure of persistence the 'slope' of the logged spectrum at the origin[4]. In fact, taking logs in both terms in *Definition 1*, we obtain as $\lambda \rightarrow 0^+$ the following representation

$$ln\left(f(\lambda)\right) \sim K - 2dln(\lambda)\,, \tag{3.9}$$

With respect to the scatterplot of the logged spectrum and $2ln(\lambda)$,the unit root case will be represented by a line with slope minus $\pi/4$ while the case $d < 1$ is represented by a flatter line. Obviously the bigger (in absolute value) the slope the greater the level of persistence. The idea expressed by (3.9) is at the core of the estimation procedure suggested by Geweke and Porter Hudak (1983) and formalized in Robinson (1995) that is briefly described in the Appendix 1 .

3.3.2 Robustness of the Barro regression

This section tries to rationalize the finding of a significant regression coefficient of β-convergence in (3.4) .

At first let us consider some back of the envelope calculations. A 2% rate of convergence superimposed as exponential over a time span of 20 - 110 years is almost observational equivalent to a parameter of fractional integration strictly between 0.5 and 1. In fact, bearing in mind the result in (3.8) a parameter of fractional integration, d, that resembles the 2% exponential rate of decay after a k period ahead shock can be obtained solving the simple equation[5]

[4]This concepts is directly derived from a well known strand in the semi-nonparametric econometrics literature Robinson 1995, Geweke and Porter Hudak 1983).

[5]Of course this is just a very simple and approximate exercise yet useful in order to understand the main intuition of the paper.

$d \sim 2\%$ exp. rate	N. of Obs.
0.912	10
0.865	20
0.821	30
0.781	40
0.742	50
0.703	60
0.667	70
0.631	80
0.596	90
0.561	100
0.527	110

Table 3.1: Parameter of Fractional Integration Corresponding to the 2 % exponential rate

$$(0.98)^k = k^{d-1}. \tag{3.10}$$

In Table 1 below we report the solutions of this simple equation, for values of k that ranges from 10 to 110. As most of empirical studies have used sample that ranges from 20 to 100 years, we can consider an underlying parameter of fractional integration strictly between 0.5 and 1 as the driving force behind the 2% rate of convergence found in the empirical literature on β-convergence.

Secondly, let us consider now the following theoretical result due to Sowell (1990), theorem 4. Regressing a variable on its lagged value , the Student t of the coefficient behaves discontinuously when the process generating the variable is an $ARFIMA(0, d, 0)$ with $d > 0$. When $d = 1$ we obtain that the asymptotic distribution of the Student t normalized at the value one, is the well known Dickey Fuller distribution (Dickey 1979). But when $d \neq 1$ one obtains very different results that is

$$t \to_p \begin{cases} \infty & , 1 < d < 3/2 \,, \\ -\infty & , 1/2 < d < 1 \,. \end{cases}$$

Let us now start to draw our conjecture.

If per capita GDP is well represented by a long memory process with parameter d with $1/2 < d < 1$, thus displaying infinite variance together with (what is important) mean-reversion, fitting the Barro regression would tend to give a significative negative Student t (actually converging to negative infinity in probability). Thus this simple inference gives exactly the same conclusion of the aforementioned regression (3.4) obtained in the literature when fitting an exponential rate of convergence.

More, the back of the envelope calculations show that superimposing an exponential rate of decay over a long memory process with $1/2 < d < 1$ gives precisely the well established 2% rate of β-convergence.

Finally, the property of long memory processes to nest the unit root case in a class that maintains a greater level of continuity rather than standard weak dependent processes motivates the empirical finding of systematically non significant unit root tests.

If our conjecture is right, we could say that the standard approach to test for β-convergence (Barro 1991), suits first step approximation despite the mispecification of the empirical model. This test tends to exhibit negative Student t in the case of mean reversion ($d < 1$), leaving nonetheless some margins of ambiguity in a particular case of lack of convergence, the unit root case ($d = 1$). On the other hand the Student t will diverge to plus infinity when $d > 1$ delivering the right answer to the issue of convergence.

At this stage, our conjecture still lacks of two elements, a purely economic one and a conclusive statistical one. We will show a possible source of the long memory feature of the data in a version of the Solow model augmented by cross sectional heterogeneity. Secondly, there is the need of a rigorous time series analysis of the data to show that the logged per capita GDP is well represented by a mean-reverting long memory process with $1/2 < d < 1$. This is done in section 6.

3.4 The Solow Model Augmented by Cross-sectional Heterogeneity

In this section we show how long memory could arise in the Solow growth model. Suppose that the economy is characterized by N units each behaving as in the standard Solow model outlined in section 2. That means that each of these units representing either different firms or sectors in the same economy are investing a fraction s_i of their output in the accumulation of capital [6]. If this is the case the dynamics of output, y_t^i, of each of these firm-

[6]Theoretically this structure could arise either in a world with imperfect capital markets where human capital is used as a collateral or because of adjustment costs (see Barro and Sala-I-Martin 1995). We decided not to model directly these frictions here because of the space constraint. Even the assumption that each units is evolving as an autoregressive process of order one is a simplifying assumption that it is is not needed to get the result as it will become clear thereafter.

sector, with steady state output $y_t^{*^i}$, is governed by

$$y_t^i - y_t^{*^i} = (1-\beta_i)[y_{t-1}^i - y_{t-1}^{*^i}] + \epsilon_t^i + \eta_t, \qquad i = 1...N,\ 0 < \beta_i < 1. \quad (3.11)$$

where ϵ_t^i, η_t represent respectively idiosyncratic and aggregate shock assumed mutually uncorrelated white noise and β_i is equal to $(1-\alpha_i)(g_i + \delta_i)$. Here α_i, g_i and δ_i are respectively the unit's specific productivity of capital, the rate of technological progress and the depreciation rate. It follows that the variable $x_t^i = y_t^i - y_t^{*,i}$ behaves like a first-order autoregressive process.

If we indicate respectively with

$$\begin{aligned}\bar{y}_t &= \frac{1}{N}\sum_{i=1}^{N} y_t^i, \\ \bar{y}_t^* &= \frac{1}{N}\sum_{i=1}^{N} y_t^{*^i},\end{aligned}$$

current and long run equilibrium aggregate output, we then have that the amount of disequilibrium in the economy evolves as

$$\bar{y}_t - \bar{y}_t^* = 1/N\sum_{i=1}^{N}(1-\beta_i)[y_{t-1}^i - y_{t-1}^{*,i} + \epsilon_t^i] + \eta_t. \qquad (3.12)$$

Let us define $\bar{x}_t = \bar{y}_t - \bar{y}_t^*$.

The above equation can behave very differently from equation (3.5) even if all the coefficients β_i are bounded between zero and one. We will show that under certain conditions on the cross sectional distribution of the coefficients β_i, $\bar{x}_t$ exhibits long memory. In fact if we assume that the aggregate η_t and idiosyncratic ϵ_t^i shocks are uncorrelated, we get that the power spectrum $f_k(\lambda)$ of x_t^k is equal to

$$f_k(\lambda) = \frac{var(\epsilon_t^k)}{2\pi \mid 1-(1-\beta_k)e^{i\lambda} \mid^2} + \frac{var(\eta_t)}{2\pi \mid 1-(1-\beta_k)e^{i\lambda} \mid^2}. \tag{3.13}$$

This implies that the power spectrum $\bar{f}(\lambda)$, $-\pi \leq \lambda < \pi$ of the aggregate $\bar{x}_t$ is equal to

$$\bar{f}(\lambda) = \bar{f}_1(\lambda) + \bar{f}_2(\lambda), \tag{3.14}$$

where

$$\bar{f}_1(\lambda) = \frac{1}{N^2}\sum_{k=1}^{N} \frac{var(\epsilon_t^k)}{2\pi \mid 1-(1-\beta_k)e^{i\lambda} \mid^2},$$
$$\bar{f}_2(\lambda) = \frac{var(\eta_t)}{2\pi N^2} \mid \sum_{k=1}^{N} \frac{1}{(1-(1-\beta_k)e^{i\lambda})} \mid^2 .$$

If we assume that the coefficients β_k are independent drawings from a distribution $F(\beta)$ and that the $var(\epsilon_t^u)$ are drawn from another distribution independent of the first, we follow Robinson (1978) and Granger (1980) to obtain

$$\bar{f}(\lambda) \simeq \frac{1}{2\pi N}(E[var(\epsilon_t^k)]) \int_{\mathcal{B}} \frac{1}{\mid 1-(1-\beta)e^{i\lambda} \mid^2} dF(\beta) + \tag{3.15}$$
$$\frac{var(\eta_t)}{2\pi} \mid \int_{\mathcal{B}} \frac{1}{(1-(1-\beta)e^{i\lambda})} dF(\beta) \mid^2, \tag{3.16}$$

where $\mathcal{B}$ denotes the support of the distribution $F(\beta)$ and $\simeq$ denotes that the relation holds approximately for N big but finite.

In general long memory arises if the integral in (3.15)

$$\int_{\mathcal{B}} \frac{1}{(1-(1-\beta))^2} dF(\beta) = E_F(1/\beta^2), \tag{3.17}$$

diverges, where $E_F(.)$ denotes the expectation over the measure $F(.)$ In fact the second integral in (3.16), viz. $E_F(1/\beta)$, diverges under stronger conditions which imply the divergence of the former integral in (3.15) but not viceversa as we will make clear in the sequel.

We can establish necessary and sufficient conditions on the distribution function $F(.)$ such that the integral (3.17) is unbounded. In general we know (e.g. in Rudin 1973) that the integral $\int_a^b h(t)dt$ for a continuous function $h(x)$ on an interval $[a, b)$ is unbounded, if $h(.)$ has at least the same order of infinity as $1/(b-x)^\alpha$ when x goes to b, that is

$$1/(b-x)^\alpha = O(h(x)), \qquad x \to b^-.$$

If we assume that the distribution function $F(\beta)$ is absolutely continuous having a density $f(\beta)$, the integrand function of (3.17) is given by $f(\beta)/\beta^2$. Thus a sufficient condition for $\bar{x}_t$ to exhibit long memory is simply given by

$$f(\beta) \geq K\beta, \quad \text{as } \beta \to 0^+ ,$$

for some positive constant K. Thus the density $f(\beta)$ might go to zero as $\beta \to 0$ but at slower rate than β [7].

The main implication is that the aggregate process might display long memory even if the aggregating elements are stationary with probability one. Also the result is valid even if the aggregating elements are ARMA processes. In this case the condition to be satisfied is that the probability of extracting

[7] Instead for the integral in (3.16) to diverge we need the stronger condition $f(\beta) \geq K$, as $\beta \to 0$ which clearly implies the former one. Moreover, the presence of the N and N^2 terms in (3.15) and (3.16) does not affect the result as we assume that the above arguments hold for a big but finite N.

a unit root in the autoregressive component dies slowly enough. Moreover it is important to stress that the result does not depend from either the nature of the idiosyncratic and common shocks given their stationarity or from the type of dependence among them. Mankiw, Romer and Weil (1992) argue that the slow speed of convergence observed empirically can be reconciled quantitatively with the the neoclassical growth model if the capital share is sufficiently high and around 0.8. This result, on the other hand, delivers a different rational for the low rate of convergence found in the empirics of the Solow growth model based on aggregation of cross-sectional heterogenous units [8].

Intuitively, long-range dependence means that shocks arbitrarily far away in time still exhibit some influence on the future dynamics of the process. Cross-sectional aggregation kills the Markovian property implicit in standard weak memory (covariance) stationary processes provided that there are some units with a sufficiently amount of persistence. In this case, to keep track of the future dynamics of the aggregate system we must recover the past history of the units of the system if we want to know the relative distribution of disequilibria in the economy.

[8] As an example we can consider Granger (1980) formulation where the coefficients β_k are drawn (independently both of the idiosyncratic shocks, ϵ_t^i, and common shocks, η_t) from a Beta(p,q) distribution. Thus we get that the integrand function (neglecting unimportant constant terms) is given by

$$\frac{(1-\beta)^{p-1}\beta^{q-1}}{\beta^2},$$

thus yielding the condition $q < 2$ which coincides with what Granger (1980) obtained by expanding the integral in terms of autocovariances. In fact in this case the aggregate process can be shown to display long memory with parameter $d = 1 - q/2$.

3.5 Generalizing the concept of Beta convergence

In this version of the Solow model augmented by cross-sectional heterogeneity, it seems reasonable to propose the following definitions of β-convergence:
(i) An economy has no tendency to converge towards either its own or the common steady state if, after fitting either a country specific or a common (linear) trend repectively, the parameter of fractional integration d of the residuals is greater or equal than one ($d \geq 1$). In the former case we say that there is no *conditional* convergence and that there is no *unconditional* convergence in the latter.
(ii) The case of the Solow model without cross sectional aggregation is represented by the absence of long memory that is d equal to zero. In this case, if we want to recover the rate of convergence of the economy, we must solve for the roots of the characteristic equation and look for the greatest solution in absolute value.
(iii) *Uniform unconditional* convergence means that if we fit a common (linear) trend across all the units in the sample, then the residuals exhibit similar parameter of fractional integration d.
(iv) *Uniform conditional* convergence means that if we fit a country specific (linear) trend for all the units in the sample, then the residuals exhibit similar parameter of fractional integration d.
We consider further evidence of the exponential 2% rate of convergence, if we find a parameter of fractional integration strictly between 0.5 and 1 (c.f. see section 3.).

In order to make inference on the parameters of long memory of the series we employ the semiparametric approach introduced by Geweke and Porter Hudak (1983) . Rigorous analysis of this estimator is given in Robinson (1995) who established consistency and asymptotic normality of the estimator. Also the result has been developed in a multivariate framework , a novel feature in this literature, which represents a crucial property in order to apply this estimator to a multicountries issue as the question of convergence. Robinson (1995) results are valid without assuming any a priori restriction on the degree of dependence in the data allowing for either antipersistence ($-1/2 < d < 0$), weak ($d = 0$) or long memory ($0 < d < 1/2$), the only restriction on the parameter space being finite variance, viz. $\mid d \mid < 1/2$. We defer to Robinson (1995) for the formal proofs of the results, describing the main features in Appendix 1.

3.6 Empirical Results

At first, to motivate our conjecture that the per capita GDP is characterized by a dynamics that is well approximated by a long memory process let us consider Figures 3.2 and 3.3. Interpreting the result according to *Definition 1*, Figures 3.2 and 3.3 show how the periodogram (i.e. an estimate of the spectrum) for each of the series in our sample displays a peak at the origin. This is what Granger (1966) defines to be the "typical spectral shape of an economic variable" and it is the main feature of a long memory process [9] .

[9]For an analysis of the behaviour of the periodogram for non stationary processes see Hurvich and Ray (1994).

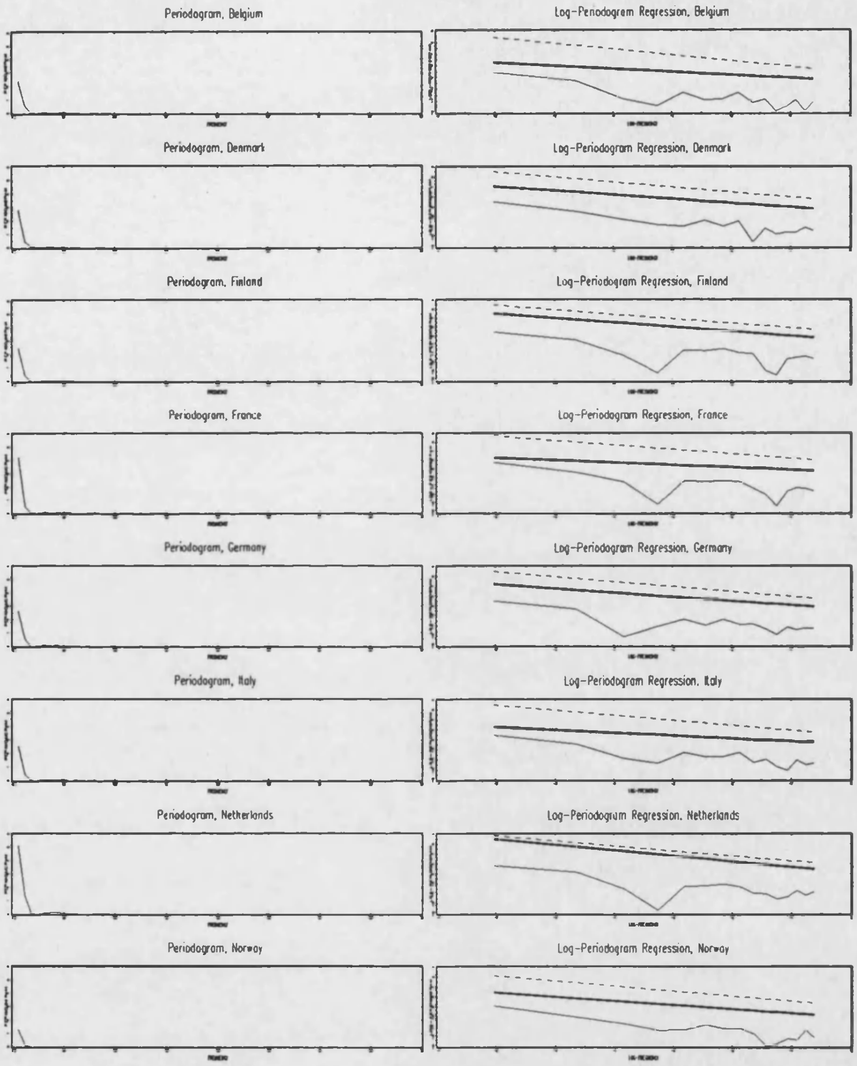

Figure 3.2: The left hand side column displays the periodogram of the logged GDP (1865-1994) for the 16 OECD countries here considered. The right hand side column displays three lines versus the logged frequency: the continuous line represents the logged periodogram ordinates, the bold line represents the OLS interpolating line (cf. Table 3.2) while the dashed line represents the unit root case (slope $\pi/4$). An interpolating line flatter than the bysector corresponds to a value of the long memory parameter d smaller than one.

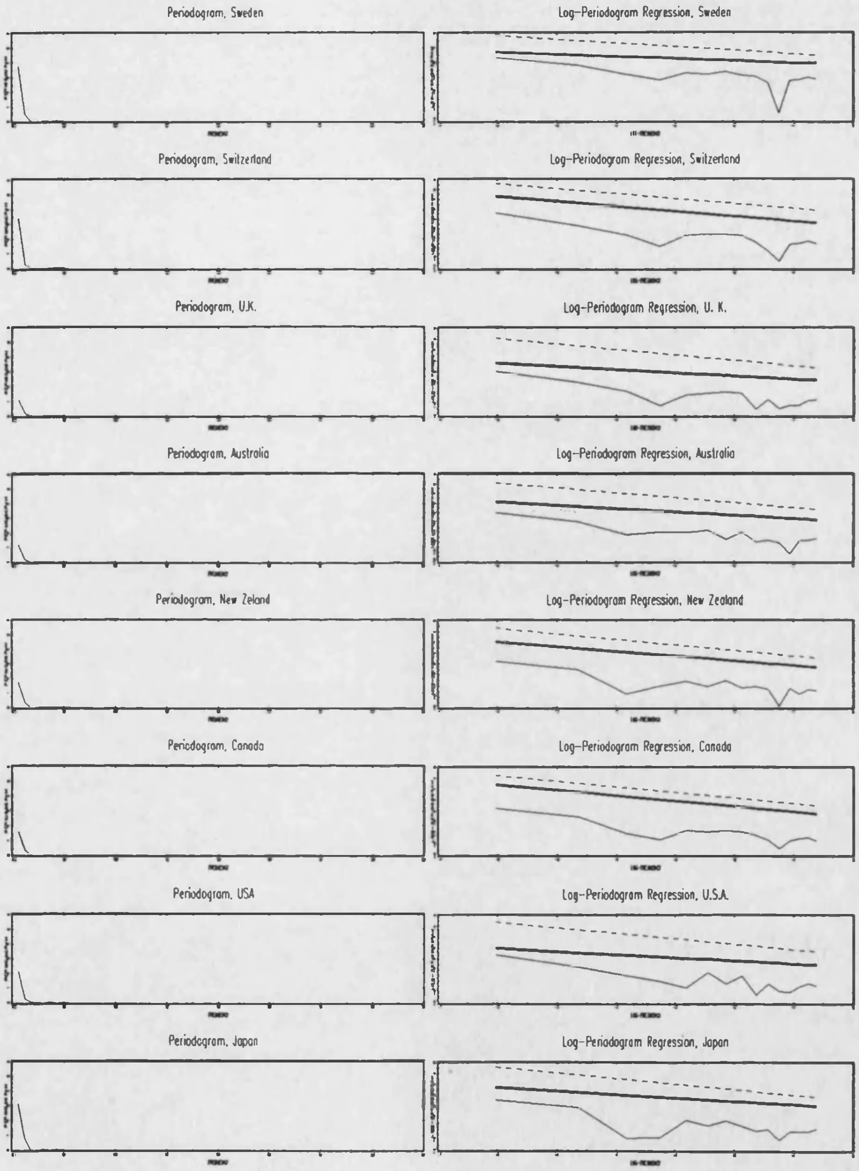

Figure 3.3: It continues previous figure.

Country	Conditional	Unconditional
Belgium	0.52	0.55
Denmark	0.84	0.55
Finland	0.99	0.98
France	0.56	0.94
Germany	0.83	0.83
Italy	0.56	0.65
Netherlands	1.11	1.26
Norway	0.81	0.82
Sweden	0.58	1.30
Switzerland	1.03	0.84
U.K.	0.58	0.58
Australia	0.69	0.75
New Zealand	0.85	0.85
Canada	0.97	0.96
U.S.A.	0.57	0.46
Japan	0.61	0.92
Asymptotic S.E.	0.177	0.177
Wald test statistic	1.24e+16 (0.0)	1.62e+16 (0.0)

Table 3.2: Log-Periodogram Estimates of d, (OECD, 1885-1994). The estimation procedure is described in the Appendix. The Wald test statistic is distributed as a χ^2 with 15 degrees of freedom under the hypothesis $H_0 : d_1 = d_2 = ... = d_{16}$. $P-$values are reported in parenthesis.

Figures 3.2 and 3.3 plot the logged periodogram against twice the logged frequency. As shown in section 3, the slope of the interpolating line expresses approximately the parameter d. The unit root case is represented by the line with slope minus $\pi/4$. It is evident how the interpolating line is always flatter than the bisector thus supporting the absence of the unit root case. Nevertheless the slope appears still positive and in particular between 1/2 and 1 .

Table 3.2 reports the estimates based on the log-periodogram estimator [10]. Most of the parameters of fractional integration,$d's$, are less than one even if with a very high standard error. As we are interested in the OECD countries as a group, we use an induced test based on the sequential Bonferroni approach[11]. We want to test for the existence of a number d_0 strictly less than one, such that all the parameter of fractional integration of the OECD countries in the sample are less or equal than d_0. For an overall level of significance of 10 percent, we examine the country with the highest ex post probability of rejecting the null hypothesis and set the significance level using the total number of countries examined.

The results of this procedure are reported in Figure 3.4. The horizontal line represents the 10 percent critical value of the t-statistics such that the null hypothesis is rejected. The x-axis represents the coefficient d_0 considered under the null. The negatively sloped line shows the actual t-statistics calculated for different null hypotheses. Figure 3.4 shows that it exists a non empty set of values of d_0, strictly less than one, such that the null hypothesis that the parameter of fractional integration of all the OECD countries are less than d_0 can not be rejected at the 10 percent significant level. We also note how this set always lays above the value 1/2.

[10]Diebold and Rudebusch (1989) has used a similar estimator, valid in a univariate case only (Geweke and Porter Hudak 1983. The multivariate framework, the gains in efficiency and computability of the Robinson (1995) estimator motivates our choice of using the latter instead of the former thus explaining the difference in the estimates of the parameter d for the US case obtained by Diebold and Rudebusch (1989). The appendix reviews the main features of the estimating procedure.

[11]This procedure yields a conservative yet consistent test (Gourieroux and Monfort, 1989, Property 19.7). The exact test for one-side multivariate hypothesis (Gourieroux and Monfort, 1989, Chapter 21) is not implementable when the number of constraints is greater than two.

The empirical results can be summarized as follows:

- GDP per capita of all the countries in the sample exhibit long memory ($d > 0$). In our framework this suggests that the economy behaves as an aggregation of Solow models rather than as a Solow model itself.
- The hypothesis that all the OECD countries are non stationary still mean reverting ($0.5 < d < 1$) can not be rejected using the induced test based on the Bonferroni procedure (Figure 3.4).
- We found the 2% rate of convergence in the form of a parameter of fractional integration strictly between 0.5 and 1.
- The rates of convergence are very low and similar across countries even if the rate of convergence is not *uniform* as the null hypothesis that the coefficients of fractional integration are constant across countries is strongly rejected (see Table 3.2) .
- As the order of integration of different OECD countries are different, time series tests of convergence based on cointegration are mispecified.
- We conclude that there is unconditional convergence across OECD countries and the rates of convergence are pretty similar even if the test reject the null of exact equality of the coefficients.

3.7 Conclusions

In this chapter we embed standard approaches to test for β-convergence in a more general framework. In order to do so we join different strands of literature, the aggregation theory of dynamic economic models, the theory of long memory processes and the literature on the empirics of growth.

We give striking evidence that the (de-trended) per capita GDP is well approximated at the low frequencies by a long memory process displaying nonstationarity together with mean reversion, stressing the importance of capturing in the very long the true rate of convergence. We then find primitive conditions under which long memory arises naturally as the result of aggregating heterogeneous units in the same economy and we then apply it to an extension of the Solow-Swan model augmented by cross-sectional heterogeneity.

Finally we draw robust inference on the possibility of conditional and unconditional β-convergence among the OECD countries and as a result we support the conclusion of the well established Barro type of regression and we reconcile both time series and cross sectional evidence.

Some questions still remain open. In particular, we stress how the degree of persistence differs among OECD countries. This drives the question of whether the underlying economic structure of OECD countries are different and asks for a further investigation of what country specific economic mechanism make long memory to arise in real world.

3.8 Appendix 1. The logperiodogram estimator

Following Robinson (1995), let us suppose that the time series under study is given by the G dimensional real valued vector $Z_t = (Z_{1,t}, \ldots, Z_{G,t})'$. The (g,h)th element of the spectral density matrix $f(\lambda)$ is denoted by $f_{gh}(\lambda)$. For (C_g, d_g) , $g = 1, ..., G$ satisfying $0 < C_g < \infty$ and $| d_g |< 1/2$ it is assumed that[12]

$$f_{gg}(\lambda) \sim C_g \lambda^{-2d_g} \quad \text{as } \lambda \to 0^+ .$$

This represent the only assumption on the shape of the spectrum which motivates the semiparametric nature of the estimator of the $G+G$ parameters $(C_g, d_g)\; g = 1, \ldots, G$ beside integrability to ensure stationarity.

The periodogram[13] for the g-th component $Z_{gt}, t = 1, \ldots, N$, N being the sample size, is denoted by

$$I_g(\lambda) = \frac{1}{2\pi N} \mid \sum_{t=1}^{N} Z_{gt} e^{it\lambda} \mid^2 , \; g = 1, \ldots, G . \tag{3.18}$$

Defining the fourier frequency $\lambda_j = \frac{2\pi j}{N}$ one has to define the log-periodogram

$$Y_{gk} = ln(I_g(\lambda_k)) \, , \; g = 1, \ldots, G \, , k = l+1, \ldots, m . \tag{3.19}$$

The positive integer m is the user-chosen bandwidth number and the positive integer $l$is the user-chosen trimming number[14]. In this context there is just the need to say that the asymptotic results require that m and l both tend to infinity with N but more slowly together with $l/m \to 0$. Then defining the unobservable random variables U_{gk} by the following set of regressions

$$Y_{gk} = c_g - d_g(2ln\lambda_k) + U_{gk}\, g = 1, \ldots, G \;\; , k = l+1, \ldots, m. \tag{3.20}$$

[12]Basically as in *Definition 1* for each component Z_{gt} .

[13]Practically one will consider the periodogram at the fourier frequencies only thus making irrelevant to demean the series by the sample mean.

[14]We refer to Robinson (1995) for a thorough discussion on the concepts and the roles played in the asymptotic theory by these two user-chosen numbers.

where $c_g = ln\, C_g + \psi(1)$ which involves the digamma function $\psi(z) = (d/dz) ln\Gamma(z)$, with $\Gamma(.)$ being the gamma function.

Then the OLS estimates of $c = (c_1, \ldots, c_G)'$ and $d = (d_1, \ldots, d_G)'$ are given by $\tilde{c}, \tilde{d}$

$$\begin{bmatrix} \tilde{c} \\ \tilde{d} \end{bmatrix} = vec(Y'X(X'X)^{-1}),$$
$$X \stackrel{def}{=} (X_{l+1}, \ldots, X_m)', \, Y \stackrel{def}{=} (Y_1, \ldots, Y_G)',$$
$$X_k \stackrel{def}{=} (1, -2ln\lambda_k)', \, Y_k = (Y_{g,l+1}, \ldots, Y_{g,m})'.$$

Denoting as usual the OLS residuals as

$$\tilde{U}_k = Y_k - \tilde{c} + \tilde{d}(2ln\lambda_k), \, k = l+1, \ldots, m, \tag{3.21}$$

and the matrix of sample variances and covariances

$$\tilde{\Omega} = \frac{1}{m-l} \sum_{i=l}^{m} \tilde{U}_k \tilde{U}_k', \tag{3.22}$$

one gets that the OLS standard errors for $\tilde{d}_g$, $g = 1, \ldots, G$ are given by the square root of the $(G+g)$th diagonal element of the matrix $(Z'Z)^{-1} \otimes \tilde{\Omega}$.

This estimating procedure allows for cross equations restrictions such as that all (or some of) the G series are characterized by a common parameter of long memory that is

$$d_g = \delta, \qquad g = 1, \ldots, G,$$

or in matrix formulation

$$d = Q\delta,$$

where $Q = (1, 1, \ldots, 1)'$ is a $G \times 1$ vector and δ is a scalar representing the unknown common long memory parameter. Thus the GLS estimator $\hat{c}$ and $\hat{d}$ is given by

$$\begin{bmatrix} \hat{c} \\ \hat{\delta} \end{bmatrix} = \left\{ \begin{bmatrix} I_G & 0 \\ 0 & Q' \end{bmatrix} (X'X \otimes \tilde{\Omega}^{-1}) \begin{bmatrix} I_G & 0 \\ 0 & Q \end{bmatrix} \right\}^{-1} \begin{bmatrix} I_G & 0 \\ 0 & Q' \end{bmatrix} vec(\tilde{\Omega}^{-1}Y'X).$$

When there are no restriction we set $Q = I_G$ and we obtain again the OLS estimator[15].

Under certain regularity conditions (Robinson 1995) among which Gaussianity of the process Z_t the following asymptotic results are obtained, which allows to perform standard inference on the OLS and GLS estimators. For the OLS Robinson established

$$\begin{bmatrix} \frac{m^{1/2}}{ln\, n}(\tilde{c} - c) \\ 2m^{1/2}(\tilde{d} - d) \end{bmatrix} \rightarrow_d N\left(0, \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes \Omega\right), \tag{3.23}$$

and for the GLS

$$\begin{bmatrix} \frac{m^{1/2}}{ln\, n}(\hat{c} - c) \\ 2m^{1/2}(\hat{d} - d) \end{bmatrix} \rightarrow_d N\left(0, \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes Q(Q'\Omega^{-1}Q)^{-1}Q'\right). \tag{3.24}$$

One obtains a consistent estimate of Ω by using (3.22) . Considering each $\tilde{d}_g$ individually the general result in (3.23) becomes

$$2m^{1/2}(\tilde{d}_g - d_g) \rightarrow_d N(0, \frac{\pi^2}{6}).$$

The results allow us to make use of all the regression theory. In particular one can build a Wald test for linear restriction expressed by

$$H_0: \qquad Pd = 0,$$

[15]Also to obtain a consistent estimate of C_g $g = 1, \ldots, G$ one has to consider the relation $C_g = exp(c_g - \psi(1))$.

where P is a $H \times G$ matrix of rank $H < G$. The test statistic is given by

$$\tilde{d}'P'[(0,P)\left\{(X'X)^{-1} \otimes \tilde{\Omega}\right\}\left(\begin{array}{c} 0 \\ P' \end{array}\right)]^{-1}P\tilde{d}, \qquad (3.25)$$

that under H_0 is asymptotically distribuited like a central χ^2 with H degrees of freedom.

Estimating Procedure.

Firstly we detrend the data fitting either a country specific or a common trend. The former has been estimated with OLS, the latter with GLS. We then evaluate the order of integration of the residuals[16]. A preliminary analysis of the parameters d_g , $g = 1, \ldots, G$ gives estimated values greater than $1/2$ thus out of the admissible region for the asymptotic results to be valid. If we first differenced the data the estimates would be totally independent on the type of β-convergence we are considering (conditional and unconditional). In fact the periodogram evaluated at the Fourier frequencies is independent of any shift of location. For this reason we prefer to difference fractionally the data before estimating, by multipying them by $(1 - L)^q$, $q = 0.5$. Obviously in doing so we have to approximate a series with finite sum. Our choice of $q = 0.5$ reflects the trade- off between differentiating "enough" (big q) to

[16]It is a reasonable question to ask if the properties of the theoretical disturbances carries over to the ones of the residuals after detrending the data with either the country specific or the common trend (see i.e. Nelson and Kang 1981). There are good reasons to believe that it does once a semi-parametric frequency domain approach is undertaken. Nelson and Kang (1981) shows that the regression of a driftless random walk against a time trend delivers residuals exhibiting a periodogram with a single peak at a period equal to 0.83 of sample size thus asymptotically at frequency zero, as one would expect. In words, the memory of the process is entirely reflected in the residuals.

obtain estimates in the stationary region and minimizing the approximation from using a sum instead of a series (small q) [17]. To initialize the fractional filter $(1 - L)^q$ we use the first 10 observations in the sample.

Choice of Trimming and Bandwidth.

In our application we choose a trimming coefficient equal to two, $l = 2$, so that we avoid the first periodogram ordinate (see also footnote 20). Unfortunately a complete theory for the optimal choice of the trimming and the bandwidth is still missing for this estimator but it seems that choosing a trimming bigger than one increases the performance of this estimator in finite samples (Hurvich and Ray 1994). Because of this reason we report estimates based on the the same criterium as the one used by Diebold and Rudebusch (1989) for their univariate analysis, that is $m = T^{0.525}$ after checking for robustness under alternative bandwidths [18]. It is nevertheless important to point out that the empirical results are very robust to changes in the the choice of the trimming and the bandwidth.

[17] Even if not formally proved, we follow the empirical literature of long memory processes conjecturing that asymptotically this approximation becomes negligible. Also the results are globally robust with respect to the choice of q.

[18] The results are available under requests. We defer to Beran (1994) for a review on parametric and semi non-parametric estimation in a long memory framework.

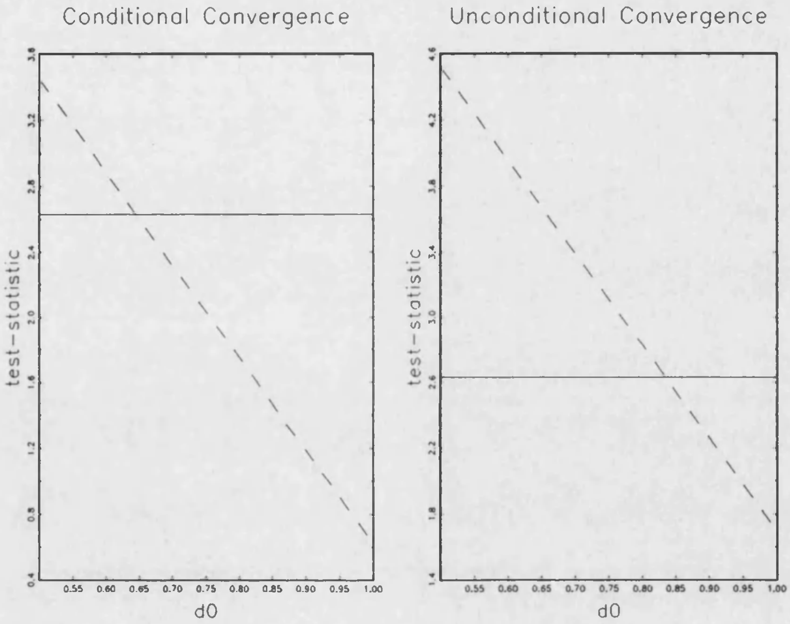

Figure 3.4: The test statistic for the null $H_0 : d_i < d_0$, $i = 1, ..., 16$ is plotted for different values of d_0. The horizontal line represents the critical value for a 10% significance level.

Chapter 4

The Macroeconometrics of Cross-Sectional Heterogeneity

Abstract

Detrended aggregate time series display a very high degree of persistence. In the frequency domain this shows up as a typical spectral shape. Empirical work has also shown that at the firm level there is a large amount of heterogeneity. This chapter claims that the two facts are closely related: cross-sectional heterogeneity is a transmission mechanism of economic shocks and has some distinctive properties. I analyse a vintage model where all the uncertainty is driven by either productivity or demand factors and it takes the form of an aggregate shock that causes reallocations of firms across technological states. As aggregate shocks create persistence without affecting either the number of firms in the market or the rate of technological progress, this degree of persistence can be attributed to cross-sectional heterogeneity. The chapter shows that this transmission mechanism is powerful, realistic, robust and general. It is *powerful* because a sufficient amount of cross-sectional heterogeneity is able to generate typical spectral shapes without

necessarily relying on technological shocks. It is *realistic* because it is shown how two of the most striking empirical regularities in economics, the typical spectral shape and Gibrat's law, might be just two faces of the same coin. It is *robust* because it is able to generate almost any kind of spectral shapes provided that aggregate shocks cleanse the economy reallocating between very low and very high technological states. It is *general* because the chapter provides a micro-foundation for the fractional cointegration of aggregate economic variables illustrating why a typical spectral shape might be such a common phenomenon.

4.1 Introduction

A well documented fact in detrended aggregate time series is persistence. This shows up, for example, in the typical spectral shape illustrated by Granger (1966). Empirical work has also shown that at the firm level there is a large amount of heterogeneity, even within narrowly defined sectors (see for example Dunne et al. 1989, Davis and Haltinwanger 1990, 1992). This chapter claims that the two facts are closely related. The heterogeneity that we observe at the micro level has macroeconomic implications and delivers a convincing explanation of aggregate persistence. Cross-sectional heterogeneity is a transmission mechanism of economic shocks and has some distinctive properties.

We consider a version of the Solow (1960) vintage model. To capture the productivity benefits of technical change, older capital vintages must be replaced with the most recent equipment. At each point in time, a firm weighs the benefits of switching to a better technology, with the the opportunity cost (in terms of forgone profits) of investing part of their capital or labour resources in technological improvements. These costs may vary across firms and thus firms adopting the same vintage can end up using different technologies. This is now a popular and plausible way of modelling the heterogeneity of an economic system (see e.g. Baily et al. 1992, Caballero and Hammour 1994, 1996, Aghion and Howitt 1994, Mortensen and Pissarides 1995, chapter one). In our model, aggregate shocks alter the opportunity cost of all firms in a similar way and cause a reallocation of firms across technological vintages. The shocks do not affect either the number of firms in the market or the rate of technological progress. Therefore any persistence can be attributed to cross-sectional heterogeneity.

Our purpose in introducing the model is to try to match a well known empirical regularity. Granger (1966) observed that most detrended macroeconomic variables exhibit a typical spectral shape. Estimated spectra look like a monotone decreasing function from low to high frequencies with a pronounced peak in the neighbourhood of the zero frequency. Frequency domain analysis decomposes the dynamics of a time series into different periodic components whose weights are given by the spectrum at the corresponding frequency. Thus, the typical spectral shape identified by Granger implies that the weight of the components with very long periods is disproportionately large, that is detrended aggregate time series display a very high degree of persistence.

The typical spectral shape might be the outcome of a process different from standard ARIMA processes. Formal empirical investigation (see Diebold and Rudebusch 1989, chapter three and Gil-Alana and Robinson

1997) has concluded that it is the result of *long memory* processes (see e.g. Robinson 1994) in which the impact of shocks vanishes at a very slow hyperbolic rate.

The search for economic mechanisms in which shocks vanish at a very slow hyperbolic rate turns out to be a formidable task. In general, the economic theory generates dynamics in which shocks either have permanent effects or vanish at the usual exponential rate[1]. The first case corresponds to a unit root in the underlying time series, the second one to an ARMA process whose spectrum is completely flat around zero frequency and thus does not exhibit a typical shape. There is a need to find economic transmission mechanisms that generate long memory processes, yielding spectral shapes similar to the ones observed in the real world. In particular this chapter finds *robust* transmission mechanisms: small alterations to the basic set-up do not shift the spectral shape of the time series from the very particular slope associated with the unit root case to the flat one corresponding to the ARMA case.

Cross-sectional heterogeneity provides such a mechanism. It is also powerful, realistic and general. It is *powerful* because a sufficient amount of cross-sectional heterogeneity is able to generate typical spectral shapes without relying on technological shocks. The Real Business Cycle tradition has argued that technology is the driving force behind the existence of the typical spectral shape, especially the particular case associated with a unit root,

[1]The Real Business Cycle tradition has often argued that technology is the driving force behind the existence of a typical spectral shape (Nelson and Plosser 1982, Rotemberg and Woodford 1996, Gali 1996). Alternatively models with strategic interaction and spillovers have shown their potential to generate multiple equilibria (see e.g. Cooper and John 1988). If so, aggregate shocks that shift the economy from one equilibrium to the other can explain persistence in aggregate fluctuations (see Durlauf 1991).

(see e.g. Rotemberg and Woodford 1996, Gali 1996). As a consequence, technology has been inferred to be a major contributor to variation in observed output[2]. In fact, the aggregate shocks analysed in this chapter, just alter the opportunity cost of firms and so they can be read as either productivity or demand shocks.

In the chapter we identify small (big) firms as the ones adopting vintages far-away from (close to) the technological frontier. In general to link productivity to size we need some assumptions on the dynamics of factor prices. If factor prices grow at the same rate as the technology frontier, a productivity ranking in terms of vintages corresponds exactly to a one in terms of size[3]. If so, cross-sectional heterogeneity is a *realistic* transmission mechanism because the assumptions required to generate typical spectral shapes match the findings of the growth-firm size empirical literature. For example, Gibrat's law claims that the expected growth rate of firms is independent of their sizes, measured by sales, employment, or assets (see e.g. Gibrat 1931 and Sutton 1996). According to our identification scheme, it means that the expected growth rate of firms is independent of the vintage currently in use. Therefore, an aggregate shock that reallocates units across different vintages, generates a permanent effect on the aggregate level of output, that is a unit root[4]. Recent studies show however, that small firms tend to have

[2]See Nelson and Plosser (1982). Lippi and Reichlin (1992) and Quah (1992), however, argue against this last inference.

[3]For example, Baily et al. (1992) and Bartelsman and Dhrymes (1994) find that employment size and productivity are positively correlated.

[4]This idea was implicitly contained in Kalecki (1945) as he claimed that "the [standard] argument [on which Gibrat's law is based] implies that as time goes by the standard deviation of the logarithm of the variate considered increases continuosly". It is well known that a distinctive feature of a random walk (a particular typical spectral shape) is

higher and more variable growth rates (see Mansfield 1962, Hall 1987, Evans 1987 and Dunne et al. 1989). We embed this evidence in the model and we assume that firms adopting vintages far-away from the technological frontier grows at similar rate but faster than firms adopting vintages close to it. If so, aggregate output exhibits a degree of persistence similar to that observed in the time series of the US aggregate GDP, that is a degree of long memory bigger than the one associated with the ARMA case but smaller than the unit root case.

In this respect cross-sectional heterogeneity generates *robust* spectral shape: small alterations to the basic set-up does not shift the spectral shape of the time series from the very particular slope associated with a unit root to the flat one corresponding to the ARMA case. It is also robust in another respect. The model can even generate long memory in the growth rate, even if the shocks affect neither the number of firms in the market nor the rate of technological progress. To generate this we posit two further assumptions. Firstly, we assume that the persistence at the top of the technological distribution is so big [5] that the steady state distribution is characterized by each firm staying with probability one at the top of the technological frontier. Secondly, we assume that aggregate shocks cleanse the economy reallocating units from very low to very high technological states (cleansing type reallocations)[6].

Finally, the chapter highlights why a typical spectral shape might be

that its variance is a linear function of time.

[5]See Baily et al. (1992) and Bartelsman and Dhrymes (1994) for evidence in this direction.

[6]See e.g. Davis and Haltinwanger (1990, 1992) and Gali and Hammour (1991) for empirical evidence in this direction.

such a *general* phenomenon. Firms are the minimal working cells of an aggregate economy where most economic decisions are taken. Then, it is not difficult to see why aggregates should be cointegrated and the chapter give a micro-foundation for the fractional cointegration of aggregate economic variables. It also shows why the residuals of the cointegrating relation can exhibit themselves arbitrary high level of persistence[7]. The decision of each firm is affected by some specific features of the cross-sectional distribution of vintages currently in use. Therefore the relation between aggregate macroeconomic variables turns out to be quite complex and linear regressions can clear only a portion of the aggregate amount of persistence in the time series.

The main contribution of the chapter can be conveniently summarized as follows. There are two independent strands of the literature. One has dealt explicitly with cross-sectional heterogeneity in order to micro-found macroeconomics solving explicit aggregation problems. The other has analysed firms dynamics, in particular the relation between growth and firms'size. This chapter notes that the two independent strands of research have important implications on the low frequency behaviour of aggregate time series once a standard vintage model is used to interpret them. Surprisingly, the implications of the model matches quite closely with the empirical evidence. Moreover, models which do not deal explicitly with cross-sectional heterogeneity seem unlikely to provide a robust explanation of aggregate persistence. Thus, the chapter concludes that cross-sectional heterogeneity is an important transmission mechanism with some distinctive properties.

The remainder of the chapter is divided into 7 sections. Section 2 reviews

[7]For evidence in this direction see Robinson and Marinucci (1997).

the meaning and the empirical evidence in favour of the typical spectral shape. Section 3 lays down the structure of a stylized vintage model where both the rate of technological progress and the size of the market are exogenous. It then analyses the effects of introducing aggregate uncertainty in the basic set-up. In doing so it generalizes results stemming from the pioneering works by Bertola and Caballero (1990, 1994), Caballero and Hengel (1990) and Caballero (1992) aimed at micro-founding macroeconomics solving explicit aggregation problems. Section 4 analyses the dynamics of the aggregate time series under the assumption that each unit will visit infinitely often all states in the system (irreducible case). Section 5 follows the empirical evidence contained in Baily et al. (1992) and Bartelsman and Dhrymes (1994) to assume that the persistence at the top of the technological distribution is so big that each firm will end up by staying with probability one at the top of the technological frontier (reducible case). In this context we prove that cleansing type reallocations can generate almost any kind of spectral shapes. Section 6 establishes the micro-foundations for the fractional cointegration of aggregate economic variables. Section 7 concludes while section 8 contains the derivation of most of the results contained in the chapter.

4.2 General typical spectral shapes: empirical evidence and meaning

This section reviews first the foundations of the frequency domain approach to time series, and then the empirical evidence in favour of the typical spectral shape of an economic variable observed by Granger (1966).

4.2.1 The Frequency Domain Approach to Time Series Analysis

Consider the spectral representation of a time series X_t

$$X_t = \int_{-\pi}^{\pi} e^{it\theta}\, dZ(\theta,\omega) = \int_{-\pi}^{\pi} [cos(t\theta)\ + i\ sin(t\theta)]\ dZ(\theta,\omega), \tag{4.1}$$

where $Z(\theta,\omega)$ is a zero-mean orthogonal increment process with the property that $E\left|dZ(\theta,\omega)\right|^2 = dF(\theta)$, where $dF(\theta)$ is called the *spectral density* of the process. Time series analysis based on representation (4.1) is called the frequency domain approach to time series, as it decomposes the variation of the time series X_t into a combination of sines and cosines of different periods. In general, the higher the value of $dF(\theta)$ the higher is the weight of the periodic component of period $\frac{2\pi}{\theta}$. The representation (4.1) is a very general one. For instance the Cramer's theorem (see e.g. Brockwell and Davis 1991) guarantees that a spectral representation like the one in (4.1) holds for any stationary process while Priestley (1965), Hurwich and Ray (1994) and Chan and Terry (1995) show how this representation can be extended for non stationary linear processes. Given a set of observations x_t with t that goes from 1 to T, the spectral density $dF(\theta)$ is usually estimated through the periodogram $I(\theta)$ (or function of it) defined as the modulus of the discrete Fourier transform of the observations[8]:

$$I(\theta) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} x_t\, e^{it\theta}\right|^2. \tag{4.2}$$

[8] In the case of a stationary process the periodogram is the sample analogue of the theoretical spectral density $dF(\theta)$.

Granger (1966) noted how the estimated spectra of most detrended economic variables had a shape that he defined to be as *typical.* It is a monotonously decreasing function with a peak in the neighbourhood of the zero frequency. Figure 1A shows the typical spectral shape for the level of detrended logged GDP per capita for the United States calculated over the period 1870-1994. A class of spectral density function able to match exactly with the spectral shape[9] that arises in the real world is given by

$$f(\theta) = dF(\theta) \sim g(\theta)\,\theta^{-2d}, \quad as\ \theta \rightarrow 0^{+}, \tag{4.3}$$

where " $\sim$ " indicates that the ratio of left- and right-hand sides tends to bounded quantity, d is a non negative constant and $g(\theta)$ is a bounded function bounded away from zero in a neighbourhood of the origin[10]. The parameter d represents the order of integration of the time series. If it is greater than zero is said to exhibit *long memory* while if the parameter is equal to 0 the process is said to exhibit *weak memory.* The parameter d

[9]The exact relation between the spectral density, $dF(\theta)$, and the periodogram (4.2) for non stationary processes ($d \geq \frac{1}{2}$ see below) is a topic that goes beyond the scope of this chapter. It seems however that the periodogram behaves without any solution of continuity in moving from the stationary region ($d < \frac{1}{2}$) to the non stationary one ($d \geq \frac{1}{2}$) (see Hurwich and Ray 1994, Velasco 1996, Robinson and Marinucci 1997). Because of this reason a possible solution consists of defining the theoretical spectral density through the periodogram like in Hurwich and Ray (1994). This is implicitly the approach pursued here, where to safeguard theoretical rigour we speak about empirical spectral shapes rather than spectra.

[10]A slightly more general definition would allow for $g(\theta) = L(\frac{1}{\theta})$ to be a slowly varying function at infinity (see e.g. Seneta 1976), that is, a positive measurable function satisfying

$$\frac{L(\kappa\theta)}{L(\theta)} \rightarrow 1, \quad as\ \theta \rightarrow \infty,\ for\ all\ \kappa > 0.$$

measures the rate of divergence of the spectrum around zero frequency and thus it measures "how typical the spectral shape is".

A time domain representation of the time series X_t, $t \geq 0$, corresponding to equation (4.3) is given by the Wold representation[11]

$$X_t = X_0 + \gamma t + \sum_{n=0}^{t} \phi_n \, \epsilon_{t-n}, \tag{4.4}$$

with Wold coefficients $\phi_n \sim \tilde{\phi}_n + d\, n^{d-1}$ as $n \uparrow \infty$, where $\tilde{\phi}_n$ is a function converging to zero at a rate at least as quick as the exponential one ($\left|\tilde{\phi}_n\right| \leq \rho^n$, $0 \leq \rho < 1$ as $n \uparrow \infty$) while d represents the order of integration of the time series[12].

A standard trend stationary process with *ARMA* disturbance exhibits Wold coefficient ϕ_n's decaying at most at an exponential rate, that is a parameter of fractional integration d equal to zero. This weak memory property of *ARMA* processes shows up in frequency domain under the form of a flat spectral shape around zero frequency (see for example Figure 1B for the $AR(1)$ case). A process with a unit root exhibits Wold coefficients asymptotically approaching a constant that is $d = 1$. Thus its spectral shape is typical yet particular as it exhibits a very specific rate of divergence around zero frequency. This set of considerations show how standard *ARIMA* processes can not generate *arbitrary* typical spectral shape because they generate shapes

[11] In particular, this will be the operational definition of fractional integration that we will use throughout the chapter.

[12] If we allowed for $g(\theta)$ to be a slowly varying function as in footnote ten, the Wold coefficients could behave as $\phi_n \sim \tilde{\phi}_n + n^{d-1}$. If so, the case $d = 0$ would correspond to Wold coefficients ϕ_n decaying as $\frac{1}{n}$. This would imply that the covariances μ_τ would behave as $\frac{\ln(\tau)}{\tau}$ for large τ, so that the spectral density $dF(\theta)$ behaves as the slowly varying function $(\ln \theta)^2$ as $\theta \to 0^+$, see Granger and Joyeux (1980).

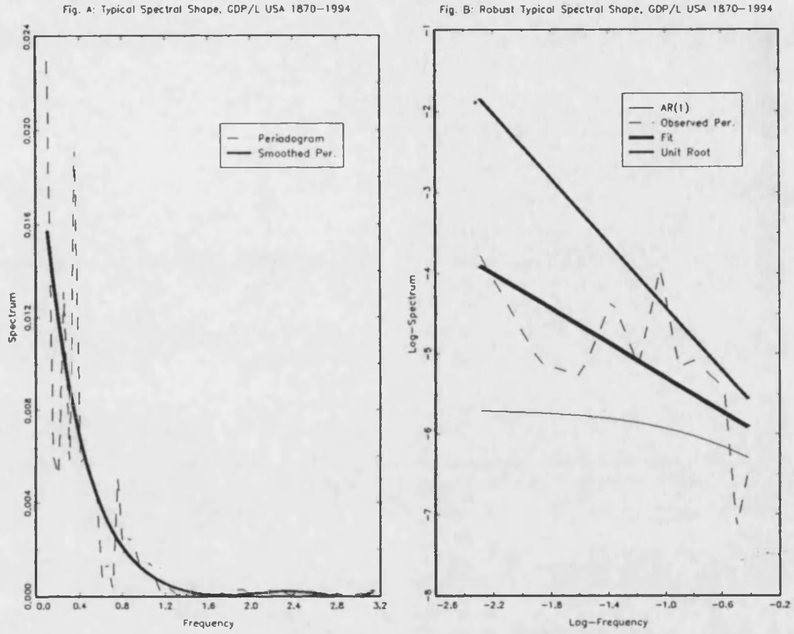

Figure 4.1: Robust Typical Spectral Shapes. The figure shows the periodogram of the US linearly detrended (OLS) logged per capita GDP (1870-1994). The periodogram in figure A is smoothed making use of an 8th order polynomial. The data are taken from Maddison (1995). The theoretical spectrum for the AR(1) process is obtained using a coefficient of first order autocorrelation equal to 0.5.

with rates of divergence equal to either the one of the unit root or a flat one. Given the representation (4.3), a reasonable way of estimating the "typicality" of the spectrum trough the parameter d consists of running a simple OLS regression of the log of the estimated spectrum (periodogram) over the log-frequency at around the zero frequency[13]

$$\ln[I(\theta)] = const. - 2d \ln \theta, \ as \ \theta \in \left[0, \bar{\theta}\right]. \tag{4.5}$$

Diebold and Rudebusch (1989) and chapter three ran this kind of regression for the GDP per capita for a set of different OECD economies and show that the rate of divergence of the periodogram is flatter than the one associated with a unit root but steeper than the *ARMA weak memory case* (see Figure 1B)[14]. The estimated rate of divergence for different bandwidths T^{α}, where T indicates the sample size and a trimming coefficient equal to one based on the log-periodogram regression proposed by Geweke and Porter-Hudak (1983) as modified by Robinson (1996) are reported in table 1[15].

This suggests how the real GDP per capita of the US is characterized by a parameter of fractional integration greater than zero, but probably less than one. That some form of very slow mean reversion actually takes place

[13]The semi-parametric nature of the estimator implies that the econometrician is left with the choice to decide when "close" is sufficiently "close", that is the size of the interval $\left[0, \bar{\theta}\right]$. The number of Fourier frequencies used in running the regression (4.5) is called the bandwidth.

[14]See also Gil-Alana and Robinson (1997) for further empirical evidence in this direction based on a different methodology.

[15]Velasco (1996) shows that the Robinson (1995) log-periodogram estimator is consistent and normal even for non stationary d's. Moreover, in this application, the results are very robust with respect to both the choice of the trimming coefficient and/or applying the log-periodogram regression on the first difference of the data and then adding unity to the obtained result. The results are available on request.

T^α $\alpha =$	d parameter	Asymptotic S.E.
0.40	0.46	0.24
0.425	0.40	0.22
0.45	0.47	0.21
0.475	0.42	0.20
0.50	0.53	0.18
0.525	0.68	0.17
0.55	0.58	0.16
0.575	0.55	0.15
0.60	0.60	0.14

Table 4.1: **Log-periodogram regression, GDP/L, US, 1870-1994.** The log-periodogram regression (5) is applied on the linearly detrended (OLS) logged GDP per capita. The trimming coefficient is equal to one while the bandwidth is set to be equal to T^α. This implies that in the log-periodogram regression (5) $T^\alpha - 1$ Fourier frequencies are used while the smallest Fourier frequency is dropped. For a discussion of the estimating procedure see Diebold and Rudebusch (1989) and chapter three. For a derivation of the theoretical properties of this estimator originally proposed by Geweke and Porter Hudak (1983) see Robinson (1995). For an analisys of the properties of the estimator in the non stationary case see Velasco (1996).

in the data is also confirmed by time domain observation. Figure 2 shows how a time trend calculated using data only from 1880 and 1929 forecasts extremely well the current level of GDP of the US economy. This implies that the new information delivered by the Wold innovations ϵ_t's, might be irrelevant for forecasting on very long horizons[16].

The empirical evidence shows that the underlying stochastic process for aggregate GDP exhibits some form of long memory even if the degree of memory might remain uncertain i.e. if the particular case of the unit root can be ruled out. As a result we would like to know if they do exist economic mechanisms able to generate typical spectral shape similar to the ones observed in the real world. In this chapter we look for robust economic mechanism able to generate typical spectral shape: we define a mechanism to be *robust* if small alterations to the basic set-up do not shift the spectral shape of the time series from the very particular slope associated with the unit root to the flat one corresponding to the weak memory case[17]. We argue that an economic system able to generate a sufficient amount of cross-sectional heterogeneity satisfies this criterion, we show that the underlying disturbances must not be necessarily technological and we highlight why a typical spectral shape might be so typical.

[16]See also Diebold and Senhadji (1996) for similar conclusions based on similar evidence.

[17]Under this respect we make use of a *boundary approach* similar to the one popularized by Sutton (1992) in industrial organization.

4.3 A stylized vintage model

This section first lays down the structure of a stylized vintage model. It then characterizes the dynamics of the system once a set of aggregate disturbances is introduced in the basic set-up.

4.3.1 The Model

Time is discrete and goes from $-\infty$ to ∞.

The rate of technological progress is exogenous at rate γ (*ongoing growth*).

The number of firms in the economy is fixed with Lebesgue measure equal to one. We think of this as a *free entry condition*. In fact this would be the equilibrium outcome in a search theoretic framework with fixed amount of resources where each operating firm requires a given amount of resources and non operating firms must wait for these resources to be freed before using them (see e.g. Pissarides 1990)[18].

The firms in the economy can be in different technological states. In particular we say that a given firm is in aggregate state $i \geq 0$ at time t if the firm is using technology $t-i$ (*cross-sectional heterogeneity)*. Firms using different technologies are able to produce different quantities of goods, in particular a firm in state i at time t produces a quantity of goods equal to $\gamma(t-i)$[19].

[18]Given these considerations, the model considers as observational equivalent the event in which technological adoption takes place through destruction and successive creation of a new firm to that in which firms live forever. Mortensen and Pissarides (1995) analyze a vintage model in which firms explicitly face a trade-off between the two events.

[19]We are implicitly assuming that all variables are denominated in logs, this implies that differences indicate growth rates while arithmetic averages are geometric ones. It is possible to work out the model in which firms in state i at time t produce a quantity of

We indicate with π_t the countable infinite dimensional vector[20] collecting the measure of firms in each state. The ith element of the vector π_t tells us the measure of firms at time t using the technology $t - i + 1$. This vector is strictly positive, bounded between zero and one and thus it has the nature of a probability measure.

This implies that the level of aggregate output at time t, Y_t is equal to

$$Y_t = \gamma t - \gamma \pi_t' O, \tag{4.6}$$

where a "$'$" indicates the transpose operator on the given vector usually taken as a column vector. The vector O indicates a column vector with the property that its ith element is exactly equal to $i - 1$.

At a given point in time t a firm in aggregate state i has two possibilities either doing nothing and using the technology $t-i$ so that in the next period the firms will be in aggregate state $i+1$, or adopting the leading technology in the economy so that in the next period the firms will be in aggregate state zero. Technological adoption, however, implies some costs. We assume, very parsimoniously, that the cost of adopting the leading technology in the economy consists of three components which enter additively. c_i is a deterministic component of the cost of adopting the leading technology in the economy while ϵ_t and λ_i represents respectively an aggregate and a firm specific state dependent stochastic opportunity cost component. As a result the total cost of adopting the leading technology in the economy for a firm

goods equal to $\exp \gamma(t-i)$. We avoided to do this here just to keep notation as simple as possible. In the footnotes we will keep track of the required modifications.

[20] Hereafter all vectors are taken to be countable infinite dimensional column vectors.

in state i at time t is given by $c_i + \epsilon_t + \lambda_i$. ϵ_t represents the realisation at time t of a zero expected value aggregate random variable common across states and assumed to be identically and independently distributed, *iid,* over time. $\{\lambda_i, i \geq 0\}$ indicates a sequence over states i's of zero expected value random variables each one with support (possibly unbounded) $Z_i \subseteq \Re$. λ_i represents an idiosyncratic shocks independently distributed across units and over time. ϵ_t, λ_i can be read indifferently as either technological or demand shocks as they simply measure the firm-specific opportunity cost (in terms of forgone profits) of investing part of its own capital or labour resources in technological improvements (see e.g. Aghion and Saint Paul 1993 and Saint Paul 1993). The value of a firm $V(i,\, t,\, \epsilon_t,\, \lambda_i)$ in state i at time t whose cost of adopting the leading technology in the economy is given by $c_i + \epsilon_t + \lambda_i$, follows the Belman equation

$$\begin{aligned} V(i,\, t,\, \epsilon_t,\, \lambda_i) &= \max_{s \in \{0,1\}} \gamma(t-i) - s\,(c_i + \epsilon_t + \lambda_i) + \\ &\quad +\beta\,(1-s)\,V^e\,(i+1,\, t+1) + \beta\, s\, V^e\,(0,\, t+1). \quad (4.7) \end{aligned}$$

$0 < \beta < 1$ is the discount factor while $V^e\,(j,\, t)$ indicates the expected value of $V(j,\, t,\, \epsilon_t,\, \lambda_i)$ taken with respect to the random variables ϵ_t and λ_i[21]. If follows from standard dynamic programming argument that the problem is well defined. In particular the operator in (4.7) defines a contraction in the function $V(i,\, t,\, \epsilon_t,\, \lambda_i)$ as the Blackwell sufficient conditions for a contraction, monotonicity and discounting, are satisfied (see e.g. Stokey and Lucas 1988,

[21] As the random variables ϵ_t's and λ_i's are independently distributed across times and across states the expected value $V^e\,(\cdot,\, \cdot)$ does not depend on past realisations of aggregate and idiosyncratic shocks.

Theorem 3.3)[22]. It also follows by standard corollaries to the contraction mapping theorem (see e.g. Stokey and Lucas 1988, Corollary 3.1), that the value function $V(i, t, \epsilon_t, \lambda_i)$ is linear in t[23],strictly decreasing in i (under the maintained assumption that $\gamma i + c_i$ is strictly increasing in i), and finally weakly decreasing in ϵ_t and λ_i[24]. In '*normal times'*, that is when the aggregate component ϵ_t is equal to zero, the firm will then decide to adjust ($s = 1$) all the times that

$$\beta \left[V^e(0, t+1) - V^e(i+1, t+1)\right] \geq c_i + \lambda_i. \tag{4.8}$$

That is the firm weights the benefits of technological adoption $\beta \left[V^e(0, t+1) - V^e(i+1, t+1)\right]$ with the costs that it implies $\lambda_i + c_i$. If we indicate with $1- p_i$ the probability that the event (4.8) occurs[25], this represents the probability that a firm in aggregate state i will be using the best technology available in the economy at time $t+1$ when $\epsilon_t = 0$. There are no theoretical

[22]In the case where a firm in state i at time t produce a quantity of goods equal to $\exp \gamma(t-i)$, the condition required to assure finite discounted values would be $\gamma < -\ln \beta$.

[23]Log-linear in t in the case where a firm in state i at time t produce a quantity of goods equal to $\exp \gamma(t-i)$.

[24]For sake of exposition only, it is useful to consider the myopic problem in which at each point in time t a firm in state i maximises

$$Max \left[\beta\gamma(t-i), \beta\gamma(t+1) - c_i - \epsilon_t - \lambda_i\right].$$

[25]It follows from the linearity in t of the value function $V(i, t, \epsilon_t, \lambda_i)$ that the probabilities $1-p_i$'s are well defined and independent of t. In the myopic problem they are defined as those corresponding to the events

$$\lambda_i \leq \beta\gamma(i+1) - c_i.$$

If the idiosyncratic shocks are *iid* with distribution function $F(x)$, p_i is equal to $F[\beta\gamma(i+1) - c_i]$.

reasons for assuming any a priori structure on the values of the probability p_i's . For example in Aghion and Howitt (1994), Caballero and Hammour (1994b) and Mortensen and Pissarides (1996), both the probability distribution of the idiosyncratic shocks λ_i and the cost of adopting the leading technology c_i, are state independent so that the probabilities p_i's are unequivocally decreasing in i, that is firms adopting obsolete technologies are more likely to end up on the technological frontier rather than firms close to it. This reflects the fact that the higher the technological gap the higher is the incentive to adjust. However, other models with switching costs and human capital specificity stress why these probabilities might be constant (chapter one) or even increasing like in Acemoglu and Scott (1995), Jones and Newman (1995) and Jovanovic and Nyarko (1996). In fact the higher the technological gap the more difficult is technological adoption so that the cost of adopting the leading technology c_i might be thought to be positively related to i[26]. This implies that the dynamics of the state of a generic firm in normal times $\epsilon_t = 0$ is characterized by the infinite dimensional Markov chain P given by

$$P = \begin{bmatrix} 1-p_0 & p_0 & 0 & 0 & 0 & 0 & \cdots \\ 1-p_1 & 0 & p_1 & 0 & 0 & 0 & \cdots \\ 1-p_2 & 0 & 0 & p_2 & 0 & 0 & \cdots \\ 1-p_3 & 0 & 0 & 0 & p_3 & 0 & \cdots \\ 1-p_4 & 0 & 0 & 0 & 0 & p_4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where the rows and columns represent the set of feasible technological

[26] Clearly a change in the deterministic component of technological adoption c_i changes also the structure of the value function, that is the left hand side of equation (4.8). Discounting, $\beta < 1$, implies however that induced changes on the left hand side are always smaller than those on impact on the right hand side of equation (4.8). As a result an increase in c_i always makes the value of $1 - p_i$ to fall.

states in the economy while the elements $1-p_i$'s indicate the probabilities that a firm in state i at any time t will end up on the technological frontier at time $t+1$. We think of the matrix P as of the *transmission mechanism* in the economy.

4.3.2 Structure of the Transmission Mechanism

Given the lack of any strong a priori theoretical restrictions on the structure of the transmission mechanism P, we merely assume that whatever it is its current aggregate state, a firm sooner or later will adjust with probability one.

Assumption 1 Indicate with β_{ji} the probability that a firms starting in aggregate state j does not adjust before i periods, so that $\beta_{ji} = \prod_{k=0}^{i-1} p_{j+k}$. Assume that $\lim_{i\uparrow\infty} \beta_{ji} = 0, \ \forall j$.

We will keep Assumption 1 throughout the whole analysis, as it seems very plausible that arbitrarily inefficient firms would be eventually be driven out of business by more efficient ones (see Jovanovic and Nyarko 1996). The side effect of this assumption is that our framework will exhibit one and only one *recurrent* (ergodic) class. That is, there does exist only one set of states each one of them will be visited infinitely often by the firms in the economy.

Lemma 1 *Under assumption 1, the transmission mechanism P has always one and exactly one recurrent class.*

Proof: See appendix.

This lemma shows how our framework rules out multiple equilibria (multiple ergodic sets) to explain persistence in aggregate fluctuations (see Durlauf 1991). Under this respect the chapter is firmly in the second generation of Neo-Keynesian economics where macroeconomics is micro-founded solving explicit aggregation problems.

The previous lemma proved the existence of one recurrent class. We are interested in distinguishing the case in which this class is of infinite dimension (*irreducible transmission mechanism*) from the case in which it is of finite dimension (*reducible transmission mechanism*). The first case constitutes a useful benchmark case and implies that each firm will visit infinitely often all the aggregate states in the economy. The second one corresponds to a situation in which firms will end up with probability one in a finite dimensional set. Despite the lack of clear empirical evidence in either directions, the observation by Baily et al. (1992) and Bartelsman and Dhrymes (1994) that the persistence at the top of the technological distribution is very high, might support the reducible nature of the transmission mechanism P.

Assumption 2 Indicate with β_i, β_{0i}, so that $\beta_i = \prod_{k=0}^{i-1} p_k$, assume that $0 < \beta_i < 1$ and $\beta_\infty = \lim_{i\uparrow\infty} \beta_{0i} = 0$.

Assumption 2 guarantees that the transmission mechanism P is irreducible.

Lemma 2 *Under assumption 2 the infinite dimensional transmission mechanism P is irreducible and recurrent.*

Proof: See appendix.

In the next section we will analyse the dynamic of the system under the condition that Assumption 2 holds (irreducible transmission mechanism*)*, while in section 4 we will analyse the dynamics of the system under the assumption that it does exist an aggregate state i such that β_i is equal to zero (reducible transmission mechanism).

In general we are interested in knowing under which conditions a steady state distribution does exist. The existence of a steady state distribution seems to be a reasonable requirement for the plausibility of the theory. The following lemma answers this question.

Lemma 3 *If either the series $\sum_{i=1}^{\infty}\beta_i$ converges or the matrix is reducible, a steady state distribution exists.*

Proof: See appendix.

4.3.3 Aggregate Shocks with Cross-sectional Heterogeneity

We now want to analyse the dynamics of the system once a certain degree of aggregate uncertainty is introduced in the system. In particular we are interested in analysing the speed with which aggregate shocks propagates in the economic system. Let us consider a sequence of zero mean aggregate shocks, ϵ_t, assumed to be identically independently distributed over time, that is *iid* events. Equations (4.7) and (4.8) show how this modifies the problem of the firm: an aggregate shocks modifies, in a similar way, the incentive to adjust of all the firms in the economy. For example a positive aggregate shock, $\epsilon_t > 0$, increases the cost of technological adoption for all the firms in the economy, so that in the next period we will be observing less firms adopting the leading technology in the economy relative to what it would have occurred in normal times, $\epsilon_t = 0$. This implies that all the times at which the aggregate shock arrives a reallocation of the technological positions of firms takes place. We indicate with δ_t the size and structure of this reallocation where δ_t has the property that the sum by column of its entries, δ_t^i, is exactly equal to 0 that is $\sum_{i=0}^{\infty} \delta_t^i = 0$ (*reallocation property of the shock)*. We think of δ_t as the *reallocation structure* of the economy. δ_t has the nature of an error term measuring the difference between the observed cross-sectional distribution π_t and what would have occurred in normal times $P'\pi_{t-1}$. As a result the dynamics of the cross-sectional distribution of vintages currently in use, π_t, is described by the equation

$$\pi_t = \delta_t + P'\pi_{t-1}. \tag{4.9}$$

In order to summarize the above discussion we introduce the following formal definition for a reallocation structure δ_t.

Definition 1 (Reallocation Structures) *Indicate with σ_t and $\{e_i,\, i > 0\}$, respectively, the natural $\sigma - field$ generated at time t by the sequence of aggregate shocks ϵ_t, and a sequence of vectors where the generic element e_i has a value minus one in place one, one in place $i+1$ and zero otherwise. A reallocation structure δ_t has the form*

$$\delta_t = \delta(\sigma_t) = \sum_{i=1}^{\infty} w^i(\epsilon_t, \sigma_{t-1})\, e_i,$$

where the weighting coefficients $w^i(\epsilon_t, \sigma_{t-1})$*'s are a non decreasing function in* $\epsilon_t, \forall i$, *and* $\forall \sigma_{t-1}$, $0 \leq w^i(\epsilon_t, \sigma_{t-1}) \leq (1-p_i)\pi_{t-1}^{i+1}$ *if* $\epsilon_t > 0$, $p_i\pi_{t-1}^{i+1} \leq w^i(\epsilon_t, \sigma_{t-1}) \leq 0$ *if* $\epsilon_t < 0$, *while* $w^i(0, \sigma_{t-1}) = 0$.

Firstly, we assume that the reallocation structure δ_t is well defined so that a reallocation has a bounded effect on aggregate output. This requirement is satisfied by the following assumption.

Assumption 3 Assume that $\forall \sigma_t$ the reallocation structure δ_t has weighting coefficients $w^i(\epsilon_t, \sigma_{t-1})$'s that are at least of order $\frac{\beta_{ji}}{i^{2+\epsilon}}$ for some $\epsilon > 0$ and some $j \geq 0$[27].

Before proving the following proposition we need to show that the infinite dimensional matrix products are well defined.

Lemma 4 *Under Assumption 3, the product* $\delta_t{}' P^n O$ *is bounded* $\forall n$ *and well defined as the matrixes associate, that is* $\delta_t{}'(P^n O) = (\delta_t{}' P^n) O$, $\forall n$.

Proof: See appendix.

This chapter focuses on the persistence of aggregate fluctuations so that we are interested in a set of simplifying assumptions implying that aggregate output behaves as a linear symmetric process[28].

[27] This assumption is trivially satisfied for "finite" reallocations, that is for reallocation structures δ_t's where only a finite number of weighting coefficients w^i's is strictly positive.

[28] Assumption 4 is unlikely to hold as a "reasonable" approximation if aggregate and idiosyncratic shocks are independent one of another, in fact Caballero, Hengel and Haltiwanger (1997) find a strong negative correlation between the second moment of the idiosyncratic distribution and aggregate shocks. The reader that believes that Assumption 4

Assumption 4 Assume that $\delta_t = \delta(\epsilon_t)$, where $\delta(\cdot)$ is a linear function[29] so that $\delta_t = \delta\,\epsilon_t$, $\delta = \sum_{i=1}^{\infty} w^i\, e_i$, where, $\forall i$, $0 \leq w^i \in \Re$, and $\sum_{i=1}^{\infty} w^i < 1$.

Proposition 1 (Heterogeneity and linearity) *Under assumption 4 the aggregate output, Y_t, is a linear process.*

Proof: In what follows we define a process to be linear if there does exist a Wold decomposition where the innovations follow a martingale difference process. That is for the process to be linear it must be that $\forall n \geq 0$

$$E(Y_{t+n} - Y_{t-1} \mid \sigma_t) = E(Y_{t+n} - Y_{t-1} \mid \epsilon_t),$$

where σ_t indicates the natural $\sigma-field$ generated at time t by the sequence of aggregate shocks. Given Assumption 4 and the *iid* property of the aggregate shocks, ϵ_t, we have that, $\forall n \geq 0$

$$\begin{aligned} E(Y_{t+n} - Y_{t-1} \mid \sigma_t) &= \gamma\,(n+1) - \gamma\pi_{t-1}{}' P^n O - \gamma\delta_t{}' P^n O + \gamma\pi_{t-1}' O = \\ &= E(Y_{t+n} - Y_{t-1} \mid \epsilon_t), \end{aligned}$$

can never hold as a reasonable approximation should think of the aggregate shocks here considered as once for all shocks. Moreover, it seems that aggregate time series do not behave as linear symmetric processes (see e.g. Neftci 1984) and well accepted measures of persistence for non-linear processes are not readily available. The main contribution of the chapter is simply to show that aggregate shocks can vanish and even get amplified at very particular rates in a simple and popular stochastic model where cross-sectional heterogeneity is explicitly taken into account.

[29] Assumption 4 can be justified formally drawing on the definition of reallocation structure. In fact $\delta_t = \delta(\epsilon_t, \sigma_{t-1})$, so that a Taylor expansion around zero implies that the approximation $\delta_t \simeq \delta(\sigma_{t-1})\epsilon_t$ holds if δ_t is differentiable with respect to ϵ_t. Assumption 4 further requires that the previous history of the system does not affect the reallocation structure. This second assumption is not necessary and only Proposition 1 would be affected by relaxing it (see footnote 29).

as δ_t is function of ϵ_t only. Q.E.D.

The above formula allows to recover the Wold coefficients for the time series of aggregate output. In fact Y_t can be written as

$$Y_t = \gamma \pi_0{}' O - \gamma \pi_0{}' P^t O + \gamma t + \sum_{n=0}^{t} \phi_n \, \epsilon_{t-n}, \tag{4.10}$$

where $\phi_n = -\gamma\delta' P^n O$, while δ indicates the reallocation structure of the economy. If a steady state distribution exists so that $\pi_0{}' P^t = \pi_0{}' \ \forall t$ and we initialize the process at this point we have that the representation (4.10) collapses to

$$Y_t = Y_0 + \gamma t + \sum_{n=0}^{t} \phi_n \, \epsilon_{t-n}, \tag{4.11}$$

where Y_0 is the level of output at the starting date normalized to be equal to 0[30].

The result contained in the previous proposition generalizes results by Bertola and Caballero (1990, 1994) and Caballero (1992). Despite the large non linearities in the process with which firms adjust to the changing of the technological environment, the aggregate time series behaves like a linear symmetric process. The reason of this depends on Assumption 4. In general, a reallocation structure δ_t is function of all the previous history of the system, that is $\delta_t = \delta(\sigma_t)$ where σ_t indicates the natural $\sigma - field$ generated at time

[30]In general, (see footnote 28) the representation for aggregate output is

$$Y_t = Y_0 + \gamma t + \sum_{n=0}^{t} \phi_n(\sigma_{t-n-1}) \, \epsilon_{t-n},$$

where $\phi_n(\sigma_{t-n-1}) = -\gamma\delta(\sigma_{t-n-1})' P^n O$.

t by the sequence of aggregate shocks. Assumption 4 imposes two strong restrictions on the reallocation structure, the first is that the reallocation that takes place is function just of the current realization of the aggregate shock, that is $\delta_t = \delta(\epsilon_t)$. If we remove this condition we obtain that aggregate time series can exhibit *non linearities.* For example this is the case if the mass of firms in a generic state that reallocates is a function of the current measure of firms in that state.

The second one is that $\delta(\cdot)$ be a linear function, for this assumption to hold it must be that $\pi_t = \delta_t + P'\pi_{t-1} \geq 0$ with probability one. If we remove this condition we have a possible source of *asymmetries* in aggregate time series. As this chapter focuses on the persistence of aggregate fluctuations, it analyses the dynamics of the linear model assuming implicitly that Assumption 4 holds as a "reasonable" approximation[31]. The main implication of this result simply states that aggregate uncertainty must enter linearly for the aggregate time series to be a linear process.

Given the representation (4.11) the following proposition shows the relation between growth in aggregate output and technological adoption.

Proposition 2 (Technological adoption and growth) *If a steady state distribution exists, the long run growth rate of the aggregate economy is equal to that of the leading technology γ.*

[31]If we relax this assumption we have at hand a possible source of non linearities and asymmetries in aggregate time series without relying on strategic complementarity. The reasonable question to ask, then, is how much non-linearities and asymmetries this simple mechanism is able to generate. See Caballero and Hengel (1994) for an explanation based on strategic complementarities.

Proof: It follows from the representation (4.11) and the fact that the sequence of aggregate shocks $\{\epsilon_t,\, t \geq 0\}$ is *iid* with zero expected value by assumption. Q.E.D.

This chapter focuses on the conditions under which typical spectral shapes arise. For this to happen it must be that the transmission mechanism P allows for a sufficient amount of cross-sectional heterogeneity.

Proposition 3 (The role of cross-sectional heterogeneity) *If the transmission mechanism P is finite dimensional, the aggregate always behaves like a trend stationary process with ARMA disturbance, that is the Wold coefficients ϕ_n's decay at least exponentially, $|\phi_n| \leq \rho^n$, $0 \leq \rho < 1$ as $n \uparrow \infty$.*

Proof: See appendix.

This very simple result might look trivial but it is not. In fact it says that if the number of states in the system are finite, the degree of integration of the aggregate variable is equal to that of the underlying stochastic disturbance[32]. Given that a variable with a finite number of states is bounded, it is not surprising that the aggregate can never be a non stationary process ($d \geq \frac{1}{2}$)[33]. This theorem however claims much more, as it states that no typical spectral shape can be generated where cross-sectional heterogeneity is uniformly

[32]This result is hardly surprising given what has been found in the representative agent framework by Real Business Cycle models. As they are, they lack a transmission mechanism (see e.g Rotemberg and Woodford 1996). The attempt of introducing search frictions in the basic RBC model (see e. g. Andolfatto 1996 and Merz 1995) goes exactly in the same direction as highlighted by this chapter.

[33]Given the representation (4.4) a process with $d \geq \frac{1}{2}$ has infinite variance and so non-stationary in second moment.

bounded. In this respect Proposition 3 allows one to generalize results by Bertola and Caballero (1990, 1994), Caballero and Hengel (1990) in the sS literature. Despite the large non linearities implied in their framework, the rate with which the expected value of the cross-sectional distribution converges to its long run value is always exponential. Moreover, the result is also consistent with the observation by Bertola and Caballero (1994), that an increase in the variance of the idiosyncratic uncertainty reduces the speed of convergence of the aggregate variable even if, in their set-up, the rate of convergence still remains exponential. In other words, this proposition illustrates the role of ongoing growth in producing the amount of cross-sectional heterogeneity required to generate a degree of integration in the aggregate variable strictly greater than that of the underlying aggregate disturbance. In fact, it might be argued that in the real world cross-sectional heterogeneity is always bounded. This is, however, not the case when the forward looking nature of the transmission mechanism P is taken into account and we allow for ongoing technological growth. In fact Proposition 3 requires merely for the possibility that a firm currently adopting a given vintage can wait for an arbitrary number of periods before switching to the best technology in the economy. This implies that at each point in time cross-sectional heterogeneity can appear to be bounded but still not uniformly bounded. This is all that is required by Proposition 3.

4.4 Irreducible Transmission Mechanism

In this section we focus on the dynamics of the aggregate system under the assumption that the transmission mechanism P of the aggregate economy is irreducible. Under Assumption 2 each firm will visit infinitely often all the states of the aggregate economic system. In this and the next section we will identify small (big) firms as the ones adopting vintages far-away from (close to) the technological frontier. For example, Baily et al. (1992) and Bartelsman and Dhrymes (1994) find that employment size and productivity are positively correlated. In general to link productivity to some measures of size (employment, sale, assets) we need some assumptions on the dynamics of factor prices. If factor prices grow at the same rate as the technology frontier, a productivity ranking in terms of vintages corresponds exactly to a one in terms of size.

The first result we show is a negative one. Under the assumption of irreducibility the reallocative structure does not matter.

Proposition 4 *If the transmission mechanism P is irreducible, the rate of absorption of the shocks as measured by the Wold coefficients ϕ_n is independent of the reallocation structure δ.*

Proof: See appendix.

We are interested in conditions such that aggregate output is trend stationary with weak memory disturbances. This is the case if for example, the chance of ending up at the top of the technological frontier is higher (or equal) for firms adopting obsolete technologies, rather than for firms already at the top of the technological frontier.

Proposition 5 (*ARMA* **processes**) *If the probabilities p_i's are weakly decreasing in i, the aggregate time series behaves as an ARMA process, that is the Wold coefficients ϕ_n's decay at least exponentially, $|\phi_n| \leq \rho^n, 0 \leq \rho < 1$* as $n \uparrow \infty$.

In this case, the steady state distribution π^* cannot have a degree of skewness greater than that associated with the geometric distribution, that is

$$\pi_i^* \leq p^i,\ 0 < p < 1,\ as\ i \to \infty,$$

where π_i^* indicates the steady state probability of ending up in state i.

Proof: See appendix.

As an application of this result we can consider the two following parametric examples:

Example 1: If $p_i = \frac{ia+p}{i+1}$, for some $0 < a < 1$, the aggregate time series behave like a trend stationary series with $AR(1)$ disturbance and first order correlation equal to a[34]. If $a = p$ we are in the case where all the probabilities p_i's are constant. In this case the steady state distribution is geometric so

[34]Indicate with $R_i^{(n)}$ the generic element in place i, of the vector $R^{(n)} = P^n O$. It follows the recursion $R_i^{(n)} = (1 - p_i)R_0^{(n-1)} + p_i\, R_{i+1}^{(n-1)}$, where $R^1 = PO$. If we guess that the solution of this system has the form $R_i^{(n)} = c_n + a^n i$, we obtain that the system is satisfied if $c_n - c_{n-1} = p_0\, a^{n-1}$, $c_1 = p_0$ so that $c_n = \frac{1-a^{n-1}}{1-a}$. This implies that

$$\delta'\, P^n O = a^n\, \delta'\, O.$$

that the upper bound in the degree of skewness permitted by Proposition 5 is reached. This is the vintage model analysed by chapter one.

Example 2: If $a = 0$, so that $p_i = \frac{p}{i+1}$,the aggregate time series behave like a trend stationary series with white noise disturbances, in this case the steady state distribution is Poisson with parameter p.

In general Proposition 5 sets a lower bound to the degree of skewness required to generate typical spectral shapes. If the transmission mechanism is irreducible, cross-sectional heterogeneity can generate typical spectral shapes only if the steady state distribution exhibits a degree of skewness greater than the one associated with the geometric distribution.

Table 1 showed that the aggregate GDP of the US economy might be characterized by a parameter of fractional integration, d, lower than one but still positive. This might arise if firms adopting older vintages, small firms, tend to grow at a similar rate but quicker than firms at the top of the technological frontier, big firms. This also implies that small firms have more volatile growth rate as recent empirical evidence seems to suggest (see Mansfield 1962, Hall 1987, Evans 1987 and Dunne et al. 1989 for empirical evidence in this direction)[35].

Proposition 6 (Long memory) *Define with g_i the expected growth rate of a unit in aggregate state i, that is $g_i = (1 - p_i)(i + 1)$. If it does exist an aggregate state s^* such that*

$$g_i = h\gamma, \ \ 1 < h, \ \ \forall i \geq s^*,$$

[35]If we consider as an index of volatility the quantities either $p_i(1-p_i)(i+1)$ or $p_i(1-p_i)$, these are in fact increasing in i under the assumptions of Proposition 6.

the order of integration of aggregate output d is equal to $2-h$, that is the Wold coefficients $\phi_n \sim n^{d-1}$, as $n \uparrow \infty$.

The steady state distribution, π^*, is approximately Pareto's, that is $\pi_i^* \sim n^{-h}$ where π_i^* indicates the steady state probability of ending up in state i.

Proof: See appendix.

The intuition of the results in Proposition 6 is simple. If the Wold coefficients decay hyperbolically the shock is absorbed at decreasing rates rather than at constant ones as it would be in the exponential case, that is $n^{d-1} \sim (1-\frac{1-d}{1})(1-\frac{d-1}{2}) \cdots (1-\frac{d-1}{n})$ (see appendix). Fast growing small firms eventually become big and so they end up by growing at the same rate as the big ones, as a result the rate at which the shock is absorbed shrinks over time. The degree of skewness in the steady state distribution required to generate typical spectral shapes, does not seem to be excessive as the Pareto distribution is often argued to possess the degree of skewness found in empirical studies (see e.g. Hall 1983 and Sutton 1995).

The following proposition shows that if the transmission mechanism is irreducible we can never generate a unit root together with a stationary distribution. We can interpret this result in either two ways. If we believe that the transmission mechanism P is irreducible, Proposition 7 makes a strong theoretical case in favour of the empirical evidence arguing that aggregate GDP exhibits some form of long memory but smaller than the one associated with a unit root. Alternatively, we can keep Proposition 7 as a hint of the fact that some additional structure must be put on both the transmission mechanism P and the reallocation structure δ in order to generate typical spectral shapes. This last point will be the consideration that we will exploit in the next section.

Proposition 7 (Killing irreducibility) *If the transmission mechanism P is irreducible and a steady state distribution does exist, the order of integration of aggregate output is always strictly smaller than one. Moreover, if the transmission mechanism P is irreducible, the order of integration of output can never be greater than one.*

Proof: See appendix.

As application of the previous proposition we can consider the following result:

Proposition 8 (Random walk and Gibrat's Law) *A necessary and sufficient condition for aggregate output to be a random walk and the transmission mechanism P to be irreducible is that*

$$g_i = h\gamma, \;\; 0 < h < 1, \;\; \forall i.$$

Proof: See appendix.

In this case firms in different technological states are growing at the same rate γh independently of their current size. This is exactly what is claimed by Gibrat's Law (see Gibrat 1931 and Sutton 1995, 1996). If there is a random walk in output, the growth rate of output is equal to γh so that it is systematically lower than γ, as $0 < h < 1$. This implies, by Proposition 3, that no steady state distribution exists. In fact the coefficients β_i are decreasing at rate i^{-h} (see appendix) that is strictly lower than the one assumed in Lemma 3. In this case the matrix is null recurrent, that is $\lim_{n\to\infty} P^n = 0$, and there is no convergence in the cross-sectional distribution so that the dispersion of technological states widen over time.

The previous proposition shows how aggregate shocks that reallocates units across vintages can generate typical spectral shapes in aggregate output under plausible conditions. However, the results are not robust in the sense that we can not generate an order of integration strictly greater than one, while a unit root in output implies (quite implausibly) that the growth rate of aggregate output γh is systematically lower than that of the technological frontier γ. The next section shows that in order to derive robust typical spectral shapes some structure is required on both the transmission mechanism P and the reallocation structure δ.

4.5 Reducible Transmission Mechanism

In this section we focus on the dynamics of the aggregate system under the assumption that the transmission mechanism P of the aggregate economy is reducible. That means that with probability one, each firm will end up by positioning at the top of the technological frontier. This implies, by Lemma 3 and Proposition 3, that a steady state distribution always exists and that the growth rate of aggregate output is always equal to that of the technological frontier γ.

The main message of this section is that, in order to generate robust ($0 < d < 2$) typical spectral shape, some structure is required both on the transmission mechanism P and on the reallocative structure δ. In what follows we show what additional structure is required.

Proposition 9 (The reallocation structure matters) *Different reallocation structures δ can generate different rate of absorption of the shocks as measured by the Wold coefficients ϕ_n's.*

Proof: See appendix.

Recent empirical evidence suggests that recessions are a period of cleansing in which marginal firms are driven out of business (see e.g. Davis and Haltinwanger 1990, 1992 and Gali and Hammour 1991). The following definition embeds this idea in our reallocation structure[36].

Definition 2 (Cleansing type reallocation) *Suppose that the transmission mechanism P is reducible where the recurrent class is identified by the first s technological states*[37]*, then we define a reallocation structure,*

$$\delta = \sum_{i=1}^{\infty} w^i e_i, \ \ 0 \leq w^i \in \Re, \ \ \forall i, \ \ \sum_{i=1}^{\infty} w^i < 1$$

of the cleansing type, if $\exists \ w^i > 0$, for some $i \geq s$[38].

[36] "Cleansing" does not have to take place necessarily in the recession. For example, if there are liquidity constraints, the best period for reallocating is the boom.

[37] For simplicity, here, we identify the recurrent class s as given by $s = \max\{i : \ p_j \neq 0 \forall j > i\}$.

[38] This implies that at time zero either we assume that we are not in steady state or

A reallocation of the cleansing type is a *drastic reallocation* in the sense that it reallocates units from the transient states to the recurrent ones. We now show that a cleansing type reallocation can generate robust spectral shapes.

Proposition 10 (Robust typical spectral shapes) *Suppose that $s < \infty$, the reallocation is of the cleansing type and the expected growth rate of a unit in aggregate state i, g_i, is such that*

$$g_i = (1 - p_i)(i + 1) = h\gamma, \quad i \geq s^*$$

for some s^* where $h > 0$, then the order of integration of aggregate output d is equal to $2 - h$, that is the Wold coefficients $\phi_n \sim n^{d-1}$, as $n \uparrow \infty$.

Proof: See appendix.

The intuition of the result is simple. In fact, if the Wold coefficients behave hyperbolically the shock propagates at decreasing rates rather than at constant ones as it would be in the exponential case, that is $n^{d-1} \sim (1 - \frac{1-d}{1})(1 - \frac{d-1}{2}) \cdots (1 - \frac{d-1}{n})$ (see appendix). As small firms eventually become big, the degree of disequilibrium induced by the shocks shrinks or gets amplified, depending of the relative growth rate of small versus big firms. Moreover, as the number of small firms shrinks over time, the rates at which the shocks propagates over time shrinks as well.

the first shock must be positive. In general Assumption 4 can not hold here. For all the reallocations to be of the cleansing type it must be that positive aggregate shocks, $\epsilon_t > 0$, cause more reallocation than negative ones.

We are particularly interested in the case in which $1 \leq h < 2$, that is $0 < d \leq 1$, because as documented in section one it is the case that seems to be consistent with the empirical evidence. If this is the case:

(i) Small firms tend to grow faster than big firms (see Mansfield 1962, Hall 1987, Evans 1987 and Dunne et al. 1989 for empirical evidence in this direction)[39].

(ii) Independent of the relative growth rate of big versus small firms there is convergence in the cross-sectional distribution, so that we are in a case of Galton's fallacy (see Quah 1993).

(iii) The growth rate of output is equal to that of the leading technology in the economy, by Proposition 2.

(iv) If $h < 1$ so that big firms grow faster than small ones (see Davies, et al. 1993 for empirical evidence in this direction), the stochastic process for the growth rate of aggregate output exhibits long memory. In the limit case, in which $h = 0$ (in this case Assumption 1 would not hold) aggregate output is an integrated process of order 2.

4.6 Fractional Cointegration

A definition of cointegration for variables integrated of order d, hereafter $I(d)$, is as follows:

[39]This also implies that small firms have more volatile growth rates as recent empirical evidence seems to suggest (see Mansfield 1962, Hall 1987, Evans 1987 and Dunne et al. 1989 for empirical evidence in this direction). If we consider as an index of volatility the quantities either $p_i(1-p_i)(i+1)$ or $p_i(1-p_i)$, these are in fact increasing in i.

Definition 3 (Fractional cointegration) *For a $p \times 1$ vector Y_t where the all $i-th$ elements $Y_{it} = I(d)$, we say that the variables are cointegrated with degree of cointegration $b = d - d'$, if there exists a $p \times 1$ vector $\alpha \neq 0$ such that $e_t = \alpha' Y_t \sim I(d')$ where $d' < d$.*

As firms are the minimal working cells of an aggregate economy where most economic decisions are taken it is not difficult to see why aggregates should be cointegrated. To fix ideas we can consider the problem of the firm of choosing the optimal quantity of labour as a function of the technology it is actually using. Suppose for example that the profits given by using an amount of labour l_i, when a firm is in aggregate state i, is equal to

$$\max_{l_i} \gamma(\epsilon_t + t - i)\, l_i - \frac{\gamma a}{2} l_i^2 - \gamma t\, l_i,$$

where $\gamma(\epsilon_t + t - i) - \gamma a\, l_i$ is the marginal revenue of labour decreasing in l_i, while γt represents the cost of using a given amount of labour assumed to grow exogenously at the same rate as the one of the leading technology in the economy. In this case, the first order condition for the firm states that

$$l_i = \frac{1}{a}(\epsilon_t - i),$$

so that in the aggregate the relation

$$\pi_t' l = \frac{1}{a}(\epsilon_t - \pi_t' O),$$

holds, where l is an infinite dimensional vector collecting the optimal decision of firms at each given aggregate state i. This implies that aggregate employment is cointegrated with aggregate output with a degree of cointegration equal to the order of integration of aggregate output ($d' = 0$).

Robinson and Marinucci (1997) show how in the real world the order of integration of the residuals is different from zero ($d' \neq 0$). In our framework the phenomenon can arise if the cost of using a given amount of labour is not just a function of the leading technology in the economy, but also of other moments of the cross-sectional distribution. If wages for instance solve some form of bargaining problem like for example in the case where some search frictions are introduced in the system (see e.g. Pissarides 1990), the wage paid by firms in aggregate state i, w_i, can be thought as equal to

$$w_i = \beta(t-i) + (1-\beta)\left[f(\pi_t) + t\right],$$

where β represents the bargaining power of the worker, while $f(\pi_t)$ captures the outside options of the workers as a function of the cross sectional distribution of vintages currently in use.

In this case

$$l_i = \frac{1}{a}[\epsilon_t - (1-\beta)i - (1-\beta)f(\pi_t)],$$

so that in the aggregate the following relation holds

$$\pi_t' l = \frac{1}{a}[\epsilon_t - (1-\beta)\pi_t' O - (1-\beta)f(\pi_t)].$$

To fix ideas, we can suppose technological adoption takes place through destruction and successive creation and that the outside option of the worker is given by the amount of creation equal to π_t^0, where π_t^0 indicates the first element of the vector π_t. It can be checked that under the assumption of Proposition 10, π_t^0 evolves as an integrated process of order $d = 1 - h$. If so a cointegrating relation between output and employment might leave an order of integration in the residual strictly positive and equal to $d' = 1 - h$. More generally under the assumptions of Proposition 10 the rth moment of the cross sectional distribution behaves as an integrated process of order $d = 1 + r - h$. Moreover, under the assumption that the Wold innovations ϵ_t are Gaussian, Taqqu's theorem (Taqqu 1975) argues that non linear transformation of the original (stationary) process can generate arbitrary rates of divergence of the spectrum around zero frequency[40].

4.7 Conclusions

Ex ante homogeneous firms end up by having very different histories. This implies that once a picture of the economic system is taken, a lot of cross-sectional heterogeneity appears. This chapter has shown that the mechanism generating heterogeneity in the real world also generates persistence in the aggregate fluctuations. Moreover, as aggregate shocks create very high degree of persistence without affecting either the number of firms in the market or the rate of technological progress, this degree of persistence is simply attributed to cross-sectional heterogeneity. Because of this reason the chapter concludes that cross-sectional heterogeneity is an important transmission mechanism with some distinctive properties. We summarize them by saying that cross-sectional heterogeneity is a powerful, realistic, robust and general transmission mechanism. It is *powerful* because a sufficient amount of cross-sectional heterogeneity is able to generate typical spectral shapes without

[40]Taqqu (1975) consider a function $g(\cdot)$ where x_t is stationary Gaussian and $E(x_t) < \infty$. He shows that the behaviour of $g(x_t)$ is governed by the Hermitian rank m of $g(\cdot)$. More precisely if $x_t = I(d)$, then $g(x_t) = I(md + \frac{1-m}{2})$ where m is the Hermitian rank of $g(\cdot)$ so a natural number greater or equal than one.

necessarily relying on technological shocks. It is *realistic* because it is shown how two of the most striking empirical regularities in economics, the typical spectral shape and Gibrat's law, might be just two faces of the same coin. It is *robust* because it is able to generate almost any kind of spectral shapes provided that reallocations cleanse the economy reallocating between very low and very high technological states. It is *general* because the chapter gives a micro-foundation for the fractional cointegration of aggregate economic variables illustrating why a typical spectral shape might be such a common phenomenon.

A lot of questions still remain open. In particular realism does not imply reality so that a careful empirical investigation is required to see what really are the sources of the persistence of aggregate fluctuations. Three obvious candidates come to the mind: the dynamics of the leading technology in the economy, γ, the size of the market (the number of firms in the market) or cross-sectional heterogeneity. This seems to be a promising basis for discriminating across different macroeconomic theories. On the one hand the Real Business Cycle tradition as well as growth theories based on learning by doing, would stress that most aggregate persistence would arise because of the dynamics of the leading technology in the economy. On the other hand Neo-Keynesian macroeconomics stressing the role of coordination failures might argue that most aggregate persistence arise because of either the dynamics of the size of the market or the fact that multiple ergodic sets exist in the economic system. This chapter has departed from both the two strands of the literature and has drawn on the recent tendency to micro-found macroeconomics solving explicit aggregation problems. It has argued that cross-sectional heterogeneity might be a powerful source of persistence and has concluded that the interplay between idiosyncratic and aggregate uncertainty shapes the dynamics of macroeconomic variables in a distinctive way.

4.8 Appendix

4.8.1 Proofs of results in section 3

Proof of Lemma 1 Two states i and j are said to communicate if it does exist a positive probability that in a finite number of transitions, state j can be reached starting from state i and vice-versa. As the concept of communication satisfies the reflexivity, the symmetry and the transitivity property is an equivalence relation. This implies that we can partition the totality of states into equivalence classes. The states in an equivalence class are those which communicate with each other. The Markov chain is *irreducible* if the equivalence relation induces only one class. A state i is *recurrent* if and only if, starting from state i, the probability of returning to state i after some finite length of time is one. A non-recurrent state is said to be *transient*. All states in an equivalence class are either recurrent or transient so that both recurrency and transience are a class property (see e.g. Karlin and Taylor 1975, 1981). Given Assumption 1, the Markov chain P has the property that starting from state i, the probability of returning to state zero is one. This implies firstly that state zero is recurrent and secondly that either state i and state zero belong to the same class or state i is transient. As state zero is recurrent, it follows that at least one recurrent class does exist and given the previous considerations this is the only one. Q.E.D.

Proof of Lemma 2 Given Assumption 2 it follows that every state can be reached starting from state zero and vice-versa so that all states communicate and the transmission mechanism P is irreducible. Given Lemma 1 a recurrent class does exist so that the transmission mechanism P is irreducible and recurrent. Q.E.D.

Proof of Lemma 3 If the matrix is reducible, given Assumption 1 and Lemma 1, the statement is trivial as eventually all units will enter with probability one in the recurrent finite dimensional set. If the Markov chain is irreducible it follows from the basic limit theorem of Markov chains (see Karlin and Taylor 1975, theorem 1.2.) that either two cases hold. Either it is null recurrent so that asymptotically P^n goes to a matrix of zeros, or it is positive recurrent so that the matrix P^n converge to a matrix whose rows are identical and equal to the steady state distribution, call that π^*. It is easy to check that π^*, if it does exist, must have the property that the element in place i π_i^* must satisfy the equality $\pi_i^* = \pi_0^* \beta_i$. This implies that a necessary and sufficient condition for the transmission mechanism P to be positive recurrent is that

$$\pi_0^* = \frac{1}{\sum_{i=1}^{\infty} \beta_i},$$

exists finite and strictly positive. Q.E.D.

Proof of Lemma 4 We first write the reallocation structure δ as equal to $\delta = \delta^+ - \delta^-$ where $\delta^+ \geq 0$ and $\delta^- \geq 0$. We then note that non-negative matrixes associate under multiplication and that the distributive property is always satisfied for denumerable matrixes (see e.g. Kemeny, Snell and Knapp 1966 proposition 1-2 and corollary 1-4). This implies that $\delta' P^n O = (\delta^+ - \delta^-)' P^n O$ is well defined provided that for each n, $(\delta^-)' P^n O$ and $(\delta^+)' P^n O$ are (not necessarily uniformly) bounded . To show this, firstly we note that the reallocation structure δ is such that $(\delta^-)' O$ and $(\delta^+)' O$ are both finite as by Assumption 3 the series $\sum_{i=1}^{\infty} |w^i(\epsilon_t, \sigma_{t-1})|\, i$ converges. Secondly we note that each element in place j of $P^n O$ has increments bounded above from one, below from $-p\beta_{jn}(j+n)$ where $0 < p \leq 1$. It then follows that $|(\delta^-)' P^n O - (\delta^-)' P^{n-1} O|$ and $|(\delta^+)' P^n O - (\delta^+)' P^{n-1} O|$ are both dominated by

$$\max \left[1,\, \tilde{\beta}_n \sum_{i=1}^{\infty} \frac{i\, w^i}{\tilde{\beta}_i} + \tilde{\beta}_n\, n \sum_{i=1}^{\infty} \frac{w^i}{\tilde{\beta}_i} \right]$$

where $\tilde{\beta}_i = \prod_{k=0}^{i-1} p_k$ after replacing $p_i = 1$ for those p_i's equal to zero. Assumption 3 then guarantees that both $(\delta^-)' P^n O$ and $(\delta^+)' P^n O$ are bounded. Q.E.D.

Proof of Proposition 1 Given in the main text.

Proof of Proposition 2 Given in the main text.

Proof of Proposition 3 (The role of cross-sectional heterogeneity) In this case the transmission mechanism P is finite dimensional and we indicate with k the number of states. If the Markov chain P is irreducible we know by Lemma 1 that it also recurrent. If all probabilities p_i's are not exactly equal to one, the Markov chain P is aperiodic. In this case we know that a steady state distribution π^* does exist and the rate of convergence is exponential and independent of the initial distribution (see e.g. Stokey and Lucas 1988, theorem 11-4) that is

$$\left|p_{ij}^{(n)} - \pi_j^*\right| \leq A\,\rho^n, \;\; \forall i, j, \tag{4.12}$$

where $A \geq 0$ and $0 \leq \rho < 1$ while $p_{ij}^{(n)}$, π_j^* indicate respectively the element in row i and column j of P^n, the nth iterated of P and the element in place jth of the column vector π^*. It follows from the triangle inequality that

$$|\delta_t{}' P^n O| \leq B\,\rho^n,$$

where B is a bounded quantity. If the matrix is reducible we know that the first $s \leq k$ states of the transmission mechanism P are recurrent while all the other $k-s$ are transient. The structure of P implies that a unit starting from one transient state will enter the recurrent class after a number of periods less or equal than $k - s$. If we indicate with π_i^* the element in place i of the steady state distribution of P, π^*, we then know that the previous reasoning applies starting from the $k - s\ th$ iterations so that

$$\left|p_{ij}^{(n)} - \pi_j^*\right| \leq A\,\rho^n, \;\; \forall i, j, \quad n \geq k - s.$$

It follows that necessarily $|\delta_t{}' P^n O| \leq B\,\rho^n$ if $n \geq k-s$ where B is a bounded quantity. Q.E.D.

4.8.2 Proofs of results in section 4

Proof of Proposition 4 Indicate with $R_i^{(n)}$ the generic element in place i

of the vector $R^{(n)} = P^n O$. For each i, $R_i^{(n)}$ follows the recursion

$$R_i^{(n)} = (1 - p_i) R_0^{(n-1)} + p_i R_{i+1}^{(n-1)}, \tag{4.13}$$

where $R^{(1)} = PO$. If we solve for $R_{i+1}^{(n-1)}$ in equation (4.13) and we substitute backwards for $R_i^{(n)}$, we obtain that, $\forall i \geq 0$,

$$R_i^{(n)} = \frac{1}{\beta_i} R_0^{(n+i)} - \frac{1-p_0}{\beta_i} R_0^{(n+i-1)} - \frac{p_0(1-p_1)}{\beta_i} R_0^{(n+i-2)} - \cdots - \frac{\beta_{i-1}(1-p_{i-1})}{\beta_i} R_0^{(n)},$$

where $\beta_i = \prod_{k=0}^{i-1} p_k$, is such that $0 < \beta_i < 1$, $\forall i$ given Assumption 2. It then follows that

$$R_i^{(n)} = \frac{\beta_0}{\beta_i} Y_{n+i-1} + \frac{\beta_1}{\beta_i} Y_{n+i-2} + \frac{\beta_2}{\beta_i} Y_{n+i-3} + \cdots + \frac{\beta_{i-1}}{\beta_i} Y_n + R_0^{(n)}, \quad \forall i \geq 0,$$

where $Y_{n+i-1} = R_0^{(n+i)} - R_0^{(n+i-1)}$. Given Assumption 4, $\delta = \sum_{i=1}^{\infty} w^i e_i$, where, $\forall i$, $0 \leq w^i \in \Re$, and $\sum_{i=1}^{\infty} w^i < 1$ so that $\sum_{i=0}^{\infty} \delta^i = 0$. As the distributive property is always satisfied for denumerable matrixes (see e.g. Kemeny, Snell and Knapp 1966 proposition 1-2 and corollary 1-4), the Wold coefficients $\phi_n = \gamma \delta' P^n O$ can be written as equal to $\phi_n = \gamma \sum_{i=1}^{\infty} w^i e_i' P^n O$ that is well defined by Assumption 3.

We now show that each element of the sum $e_i' P^n O$ behaves asymptotically behaves as Y_n. Firstly we show a limit for Y_n does exist. In fact, by the definition of Y_n we obtain that

$$Y_n = R_0^{(n+1)} - R_0^{(n)} = \sum_{i=0}^{n+1} [p_{0i}^{(n+1)} - p_{0i}^{(n)}]\, i,$$

where $p_{0i}^{(n)}$ indicates the element in row 0 and column j of P^n, the nth iterated of P. As $0 < 1 - p_i < 1$, the transmission mechanism P is aperiodic, so that

$p_{0i}^{(n)}$ always converges as $n \uparrow \infty$ by the basic limit theorem of Markov chains (see e.g. Karlin and Taylor 1975, theorem 1.2.). It follows that a limit for Y_n always exists. Moreover we know that this limit will be greater or equal than zero by recurrency and will be less than one as the first difference of $R_0^{(n)}$ is uniformly bounded by one. It follows that

$$\lim_{n\to\infty} Y_n = a, \quad 0 \leq a \leq 1.$$

Suppose now that $\lim_{n\to\infty} Y_n = 0$, we want to show that, $\forall i$, $e_i' P^n O = \frac{\beta_0}{\beta_i} Y_{n+i-1} + \frac{\beta_1}{\beta_i} Y_{n+i-2} + \frac{\beta_2}{\beta_i} Y_{n+i-3} + \cdots + \frac{\beta_{i-1}}{\beta_i} Y_n \sim Y_n$. To prove this we note that $\lim_{n\to\infty} \frac{Y_n+i}{Y_n} = A_i \leq 1, \forall i$ where A_i is positive as Y_n is positive for big n because of recurrency. Let us argue by contradiction and suppose that $A_i > 1$. Then $\forall \epsilon$ there does exist $N^* \ni \forall n > N^*$ $A_i - \epsilon < \frac{Y_{n+i}}{Y_n} < A_i + \epsilon$. If so $Y_m > (A_i - \epsilon) Y_{m-i} \geq (A_i - \epsilon)^{\frac{m-n}{i}} Y_n$, that is eventually unbounded when m increases if $A_i > 1$, but this implies a contradiction as Y_n is uniformly bounded by one. As a result we obtain that $e_i' P^n O \sim Y_n$. Then we note that $\phi_n = \gamma \sum_{i=1}^{\infty} w^i e_i' P^n O$ is written as the sum with non negative weights of quantities each one behaving asymptotically as Y_n, then it follows that ϕ_n itself behaves asymptotically as Y_n (see e.g.g Davidson 1994). For the general case where $\lim_{n\to\infty} Y_n = a > 0$, we can follow the same steps as before to show

$$e_i' P^n O - B_i = \frac{\beta_0}{\beta_i} Y_{n+i-1} + \frac{\beta_1}{\beta_i} Y_{n+i-2} + \frac{\beta_2}{\beta_i} Y_{n+i-3} + \cdots + \frac{\beta_{i-1}}{\beta_i} Y_n - B_i \sim Y_n - a,$$

where $B_i = \frac{\sum_{j=0}^{i-1} \beta_j}{\beta_i} a$. From the previous reasoning it follows that $\phi_n - B \sim Y_n - a$ where $\lim_{n\to\infty} \phi_n = B = \sum_{i=1}^{\infty} w^i B_i$, where B is bounded by Assumption 3 as $\sum_{j=0}^{i-1} \beta_j \leq i$. Q.E.D.

In order to prove the next three propositions we will make use of the following lemma:

Lemma 5 $Y_n = R_0^{(n+1)} - R_0^{(n)}$ satisfies the recursive relation given by

$$Y_n = -\sum_{i=0}^{\infty} \beta_{i+1} Y_{n-i-1} + \beta_n n, \ \forall n > 0, \tag{4.14}$$

where $Y_n = 0$ if $n \leq 0$.

Moreover, if the transmission mechanism is irreducible and a steady state distribution does exist $\lim_{n\to\infty} Y_n = 0$.

Proof of Lemma 5 In the course of the proof of Proposition 4 we have shown that a limit for Y_n does exist so that

$$\lim_{n\to\infty} Y_n = a, \ \ 0 \leq a \leq 1.$$

$R_0^{(n)}$ indicates the expected position after n iterations of a unit starting in state 0. By the law of iterated expectations, it follows that

$$R_0^{(n)} = \sum_{i=0}^{\infty} \beta_i (1 - p_i) R_0^{(n-i)} + \beta_n n, \ n \geq 0,$$

where $R_0^{(n)} = 0$ if $n \leq 0$,while $\beta_i = \prod_{k=0}^{i-1} p_k$. It follows from the definition of Y_n that

$$Y_n = R_0^{(n+1)} - R_0^{(n)} = \sum_{i=0}^{\infty} \beta_i (1-p_i) Y_{n-i-1} + \beta_n n - \beta_{n-1}(n-1), \ n \geq 0, \tag{4.15}$$

where $Y_n = 0$ if $n \leq 0$. We can now use equation (4.15) to prove by recursion that

$$\sum_{i=0}^{\infty} \beta_i Y_{n-i-1} - \beta_{n-1}(n-1) = 0, \; \forall n > 0,$$

so that

$$Y_n = -\sum_{i=0}^{\infty} \beta_{i+1} Y_{n-i-1} + \beta_n n, \; \forall n > 0,$$

where $Y_n = 0$ if $n \leq 0$. This proves the first part of the lemma.

We now show that under the assumption that a steady state distribution does exist $\lim_{n\to\infty} Y_n = 0$. By step one we know that the limit is bounded. Let us argue by contradiction and suppose that $\lim_{n\to\infty} Y_n = a > 0$. If so $\forall \epsilon$, $\exists N^*$, $\ni \forall n > N^*$ the following inequality is satisfied

$$\begin{aligned} |Y_n - a| &= \left| -\sum_{i=0}^{\infty} \beta_{i+1}(Y_{n-i-1} - a) - a\sum_{i=0}^{\infty} \beta_i + \beta_n n \right| \\ &< \frac{\epsilon}{\sum_{i=0}^{\infty} \beta_{i+1}}, \end{aligned} \tag{4.16}$$

that is well defined as $\sum_{i=0}^{\infty} \beta_{i+1}$is bounded by the assumption that a steady state distribution does exist and Lemma three. We now note that given any two quantities A and B,

$$|A| - |B| \leq |A + B| \leq |A| + |B|.$$

We apply this result to equation (4.16) with

$$\begin{aligned} A &= -a \sum_{i=0}^{n-N^*} \beta_i, \\ B &= -\sum_{i=0}^{n-N^*} \beta_{i+1}[Y_{n-i-1} - a] - \sum_{i=n-N^*+1}^{n} \beta_{i+1} Y_{n-i-1} + \beta_n n, \end{aligned}$$

so that $|Y_n - a| = |A + B|$. We now show that $|B|$ can be made arbitrarily small so that equation (4.16) can be satisfied only if $|A| = 0$ that is $a = 0$. In fact, the triangle inequality implies

$$
\begin{aligned}
|B| &\leq \left| -\sum_{i=0}^{n-N^*} \beta_{i+1}(Y_{n-i-1} - a) \right| + \left| \sum_{i=n-N^*+1}^{n} \beta_{i+1} Y_{n-i-1} \right| + \beta_n n \leq \\
&\leq \epsilon + (N^*)^2 \sup_{n-N^* \leq i \leq n} \beta_i + \beta_n n,
\end{aligned}
$$

as Y_{n-i}, for $i \geq 0$ can have a jump of size at most equal to n. As $\lim_{n\to\infty} \beta_n n = 0$ by Lemma three, $|B|$ can be made arbitrarily small so that equation (4.16) can be satisfied only if $|A| = 0$ that is $a = 0$. Q.E.D.

Proof of Proposition 5 (*ARMA* processes) It follows from Proposition 4 that the Wold coefficients $\phi_n = \gamma\delta' P^n O \sim Y_n$ where Y_n is the first difference of $R_0^{(n)}$, that is $Y_{n-1} = R_0^{(n)} - R_0^{(n-1)}$. We want to prove that $|Y_n| \leq \rho^n$, $0 \leq \rho < 1$ as $n \uparrow \infty$. If the probability p_i are decreasing in i, β_i falls at a rate that is quicker than the exponential one as $\beta_i \leq p_0^i$, so that by Lemma 3 a steady state distribution does exist. This implies by Lemma 5 $\lim_{n\to\infty} Y_n = 0$. We want to show that there does exist a number $\frac{1}{\rho} > 1$ such that $\lim_{n\to\infty} \frac{Y_n}{\rho^n} = 0$. We choose ρ such that $\frac{p_0}{\rho} < 1$ that is well defined as by Assumption 2, $0 < p_0 < 1$. Given equation (4.14) it follows that

$$
\frac{Y_n}{\rho^n} = \tilde{Y}_n = -\sum_{i=0}^{\infty} \tilde{\beta}_{i+1}\, \tilde{Y}_{n-i-1} + \tilde{\beta}_n n, \ \forall n > 0,
$$

where $\tilde{Y}_n = \frac{Y_n}{\rho^n}$, $\tilde{\beta}_{i+1} = \frac{\beta_{i+1}}{\rho^{i+1}} \leq \left(\frac{p_0}{\rho}\right)^{i+1}$. We can then follow the same procedure as in the proof of Lemma 5 to show that $\lim_{n\to\infty} \tilde{Y}_n = 0$ as the series $\sum_{i=0}^{\infty} \tilde{\beta}_{i+1}$ converges.

Finally, by Lemma 3 we know that if a steady state distribution, π^*, does exist it must have the property that its element in place i, π_i^*, is such that $\pi_i^* = \pi_0^* \beta_i$ where $\beta_i \leq p_0^i$. This proves the last assertion of the proposition. Q.E.D.

Proof of Proposition 6 (Long memory) It follows from Proposition 4 that the Wold coefficients $\phi_n = \gamma\delta' P^n O$ asymptotically behaves as Y_n where Y_n is the first difference of $R_0^{(n)}$, that is $Y_{n-1} = R_0^{(n)} - R_0^{(n-1)}$. In order to prove the assertion we want to show that $Y_n \sim n^{d-1} = n^{1-h}$, where d is the order of integration of aggregate output.

If $n > s^*$, $\beta_n = \beta_{s^*}\,\beta_{s^*(n-s^*)}$. We now show that $\beta_{s^*(n-s^*)} \sim n^{-h}$ as $n \uparrow \infty$. In fact $\beta_{s^*(n-s^*)} = \prod_{k=0}^{n-s^*-1} p_{s^*+k}$ is equal to

$$\beta_{s^*(n-s^*)} = (1 - \frac{h}{s^*+1})(1 - \frac{h}{s^*+2}) \ \cdots \ (1 - \frac{h}{n}). \tag{4.17}$$

For large i, we have the approximation

$$\left[1 - \frac{h}{i+1}\right] \cong \left[1 + \frac{1}{i+1}\right]^{-h}, \tag{4.18}$$

as it follows from a Taylor expansion around zero of the function $(1+x)^{-h}$. Substituting (4.18) into (4.17) we obtain that

$$\begin{aligned} \beta_{s^*(n-s^*)} &\sim \left[\frac{n+1}{n}\right]^{-h} \left[\frac{n}{n-1}\right]^{-h} \cdots \left[\frac{i+2}{i+1}\right]^{-h} \cdots \left[\frac{s^*+3}{s^*+2}\right]^{-h} \left[\frac{s^*+2}{s^*+1}\right]^{-h} = \\ &= B(n+1)^{-h}, \end{aligned}$$

where B is a bounded quantity. It follows that β_n behaves asymptotically as n^{-h}, $\beta_n \sim n^{-h}$. As $1 < h$, it follows from Lemma 3 that a steady state distribution does exist so that $\lim_{n\to\infty} Y_n = 0$, by Lemma 5.

We now show $\lim_{n\to\infty}(n+b)^{h-1-\epsilon}Y_n = 0$, $\forall\epsilon > 0$, where b is an arbitrary positive quantity. In fact, it follows from equation (4.14) that

$$(n+b)^{h-1-\epsilon}Y_n = \tilde{Y}_n = -\sum_{i=0}^{\infty}\tilde{\beta}_{i+1}\, g(n,i)\ \tilde{Y}_{n-i-1} + \tilde{\beta}_n\, n,\ \forall n > 0,$$

where

$$\begin{aligned}\tilde{Y}_n &= (n+b)^{h-1-\epsilon}Y_n,\\ \tilde{\beta}_{i+1} &= \beta_{i+1}(i+1)^{h-1-\epsilon},\\ g(n,i) &= \frac{(n+b)^{h-1-\epsilon}}{(i+1)^{h-1-\epsilon}\,(n-i-1+b)^{h-1-\epsilon}}.\end{aligned}$$

We now show that the positive quantity

$$\lim_{n\to\infty}\sum_{i=0}^{n-1}\tilde{\beta}_{i+1}g(n,i) \tag{4.19}$$

is bounded and bounded away from zero. In fact, $\min_{1\leq i\leq n-1}(i+1)\,(n-i-1+b) = 2(n+b)$ if b is sufficiently big. This implies that

$$\sum_{i=0}^{n}\tilde{\beta}_{i+1}g(n,i) \leq 2\sum_{i=0}^{n}\tilde{\beta}_{i+1},$$

that is bounded as $\tilde{\beta}_{i+1} \sim (i+1)^{-1-\epsilon}$. We can then follow the same procedure as in the proof of Lemma 5 to show that $\lim_{n\to\infty}\tilde{Y}_n = 0$.

This implies that the rate of convergence of Y_n is at least equal to n^{1-h}. We now show that, the rate is exact. Let us argue by contradiction and suppose that $\lim_{n\to\infty}(n+b)^{h-1}Y_n = 0$ where b is an arbitrary positive quantity. It follows from equation (4.14)that

$$(n+b)^{h-1}Y_n = \tilde{Y}_n = -\sum_{i=0}^{n-1} \beta_{i+1}\, h(n,i)\ \tilde{Y}_{n-i-1} + (n+b)^{h-1}\beta_n\, n,\ \forall n > 0,$$

where $h(n,i) = \frac{(n+b)^{h-1}}{(n-i-1)^{h-1}}$. This implies that

$$(n+b)^{h-1}\beta_n\, n - \left| -\sum_{i=0}^{\infty} \beta_{i+1}\, h(n,i)\ \tilde{Y}_{n-i-1} \right| \leq \left|\tilde{Y}_n\right| \leq$$

$$\leq (n+b)^{h-1}\beta_n\, n + \left| -\sum_{i=0}^{\infty} \beta_{i+1}\, h(n,i)\ \tilde{Y}_{n-i-1} \right|. \tag{4.20}$$

We now show that the positive quantity

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} \beta_{i+1} h(n,i) \tag{4.21}$$

is bounded and bounded away from zero. In fact

$$\sum_{i=0}^{n-1} \beta_{i+1} h(n,i) \sim \sum_{i=0}^{n-1} \frac{(n+b)^{h-1}}{(n-i-1+b)^{h-1}(i+1)^h} =$$

$$= \sum_{i=0}^{\frac{n}{2}} \frac{(n+b)^{h-1}}{(n-i-1+b)^{h-1}(i+1)^h} + \sum_{i=\frac{n}{2}+1}^{n-1} \frac{(n+b)^{h-1}}{(n-i-1+b)^{h-1}(i+1)^h} \leq$$

$$\leq (\frac{1}{2}+\frac{b}{n})^{1-h} \sum_{i=0}^{\frac{n}{2}} \frac{1}{(i+1)^h} + (\frac{n}{2})^{-h} \sum_{i=\frac{n}{2}+1}^{n-1} \left(1-\frac{i}{n}+\frac{b}{n}\right)^{1-h},$$

where the first term is bounded as $h > 1$, while

$$(\frac{n}{2})^{-h} \sum_{i=\frac{n}{2}+1}^{n-1} \left(1-\frac{i}{n}+\frac{b}{n}\right)^{1-h} \sim (\frac{n}{2})^{-h} \left[\sum_{i=0}^{\frac{n}{2}} \left(\frac{i}{n}\right)^{1-h} + B \right] \sim (\frac{n}{2})^{-h} \frac{n^{2-h}}{n^{1-h}},$$

so that it goes to zero when $n \uparrow \infty$. We can then follow the same procedure as in the proof of Lemma 5 to show that under the assumption that $\lim_{n\to\infty} \tilde{Y}_n = 0$, $\forall \epsilon, \exists N^* \ni \forall n > N^*$

$$(n+b)^{h-1}\beta_n n - \epsilon \le \left|\tilde{Y}_n\right| \le (n+b)^{h-1}\beta_n n + \epsilon,$$

that is a contradiction.

Finally we prove the last assertion of the proposition. By Lemma 3 we know that if a steady state distribution π^* does exist, it must have the property that its element in place i, π_i^*, is such that $\pi_i^* = \pi_0^* \beta_i$ where $\beta_i \sim n^{-h}$ as $n \uparrow \infty$. Q.E.D.

Proof of Proposition 7 (Killing irreducibility) For aggregate output to exhibit an order of integration greater or equal than one it must be that as $\lim_{n \longrightarrow \infty} Y_n = a \neq 0$ possibly unbounded, where Y_n is the first difference of $R_0^{(n)}$, where $R_0^{(n)}$ indicates the expected position after n iterations of a unit starting in state 0. It follows by Lemma 5 that if a steady state distribution does exist the order of integration must be strictly less than one. To prove the second part of the proposition, we recall that for aggregate output to exhibit an order of integration greater than one it must be that the Wold coefficients $\phi_n \sim n^{d-1}$ where $d > 1$, so that they are eventually unbounded. This, however, can not be the case. In fact the Wold coefficients $\phi_n \sim Y_n$, but the $\lim_{n \longrightarrow \infty} Y_n$ is greater or equal than zero by recurrency and is less than one as the first difference of $R_0^{(n)}$ is uniformly bounded above by one. Q.E.D.

Proof of Proposition 8 (Random walk and Gibrat's law) For aggregate output to be a random walk it must be that the Wold coefficients $\phi_n = \gamma\delta' P^n O$ be equal to a constant for all n. If we indicate with $R_i^{(n)}$ the generic element in place i of the vector $R^{(n)} = P^n O$, we note that a necessary and sufficient condition for aggregate output to be a random walk is that

$$R_i^{(n)} = i + c_n, \; \forall i, \forall n, \tag{4.22}$$

where c_n is function of n only. It can be checked that

$$R_i^{(n)} = (1-p_i)R_0^{(n-1)} + p_i R_{i+1}^{(n-1)}, \; \textit{where } R^{(1)} = PO, \; \forall i. \tag{4.23}$$

Putting together equation (4.22) and equation (4.23) we obtain that

$$p_i = \frac{(c_n - c_{n-1} + i)}{i+1},$$

while c_n is such that $c_n = c_{n-1} + p_0$. This implies that $p_i = (1 - \frac{1-p_0}{i+1})$, so that $g_i = \gamma(1-p_0)$, where $0 < p_0 < 1$ and $h = 1 - p_0$. Q.E.D.

4.8.3 Proofs of results in section 5

Proof of Proposition 9 (The reallocation structure matters) We are assuming that the transmission mechanism P is reducible where the recurrent class is identified by the first s technological states. Following the same steps as in the proof of Proposition 3 it can be shown that $P^n O$ has the structure

$$P^n O = \begin{bmatrix} c_0^n \\ c_1^n \\ \vdots \\ c_s^n \\ c_{s+1}^n + p_{s+1}p_{s+2}\cdots p_{s+n}\,(s+1+n) \\ c_{s+2}^n + p_{s+2}p_{s+3}\cdots p_{s+n+1}\,(s+2+n) \\ \vdots \end{bmatrix}, \tag{4.24}$$

where c_i^n's are such that $c_i^n \leq \rho^n c_i$, where $0 \leq \rho < 1$ while c_i's are properly defined positive bounded quantities such that, $\forall i$, $c_i \leq s$, see for example equation (4.12). Given Assumption 4 $\delta = \sum_{i=1}^{\infty} w^i e_i$, where, $\forall i$, $0 \leq w^i \in \Re$, and $\sum_{i=1}^{\infty} w^i < 1$. It follows that

$$\begin{aligned} \delta' P^n O \quad &\sim \quad \rho^n \sum_{i=1}^{\infty} w^i (c_i - c_0) + \tilde{\beta}_{s+1\,n} \sum_{i=1}^{\infty} \frac{s+i}{\tilde{\beta}_{s+1\,s+i}} \tilde{w}^{s+i} + \\ &+ n\, \tilde{\beta}_{s+1\,n} \sum_{i=1}^{\infty} \frac{\tilde{w}^{s+i}}{\tilde{\beta}_{s+1\,s+i}}, \end{aligned} \tag{4.25}$$

where $\tilde{\beta}_{ji} = \prod_{k=0}^{i-1} p_{j+k}$ after replacing $p_i = 1$ for those p_i's equal to zero, where $\tilde{w}^{s+i}$ is equal to w^{s+i} if $\min_{0 \leq j < n} p_{s+i+j} \neq o$, equal to zero otherwise while the quantities $\sum_{i=1}^{\infty} \frac{s+i}{\tilde{\beta}_{s+1\,s+i}} \tilde{w}^{s+i}$ and $\sum_{i=1}^{\infty} \frac{\tilde{w}^{s+i}}{\tilde{\beta}_{s+1\,s+i}}$ are both bounded by Assumption 3. It follows that the reallocation structure might matter. In fact, a reallocation structure that reallocates just inside the recurrent class, that is $w^i = 0, \forall i \geq s$, generate Wold coefficients, $\phi_n = \gamma\, \delta' P^n O$, that will decay exponentially. A reallocation structure that reallocates also out of the recurrent class, that is $\exists\, w^i > 0$, for some $i \geq s$ will generate rates of decaying that will be equal to the minimum between the exponential one and that of $n\, \beta_{s+1\,n}$ if $p_i \neq 0$, $\forall i > s$, where $\beta_{s+1\,n} = \prod_{k=0}^{n-1} p_{s+1+k}$. Q.E.D.

Proof of Proposition 10 (Robust typical spectral shapes) In order to prove the assertion we want to show that $\phi_n = \gamma \delta' P^n O \sim n^{d-1} = n^{1-h}$, where d is the order of integration of aggregate output while the symbol " $\sim$ " indicates that the ratio of left- and right-hand sides tends to a bounded quantity as $n \uparrow \infty$. As the transmission mechanism is reducible the representation (4.25) for $\delta' P^n O$ holds. Following the same reasoning as the one in Proposition 6, it can be shown that under the stated assumptions

$\beta_{in} \sim n^{-h}$ for big i. This implies that a reallocation structure of the cleansing type will generate Wold coefficients, $\phi_n \sim n^{1-h}$. Q.E.D.

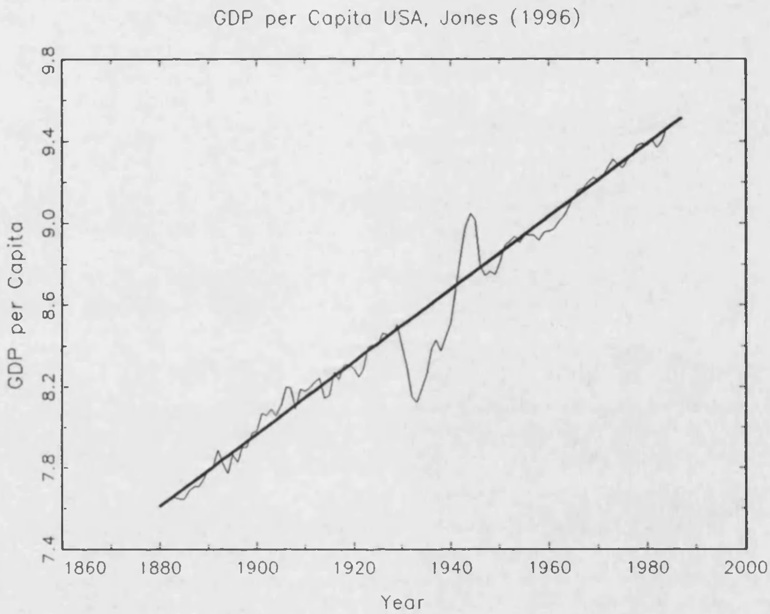

Figure 4.2: Per Capita GDP in the United States, 1880-1987 (Natural Logarithm). The Data are from Maddison (1982,1989) as used in Jones (1996). The solid bold line represents the time trend calculated using data only from 1880 to 1929.

Bibliography

[1] Acemoglu, D. and Scott, A. (1995), "Asymmetric Business Cycles: Theory and Time Series Evidence", *MIT Working Chapter*, 95-24, August.

[2] Acemoglu, D. (1996), "A Micro-foundation for Social Increasing Returns in Human Capital Accumulation", *Quarterly Journal of Economics,* Vol. 111(3), August, pp.779-804.

[3] Adams, J. (1990),"Fundamental Stocks of Knowledge and Productivity Growth", *Journal of Political Economy,* Vol. 98, August, pp. 673-702.

[4] Adams, J. (1993),"Science, R&D, and Invention Potential Recharge: U.S. Evidence", *American Economic Review Papers and Proceedings,* Vol. 83(2), May, pp. 458-462.

[5] Aghion, P. and Howitt, P. (1992), "A Model of Growth Through Creative Destruction", *Econometrica* Vol.60, March, pp.323-351.

[6] Aghion, P. and Howitt., P. (1994), "Growth and Unemployment", *Review of Economic Studies*, July, 61(3), pp. 477-494.

[7] Aghion, P. and Howitt., P. (1996), "On the Macroeconomic Effects of Major Technological Change", *Mimeo, London School of Economics.*

[8] Aghion, P. and Tirole, J. (1994), "Opening the Black Box of Innovation", *European Economic Review* 38, pp.701-710.

[9] Aghion, P. and Tirole, J. (1994), "The Management of Innovation", *Quarterly Journal of Economics*, pp.1185-1209.

[10] Aghion, P. and Saint-Paul, G. (1993), "Uncovering some Causal Relationships between Productivity Growth and the Structure of Economic Fluctuations: a Tentative Survey",*NBER Working Paper Series*, No. 4603, December.

[11] Aghion, P. and Gilles Saint-Paul, (1991), "On the Virtue of Bad Times: an Analysis of the Interaction between Economic Fluctuations and Productivity Growth", *CEPR Discussion Paper* No. 578, October.

[12] Andolfatto, D. (1996), "Business Cycles and Labour-Market Search", *American Economic Review, Vol.86,* NO.1, March., pp.112-132.

[13] Bagliano, F. Brandolini, A. and Dalmazzo, A. (1991) "The OECD-CEP Data Set", *Centre for Economic Performance Working Paper n.118,* June.

[14] Baily, M.N. Hulten, C. and Campbell, D. (1992), "Productivity Dynamics in Manufacturing Plants", *Brookings Papers on Economic Activity: Microeconomics*, pp. 187-249.

[15] Barro, J. (1991), "Economic Growth in a Cross Section of Countries", *Quarterly Journal of Economics*, 2, pp. 407-443.

[16] Barro, J. and Sala-I-Martin, X. (1991), "Convergence across States and Regions", *Brookings Paper on Economic Activity*, 1, pp. 107-182.

[17] Barro, J. and Sala-I-Martin, X. (1995), *Economic Growth*, McGraw-Hill, Advanced Series in Economics.

[18] Bartelsman, E. and Dhrymes, P. (1994), "Productivity Dynamics: US Manufacturing Plants, 1972-1986", *Division of Monetary Affairs Federal Reserve Board, Washington,* Finance and Economics Discussion Series no. 94-1, January.

[19] Baumol, W. (1990), "Entrepreneurship: Productive, Unproductive, and Destructive", *Journal of Political Economy,* Vol. 98(5), pp.893-921.

[20] Bean, C. (1990),"Endogenous Growth and the Procyclical Behaviour of Productivity", *European Economic Review* 34, pp.355-363.

[21] Bean, C. and Pissarides, C. (1993), "Unemployment, Consumption and Growth", *European Economic Review* 37(4), May, pp. 837-854.

[22] Benhabib, J., Rogerson, R. and Wright, R. (1991), "Homework in Macroeconomics: Household Production and Aggregate Fluctuations", *Journal of Political Economy,* Vol. 99(6), pp.1166-1187.

[23] Bentolila, S. and Bertola, G. (1990), "Firing Costs and Labour Demand: How Bad is Eurosclerosis?", *Review of Economic Studies*, 57, pp.381-402.

[24] Beran, J. (1994) *Statistics for Long Memory Processes*, Chapman and Hall.

[25] Bernard, A. and Durlauf, S. (1996), "Interpreting Tests of the Convergence Hypothesis", *Journal of Econometrics*, 71(1-2), pp. 161-173.

[26] Bertola, G. and Caballero, R. (1990), "Kinked Adjustment Costs and Aggregate Dynamics", in Blanchard, O.J. and Fischer, S. (eds) *NBER Macroeconomics Annual 1990*,(Cambridge (MA): MIT Press).

[27] Bertola, G. and Caballero, R. (1994a), "Irreversibility and Aggregate Investment", *Review of Economic Studies*, 61, pp.223-246.

[28] Bertola, G. (1994b), "Flexibility, Investment, and Growth", *Journal of Monetary Economics* 34 (2), October, pp. 215-38.

[29] Blanchard, O.J. and Quah, D. (1989), "The Dynamic Effects of Aggregate Demand and Supply Disturbances",*American Economic Review*, 79, pp. 655-673.

[30] Blanchard, O.J. and Diamond, P. (1989), "The Beveridge Curve", *Brookings papers on Economic Activity*, pp.1-60.

[31] Blanchard, O. (1990), "Why Does Money Affect Output? A Survey", in *Handbook of Monetary Economics* edited by Friedman B. and Frank H.,Volume 2, chapter 19, pp.779-835

[32] Bradford De Long, J. (1990), "Liquidation Cycles: Old-Fashioned Real Business Cycle Theory and the Great Depression", *NBER Working Paper Series* No. 3546, December.

[33] Brock, W. and Sayers, C. (1988), "Is the Business Cycle Characterized by Deterministic Chaos?", *Journal of Monetary Economics* 22, pp. 71-90.

[34] Brockwell P., and Davis, R. (1991), *Time Series: Theory and Methods*, Springer Verlag New York.

[35] Caballero, R. and Engel, E. M.R.A. (1991), "Dynamic (S,s) Economies", *Econometrica*, 59, pp. 1659-1686.

[36] Caballero,R. (1992), "A Fallacy of Composition", *American Economic Review,* December, pp.1279-1292.

[37] Caballero, R. and Engel, E. M.R.A. (1993), "Heterogeneity and Output Fluctuations in a Dynamic Menu-Cost Economy", *Review of Economic Studies*, 60, pp. 95-119.

[38] Caballero, R. (1993), "Unemployment, Consumption and Growth: Comment", *European Economic Review* 37(4), May, pp.854-861.

[39] Caballero, R. and Jaffe, A. (1993), "How High are the Giants Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a model of Economic Growth". In *NBER Macroeconomics Annual 1993*, edited by Blanchard, O. and Stanley, F. Cambridge Mass.: MIT Press.

[40] Caballero, R. and Hammour, M. (1994a), "On the Timing and Efficiency of Creative Destruction", *NBER Working Paper Series*, No 4768, June.

[41] Caballero, R. and Hammour, M. (1994b), "The Cleansing Effect of Recessions", *American Economic Review* Vol. 84 NO.5, December, pp. 1350-1368.

[42] Caballero, R. and Hammour, M. (1996), "The Fundamental Transformation in Macroeconomics", *NBER Working Paper n. 5471*, February.

[43] Campell, J. and Mankiw, N. (1987), "Are Output Fluctuations Transitory?", *Quarterly Journal of Economics*, 4, pp. 857-880.

[44] Chan, N. and Terrin, N. (1995), "Inference for Unstable Long-Memory Processes with Applications to Fractional Unit Root Autoregressions", *Annals of Statistics,* Vol. 23 No. 5, pp. 1662-1683.

[45] Chow, G. C. (1979),"Optimum Control of Stochastic Differential Equation Systems", *Journal of Economic Dynamics and Control* 1 1 pp. 143-175.

[46] Cochrane, J. (1988), "How Big is the Random Walk in GNP?", *Journal of Political Economy,* 96, pp. 893-920.

[47] Coe, T. and Helpman, E. (1993), "International *R&D* Spillovers", *NBER Working Paper n.4444*, August.

[48] Cooper,R. and John, A. (1988), "Coordinating Coordination Failures in Keynesian Models", *Quarterly Journal of Economics,* 13, pp. 441-465.

[49] Cornish, W.R. (1989), *Intellectual Property: patents, copyright, trade marks and allied rights,* Sweet & Maxwell, second edition.

[50] Davidson, J. (1994) *Stochastic Limit Theory,* Oxford University Press 1994.

[51] Davis, S. and Haltiwanger, J. (1990),"Gross Job Creation and Destruction: Microeconomic Evidence and Macroeconomics Implications", *NBER Macroeconomics Annual* , Vol.5,pp. 99-113.

[52] Davis, S. and Haltiwanger, J. (1992), "Gross Job Creation, Gross Job Destruction and Employment Reallocation", *Quarterly Journal of Economics* 107, pp. 819-63.

[53] Dickey, D.A. and Fuller, W. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of American Statistical Association* , 74, pp. 427-431.

[54] Diebold, F. and Rudebusch, G. (1991),"Long Memory and Persistence in Aggregate Output",*Journal of Monetary Economics,* 24, pp. 189-209.

[55] Diebold, F. and Rudebusch, G. (1989),"Long Memory and Persistence in Aggregate Output",*Journal of Monetary Economics,* 24, pp. 189-209.

[56] Diebold, F. and Senhadji, A. (1996),"Deterministic vs. Stochastic Trend in U.S. GNP, yet Again",*American Economic Review, 86*,pp. 1291-1298.

[57] Dunne, T. Roberts, M. and Samuelson, L. (1989) , "The Growth and Failure of U.S. Manufacturing Plants", *Quarterly Journal of Economics,* 104(4), November, pp. 671-98.

[58] Durlauf, S. (1991), "Multiple Equilibria and Persistence in Aggregate Fluctuations", *American Economic Review Papers and Proceedings,* May, Vol. 81 no.2, pp. 70-74.

[59] Evans, D.S. (1987), "Tests of Alternative Theories of Firm Growth", *Journal of Political Economy,* vol.87, no.4, pp.657-674.

[60] Evans, D. and Leighton, L. (1989), "Some Empirical Aspects of Entrepreneurship", *American Economic Review,* Vol. 79(3), June, pp. 519–535.

[61] Evans, D. and Jovanovic, B. (1989), "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints", *Journal of Political Economy,* Vol. 97(4), pp.808-827.

[62] Forni, M. and Reichlin, L. (1995), "Let's Get Real: a Dynamic factor analytical approach to Disaggregated Business Cycles", *CEPR Discussion Paper Series* no. 1244.

[63] Gali, J. and Hammour, M. (1991), "Long Run Effects of Business Cycles", *Mimeo,* (Columbia University, New York).

[64] Gali, J. (1996), "Technology, Employment and the Business Cycle: do technology shocks explain aggregate fluctuations?", *CEPR Discussion Paper Series* no. 1499.

[65] Geroski, P.A. and Walters, C.F. (1995), "Innovative Activity over The Business Cycle", *The Economic Journal, 105,* July, pp. 916-928.

[66] Geweke, J. and Porter Hudak, S. (1983), "The Estimation and Application of Long Memory Time Series Models", *Journal of Time Series Analisys*, 4, pp.221-238.

[67] Gibrat,R. (1931), "Les inégalités économiques applications: aux inégalités des richesses, a la concetration des entreprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle la loi de l'effet proportionel", *Paris: Librarie du Recueil Sirey.*

[68] Gil-Alana, L. and Robinson, P. M. (1997), "Testing Unit Root and other Non-Stationary Hypotheses in Macroeconomic Time Series", *Journal of Econometrics,* Vol.80-2, October, pp.223-239.

[69] Gourieroux, C. and Monfort, A. (1989), *Statistics and Econometric Models*, Cambridge Press.

[70] Griliches, Z. (1979), "Patents: Recent Trends and Puzzles", *Brookings Papers on Economic Activity, Microeconomics*, pp. 291-330.

[71] Griliches, Z. (1988), "Productivity Puzzles and R &D, Another Non-explanation", *Journal of Economic Perspectives*, 28, pp. 1661-1707.

[72] Griliches, Z. (1990), "Patent Statistics as Economic Indicators: a Survey", *Journal of Economic Literature*, 28,December, pp. 1661-1707.

[73] Granger, C.W.J. (1966), "The Typical Spectral Shape of an Economic Variable", *Econometrica* 34,no.1, pp. 150-161.

[74] Granger, C.W.J. (1980), "Long Memory Relationships and the aggregation of Dynamic Models", *Journal of Econometrics* 14, pp. 227-238.

[75] Granger, C.W.J. and Joyeux, R. (1980), "An Introduction to Long-Memory Time Series Models and Fractionally Differencing", *Journal of Time Series Analysis* Vol.1, no.1, pp. 15-29.

[76] Grossman, G. and Helpman, E. (1991), "Quality Ladder in the Theory of Growth", *Review of Economic Studies,* January, 43(1), pp.43-62.

[77] Hall, R. (1988), "The Relation between Price and Marginal Cost in U.S. Industry", *Journal of political Economy,* vol. 96-5, pp. 921-947.

[78] Hall, B. (1987), "The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector", *The Journal of Industrial Economics,* Vol. 35-4, June, pp. 583-606

[79] Harrod, R. (1939),"An Essay in Dynamic Theory", *The Economic Journal* 49, March, pp.14-33.

[80] Helpman, E. and Trajtenberg, M. (1994), "A Time to Sow and a Time to Reap: Growth Based on General Purpose Technologies", *NBER Working Paper Series,* No 4854, September.

[81] Hosios, A. (1990), "On the Efficiency of Matching and Related Models of Search and Unemployment", *The Review of Economic Studies*, 57,pp. 279-298.

[82] Hurwich, C.M. and Ray, B.K. (1995), "Estimation of the memory Parameter for Non-stationary or Non-invertible Fractionally Integrated Processes", *Journal of Time Series Analysis,* Vol.16, No.1, pp. 17-41.

[83] Jones,C. (1995a), "Time Series Tests of Endogenous Growth Models", *Quarterly Journal of Economics*,110, pp. 495-525.

[84] Jones,C. (1995b), "*R&D* Based Models of Economic Growth", *Journal of Political Economy*, 103-4, pp. 759-783.

[85] Jones, R. and Newman, G. (1995), "Adaptive Capital, Information Depreciation and Schumpeterian Growth", *Economic Journal 105,* July pp. 897-915.

[86] Jovanovic, B. and Nyarko, Y. (1996), "Learning by Doing and the Choice of Technology", *Econometrica*, Vol. 64, No. 6, November, pp. 1299-1310.

[87] Kalecki, M.(1945), "On the Gibrat Distribution", *Econometrica*, April, 13(2), pp.161-170.

[88] Karlin, S. and Taylor, H.M. (1975), *A First Course in Stochastic Processes*, Academic Press, Inc. (London) LTD.

[89] Karlin, S. and Taylor, H.M. (1981), *A Second Course in Stochastic Processes*, Academic Press, Inc. (London) LTD.

[90] Kemeny, J.G. Snell, J.L. and Knapp, A. W. (1966), *Denumerable Markov Chains*, The University Series in Higher Mathematics, D.Van Nostrand Company, Inc., Princeton New Jersey.

[91] Keynes, J.M. (1936), *The General Theory of Employment Interest and Money*, Reprinted Harbinger, Harcourt Brace and World 1964.

[92] Kortum, S. (1993), "Equilibrium R&D and the Decline in the Patent-R&*D* Ratio", *American Economic Review: Papers and Proceedings*,Vol. 83 2, pp. 450-457.

[93] Kortum, S. (1994), "A Model of Research, Patenting and Productivity Growth", *NBER Working Paper n. 4646*, February.

[94] Laing, D. Palivos, T. and Wang, P. (1995a), "Learning, Matching and Growth", *Review of Economic Studies*, vol. 62(1) n. 210, January, pp. 115-130.

[95] Laing, D. Palivos, T. and Wang, P. (1995b), "R&D in a Model of Search and Growth", *American Economic Rewiev Papers and Proceedings*, Vol. 85 n.2, May, pp291-295.

[96] Levin, A. and Lin, C.F. (1992), "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties",*Discussion Paper 92-92, UCSD.*

[97] Levine, R. and Renelt, D. (1992), "A Sensitivity Analysis of Cross-Country Growth Regressions", *American Economic Review*, Vol. 82, No.4, September.

[98] Lippi, M. and Reichlin, L. (1992),"Permanent and Transitory Components in Macroeconomics" in Thygesen,N., Velupilai,K. and Zambelli, S. eds. *Business Cycles:Theories, Evidence and Analisys,* Macmillan Academic and Professional Ltd.

[99] Lippi, M. and Reichlin, L. (1994),"Diffusion of Technical Change and the decomposition of output into Trend and Cycles", *Review of Economic Studies,* 61, pp.19-30.

[100] Lucas, R. (1988), "On the Mechanics of Economic Development", *Journal of Monetary Economics*, July, 22(1), pp.3-42.

[101] Malthus, T.R. (1798), *An Essay on the Principle of Population*, London, W.Pickering, 1986.

[102] Mankiw, N. Romer, R. and Weil, D. (1992),"A Contribution to the Empirics of Economic Growth",*Quarterly Journal of Economics,* 107-2, pp. 407-437.

[103] Mansfield, E. (1962), "Entry, Gibrat's Law, Innovation, and the Growth of Firms", *American Economic Review,* Vol.52-5, December, pp.1023-1051.

[104] Merz, M. (1995), "Search in the Labour Market and the Real Business Cycle ", *Journal of Monetary Economics,* 36, pp. 269-300.

[105] Mortensen, D. (1991), "The Discovery Process and Endogenous Growth", *Mimeo*, August, Northwestern University.

[106] Mortensen, D.T. and Pissarides, C. (1994), "Job Creation and Job Destruction in the Theory of Unemployment", *Review of Economic Studies*, 61, July, 397-415.

[107] Mortensen, D.T. and Pissarides, C. (1995), "Technological Progress, Job Creation and Job Destruction", *Centre for Economic Performance, Discussion Paper* No. 264, September.

[108] Murphy, K. Shleifer, A. and Vishny, R. (1991), "The Allocation of Talent: Implications for Growth", Vol.56, pp.503-530.

[109] Nelson, C. and Plosser, C. (1982), "Trends and Randon Walks in Macroeconomic Time Series: some Evidence and Implications", *Journal of Monetary Economics*,10, pp. 139-162.

[110] Nickell, S. Wadhwani, S. and Wall, M. (1992), "Productivity Growth in U.K. Companies, 1975-1986", *European Economic Review*, 36, pp. 1055-1091.

[111] Pissarides, C. A. (1985), "Short-Run Equilibrium Dynamics of Unemployment Vacancies and Real Wages", *American Economic Review*, 75, pp. 676-90.

[112] Pissarides, C. A. (1990), *Equilibrium Unemployment Theory*, Oxford: Basil Blackwell 1990.

[113] Priestley, M. (1965), "Evolutionary Spectra and Non-Stationary Processes", *Journal of the Royal Statistical Association,* No. 27, pp. 204-237.

[114] Quah, D. (1992a), "The relative Importance of Permanent and Transitory Components: Identification and some Theoretical Bounds", *Econometrica 60,* pp. 107-118.

[115] Quah, D. (1992b), "International Patterns of Growth: Persistence in Cross-Country Disparities", *Working Paper, London School of Economics.*

[116] Quah, D. (1993a), "Empirical Cross-Section Dynamics in Economic Growth", *European Economic Review,* 37-2/3, pp. 426-434.

[117] Quah, D. (1993b), "Galton's fallacy and tests of the convergence hypothesis", *Scandinavian Journal of Economics,* 95(4), December, pp. 427-443.

[118] Quah, D. (1995), "Empirics for Economic Growth and Convergence ", *Discussion Paper 253, Centre for Economic Performance* .

[119] Ramey, G. and Ramey, V. (1991), "Technology Commitment and the Cost of Economic Fluctuations", *NBER Working Paper Series*, no 3755, June.

[120] Ramey, G. and Ramey, V. (1994), "Cross-Country Evidence on the Link between Volatility and Growth", *American Economic Review,* December, Vol. 85 no. 5, pp.1138-1151.

[121] Ricardo, D. (1817), *On the Principles of Political Economy and Taxation*, Cambridge, Cambridge University Press, 1951.

[122] Rivera-Batiz, L. and Romer, P.M. (1991), "Economic Integration and Endogeneous Growth", *Quarterly Journal of Economics*, 106, 2(May), pp.531-555.

[123] Robinson, P.M. (1978),"Statistical Inference for a Random Coefficient Autoregressive Model", *Scandinavian Journal of Statistics* 5, pp. 163-168.

[124] Robinson, P.M. (1986),"On a Model for a Time Series of Cross-Sections", *Journal of Applied Probability*,23A, pp. 113-126.

[125] Robinson, P. (1994), "Time Series with Strong Dependence", in *Advances in Econometrics, Sixth World Congress*, ed. by C.A. Sims, Vol. 1, pp. 47- 95, Cambridge University Press.

[126] Robinson, P.M. (1995), "Log-Periodogram Regression of Time Series with Long Range Dependence", *Annals of Statistics* , 23-3, pp. 1048-1072.

[127] Robinson, P.M. and Marinucci, D. (1997), "Semiparametric Frequency-Domain Analysis of Fractional Cointegration", *Mimeo, London School of Economics.*

[128] Romer, P. (1986), "Increasing Returns and Long Run Growth", *Journal of Political Economy*, October, 94(5), pp. 676-90.

[129] Romer, P. (1990a), "Are Non Convexities Important for Understanding Growth?", *NBER Working Paper*, no. 3271, February.

[130] Romer, P. (1990b), "Endogeneous Technological Change", *Journal of Political Economy*, 98(5), pp. 71-102.

[131] Rosen, S. (1981), "The Economics of Superstars", *American Economic Review,* Vol. 71(5), December, pp.845-858.

[132] Rotemberg, J. and Woodford, M. (1996), "Real-Business-Cycle Models and the Forecastable Movements in Output, Hours, and Consumption", *American Economic Review,* Vol. 86-1, March, pp.71-89.

[133] Rudebusch, G. (1993), "The Uncertain Unit Root in Real GNP", *American Economic Review,* 83, pp. 264-272.

[134] Rudin, W. (1973), *Functional Analysis*, McGraw-Hill New York.

[135] Saint-Paul, G. (1993),"Productivity Growth and the Structure of the Business Cycle", *European Economic Review* 37, pp.861-890.

[136] Schankerman, M. and Pakes, A. (1986), "Estimates of the value of Patent Rights in European Countries during the post-1950 Period", *The Economic Journal*, 96, December, pp. 1052-1076.

[137] Schumpeter, J. (1934), "Depressions", in Douglas Brown *et al. The Economics of the Recovery Program*, New York:McGraw-Hill, pp.3-21.

[138] Schumpeter, J. (1939), *Business Cycles: a Theoretical, Historical and Statistical Analysis of the Capitalist Process*, New York McGraw-Hill.

[139] Schumpeter, J. A. (1943), "Capitalism in the Postwar World", in *Postwar Economic Problems*, Seymour E. Harris, ed., New York:McGraw-Hill, pp. 113-126.

[140] Schumpeter, J. A. (1946), "Capitalism", in *Encyclopedia Britannica* , Vol. IV, pp. 801-807.

[141] Schumpeter, J. A. (1947), "The Creative Response in Economic History", *Journal of Economic History,* November, pp. 149-159.

[142] Schumpeter, J.A. (1949), "Economic Theory and Entrepreneurial History", in *Change and the Entrepreneur.* Prepared by the Research Center in Entrepreneurial History, Harvard University, Cambridge, Mass.: Harvard University Press, pp.63-84.

[143] Seneta, E. (1976), *Regularly Varying Functions,* Springer, Berlin.

[144] Shleifer, A. (1986), "Implementation Cycles", *Journal of Political Economy,* 94, pp.1163-1190.

[145] Solow, R. (1960), "Investment and Technical Progress", in *Mathematical Methods in the Social Sciences, 1959,* edited by Arrow, K.J., Karlin, S. and Suppes P., Stanford University Press, Stanford California.

[146] Sowell, F. (1990), "The Fractional Unit Root Distribution", *Econometrica,* 58-2, pp. 495-505.

[147] Stiglitz, J. (1993), "Endogenous Growth and Cycles", *NBER Working Paper Series,* No 4286, March.

[148] Stokey, N. Lucas, R. and Prescott, E. (1988), *Recursive Methods in Economic Dynamics*, Cambridge MA, Harvard University Press.

[149] Summers, R. and Heston, A. (1991), "The Penn World Table (Mark 5): An Expanded set of International Comparisons, 1950-1988", *Quarterly Journal of Economics*, pp. 327-368.

[150] Sutton, J. (1992), *Sunk Costs and Market Structure* , The MIT Press.

[151] Sutton, J. (1995), "The Size Distribution of Business Part I: a Benchmark Case", *The Economics of Industry Group, Discussion Paper Series EI/9,* December.

[152] Sutton, J. (1996), "Gibrat's Legacy", *The Economics of Industry Group, Discussion Paper Series EI/14,* October.

[153] Tirole, J. (1988), *The Theory of Industrial Organization*, Cambridge, MA: M.I.T. Press.

[154] Velasco, C. (1996), "Non-Stationary Log-Periodogram Regression", *Mimeo,London School of Economics.*

[155] Yong, C. (1974), *Trigonometric Series*, Cambridge University Press.

[156] Young, A. (1995a), "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience", *Quarterly Journal of Economics*, August, pp.641-680.

[157] Young, A. (1995b), "Growth Without Scale Effects", *NBER Working Paper n. 5211*, August.