# Cross-Sectional Heterogeneity and the Persistence of Aggregate Fluctuations

Claudio Michelacci[1]
*CEMFI* [2]

First Preliminary Draft: December 1996
This Version: November 1998

[1]I would like to thank Cristopher Bliss, Ricardo Caballero, Liudas Giraitis, Marco Lippi, Domenico Marinucci, Chris Pissarides, Peter M. Robinson and Paolo Zaffaroni for very useful conversations and suggestions. The usual disclaimer applies.

[2]Correspondence to: Claudio Michelacci, CEMFI, Casado del Alisal 5, 28014 Madrid, Spain, Tel: ++34-91-4290551, Fax: ++34-91-4291056, Email: c.michelacci@cemfi.es.

## Abstract

It is well known from time series analysis that shocks to aggregate output have very persistent effects. This paper argues that the relation between the expected growth rate of a firm and its size provides a microfoundation for such aggregate persistence. The empirical evidence indicates that small firms grow faster than big ones. If this is true, a shock that reallocates units across sizes will be absorbed, yet at very low decreasing rates. Once the shock hits the system, firms are reallocated across sizes. If small firms grows faster than big ones, the shock will then be absorbed. However, fast growing small firms eventually become big and grow as big firms. Thus the number of small firms shrinks over time as well as the rate at which the shock is absorbed. This transmision mechanism reconciles the micro evidence with the observed degree of aggregate persistence. It requires changes in neither the number of firms in the market nor the rate of technological progress. It is merely the result of the cross-sectional heterogeneity that we observe in real economies.



*Keywords:* Persistence, Gibrat´s law, cross-sectional heterogeneity, fractional integration, vintage model.

*JEL Classification:* C43, E1, E32, L11.

# 1 Introduction

The time series of (detrended) aggregate output exhibits considerable persistence. Indeed, Nelson and Plosser (1982) have argued that GDP exhibits a unit root, and therefore that temporary shocks have permanent effect on the level of output. More generally, persistence is taken to mean that aggregate shocks propagate at very low rates. Many studies have argued about the exact degree of aggregate persistence as well as about its driving force. But what type of firm behavior is consistent with the observed degree of aggregate persistence? This is the question addressed in this paper.

The main claim of the paper is that the observed empirical relation between the expected growth rate of a firm and its size provides a microfoundation for aggregate persistence. Gibrat (1931) first investigated the relationship between expected growth rate and firm size measured by either sales, employment or assets. He claimed the existence of a law, from then on called Gibrat's, according to which the expected growth rate of a firm is independent of its size[1]. Recent, more comprehensive studies, however, question Gibrat´s law and show that small firms tend to have higher and more variable growth rates[2].

At first suppose that Gibrat's law holds. If this is true, an aggregate shock that reallocates units across sizes has a permanent effect on the level of output, inducing a unit root in its time series formulation[3]. Once the shock hits the system, firms are reallocated across sizes. Then, given Gibrat's law, they keep growing at the same rate, perpetuating forever the impact effect of the shock. On the contrary, in a world where Gibrat's law fails and small firms grow faster than big ones, the same shock will be absorbed, yet at very low decreasing rates. Fast growing small firms eventually become

[1] The early literature focused mainly on big and listed firms. See Sutton (1996) for a survey on this debate.

[2] See Mansfield (1962), Hall (1987), Evans (1987) and Dunne, Roberts and Samuelson (1989). See also Davis, Haltiwanger and Schuh (1993) for empirical evidence to the contrary. I rationalize this discrepancy in section 5.

[3] 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". A distinctive feature of a random walk is indeed that its variance is a linear function of time.

big and grow as big firms. Thus the number of small firms shrinks over time as well as the rate at which the shock is absorbed. If we keep the empirical findings on the relation between the expected growth rate of a firm and its size as maintained assumptions, we conclude that the persistence of aggregate fluctuations is very high, that shocks are absorbed and that the rate of absorption is decreasing over the adjustment process.

To give economic content to the claim and explore its theoretical implications we consider a model and a measure of aggregate persistence. We analyze 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 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 using the same vintage can end up adopting different technologies. This is now a popular and plausible way of modelling the heterogeneity of an economic system[4]. 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.

In the model firms using vintages far away from (close to) the technological frontier are small (big). Some assumptions on factor allocation are required to link productivity to size measured by either sales, employment or assets. In general, if factor markets are not segmented and productivity increases the marginal revenue of each factor, a productivity ranking corresponds one for one to a size ranking[5].

We then introduce a measure of persistence. It is taken from time series econometrics and is based on the notions of *long memory* and order of integration of a stochastic process[6]. In fact formal empirical investigation has concluded that the low frequency behavior of aggregate time series is the

[4]See Baily, Hulten and Campbell (1992), Caballero and Hammour (1994, 1996), Aghion and Howitt (1994) and Mortensen and Pissarides (1998).

[5]For example, Baily, Hulten and Campbell (1992) and Bartelsman and Dhrymes (1994) find that employment size and productivity are positively correlated.

[6]See Robinson (1994) for a survey on the topic.

result of long memory processes in which the impact of shocks vanishes at a very slow hyperbolic rate[7]. 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 propagate at constant rates. That is, shocks either have permanent effects or vanish at the usual exponential rate. Long memory implies, instead, that shocks propagate at decreasing rather than constant rates and that the rate of absorption of the shock at each stage $n$ of the adjustment process is a decreasing function of $n$.

We show that our model is able to replicate the observed degree of aggregate persistence. What drives the result is the process of ongoing churning and catching up that takes place in the model as well as in the real economy. Once the shock hits the system, firms are reallocated across sizes (vintages). If small firms grow faster than big ones, the shock will be absorbed. However, fast growing small firms eventually become big and grow as big firms. The shock will then be absorbed, yet at very low decreasing rates thus replicating the long memory feature of the data.

The model says also something about the driving force of aggregate persistence. The Real Business Cycle tradition has often argued it is technology[8]. That means either that the aggregate shock itself is a technological shock with a sufficient amount of persistence or that the shock exhibits persistence because it directly affects technology. Neither is the case in the model analysed in this paper. The aggregate shocks just alter the opportunity cost of firms and so they can be read as either productivity or demand shocks. The shocks affect neither the number of firms in the market nor the rate of technological progress. Any persistence can therefore be attributed to the cross-sectional heterogeneity generated by the model.

The main contribution of the paper 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 provide micro-

[7]See Diebold and Rudebusch (1989), Gil-Alana and Robinson (1997) and Michelacci and Zaffaroni (1998).

[8]See for example Nelson and Plosser (1982), Rotemberg and Woodford (1996) and Gali (1996).

foundations of macroeconomics solving explicit aggregation problems[9]. The other has analysed firm dynamics, in particular the relation between growth and firm size. This paper notes that the two independent strands of research have important implications for the low frequency behavior of aggregate time series once a standard vintage model is used to combine them. This approach is able to reconcile the macro and micro evidence. Moreover, models which do not deal explicitly with cross-sectional heterogeneity seem incapable of replicating the observed degree of aggregate persistence. Thus, the paper concludes that the process of ongoing churning and catching-up that takes place in the economy is a key factor in explaining aggregate persistence.

The remainder of the paper is divided into 4 sections. Section 2 introduces and justifies our metrics for aggregate persistence. 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 introduces an aggregate shock in the model and generalizes some results in the literature on irreversibilities and $Ss$ adjustment processes[10]. Section 4 shows that the model can replicate the observed degree of aggregate persistence. Section 5 discusses the roles of each assumption while section 6 concludes. The appendix contains the derivation of most of the results contained in the paper.

## 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 favor of the typical spectral shape of an economic variable observed by Granger (1966). We draw on this empirical evidence to build a measure of aggregate persistence based on the notion of order of integration of a time series.

[9] See for example Bertola and Caballero (1990, 1994), Caballero and Engel (1991, 1993, 1994) and Caballero (1992) on the theoretical side and on the empirical side Davis and Haltiwanger (1990, 1992), Davis, Haltiwanger and Schuh (1993) and Caballero, Engel and Haltiwanger (1997).

[10] See for example Bertola and Caballero (1990, 1994), Caballero and Engel (1990) and Caballero (1992).

## 2.1 The Frequency Domain Approach to Time Series Analysis

It is sometimes useful to decompose the dynamics of a time series into different periodic components. Consider for example the spectral representation of a time series $X_t$

$$X_t = \int_{-\pi}^{\pi} e^{it\theta}\, dZ(\theta,\omega) = \int_{-\pi}^{\pi} [cos(t\theta)\ + i\ sin(t\theta)]\ dZ(\theta,\omega), \tag{1}$$

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 the *spectral density* of the process. Time series analysis based on representation (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 (1) is a very general one. For instance, Cramer's theorem[11] guarantees that a spectral representation like the one in (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 $x_t$, where $t$ runs from 1 to $T$, the spectral density $dF(\theta)$ is usually estimated through the periodogram $I(\theta)$ (or some function thereof) defined as the modulus of the discrete Fourier transform of the observations[12]:

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

Granger (1966) noted that the estimated spectra of most detrended economic variables had a shape that he defined to be as *typical.* It is a monotonically decreasing function with a very pronounced peak in the neighborhood of the zero frequency. For instance, Figure 1A shows that the level of the US detrended logged GDP per capita calculated over the period 1870-1994 exhibits the typical spectral shape identified by Granger. Frequency domain

[11] See for example Brockwell and Davis (1991).

[12] In the case of a stationary process the periodogram is the sample analogue of the theoretical spectral density $dF(\theta)$. Therefore the periodogram has a simple method of moments interpretation.

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.

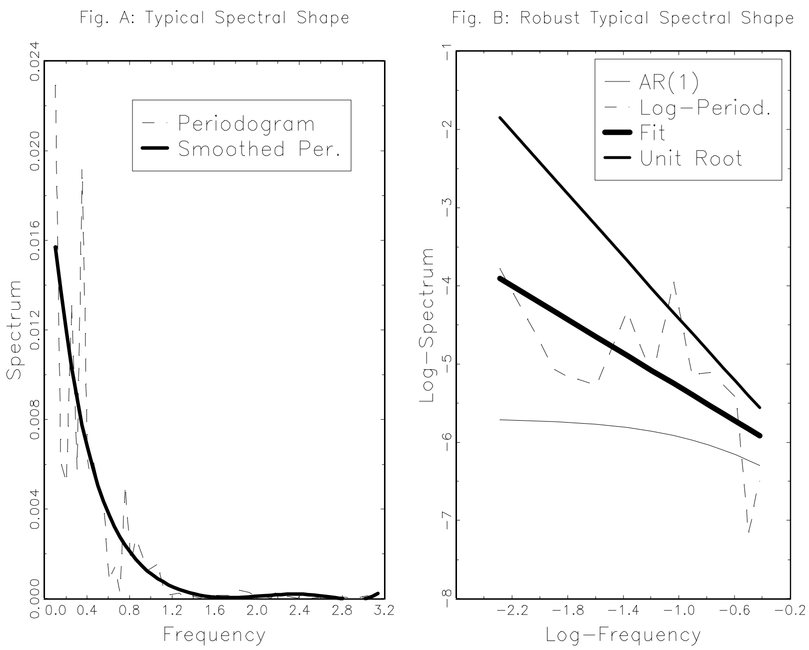


Figure 1: **Robust Typical Spectral Shapes (US Per Capita GDP, 1880-1994).** The figure shows the periodogram of the US linearly detrended (OLS) logged per capita GDP (1870-1994). The periodogram in figure A is smoothed making use of an 8th order polynomial. The data are taken from Maddison (1995). The theoretical spectrum for the AR(1) process is obtained using a coefficient of first order autocorrelation equal to 0.5.

A class of spectral density functions able to match exactly the spectral shape[13] that arises in the real world is given by

[13]The exact relation between the spectral density, $dF(\theta)$, and the periodogram (2) for

$$f(\theta) = dF(\theta) \;\sim\; g(\theta)\,\theta^{-2d}, \;\; as\;\theta \to 0^+, \tag{3}$$

where '$\sim$' indicates that the ratio of the left- and right-hand side tends to a bounded quantity, $d$ is a non-negative constant and $g(\theta)$ is a bounded function bounded away from zero in a neighborhood of the origin[14]. The parameter $d$ represents the order of integration of the time series. If it is greater than zero, the time series exhibits *long memory,* while it exhibits *weak memory* if the parameter is equal to 0. The parameter $d$ measures the rate of divergence of the spectrum around zero frequency and thus it measures how 'typical' is the spectral shape. $d$ is our measure of persistence: the higher the value of $d$, the higher is the weight of the components with very long periods and thus the higher is the persistence of the process.

A time domain representation of the time series $X_t$, $t \geq 0$, corresponding to equation (3) is given by the Wold representation[15]

$$X_t = X_0 + \gamma\,t + \sum_{n=0}^{t} \phi_n\,\epsilon_{t-n}, \tag{4}$$

with Wold coefficients $\phi_n = \tilde{\phi}_n + d\,n^{d-1} + o\left(\tilde{\phi}_n\right)$, where $\tilde{\phi}_n$ is a function converging to zero at a rate at least as high as the exponential one (that is $\left|\tilde{\phi}_n\right| \leq K\rho^n$, $K$ a bounded quantity, $0 \leq \rho < 1$), $d$ is the order of integration of the time series[16] while $o\left(\tilde{\phi}_n\right)$ indicates a quantity of lower order than $\tilde{\phi}_n$,

non stationary processes ($d \geq \frac{1}{2}$ see below) is a topic that goes beyond the scope of this paper. 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). A possible solution would therefore consist of defining the theoretical spectral density through the periodogram, as in Hurwich and Ray (1994). This is implicitly the approach pursued here, where to safeguard theoretical rigour we speak of empirical spectral shapes rather than spectra.

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

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

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

[16] If we allowed for $g(\theta)$ to be a slowly varying function as in footnote (15), the Wold

that is $\lim_{n\to\infty}\frac{o(\tilde{\phi}_n)}{\tilde{\phi}_n}=0$. The Wold coefficient $\phi_n$ gauges the fraction of the shock $\epsilon_{t-n}$, $n$ periods ahead, which has not yet been absorbed. Therefore, the rate of decay of the Wold coefficients measures the persistence of shocks.

A standard trend stationary process with $ARMA$ disturbance exhibits Wold coefficients $\phi_n$'s decaying no more slowly than an exponential rate. This implies that the persistence is low and that the parameter of fractional integration is equal to zero. This weak memory property of $ARMA$ processes shows up in the frequency domain under the form of a flat spectral shape around zero frequency (see for example Figure 1B for the $AR(1)$ case). In a process with a unit root, temporary shocks have permanent effects on the level of the time series. That means that the Wold coefficients asymptotically approach a constant and $d$ is equal to 1. Thus the spectral shape of a unit root is typical yet particular as it exhibits a very specific rate of divergence around zero frequency. This set of considerations shows how standard $ARIMA$ processes can not generate *arbitrary* typical spectral shapes, because they generate shapes with rate of divergence equal to either that of the unit root or the flat one.

In fact, in the $ARIMA$ processes shocks can propagate only at constant (or increasing) rates. For example, an exponential rate of absorption, $\phi_n\sim\rho^n$, means that, at each stage $n$ of the adjustment process, a constant fraction $1-\rho$ of the amount of shock still unabsorbed will be absorbed at stage $n+1$. A unit root is just a particular case of propagation at constant rates, one where asymptotically shocks are not absorbed at all so that $1-\rho=0$. In contrast, long memory allows for the possibility of decreasing rates of absorption. In fact when the rate of absorption is hyperbolic, the Wold coefficients $\phi_n$ behave like $n^{d-1}\sim(1-\frac{1-d}{1})(1-\frac{1-d}{2})\;\cdots\;(1-\frac{1-d}{n})$[17]. That means that at each stage $n$ of the adjustment process a fraction $\frac{1-d}{n+1}$ of the still unabsorbed part of the shock will be absorbed by stage $n+1$, that is the rate of absorption of the shock is a decreasing function of $n$. In doing so long memory allows for a

coefficients could behave as $\phi_n=\tilde{\phi}_n+n^{d-1}+o\left(\tilde{\phi}_n\right)$. If so, the case $d=0$ would correspond to Wold coefficients $\phi_n$ decaying as $\frac{1}{n}$. This would imply that the covariances $\mu_\tau$ would behave as $\frac{\ln(\tau)}{\tau}$ for large $\tau$, so that the spectral density $dF(\theta)$ behaves as the slowly varying function $(\ln\theta)^2$ as $\theta\to0^+$, see Granger and Joyeux (1980).

[17]See for example equation (29) in the appendix for a derivation of the result.

variety of intermediate cases, and thus smoothly bridges the gap between the spectral shape with the particular slope of the unit root and the flat spectral shape associated with the absolute lack of memory.

Given the representation (3), a 'reasonable'[18] way of estimating the 'typicality' of the spectrum through 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

$$\ln[I(\theta)] = const. - 2d \ln \theta, \ as \ \theta \in [\theta_l, \theta_u]^{19}. \tag{5}$$

Diebold and Rudebusch (1989) and Michelacci and Zaffaroni (1998) ran regression (5) for the GDP per capita for a set of different OECD economies and show that the parameter $d$, the rate of divergence of the periodogram, is between zero and one. That means that the degree of aggregate persistence is lower than that associated with a unit root but greater than in the *ARMA weak memory case* (see Figure 1B)[20]. The estimated value of $d$ based on the log-periodogram regression (5) is reported in table 1[21].

This suggests that the real GDP per capita of the US is characterized by a parameter of fractional integration greater than zero and (probably) less than one. That some form of very slow mean reversion actually takes place in the data is also confirmed by time domain observation. Jones (1995) shows how a time trend calculated using data only from 1880 to 1929 forecasts extremely well the current level of GDP of the US economy (see Figure 2). This implies that the new information delivered by the Wold innovations $\epsilon_t$

[18] Robinson (1995) proves consistency and asymptotic normality of this estimator originally proposed by Geweke and Porter Hudak (1983). Giraitis, Robinson and Samarov (1997) proves that the estimator is rate optimal.

[19] 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 intervals $[0, \theta_u]$ and $[0, \theta_l)$. The number of Fourier frequencies contained in the first interval is called the trimming coefficient, while the number of Fourier frequencies contained in the second is called the bandwidth.

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

[21] Velasco (1996) shows that the Robinson (1995) log-periodogram estimator is consistent and normal even for non stationary $d$. 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 upon request.

| $T^\alpha$ $\alpha =$ | 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 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 $T^\alpha$, where $T$ is the sample size. 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 Michelacci and Zaffaroni (1998). For a derivation of the theoretical properties of this estimator, originally proposed by Geweke and Porter Hudak (1983), see Robinson (1995). For an analysis of the properties of the estimator in the non stationary case see Velasco (1996).

is irrelevant for forecasting on very long horizons and is incompatible with a unit root in output[22].

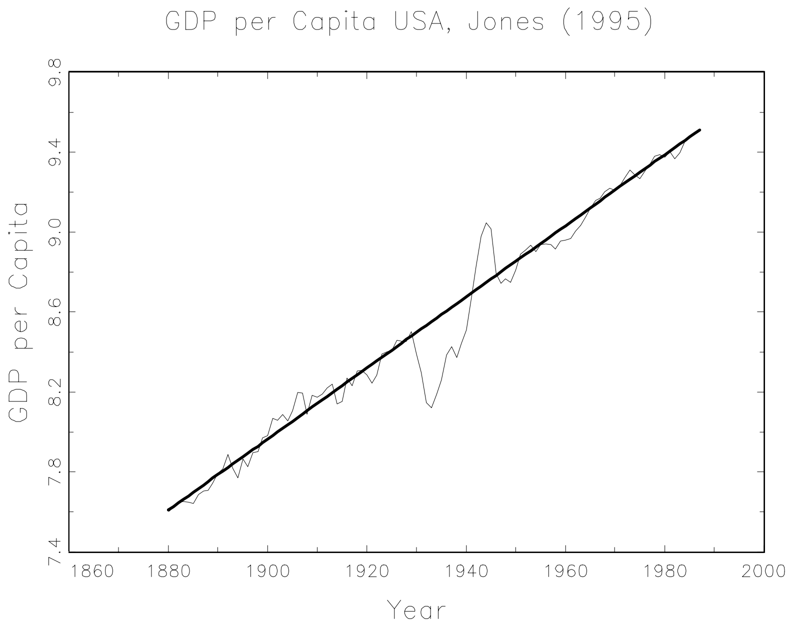


Figure 2: **Per Capita GDP in the United States, 1880-1987** (Natural Logarithm). The Data are from Maddison (1982,1989) as used in Jones (1996). The solid bold line represents the time trend calculated using data only from 1880 to 1929.

The empirical evidence shows that the underlying stochastic process for aggregate GDP exhibits some form of long memory. As a result we would like to know what economic mechanism can generate typical spectral shapes similar to the ones observed in the real world. Moreover, as the exact degree of memory is still uncertain, we would like to find 'robust' economic mechanisms: 'small' alterations to the basic set-up must not shift the slope of the spectral shape from the very particular slope associated with the unit root to

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

the flat one corresponding to the weak memory case[23]. In this paper we draw on the observed empirical relation between the expected growth of a firm and its size to provide a robust microfoundation for the typical spectral shape of an economic variable. In particular, we argue that the process of ongoing churning and catching up that takes place in the real economy slows down the propagation of the shocks and is a key element in explaining aggregate persistence.

# 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 in response to an aggregate shock.

## 3.1 The Model

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

The rate of technological progress is exogenous at rate $\gamma$.

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[24].

The firms in the economy can be in different technological states. In particular a firm is in state $i \geq 0$ at time $t$ if it is using technology $t - i$.

[23] The Real Business Cycle tradition has often argued that technology is the driving force of the 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 for example Cooper and John 1988). If so, aggregate shocks that shift the economy from one equilibrium to the other can generate a typical spectral shape (see Durlauf 1991). In both cases, a temporary shock has a permanent effect on the level of output. In general standard transmission mechanisms generate the very particular typical spectral shape associated with the unit root.

[24] See for example Pissarides (1990). Given these considerations, the model considers as observationally 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 (1998) analyze a vintage model in which firms explicitly face a trade-off between the two events.

Firms using different technologies are able to produce different quantities of goods, more exactly a firm in state $i$ at time $t$ produces a quantity of goods equal to $\gamma(t-i)$[25].

We indicate with $\pi_t$ the vector of countably infinite dimension collecting the measure of firms in each state. The $ith$ element of the vector $\pi_t$ measures the number of firms using technology $t-i+1$ at time $t$. $\pi_t$ is strictly positive, bounded between zero and one, with elements summing up to one and therefore is a probability measure.

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

$$Y_t = \gamma\, t - \gamma\, \pi_t'\, O,$$

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 state $i$ has two possibilities: either doing nothing and using the technology $t-i$ so that in the next period the firm will be in state $i+1$, or adopting the leading technology in the economy so that in the next period the firm will be in 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 two components, $c_i$ and $\lambda$, which enter additively. $c_i$ is a deterministic component function of the state $i$ of the firm. $\lambda$ is a random variable identically independently distributed across units and over time with common distribution $F$ over the support (possibly unbounded) $\Lambda \subseteq \Re$ and zero expected value. $\lambda$ gauges the firm-specific opportunity cost (in terms of forgone profits) of investing part of its own capital or labour resources in technological improvements. Therefore it can be read indifferently as either a technological or a

[25] All variables are denominated in logs, implying that differences indicate growth rates while arithmetic averages indicate the logarithm of geometric ones. We are implicitly assuming that a firm in state $i$ at time $t$ produces a quantity of intermediate goods equal to $\exp \gamma(t-i)$ and that, as in Grossman and Helpman (1991), final output is given by the aggregate production function $\exp \int_0^1 q_i di$ where $q_i$ is the amount of intermediate goods produced by firm $i$. We do not make these assumptions explicit both because of the space constraint and to keep notation as simple as possible. In a related paper, we show how the results extend under the alternative assumption that the intermediate goods produced by each firm are perfectly substitutable so that aggregate output $Y_t$ is equal to $\int_0^1 q_i di$. See Michelacci (1998).

demand shock[26]. The value of a firm $V(i, t, \lambda)$ in state $i$ at time $t$, whose cost of adopting the leading technology is $c_i + \lambda$, follows the Bellman equation

$$\begin{aligned} V(i, t, \lambda) &= \max_{s \in \{0,1\}} \gamma(t-i) - s(c_i + \lambda) + \\ &\quad + \beta(1-s) V^e(i+1, t+1) + \beta s V^e(0, t+1). \end{aligned} \tag{6}$$

$0 < \beta < 1$ is the discount factor while $V^e(j, t)$ indicates the expected value of $V(j, t, \lambda)$ taken with respect to the random variable $\lambda$[27]. It follows from dynamic programming arguments that the problem is well defined[28]. In particular the value function $V(i, t, \lambda)$ is linear in $t$, weakly decreasing in $\lambda$ and finally strictly decreasing in $i$ if $\gamma i + c_i$ is strictly increasing in $i$.

In general the firm decides to adjust and chooses $s = 1$ whenever the realization of the idiosyncratic shock $\lambda$ is such that

$$\beta [V^e(0, t+1) - V^e(i+1, t+1)] \geq c_i + \lambda. \tag{7}$$

That is, the firm weights the benefits of technological adoption $\beta [V^e(0, t+1) - V^e(i+1, t+1)]$ with the associated costs $\lambda + c_i$. We indicate with $1 - p_i$ the probability that the event (7) occurs[29], that is $1 - p_i$ is the probability that a firm in state $i$ will be using the best technology available in the economy in the next period. Given the assumption that the idiosyncratic shocks are *iid* with distribution function $F(\cdot)$, we obtain that

$$1 - p_i = F(\beta [V^e(0, t+1) - V^e(i+1, t+1)] - c_i), \quad \forall i. \tag{8}$$

As a result, the dynamics of the state of a generic firm is fully described by the infinite dimensional Markov chain $P$ given by

[26] See for example Aghion and Saint Paul (1993) and Saint Paul (1993).

[27] As the random variable $\lambda$ is independently distributed over time, the expected value $V^e(\cdot, \cdot)$ does not depend on past realisations of the idiosyncratic shocks.

[28] Despite the unbounded returns, the linearity of the technological frontier together with discounting guarantee that there is a one to one correspondence between the solution to the functional equation (6) and the corresponding sequential problem.

[29] It follows from the linearity in $t$ of the value function $V(i, t, \epsilon_t, \lambda_i)$ that the probabilities $1 - p_i$ are well defined and independent of $t$.

$$P=\begin{bmatrix} 1-p_0 & p_0 & 0 & 0 & 0 & 0 & \cdots \\ 1-p_1 & 0 & p_1 & 0 & 0 & 0 & \cdots \\ 1-p_2 & 0 & 0 & p_2 & 0 & 0 & \cdots \\ 1-p_3 & 0 & 0 & 0 & p_3 & 0 & \cdots \\ 1-p_4 & 0 & 0 & 0 & 0 & p_4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{9}$$

where the rows and columns represent the set of feasible technological states in the economy while the elements $1-p_i$ indicate the probabilities that a firm in state $i$ at time $t$ will be on the technological frontier at time $t+1$. $P$ is the *transmission mechanism* in the economy: $P$ maps the cross sectional distribution $\pi_t$ at time $t$ into the cross-sectional distribution $\pi_{t+1}$ at time $t+1$.

The next sub-section focus on the steady state properties of the system[30], while the following one introduces an aggregate shock and analyses dynamics. Some results are of independent interest. In fact, we develop a general way to model dynamics in an environment with cross-sectional heterogeneity. The advantage of the approach derives from dealing directly with the moving average representation of the process, with the impulse response of the economy to the shock then arising as a natural outcome of the analysis. In the next section we draw on this analysis and show under which conditions the model can replicate the observed degree of aggregate persistence, a parameter of fractional integration between zero and one.

## 3.2 Structure of the Transmission Mechanism

To characterize both the dynamics and the steady state properties of the system we focus directly on the structure of the transmission mechanism $P$, rather than on the structural parameters of the model given by the distribution function $F(\cdot)$, the parameters $\gamma$ and $\beta$, and the sequence of structural costs, $\{c_i, i \geq 0\}$. This exercise is sensible only if any given arbitrary transmission mechanism $P$ can be read, for some structural parameters, as a solution of the firm problem, defined by equations (6) and (8). Lemma 1 guarantees the validity of this 'semi-structural' approach: any assumption on

[30] The next sub-section is technical, and it might be skipped on first reading.

the transmission mechanism $P$ is the result of a corresponding assumption on the structural parameters of the model.

**Lemma 1 (Validity of the 'semi-structural' approach)** *Given a distribution function* $F(\cdot)$*, the parameters* $\gamma$ *and* $\beta$*, and an arbitrary sequence of probabilities* $\{p_i, i \geq 0\}$ *, then there does exist a sequence of structural costs* $\{c_i, i \geq 0\}$ *, whose solution to the firm problem, defined by equations (6) and (8), delivers the given sequence of probabilities* $\{p_i, i \geq 0\}$*. The sequence of adjustment costs* $\{c_i, i \geq 0\}$ *is uniquely defined given an arbitrary initial condition* $c_0$*.*

*Proof:* See appendix.

In general there are no strong theoretical reasons for assuming any a priori structure on the values of the probability $1-p_i$ defined by equation (8). They are indeed the outcome of two contrasting forces. On the one hand, the bigger the technological gap, the greater is the gain to adopt new technologies. On the other hand, the bigger the technological gap the costlier is technological adoption. If the first effect dominates, the probabilities $1-p_i$ are increasing in $i$, and firms using obsolete technologies are more likely to end up on the technological frontier rather than firms close to it[31]. Models with switching costs and human capital specificity suggest, however, why the probabilities $1-p_i$ may be decreasing in $i$[32]. The cost of adopting the leading technology $c_i$ is in general positively related to $i$[33] and firms using new technologies are more likely to end up on the technological frontier rather than firms far away from it. This reflects the fact that the higher the technology gap the more

[31] For example in Aghion and Howitt (1994), Caballero and Hammour (1994b) and Mortensen and Pissarides (1996), both the probability distribution of the idiosyncratic shocks $\lambda$ and the cost of adopting the leading technology $c_i$, are state independent. As the gains from technological adoption are always increasing in $i$, the monotonicity of $F(\cdot)$ implies that the probabilities $1-p_i$ are unequivocally increasing in $i$.

[32] See for example Acemoglu and Scott (1995), Jones and Newman (1995) and Jovanovic and Nyarko (1996).

[33] Clearly a change in the deterministic component of technological adoption $c_i$ modifies the structure of the value function, that is the left hand side of equation (7). 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 (7). As a result an increase in $c_i$ always reduces the value of $1-p_i$. For a formal proof see equation (20) in the appendix.

difficult is technological adoption. In order to characterize the transmission mechanism $P$ we will then draw on a combination of empirical evidence and theoretical arguments. The ultimate task is to show both that the model can replicate the observed degree of aggregate persistence and that the required conditions have a reasonable theoretical and empirical content.

Firms can wait an arbitrarily long time before adjusting, but sooner or later they must adjust in order to remain in the market. In fact, arbitrarily inefficient firms would eventually be driven out of business by more efficient ones[34]. Therefore, we impose that, whatever its current state, a firm sooner or later will adjust with probability one.

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

The side effect of this assumption is that our framework will exhibit one and only one *recurrent* (ergodic) class. That is, there exists only one set of states, each one of which will be visited infinitely often by the firms in the economy. The existence of a unique recurrent class is the counterpart in a stochastic set-up to a unique stable equilibrium. If we start from a situation where all units are in the set of recurrent states and we perturb the system, it will converge back to the initial situation with all units being in the initial set of recurrent states.

**Lemma 2 (Uniqueness and Stability of the equilibrium)** *Under Assumption 1, the transmission mechanism $P$ has always one and exactly one recurrent class containing the state zero.*

*Proof:* See appendix.

Lemma 2 shows how our framework rules out multiple equilibria (multiple ergodic sets) to explain persistence in aggregate fluctuations. In this model a shock can not move units from one ergodic set to the other and therefore the persistence generated by the model is not caused by a shift in the 'equilibrium'

[34] See Jovanovic and Nyarko (1996) for analogous considerations.

of the economy[35].

We are interested in further characterizing the equilibrium properties of the system and the structure of the recurrent class. In particular we distinguish the case in which the recurrent class consists of an infinite number of states (*irreducible transmission mechanism)* from the case in which the class consists of a finite number of states (*reducible transmission mechanism).* The first case implies that each firm will visit infinitely often all the states in the economy. The second corresponds to a situation where, in the steady state, firms will end up with probability one in a finite dimensional set close to the technological frontier.

**Assumption 2** $i^* = \min\left\{i : \beta_i = \beta_i^0 = \prod_{k=0}^{i-1} p_k = 0,\ i > 0\right\}$. Assume that $i^* < \infty$.

Assumption 2 means that firms adopting new technologies will adjust in a finite number of periods and therefore will remain close to the technological frontier. The observation by Baily, Hulten and Campbell (1992) and Bartelsman and Dhrymes (1994) that the persistence at the top of the technological distribution is particularly high, might support this assumption. However, the empirical evidence is not conclusive on this point and we will discuss further the role of the assumption in section 5. Assumption 2 guarantees that the transmission mechanism $P$ is reducible.

**Lemma 3 (Reducibility)** *Under Assumption 1, the transmission mechanism $P$ is reducible if and only if Assumption 2 holds.*

*Proof:* See appendix.

In the paper we consider a technical modification of Assumption 2 and we call it Assumption 2'. It ensures that, once entered in the recurrent class, units do not jump deterministically from one state to the other[36].

[35] See Durlauf (1991) for an explanation of aggregate persistence based on multiple equilibria.

[36] Relaxing this additional assumption would not affect any results of the paper, except the one concerning the existence of a steady state distribution.

**Assumption 2'** Assume that Assumption 2 holds and that if $i^* > 1$, $\exists$ $1 \leq i < i^* \ni \beta_i \neq 1$.

Lemma 2 guarantees that if a steady state distribution exists, it is unique and stable. We are also 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 4 (Existence of a Steady-State distribution)** *Assumption 1 is given. If Assumption 2' holds, the transmission mechanism is reducible and aperiodic and a steady state distribution always exists. If Assumption 2 does not hold, the transmission mechanism is irreducible and a steady state distribution exists if and only if the series* $\sum_{i=1}^{\infty} \beta_i$ *converges, where* $\beta_i = \beta_i^0 = \prod_{k=0}^{i-1} p_k$.

*Proof:* See appendix.

## 3.3 An Aggregate Shock with Cross-sectional Heterogeneity

We now introduce an aggregate shock $\epsilon_t$ that hits the system at time $t$ and we characterize the dynamic response of the system. In particular we gauge the rates at which the aggregate shock propagates in the economic system.

The aggregate shock, $\epsilon_t$, modifies, in a similar way, the opportunity cost of adjusting for all the firms in the economy. That means that the cost of adopting the leading technology for a firm in state $i$ with idiosyncratic component equal to $\lambda$ becomes equal to $c_i + \lambda + \epsilon_t$. Equations (6) and (8) show how this modifies the problem of the firm. For example, when $\epsilon_t > 0$ ($\epsilon_t < 0$) the cost of technological adoption is bigger (smaller ) and in the next period we will observe fewer (more) firms adopting the leading technology relative to the number that would have done so in the absence of the shock ($\epsilon_t = 0$). More formally, when $\epsilon_t \neq 0$ a firm in state $i$ will decide to adjust whenever the realization of the idiosyncratic component $\lambda$ is such that[37]

$$\beta \left[V^e(0,\, t+1) - V^e(i+1,\, t+1)\right] \geq c_i + \lambda + \epsilon_t. \tag{7'}$$

We indicate with $1-\ p_i(\epsilon_t)$ the probability that the event (7') occurs, that is

$$1 - p_i(\epsilon_t) = F\left(\beta \left[V^e(0,\, t+1) - V^e(i+1,\, t+1)\right] - c_i - \epsilon_t\right),\ \ \forall i, \tag{8'}$$

is the probability that, when the aggregate shock is equal to $\epsilon_t$, a firm in state $i$ will be using the best technology available in the economy in the next period.

We indicate with $\bar{P}(\epsilon_t)$ the Markov chain analogous to (9) collecting the probabilities $p_i(\epsilon_t)$. The dynamics of the cross-sectional distribution of vintages currently in use, $\pi_t$, are therefore described by the equation

$$\pi_t = \delta_t + P^{'}\pi_{t-1} \tag{10}$$

where $\delta_t = \left(\bar{P}(\epsilon_t) - P\right)^{'} \pi_{t-1}$. In the absence of the aggregate shock, $\bar{P}(\epsilon_t) = P$ and $\delta_t = 0$. In this case the transmission mechanism $P$ maps the cross sectional distribution $\pi_{t-1}$ at time $t-1$ into the cross-sectional distribution $\pi_t$ at time $t$. The infinite dimensional column vector $\delta_t$ is simply an error term: it is equal to the difference between the observed cross-sectional distribution given by $\pi_t = \bar{P}(\epsilon_t)^{'} \pi_{t-1}$ and that which would have occurred in the absence of the aggregate shock, equal to $P^{'}\pi_{t-1}$.

In this model an aggregate shock drives a reallocation of the technological positions of firms. The vector $\delta_t$ measures the size and structure of the reallocation and has two general properties. Firstly, the sum by column of its entries, $\delta_t^i$, $i \geq 1$, is exactly equal to 0, that is

$$1^{'}\delta_t = 1^{'}\left(\bar{P}(\epsilon_t) - P\right)^{'} \pi_{t-1} = 0, \tag{11}$$

where 1 is a vector of ones. That means that a shock simply *reallocates* units across technological vintages[38]. Secondly, a negative (positive) aggregate

[37] The aggregate shock considered here is a once and for all shock. In a related paper I analyze the role of ongoing uncertainty in the model; nothing is modified substantially if aggregate and idiosincratic shocks are independent. See Michelacci (1998).

[38] Of course, in the real world aggregate shocks might affect both the number of firms in the market and the rate of technological progress. The assumption that they are both exogenous, allows us to isolate the contribution of the process of ongoing churning and catching up in the propagation of the aggregate shock.

shock fosters (harms) technological adoption. More formally, if we indicate by $I(\cdot)$ the characteristic or indicator function, it follows from the monotonicity of $p_i(\epsilon_t)$ with respect to $\epsilon_t$ that

$$I\left(\delta_t^1 > 0\right) = I\left(-\delta_t^j > 0\right) = I\left(-\epsilon_t > 0\right), \; \forall \left|\delta_t^j\right| \neq 0, \; j \neq 1. \tag{12}$$

As in Caballero and Hammour (1994b) and Mortensen and Pissarides (1996) a negative aggregate shock 'cleanses' the economy[39], fosters technological adoption and increases the level of output.

We assume that the reallocation structure $\delta_t$ has on impact a bounded effect on the level of aggregate output. For example this is always the case for 'finite' reallocations, that is for reallocation structures $\delta_t$ where only a finite number of coefficients $\delta_t^i$ are strictly positive.

**Assumption 3** Assume that $\delta_t' O < \infty$.

Before characterizing the dynamic response of aggregate output, we need to show that the infinite dimensional matrix products $\delta_t{}' P^n O$ are well defined for all $n$.

**Lemma 5 (Pseudo Wold Coefficients)** *Under Assumption 3, the product $\delta_t{}' P^n O$ is bounded $\forall n$ and well defined as the matrices associate, that is $\delta_t{}'(P^n O) = (\delta_t{}' P^n)O$, $\forall n$.*

*Proof:* See appendix.

We now turn to the question of characterizing the dynamic response of the economic system to the aggregate shock. The effect of the shock $n$ periods ahead is given by the difference between the level of output $n$ periods after the realization of the shock and that which would have occurred in the absence of the shock. Given an initial distribution $\pi_{t-1}$ at time $t-1$, the level of output at time $t+n$ is equal to

$$Y_{t+n} = \gamma(t+n) - \gamma \pi_{t-1}{}' P^n O - \gamma \delta_t{}' P^n O \tag{13}$$

[39] 'Cleansing' does not take place necessarily in the recession. For example, if there are liquidity constraints, the best period for adopting new technologies might be a boom.

while it would have been equal to

$$Y_{t+n} = \gamma\,(t+n) - \gamma {\pi_{t-1}}'\, P^n O \tag{14}$$

in the absence of the shock. The difference between (13) and (14) gauges the dynamic response of the aggregate economy to the shock $\epsilon_t$. In other words the quantities

$$\phi_n = -\,\gamma \delta_t' P^n O,\ \forall n \tag{15}$$

are analogous to the Wold coefficients analysed in the previous section: $\phi_n$ gauges the effect in the economic system of the shock $\epsilon_t$, $n$ periods ahead. Therefore as in equation (4) the rate of decay of $\phi_n$ measures the persistence of the shock in the model.

It is interesting to extend the comparison between the Wold coefficients in (4) and the 'pseudo' Wold coefficients in (15). For this purpose only, assume that aggregate shocks were to occur at different times. If so, the level of output at time $t$ $Y_t$ may be written as

$$Y_t = \gamma {\pi_0}'\, O - \gamma {\pi_0}'\, P^t O + \gamma\, t + \sum_{n=0}^{t} \phi_n^t\,, \tag{16}$$

where $\phi_n^t = -\,\gamma \delta_{t-n}' P^n O$, while $\delta_{t-n}$ indicates the reallocation structure of the economy at time $t-n$. If a steady state distribution exists so that ${\pi_0}'\, P^t = {\pi_0}'$, $\forall t$ and we initialize the process at this point, the representation (16) collapses to

$$Y_t = Y_0 + \gamma\, t + \sum_{n=0}^{t} \phi_n^t\,, \tag{17}$$

where $Y_0$ is the level of output at the starting date normalized to be equal to 0. The representation (17) resembles the Wold representation of a time series (4). The two representations would indeed be equivalent if for example the quantity $\delta_{t-n}$ could be written as a linear function, that is $\delta_{t-n} = \delta \epsilon_{t-n}$. This would imply firstly that positive and negative shocks have symmetric effects on the level of output. Secondly, it would mean that the previous history of the system synthesized by the current cross sectional distribution $\pi_{t-n-1}$ is irrelevant in characterizing the dynamic response of the economy to the shock $\delta_{t-n}$. In general, neither is the case in the model analysed here, and an interesting but independent issue is in what respect the representation (17) differs from that of a standard linear process[40].

# 4 The Persistence of Aggregate Fluctuations

In this section we assume that some versions of Assumptions 1, 2' and 3 hold, we measure the degree of persistence of the shock $\epsilon_t$ and we show under which conditions the model can replicate the observed degree of aggregate persistence, a value of $d$ between zero and one. Just as in the case of the Wold coefficients in (4), the rate of decay of the pseudo Wold coefficients $\phi_n$ in (15) gauges the persistence of the shock in the model. In particular we say that the transmission mechanism $P$, together with the reallocation structure $\delta_t$, generates a typical spectral shape with order of integration $d$ if $\phi_n \sim n^{d-1}$ as in (4).

Firstly we show under which conditions the model cannot generate a typical spectral shape in output. If the probabilities $1- p_i$ are increasing in $i$, firms using obsolete technologies are more likely to end up on the technological frontier than firms currently in the technological lead. In this case the hazard function is increasing in the size of technological gap and Proposition 1 shows that no typical spectral shape can be generated by the model.

**Proposition 1 (Increasing hazard functions)** *Suppose that Assumption 2' and 3 hold and that there does exist a state $\kappa \geq i^*$ such that $p_\kappa < 1$ and that the probabilities $p_i$ are weakly decreasing for any $i \geq \kappa$. Then aggregate output always behaves like a trend stationary process with ARMA disturbances, that is the quantities $\phi_n$ in (15) decay at least exponentially, i.e. $|\phi_n| \leq K\rho^n$, $0 \leq \rho < 1$, where $K$ is a positive bounded quantity.*

*Proof:* See appendix.

[40] A linear process is defined as a process for which the representation (4) holds where the shock $\epsilon_t$ are *iid*. In a companion paper, we analyze the relation between the two representations and we characterize more formally the properties of the coefficients $\phi_n^t$. In particular we show under which conditions asymmetries and conditional heteroscedasticity arise in the model. See Michelacci (1998). See also Caballero and Engel (1998).

In fact, Baily, Hulten and Campbell (1992) find that the probability of being a relatively high productivity firm in 5 or 10 years time is strongly increasing in the current level of relative productivity. High rather than low productivity firms are more likely to be in the technological lead in the next period, that is the hazard function relative to adopting new technologies seems to be decreasing in the size of technological gap[41]. This reflects the fact that the bigger the technological gap the more difficult is technological adoption.

We now draw on the relation between the expected growth of a firm and its size to show the model may generate typical spectral shapes in output. In the model firms using obsolete technologies are 'smaller' (produce less output) than firms operating close to the technological frontier. Some assumptions on factor allocation are required to link productivity to size measured by either sales, employment or assets. In general, if factor markets are not segmented and productivity increases the marginal revenue of each factor, a productivity ranking corresponds one for one to a size ranking[42]. We classify as 'big' the firms that are adopting vintages close to the technological frontier, while all other firms are 'small'. More specifically, a firm is big if it is in state $i < i^*$ where $i^* = \min \left\{ i : \beta_i = \beta_i^0 = \prod_{k=0}^{i-1} p_k = 0,\ i > 0 \right\}$ (see Assumption 2): a big firm is in the technological lead, it is in 'steady state' and it is growing at the same rate $\gamma$ as the technological frontier. Small firms might grow at rates different from $\gamma$. In particular a firm in state $i$ either raises output by $\gamma\,(i+1)$ with probability $1 - p_i$ or it keeps output constant with probability $p_i$. Therefore $g_i = (1 - p_i)(i + 1)$ is the expected growth of a firm in state $i$[43]. Proposition 2 shows that if all small firms are growing in the same way, i.e. $g_i = h\gamma$ independent of $i$, the model is capable at generating typical spectral

[41] Estimated hazard functions relative to capital and labour are in general increasing. That means that the probability of adjusting capital or labour is increasing in the difference between actual and desired capital or labour, see for example Caballero, Engel, and Haltiwanger (1997). To the extent that the desired amount of capital and labour is a function of the technology currently adopted, estimated hazard functions do not address the question of what is the hazard function relative to adopting new technologies: estimated hazard functions are distribution functions conditional to a given technology.

[42] For example, Baily, Hulten and Campbell (1992) and Bartelsman and Dhrymes (1994) find that employment size and productivity are positively correlated.

[43] Under the conditions analyzed in footnote (25), $g_i$ is the expected growth rate of a firm in state $i$ just as $\gamma$ is the growth rate of the aggregate economy.

shapes in output. $h$ then measures the relative growth of small versus big firms. Small firms are growing faster (more slowly) than big ones if $h >(<)1$, while all firms grow at the same rate if $h = 1$.

**Proposition 2 (Robust typical spectral shapes)** *Suppose that Assumptions 2' and 3 hold. Assume also that*

$$0 < \sum_{i=0}^{\infty} \frac{\left|\delta_t^{s+i+1}\right|}{\beta_i^s} < \infty, \ \beta_0^s = 1 \tag{A4}$$

*where* $s = \max\{i : i \geq 0, p_i = 0\} + 1$ *and that the expected growth of a unit in state* $i$*,* $g_i$*, is such that, for some arbitrary* $s^*$*,*

$$g_i = (1 - p_i)(i + 1) = h\gamma, \ \ \forall i \geq s^* \ \ h > 0. \tag{A5}$$

*Then the order of integration of aggregate output* $d$ *is equal to* $2 - h$*, that is the quantities* $\phi_n$ *in (15) are such that* $\phi_n \sim n^{d-1}$*, where* $d = 2 - h$*.*

*Proof:* See appendix.

Assumption $(A4)$ requires that some units end up in the non recurrent set ($\delta_t^{s+i+1} \neq 0$ for some $i$). Moreover it puts some boundaries (in addition to Assumption 3) on the amount of reallocation that is driven by the aggregate shock. We will discuss further its role in the next section. For example $(A4)$ is satisfied in the case of 'finite' reallocations, that is for reallocation structures $\delta_t$ where only a finite number of coefficients $\delta_t^{s+i+1}$ are strictly different from zero.

What drives Proposition 2 is the process of initial churning and subsequent catching up that takes place in the model. If the Wold coefficients $\phi_n \sim n^{d-1}$, the shock propagates at decreasing rates, in contrast to the constant rates that would arise in the exponential case, that is $n^{d-1} \sim (1 - \frac{1-d}{1})(1 - \frac{1-d}{2}) \ \cdots \ (1 - \frac{1-d}{n})$. Once the aggregate shock $\epsilon_t$ hits the system, firms are reallocated across sizes according to the reallocation structure $\delta_t$. Whether the shock will be absorbed or not depends then on the relative growth of small versus big firms[44]. However, fast growing small firms

[44] It can be noted that Proposition 2 is a case of Galton's fallacy (see Quah 1993): the

eventually become big, and grow as big firms. Thus the number of small firms shrinks over time as well as the rate at which the shock propagates in the economy. The process of ongoing churning and catching up that takes place in the model slows down the propagation of the shock and allows the model to replicate the typical spectral shape in aggregate output observed by Granger.

Proposition 2 can be summarized as follows:

($i$) If small firms grow faster than big ones, $1 \leq h < 2$, the model replicates the order of integration $d$ between 0 and 1 observed in aggregate output[45].

($ii$) If $h < 1$, big firms grow faster than small ones and the first difference of aggregate output exhibits long memory[46]. 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.

($iii$) A particular case arises if 'Gibrat's law' holds and all firms grow in the same way, $h = 1$. In this case the shock has permanent and bounded effect on the level of output, this is equivalent to a unit root in output, $d = 1$.

($iv$) We conclude that the transmission mechanism $P$ is *robust* because it generates typical spectral shapes independently of the relative growth of small versus big firms.

# 5 Interpretation of the Assumptions

Proposition 2 is particularly dependent upon Assumptions 2', ($A4$) and ($A5$). We now discuss in more detail their roles.

relative growth of small versus big firms does not say anything on the presence of convergence rather than divergence. In fact, independently of whether small firms are growing faster or slower than big ones, there is convergence in the cross-sectional distribution of vintages.

[45] See Mansfield (1962), Hall (1987), Evans (1987) and Dunne, Roberts and Samuelson (1989) for empirical evidence in this direction.

[46] See Davis, Haltiwanger and Schuh (1993) for empirical evidence in this direction.

Assumption 2' implies that all firms end up eventually in a finite dimensional set close to the technological frontier, so that in the long run there is 'convergence in size'; see Lemma 3. Once we relax this assumption, the recurrent class is infinite dimensional and firms keep wandering across all possible states in the economy. In this case it is still true that fast-growing small firms eventually become big, and grow as big firms, but it is also the case that big firms eventually become small and grow as small firms. If so, the impact effect of the shock cannot be amplified without limit and the model is unable to generate orders of integration greater than one. In other words Assumption 2' plays the role of increasing the *robustness* of the transmission mechanism[47].

Assumption $(A4)$ requires that some units end up in the non recurrent set ($\delta_t^{s+i+1} \neq 0$ for some $i$). For example if at time $t-1$ the system is in steady state, assumption $(A4)$ implies that the shock $\epsilon_t$ is positive. Some units do not adjust and they end up in the non-recurrent class. This condition does not look very restrictive. Eventually a positive aggregate shock hits the economy and some units end up in the non recurrent set. Moreover, assumption $(A4)$ limits the size of the reallocation driven by the aggregate shock and it bounds the persistence generated by the model. $(A4)$ is always satisfied in the case of 'finite' reallocations, but the number of entries in the reallocation structures $\delta_t$ strictly different from zero does not have to be finite. We consider now a counterexample where $(A4)$ fails. We assume for example that all firms are initially in the tail of the distribution of vintages, $\pi_{t-1}^i \sim \beta_i^s$ as $i \uparrow \infty$. We consider the effect of a negative aggregate shock $\epsilon_t < 0$, so large in absolute value that $\delta_t^{s+i} = -a\,\beta_i^s$, $a > 0$. Proposition 3 shows that in this case the model generate a degree of persistence greater than before. As a corollary it follows that the degree of persistence generated by the model is dependent upon features of the reallocation structure $\delta_t$.

[47]In particular, if the transmission mechanism $P$ is irreducible, $h \leq 1$ implies that $\beta_i \sim i^{-h}$ as $i \uparrow \infty$ and no steady state distribution exists by Lemma 4. If $1 < h < 2$, the model still replicates the order of integration $d$ between 0 and 1 observed in aggregate output. I analyze this case in my Phd dissertation and in a companion paper. See Michelacci (1998).

**Proposition 3 (Counter-example)** *Suppose that Assumptions 2' holds. Assume also that*

$$\delta_t^{s+i} = -a\,\beta_i^s,\ 0 < a < 1,\ \forall i \tag{A4'}$$

*where* $s = \max\{i : i \geq 0,\, p_i = 0\} + 1$ *is bounded and that the expected growth of a unit in state* $i$*,* $g_i$*, is such that*

$$g_i = (1 - p_i)(i + 1) = h\gamma,\ \ \forall i \geq s, \quad h > 2. \tag{A5'}$$

*Then the order of integration of aggregate output* $d$ is equal to $3 - h$, *that is the quantities* $\phi_n$ *in (15) are such that* $\phi_n \sim n^{d-1}$, *where* $d = 3 - h$.

*Proof:* See appendix.

**Corollary (The reallocation structure matters)** *Assumption 1, 2' and 3 hold. For a given transmission mechanism P, different reallocation structures* $\delta_t$ *can generate different rate of absorption of the shocks as measured by the quantities* $\phi_n$ *in (15).*

*Proof:* See appendix.

Proposition 3 is interesting for three reasons. Firstly, it shows that the model can generate asymmetric responses to shocks. In fact $\delta_t = \left(\bar{P}\left(\epsilon_t\right) - P\right)' \pi_{t-1}$, and if $p_i = 1 - \frac{\gamma h}{1+1}$, only a negative shock can generate a difference between $p_i\left(\epsilon_t\right)$ and $p_i$ equal to a constant, which is why $a$ in ($A4$') may only be negative. Secondly, Proposition 3 is an application of the 'folk wisdom' claiming a positive relation between amplification and propagation mechanisms: large shocks are more persistent than small shocks. Finally, Proposition 3 delivers an alternative and suggestive characterization of aggregate persistence. According to this view most shocks generate low persistence[48]; sometimes, however, large negative shocks hit the system when the average productivity of the economy is low: these are the cause of the spectral shape that we observe in the real world. This view is not new. Many have noticed that once we allow for some structural breaks in the time series of US aggregate GDP,

[48] For example $h > 2$ does not generate a typical spectral shape under the conditions assumed in Proposition 2.

it may be well represented by a standard weak memory process[49]. What is suggestive is that the model might explain why most structural breaks identified in the literature (the 'big recession', World War II, the oil price shock) are associated with large negative shocks.

Assumption $(A5)$ seems to be very specific. It implies that the adjustment cost $c_i$ must grow sufficiently fast when the tecnology gap $i$ increases. A possible motivation would be that $(A5)$ is consistent with what has been observed in the data[50]. That is why we argue that the empirical relation between expected growth and firm size might provide a microfoundation for aggregate persistence.

Finally, the literature on the relation between growth and firm size is itself controversial. We feel however that our reading of the literature is the closest to the macro approach we are pursuing here. In fact, what makes the results controversial is the treatment of the sample selection problem[51]. Conditional on survival, small firms grow faster but also tend to die more often. The question then is what growth we should attribute to a dead firm. In fact dead firms free resources that potential investors can now exploit, and therefore in general equilibrium death might be associated with high growth. Once this concern is taken into account, the empirical evidence tends to conclude that small firms grow faster than big ones[52]: this is what is required to replicate an order of integration between zero and one. If so, micro and macro evidence match quite closely.

# 6 Conclusions

Ex ante homogeneous firms end up by having very different histories. This implies that once we take a picture of the economic system, a lot of cross-sectional heterogeneity appears. This paper has shown that the mechanism generating heterogeneity in the real world also generates persistence in aggre-

[49] See for example Perron (1989).

[50] See also footnote (25) and (43).

[51] See Sutton (1996).

[52] Davis, Haltiwanger and Schuh (1993) attribute a rate of growth of minus two to dead establishments and conclude that big establishments grow faster. Mansfield (1962), Hall (1987), Evans (1987) and Dunne, Roberts and Samuelson (1989) address the sample selection problem and conclude that small firms grow faster than big ones.

gate fluctuations. In our model aggregate shocks affect neither the number of firms in the market nor the rate of technological progress. Therefore any persistence can be attributed to cross-sectional heterogeneity. The paper concludes that cross-sectional heterogeneity is an important transmission mechanism with some distinctive properties. We might summarize them by saying that cross-sectional heterogeneity is a powerful, realistic and robust transmission mechanism. 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 the observed empirical relation between expected growth of a firm and its size provides a microfoundation for the typical spectral shape of an economic variable. It is *robust* because 'small' alterations to the basic set-up do not shift the slope of the spectral shape from the very particular slope associated with a unit root to the flat one corresponding to the weak memory case.

A lot of questions still remain open. In particular, realism does not imply reality so that further, careful empirical investigation is required to see what really are the sources of the persistence of aggregate fluctuations. Three obvious candidates come to mind: the dynamics of the leading technology in the economy (here $\gamma$), 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 paper has departed from both these strands of the literature and has drawn on the recent tendency to provide micro-foundations for macroeconomics by solving explicit aggregation problems. It has argued that cross-sectional heterogeneity might be a key element in explaining aggregate persistence and has concluded that the interplay between idiosyncratic and aggregate uncertainty shapes the dynamics of macroeconomic variables in a distinctive way.

# 7 Appendix

## 7.1 Proofs of results in section 3

**Proof of Lemma 1** As the value function $V(i,\, t,\, \lambda)$ is linear in $t$, it can be written as

$$V(i,\, t,\, \lambda) = at + \tilde{V}(i, \lambda)$$

where $a$ is equal to $\frac{\gamma}{1-\beta}$. It follows that

$$1 - p_i = F\left(\beta\,[\tilde{V}^e\,(0) - \tilde{V}^e\,(i+1)] - c_i\right), \quad \forall i,$$

where $\tilde{V}^e\,(j)$ indicates the expected value of $\tilde{V}\,(j,\, \lambda)$ taken with respect to the random variable $\lambda$. We indicate with $\lambda_R^i$ the solution to the equation

$$\beta\,[\tilde{V}^e\,(0) - \tilde{V}^e\,(i+1)] = c_i + \lambda_R^i, \quad \forall i. \tag{18}$$

Therefore, $\forall i$, $1 - p_i = F\left(\lambda_R^i\right)$ so that there is a one to one correspondence between $p_i$ and $\lambda_R^i$. From the monotonicity of $\tilde{V}(i, \lambda)$ with respect to $\lambda$ it follows that

$$\begin{aligned} \tilde{V}(i, \lambda) &= -\gamma i + \beta \tilde{V}^e\,(i+1), & if\ \lambda \geq \lambda_R^i, \\ &= -\gamma i - c_i - \lambda + \beta \tilde{V}^e\,(0), & if\ \lambda < \lambda_R^i. \end{aligned}$$

Taking expectations we obtain

$$\tilde{V}^e\,(i) = -\gamma i - p_i \lambda_R^i - c_i - \int_{\lambda_R^i}^{\infty} s\; dF\,(s) + \beta \tilde{V}^e\,(0), \quad \forall i. \tag{19}$$

(18) together with (19) implies that

$$\begin{aligned} c_{i+1} &= p_0 \lambda_R^0 + c_0 + \left(c_i + \lambda_R^i\right)\frac{1}{\beta} + \int_{\lambda_R^0}^{\infty} s\; dF\,(s) - \gamma\,(i+1) + \\ &\quad -p_{i+1}\lambda_R^{i+1} - \int_{\lambda_R^i}^{\infty} s\; dF\,(s)\,, \quad \forall i. \end{aligned} \tag{20}$$

For any given sequence of probabilities $\{p_i, i \geq 0\}$, equation (20) defines a difference equation of the first order in $c_i$, whose solution is unique once an initial condition for $c_0$ is set. Moreover, given equations (18) and (19) a sequence of adjustment costs $\{c_i, i \geq 0\}$ that solves (20) delivers the given sequence of probabilities $\{p_i, i \geq 0\}$ as a solution of the firm problem. Lastly, we note that as $c_0$ is arbitrary and $\beta < 1$, adjustment costs can always be chosen to be strictly positive. Q.E.D.

**Proof of Lemma 2** We start with reviewing some basic concepts in the theory of Markov chains (see e.g. Karlin and Taylor 1975, 1981). Two states $i$ and $j$ *communicate* if there exists 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 reflexivity, symmetry and 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. The Markov chain is *reducible* if it is not irreducible. 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 *transient.* All states in an equivalence class are either recurrent or transient so that both recurrency and transience are class properties.

Given Assumption 1, the Markov chain $P$ has the property that starting from any 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 3** Given Assumption 1, the probability of returning to state zero starting from any state $i$ is one, that is all states communicate with state zero. State zero communicates with a given state $i$ if and only if $\beta_i > 0$. This implies that all states can be reached from state zero and the transmission mechanism $P$ is irreducible if and only if Assumption 2 does not hold. Q.E.D.

**Proof of Lemma 4** Given Assumption 1 and Lemma 1, there exists one and only one recurrent class and all units will enter the set of recurrent states with probability one. Given Assumption 2' and Lemma 3 the matrix is reducible and the recurrent class is finite dimensional. We indicate with $i^*$ the number of states of the recurrent class. Given Assumption 2', there exists a state $0 \leq i < i^* - 1$ such that $p_i \neq 1$ and the Markov chain $P$ is aperiodic. Therefore the unique finite dimensional recurrent class is aperiodic and a steady state distribution always exists.

If Assumption 1 holds and Assumption 2 does not, the Markov chain is irreducible and recurrent by Lemma 3. The basic limit theorem of Markov chains (see Karlin and Taylor 1975, theorem 1.2) implies that one of two cases must hold. Either the Markov chain is null recurrent so that asymptotically

$P^n$ goes to a matrix of zeros, or it is positive recurrent so that the matrix $P^n$ converges to a matrix whose rows are identical and equal to the steady state distribution, call that $\pi^*$. It can be checked that $\pi^*$ is such that its element in place $i$, $\pi_i^*$, satisfies the relation $\pi_i^* = \pi_1^* \beta_{i-1}$, where $\beta_0 = 1$. This implies that a necessary and sufficient condition for the transmission mechanism $P$ to be positive recurrent is that

$$\pi_1^* = \frac{1}{\sum_{i=0}^{\infty} \beta_i},$$

is finite and strictly positive (see Billingsley 1986, theorem 8.8. and example 8.13). Q.E.D.

**Proof of Lemma 5** We first write the reallocation structure $\delta_t = \delta$ 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 matrices (see 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 . By Assumption 3 the reallocation structure $\delta$ is such that $(\delta^-)' O$ and $(\delta^+)' O$ are both finite as either $\delta^+$ or $\delta^-$ has just one element strictly positive by equation (12). We then note that each element in place $j$ of $P^n O$ has increments bounded above from one. Therefore, if we indicate with 1 a vector of ones we have that

$$0 \leq (\delta^-)' P^n O \leq (\delta^-)' (O + n\,1) < \infty$$

and analogously for $(\delta^+)' P^n O$

$$0 \leq (\delta^+)' P^n O \leq (\delta^+)' (O + n\,1) < \infty,$$

so that $(\delta^-)' P^n O$ and $(\delta^+)' P^n O$ are bounded $\forall n$. Q.E.D.

## 7.2 Proofs of results in section 4

The next Lemma is used throughout the proofs.

**Lemma 6 (Technical lemma)** $\kappa = \max\{i : p_i = 0, i \geq 0\} + 1$ while

$$s = \begin{cases} \kappa \;\; if \;\; \kappa < \infty \\ i^* \;\; if \;\; \kappa = \infty \end{cases} \tag{21}$$

where $i^* = \min\left\{ i : \beta_i = \beta_i^0 = \prod_{k=0}^{i-1} p_k = 0, i \geq 1 \right\}$.

Then under Assumptions 1, 2' and 3,

$$\delta_t' P^n O = A + B + C - D, \tag{22}$$

where $\forall n$, $A$, $B$, $C$ and $D$ are equal to

$$\begin{aligned}
A &= \sum_{i=1}^{\infty} \delta_t^i \, c_{i-1}^n, && (A)\\
B &= n \sum_{i=0}^{\infty} \delta_t^{s+i+1} \beta_n^{s+i} && (B)\\
C &= \sum_{i=0}^{\infty} \delta_t^{s+i+1} \beta_n^{s+i} \left(s+i\right), && (C)\\
D &= S \sum_{i=0}^{\infty} \delta_t^{s+i+1} \beta_{n-1}^{s+i} && (D)
\end{aligned}$$

where $\beta_i^j = \prod_{k=0}^{i-1} p_{j+k}$, the quantities $c_i^n$'s are such that $0 \leq c_i^n \leq \rho^n K$, $0 < \rho < 1$, $K$ is a bounded quantity while $0 \leq S < i^*$, is the expected value of the steady state distribution.

**Proof of Lemma 6** By definition (21) and Assumption 2' $s$ is always bounded. We consider the submatrix $\tilde{P}$ of the transmission mechanism $P$ identified by the first $s$ rows and columns of the matrix $P$. $\tilde{P}$ is a positive square matrix whose rows sum up to one, that is a stochastic matrix. It follows from the same reasoning as in the proof of Lemmas 2, 3, and 4 that $\tilde{P}$ is irreducible if $i^* = s$, while it is reducible if $i^* < s$. By Lemmas 2 and 3 the recurrent class and the steady state distribution of $\tilde{P}$ are the same as those of $P$. We can then partition the matrix $P$ as follows

$$P = \left[\begin{array}{c|cccccc}
\underset{s\times s}{\tilde{P}} & \underset{s\times 1}{0} & \underset{s\times 1}{0} & \underset{s\times 1}{0} & \underset{s\times 1}{0} & \cdots \\
\hline
\underset{1\times s}{r_s^1} & 0 & p_s & 0 & 0 & \cdots \\
\underset{1\times s}{r_{s+1}^1} & 0 & 0 & 0 & p_{s+1} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}\right],$$

where $r_j^1$ indicates a row vector of dimension $1 \times s$ corresponding to the first $s$ elements of the row $j+1$ (state $j$) of the matrix $P$. For example in this notation $\tilde{P}$ is equal to

$$\underset{s \times s}{\tilde{P}} = \begin{bmatrix} r_0^1 \\ r_1^1 \\ \vdots \\ r_{s-1}^1 \end{bmatrix}.$$

It follows that $P^n$, the $nth$ iterate of $P$ is equal to

$$P^n = \left[\begin{array}{c|cccccccc} \tilde{P}^n & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots \\ \hline r_s^n & 0 & \cdots & 0 & \beta_n^s & 0 & \cdots & 0 & \cdots \\ r_{s+1}^n & 0 & \cdots & 0 & 0 & \beta_n^{s+1} & \cdots & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{s+i}^n & 0 & \cdots & 0 & 0 & 0 & \cdots & \beta_n^{s+i} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}\right],$$

where $r_j^n$ indicates a row vector of dimension $1 \times s$ corresponding to the first $s$ elements of the row $j+1$ (state $j \geq 0$) of the matrix $P^n$, while the element $\beta_n^{s+i}$, $i \geq 0$, are in row $s+i+1$ and column $s+i+1+n$. If we indicate with $e_1$ a vector of dimension $s \times 1$ equal to the first column of an identity matrix of dimension $s \times s$, it can be proved by recursion that for $i \geq s$

$$\begin{aligned} r_i^n &= (1-p_i)\, r_0^{n-1} + p_i\,(1-p_{i+1})\, r_0^{n-2} + \cdots + \\ &+ \beta_{n-2}^i\,(1-p_{i+n-2})\, r_0^1 + \beta_{n-1}^i\,(1-p_{i+n-1})\, e_1' = \\ &= \sum_{j=0}^{n-1} \beta_j^i\,(1-p_{i+j})\; r_0^{n-1-j}, \end{aligned} \tag{23}$$

where $\beta_0^i = 1$, while $r_0^0 = e_1'$.

To prove (22) we proceed as follows. Firstly we show that $r_i^n$, $0 \leq i < s$ converges at least exponentially to $\pi$, the vector of dimension $s \times 1$ corresponding to the steady state distribution of $\tilde{P}$, that is

$$|r_i^n - \pi'| \leq H\,\rho^n 1', \;\; \forall i < s, \;\; \forall n > 0 \tag{24}$$

where $1 \geq H \geq 0$, $0 \leq \rho < 1$ and $1'$ is a vector of dimension $1 \times s$ whose elements are all equal to one.

Secondly we will show that $\forall i \geq s$, $\forall n$, $r_i^n$ is equal to

$$r_i^n = \left(1 - \beta_{n-1}^i\right)\pi' + \beta_{n-1}^i\,(1-p_{i+n-1})\, e_1' + a_i^n, \tag{25}$$

where the $a_i^n$ are such that $\forall n$, $\forall i$, $0 \leq a_i^n \leq \rho^n H\, 1'$, $0 \leq \rho < 1$, $H$ being a bounded quantity independent of $i$ and $n$. (24) together with (25) will then imply that

$$P^n O = \begin{bmatrix} c_0^n + S \\ \vdots \\ c_{s-1}^n + S \\ c_s^n + S + \beta_n^s \, (s+n) - S \, \beta_{n-1}^s \\ c_{s+1}^n + S + \beta_n^{s+1} \, (s+1+n) - S \, \beta_{n-1}^{s+1} \\ \vdots \\ c_{s+i}^n + S + \beta_n^{s+i} \, (s+i+n) - S \, \beta_{n-1}^{s+i} \\ \vdots \end{bmatrix}. \quad (26)$$

where the quantities $c_i^n$ are such that $0 \leq c_i^n \leq \rho^n K$, $0 \leq \rho < 1$, $K$ being a bounded quantity independent of $i$ and $n$ ,while $S$ is the expected value of the steady state distribution $\pi$ and so bounded below from zero and above from $i^*$. From (26), Assumption 3 and the properties (11) and (12) of $\delta_t$ it will then follow that $\delta_t' P^n O$ is equal to (22) and the proof will be complete.

We now prove (24). We consider first the case $i^* = s$. Given Assumption 2', there exists a state $0 \leq i < i^* - 1$ such that $p_i \neq 1$ and the Markov chain $\tilde{P}$ is both irreducible and aperiodic. In this case we know that a steady state distribution $\pi$ of dimension $s \times 1$ does exist and that the rate of convergence is exponential and independent of the initial distribution (see e.g. Stokey and Lucas 1988, theorem 11-4), that is (24) holds.

We now consider the case where $i^* < s$. In this case the matrix $\tilde{P}$ is reducible and we know that the first $i^* < s$ states of the Markov chain $\tilde{P}$ are recurrent while all the other $s - i^*$ are transient. The structure of $\tilde{P}$ implies that a unit starting from one transient state $i^* < i \leq s$ will enter the recurrent class after a number of periods less or equal than $s - i^*$ . We then know that (24) holds starting from the $s - i^*$ $th$ iterations of $\tilde{P}^n$ so that there can always be found a bounded constant $H$ such that

$$|r_i^n - \pi'| \leq H \, \rho^n 1', \;\; \forall i < s, \;\; \forall n > 0, \;\; 0 \leq \rho < 1.$$

This proves (24).

We now prove (25). From (23) and (24) we obtain that $\forall i \geq s$, $\forall n > 0$

$$\left| r_i^n - \pi' \left(1 - \beta_{n-1}^i\right) - \beta_{n-1}^i \left(1 - p_{i+n-1}\right) e_1' \right| \leq H \left(1 - \beta_{n-1}^i\right) \rho^n 1' \leq H \, \rho^n 1',$$

$0 \leq \rho < 1$. This proves (25). Q.E.D.

**Proof of Proposition 1 (Increasing hazard functions)** To prove the assertion we proceed as follows. Firstly we show that Assumptions 1, 2' and 3 hold. We then apply Lemma 6, so that $\delta_t' P^n O$ is equal to (22). We then show that $A$, $B$, $C$ and $D$ are all $O(\rho^n)$, $0 < \rho < 1$ where $O(\rho^n)$ indicates a quantity at most of order $\rho^n$ that is $\lim_{n\to\infty} \frac{O(\rho^n)}{\rho^n} < \infty$. This will imply that $\delta_t' P^n O$ is also $O(\rho^n)$ concluding the proof.

If the probabilities $p_i$ are decreasing in $i$, for $i \geq \kappa$, $\beta_j^i$ falls at a rate that asymptotically is at least as great as the exponential one, $\forall i$. In fact for large $j$

$$\beta_j^i = \beta_\kappa^i \beta_{j-\kappa}^\kappa \leq \beta_\kappa^i (p_\kappa)^{j-\kappa} \leq (p_\kappa)^{j-\kappa} \tag{27}$$

so that Assumption 1 holds.

As assumptions 1, 2' and 3 hold, Lemma 6 guarantees that $\delta_t' P^n O$ is equal to (22). We know that $A$ is always $O(\rho^n)$, $0 \leq \rho < 1$. Assumption 3 together with (27) guarantees that also $B$, $C$ and $D$ are $O(\rho^n)$, $0 \leq \rho < 1$. Q.E.D.

**Proof of Proposition 2 (Robust typical spectral shapes)** In order to prove the assertion we want to show that $\phi_n = \gamma \delta_t' P^n O \sim n^{1-h} = n^{d-1}$, $d = 2 - h$. To prove the assertion we proceed as follows. Firstly we show that Assumptions 1, 2' and 3 hold. We then apply Lemma 6, so that $\delta_t' P^n O$ is equal to (22). We know that $A = O(\rho^n)$, $0 \leq \rho < 1$. We then show that $B \sim n^{1-h}$, $C = O(n^{1-h})$ while $D \sim n^{-h}$. Consequently $\phi_n = \gamma \delta_t' P^n O \sim n^{1-h}$, and it will conclude the proof.

Assumptions 2' and 3 hold by hypothesis. We now show that $\beta_n^{s^*} \sim \beta_n^s \sim n^{-h}$ so that Assumption 1 is also satisfied. In fact, from the recursion of the Gamma function (see Abramowitz and Stegun, 1972, formula 6.1.15)

$$\Gamma(\alpha+1) = \alpha \Gamma(\alpha), \tag{28}$$

it follows that

$$\begin{aligned}
\beta_n^{s^*} &= \left(1 - \frac{h}{s^*+1}\right)\left(1 - \frac{h}{s^*+2}\right)\cdots\left(1 - \frac{h}{s^*+n}\right) = \\
&= \frac{(s^*)!}{(s^*+n)!}(s^*+1-h)(s^*+2-h)\ldots(s^*+n-h) = \\
&= \frac{\Gamma(s^*+1)}{\Gamma(s^*+n+1)}\frac{\Gamma(s^*+n+1-h)}{\Gamma(s^*+1-h)} \sim \beta_n^s \sim n^{-h} \; as \; n \uparrow \infty, \quad (29)
\end{aligned}$$

where the last asymptotic equivalence used the fact that

$$\lim_{n\to\infty} n^{b-a} \frac{\Gamma(n+a)}{\Gamma(n+b)} = 1. \tag{30}$$

See Abramowitz and Stegun (1972), formula 6.1.46.

As Assumptions 1, 2' and 3 hold, by Lemma 6, $\delta_t' P^n O$ is equal to (22).

We now show that $B \sim n\,\beta_n^s$ so that $B \sim n^{1-h}$ by (29). From equation $(B)$ and the identity $\frac{\beta_{n+i}^s}{\beta_i^s} = \beta_n^{s+i}$ we obtain

$$B = n\,\beta_n^s \sum_{i=0}^{\infty} \frac{\delta_t^{s+i+1}}{\beta_i^s} \beta_i^{s+n} \ ,$$

which by Assumption $(A4)$ we know to be bounded for all $n$. $\beta_i^{s+n}$ is a probability and thus is a strictly positive quantity bounded above by one and below by zero. From the same reasoning as that which led to (29) we obtain that for $n \geq s^* - s$

$$\beta_i^{s+n} = \frac{\Gamma(n+s+1)}{\Gamma(n+s+1-h)} \frac{\Gamma(n+s+i+1-h)}{\Gamma(n+s+1+i)}. \tag{31}$$

Condition $(A4)$, the fact that by (31) and (30) $\lim_{n\to\infty} \beta_i^{s+n} = 1, \forall i$, together with the Lebesgue dominated convergence theorem guarantees that

$$\lim_{n\to\infty} \sum_{i=0}^{\infty} \frac{\left|\delta_t^{s+i+1}\right|}{\beta_i^s} \beta_i^{s+n} = \sum_{i=0}^{\infty} \frac{\left|\delta_t^{s+i+1}\right|}{\beta_i^s} < \infty. \tag{32}$$

Given (12) it follows that

$$\lim_{n\to\infty} \frac{|B|}{n\,\beta_n^s} = Z$$

where $0 < Z = \sum_{i=0}^{\infty} \frac{\left|\delta_t^{s+i+1}\right|}{\beta_i^s} < \infty$ by condition $(A4)$. It follows that $B \sim n\,\beta_n^s$.

We now show that $C = O(n^{1-h})$ where $O(n^{1-h})$ indicates a quantity at most of order $n^{1-h}$ that is $\lim_{n\to\infty} \frac{O(n^{1-h})}{n^{1-h}} < \infty$. To prove it we show two preliminary results.

Firstly from $(C)$ and the identity $\frac{\beta_{n+i}^s}{\beta_i^s} = \beta_n^{s+i}$ we obtain that

$$\begin{aligned} C &= \sum_{i=0}^{\infty} \delta_t^{s+i+1} \beta_n^{s+i} (s+i) = \\ &= \beta_n^s \sum_{i=0}^{\infty} \frac{\delta_t^{s+i+1}}{\beta_i^s} \beta_i^{s+n} \ (s+i) . \end{aligned} \tag{$C$'}$$

Secondly, we prove the following result concerning sums of Gamma functions

$$\sum_{i=1}^{\infty} \frac{\Gamma(b-1-a+i)}{\Gamma(b+i)} = \frac{\Gamma(b-a)}{a\Gamma(b)}, \ \forall b > a > 0, \tag{33}$$

for any integer $b$. If $a$ is an integer, (33) can be proved by induction after using the identity

$$\sum_{i=1}^{\infty}\frac{(i-1)!}{(a+i)!}=\frac{1}{a\,a!}$$

(see Gradshteyn and Ryzhik , 1997, formula 0.247) and the fact that

$$\sum_{i=b-a}^{\infty}\frac{(i-1)!}{(a+i)!}=\sum_{i=1}^{\infty}\frac{\Gamma\left(b-1-a+i\right)}{\Gamma\left(b+i\right)}.$$

If $a$ is not an integer, from the series expansion of the function $(1-L)^{a}$ one obtains that for any $L$ (see Granger and Joyeux, 1980, p. 18)

$$(1-L)^{a}=\sum_{i=0}^{\infty}\frac{\Gamma\left(i-a\right)}{\Gamma\left(-a\right)\Gamma\left(i+1\right)}L^{i},\ \forall a>0$$

so that for $L=1$ we obtain that

$$\sum_{i=0}^{\infty}\frac{\Gamma\left(i-a\right)}{\Gamma\left(i+1\right)}=0.$$

Therefore

$$\sum_{i=1}^{\infty}\frac{\Gamma\left(b-1-a+i\right)}{\Gamma\left(b+i\right)}=-\sum_{i=0}^{b-1}\frac{\Gamma\left(i-a\right)}{\Gamma\left(i+1\right)}.$$

We now show by induction that

$$-\sum_{i=0}^{b-1}\frac{\Gamma\left(i-a\right)}{\Gamma\left(i+1\right)}=\frac{\Gamma\left(b-a\right)}{a\Gamma\left(b\right)}$$

which is what is required to prove (33). In fact for $b=1$, we obtain

$$-\sum_{i=0}^{b-1}\frac{\Gamma\left(i-a\right)}{\Gamma\left(i+1\right)}=-\frac{\Gamma\left(-a\right)}{\Gamma\left(1\right)}=\frac{\Gamma\left(1-a\right)}{a\Gamma\left(1\right)}.$$

If we assume that the formula holds for $b=b$ we then obtain that for $b=b+1$

$$-\sum_{i=0}^{b}\frac{\Gamma\left(i-a\right)}{\Gamma\left(i+1\right)}=\frac{\Gamma\left(b-a\right)}{a\Gamma\left(b\right)}-\frac{\Gamma\left(b-a\right)}{\Gamma\left(b+1\right)}=\frac{\left(b-a\right)\Gamma\left(b-a\right)}{a\Gamma\left(b+1\right)}=\frac{\Gamma\left(b+1-a\right)}{a\Gamma\left(b+1\right)},$$

which completes the proof by induction of (33).

To prove that $C = O(n^{1-h})$ we distinguish the case where $h > 1$ from that in which $1 \leq h$. Consider first the case in which $h > 1$. From assumption $(A4)$ it follows that

$$\frac{\left|\delta_t^{s+i+1}\right|}{\beta_i^s}\ (s+i) < H,\ \forall i$$

and

$$\exists i \ni \frac{\left|\delta_t^{s+i+1}\right|}{\beta_i^s}\ (s+i) > K$$

for some strictly positive bounded constants $H$ and $K$. From this together with ($C$'), (31), (33) and (30) it follows that

$$\begin{aligned}|C| &\sim \beta_n^s \frac{\Gamma(n+s+1)}{\Gamma(n+s+1-h)} \sum_{i=1}^{\infty} \frac{\Gamma(n+s+i-h)}{\Gamma(n+s+i)} = \\ &= \beta_n^s \frac{\Gamma(n+s+1)}{\Gamma(n+s+1-h)} \frac{\Gamma(n+s+1-h)}{(h-1)\Gamma(n+s)} \sim n\beta_n^s.\end{aligned}$$

We now consider the case $h \leq 1$. In this case

$$C = (s+n)\ \beta_n^s \sum_{i=0}^{\infty} \frac{\delta_t^{s+i+1}}{\beta_i^s} \frac{\beta_i^{s+n}}{(s+n)}\ (s+i)\,,$$

which by Assumption 3 and Lemma 4 we know to be bounded for all $n$. We now show that if $h \leq 1$, the quantity $\frac{\beta_i^{s+n}}{(s+n)}$ is always decreasing in $n$. (12), the fact that $\lim_{n\to\infty} \frac{\beta_i^{s+n}}{(s+n)} = 0$, together with the Lebesgue dominated convergence theorem will then imply that if $h \leq 1$, $C = o\left(n\,\beta_n^s\right)$, that is $\lim_{n\to\infty} \frac{C}{n\,\beta_n^s} = 0$. So that in general for any $h$, $C = O(n\,\beta_n^s) = O(n^{1-h})$ by (29).

Therefore we now show that the derivative of

$$\frac{\beta_i^{s+n}}{s+n} = \frac{\Gamma(n+s+1)}{(s+n)\,\Gamma(n+s+1-h)} \frac{\Gamma(n+s+i+1-h)}{\Gamma(n+s+1+i)} \tag{34}$$

with respect to $n$ is negative. In fact, once we define the Psi-function $\psi(\cdot)$

$$\psi(x) = \frac{d\left[\ln \Gamma(x)\right]}{dx} = \frac{\Gamma'(x)}{\Gamma(x)}$$

we obtain that

$$\frac{d\left(\ln\frac{\beta_i^{s+n}}{s+n}\right)}{dn} = [\psi(n+s+1)+\psi(n+s+i+1-h)-\psi(n+s+1-h)$$
$$- \psi(n+s+i+1)-\frac{1}{s+n}] \quad (35)$$

which is equal to

$$[\psi(n+s)-\psi(n+s+1-h)+\psi(n+s+i+1-h)-\psi(n+s+i+1)] \quad (36)$$

after using the recursion formula $\psi(z+1)-\frac{1}{z}=\psi(z)$ (see Abramowitz and Stegun, 1972, formula 6.3.5). $h\leq 1$, together with the strictly increasing nature of $\psi(\cdot)$ over the positive real line (see Abramowitz and Stegun, 1972, formula 6.3.16), implies that (36) and consequently (35) are negative.

Finally we show that $D\sim\beta_n^s$ so that $D\sim n^{-h}$ by (29). From equation $(D)$ we obtain

$$D=\beta_{n-1}^s\sum_{i=0}^{\infty}\frac{\delta_t^{s+i+1}}{\beta_i^s}\beta_i^{s+n-1},$$

which by Assumption 3 and Lemma 4 we know to be bounded for all $n$. $\beta_i^{s+n-1}$ is a probability and thus is a strictly positive quantity bounded above by one and below from zero. Condition $(A4)$, (12), the fact that by (31) and (30), $\lim_{n\to\infty}\beta_i^{s+n}=1$, $\forall i$, together with the Lebesgue dominated convergence theorem imply that

$$\lim_{n\to\infty}\frac{|D|}{\beta_n^s}=Z$$

where $0<Z=\sum_{i=0}^{\infty}\frac{\left|\delta_t^{s+i+1}\right|}{\beta_i^s}<\infty$ by condition $(A4)$. Q.E.D.

## 7.3 Proofs of results in section 5

**Proof of Proposition 3 (Counter-example)** We want to show that $\phi_n=\gamma\delta_t'P^nO\sim n^{d-1}$, where $d=3-h$. To prove the assertion we proceed as follows. Firstly we show that Assumptions 1, 2' and 3 hold. We then apply Lemma 6, so that $\delta_t'P^nO$ is equal to (22). We know that $A=O(\rho^n)$ $0\leq\rho<1$. We then show that $B\sim n^{2-h}$, $C\sim n^{2-h}$ while $D\sim n^{1-h}$. Consequently $\phi_n=\gamma\delta_t'P^nO\sim n^{2-h}$, and completing the proof.

We now show that Assumptions 1 and 3 hold (Assumption 2' holds by hypothesis). We first note that, given $(A4')$

$$\beta_i^s = \frac{\Gamma(s+1)}{\Gamma(s+1-h)} \frac{\Gamma(s+i+1-h)}{\Gamma(s+1+i)} \tag{37}$$

and then that

$$\beta_n^{s+i} = \frac{\Gamma(i+s+1)}{\Gamma(i+s+1-h)} \frac{\Gamma(i+s+n+1-h)}{\Gamma(i+s+1+n)}. \tag{38}$$

As $h > 2$, (30) together with (37) and (38) implies that Assumptions 1 and 3 are satisfied. By Lemma 6 $\delta_t' P^n O$ is then equal to (22). (37) and (38) imply that

$$\beta_i^s \beta_n^{s+i} = \frac{\Gamma(s+1)}{\Gamma(s+1-h)} \frac{\Gamma(i+s+n+1-h)}{\Gamma(i+s+1+n)},$$

so that from ($A4$')

$$B = -a \frac{\Gamma(s+1)}{\Gamma(s+1-h)} n \sum_{i=0}^{\infty} \frac{\Gamma(i+s+n+1-h)}{\Gamma(i+s+1+n)}, \tag{$B$"}$$

$$C = -a \frac{\Gamma(s+1)}{\Gamma(s+1-h)} \sum_{i=0}^{\infty} \frac{\Gamma(i+s+n+1-h)}{\Gamma(i+s+1+n)} (s+i), \tag{$C$"}$$

$$D = -aS \frac{\Gamma(s+1)}{\Gamma(s+1-h)} \sum_{i=0}^{\infty} \frac{\Gamma(i+s+n-h)}{\Gamma(i+s+n)}. \tag{$D$"}$$

We now show that $B \sim n^{2-h}$. In fact (33), (30) and ($B$") imply that

$$B = -a \frac{\Gamma(s+1)}{\Gamma(s+1-h)} \frac{\Gamma(s+n+1-h)}{(h-1)\Gamma(s+n)} n \sim n^{2-h}.$$

We now show that $C \sim n^{2-h}$. To do so, firstly we proove that for any integer $b$

$$\sum_{i=1}^{\infty} i \frac{\Gamma(b-2-a+i)}{\Gamma(b+i)} = \frac{\Gamma(b-1-a)}{a(1+a)\Gamma(b-1)}, \ \forall a > 0. \tag{39}$$

Indeed (33) implies that $\sum_{i=1}^{\infty} i \frac{\Gamma(b-2-a+i)}{\Gamma(b+i)}$ is equal to

$$\sum_{i=1}^{\infty} \frac{\Gamma(b-2-a+i)}{\Gamma(b+i-1)} - (b-1) \sum_{i=1}^{\infty} \frac{\Gamma(b-2-a+i)}{\Gamma(b+i)} = \frac{\Gamma(b-1-a)}{a(1+a)\Gamma(b-1)},$$

which proves (39).

(30), (33) and ($C$") together with (39) imply that

$$C = -a \frac{\Gamma(s+1)}{\Gamma(s+1-h)} \frac{(h-1)(s-1)+n}{(h-2)(h-1)} \frac{\Gamma(-h+s+n+1)}{\Gamma(s+n)} \sim n^{2-h}.$$

Finally we show that $D \sim n^{1-h}$. In fact (33), (30) and ($D$") imply that

$$D = -aS \frac{\Gamma(s+1)}{\Gamma(s+1-h)} \frac{\Gamma(s+n-h)}{(h-1)\Gamma(s+n-1)} \sim n^{1-h}.$$

Q.E.D.

**Corollary (The reallocation structure matters)** Given Assumptions 1, 2' and 3, the assertion follows from Lemma 6 and Proposition 2 and 3. For example a reallocation structure $\delta_t$ that reallocates only inside the recurrent class, that is $\delta_t^i = 0, \forall i > s$, given (22) produces $\delta_t' P^n O = A = O(\rho^n)$, $0 \leq \rho < 1$, independent of the structure of the transmission mechanism $P$. On the other hand a reallocation structure that also reallocates outside the recurrent class, that is $\exists\ \delta_t^i \neq 0$, for some $i > s$, can generate a rate of decay in the quantities $\phi_n = \gamma\, \delta_t' P^n O$ that might be slower than the exponential rate. See Propositions 2 and 3. Q.E.D.

## REFERENCES

**Abramowitz, M. and Stegun I. (1972)**, *Handbook of Mathematical Functions,* Dover Publications, Inc., New York.

**Acemoglu, D. and Scott, A. (1995)**, "Asymmetric Business Cycles: Theory and Time Series Evidence", *MIT Working Paper*, 95-24, August.

**Aghion, P. and Howitt, P. (1994),** "Growth and Unemployment", *Review of Economic Studies*, July, 61-3, 477-494.

**Aghion, P. and Saint-Paul, G. (1991)**, "On the Virtue of Bad Times: an Analysis of the Interaction between Economic Fluctuations and Productivity Growth", *CEPR Discussion Paper* No. 578, October.

**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.

**Baily, M.N. Hulten, C. and Campbell, D. (1992)**, "Productivity Dynamics in Manufacturing Plants", *Brookings Papers on Economic Activity: Microeconomics*, 187-249.

**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.

**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).

**Bertola, G. and Caballero, R. (1994)**, "Irreversibility and Aggregate Investment", *Review of Economic Studies*, 61, 223-246.

**Billingsley,P. (1986)**, *Probability and Measure*, 2nd ed. New York: John Wiley & Sons.

**Brockwell, P. and Davis, R. (1991)**, *Time Series: Theory and Methods*, Springer Verlag New York.

**Chan, N. and Terrin, N. (1995)**, "Inference for Unstable Long-Memory Processes with Applications to Fractional Unit Root Autoregressions", *Annals of Statistics,* 23-5, 1662-1683.

**Caballero, R. (1992)**, "A Fallacy of Composition", *American Economic Review,* December, 1279-1292.

**Caballero, R. and Engel, E. M.R.A. (1991)**, "Dynamic (S,s) Economies", *Econometrica*, 59, 1659-1686.

**Caballero, R. and Engel, E.M.R.A. (1993)**, "Heterogeneity and Output Fluctuations in a Dynamic Menu-Cost Economy", *Review of Economic Studies*, 60, 95-119.

**Caballero, R. and Engel, E. M.R.A. (1998)**, "Nonlinear Aggregate Investment Dynamics: Theory and Evidence", *NBER Working Paper Series*, No. 6420, February.

**Caballero, R. and Hammour, M. (1994)**, "The Cleansing Effect of Recessions", *American Economic Review,* 84-5, December, 1350-1368.

**Caballero, R.and Hammour, M. (1996)**, "On the Timing and Efficiency of Creative Destruction", *Quarterly Journal of Economics*, 111-3, August, 805-852.

**Caballero, R., Engel, E. M.R.A. and Haltiwanger, J (1997)**, "Aggregate Employment Dynamics: Building from Microeconomic Evidence", *American Economic Review*, 87-1, March, 115–137.

**Cooper, R. and A. John, (1988)**, "Coordinating Coordination Failures in Keynesian Models", *Quarterly Journal of Economics,* 13, 441-465.

**Davis, S. and Haltiwanger, J. (1990)**, "Gross Job Creation and Destruction: Microeconomic Evidence and Macroeconomics Implications", *NBER Macroeconomics Annual* , 5, 99-113.

**Davis, S. and Haltiwanger, J. (1992)**, "Gross Job Creation, Gross Job Destruction and Employment Reallocation", *Quarterly Journal of Economics* 107, 819-63.

**Davis, S., Haltiwanger, J. and Schuh, S. (1993)**, "Small Business and Job Creation: Dissecting the Myth and Reassessing the Facts", *NBER Working Paper Series,* no. 4492, October.

**Diebold, F. and Rudebusch, G. (1989)**,"Long Memory and Persistence in Aggregate Output", *Journal of Monetary Economics,* 24, 189-209.

**Diebold, F. and Senhadji, A. (1996)**,"The Uncertain Unit Root in real GNP: Comment", *American Economic Review, 86,* 1291-1298.

**Dunne, T. Roberts, M. and Samuelson, L. (1989)**, "The Growth and Failure of U.S. Manufacturing Plants", *Quarterly Journal of Economics,* 104-4, November, 671-98.

**Durlauf, S. (1991)**, "Multiple Equilibria and Persistence in Aggregate Fluctuations", *American Economic Review Papers and Proceedings,* May, 81-2, 70-74.

**Evans, D.S. (1987)**, "Tests of Alternative Theories of Firm Growth", *Journal of Political Economy*, 87-4, 657-674.

**Gali, J. (1996)**, "Technology, Employment and the Business Cycle: do technology shocks explain aggregate fluctuations?", *CEPR Discussion Paper Series* no. 1499.

**Geweke, J. and Porter Hudak, S. (1983)**, "The Estimation and Application of Long Memory Time Series Models", *Journal of Time Series Analisys*, 4, 221-238.

**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.*

**Gil-Alana, L. and Robinson, P. M. (1997)**, "Testing Unit Root and other Non-Stationary Hypotheses in Macroeconomic Time Series", *Journal of Econometrics,* 80-2, October, 223-239.

**Giraitis, L. Robinson, P M and Samarov A. (1997)**, "Rate Optimal Semiparametric Estimation of the Memory Parameter of the Gaussian Time Series with Long-Range Dependence", *Journal of Time Series Analysis, 18-1*, 49 -60.

**Gradshteyn, I.S. and Ryzhik I. M. (1997)**, *Table of Integrals, Series, and Products,* Fifth Edition, Academic Press Limited, London UK.

**Granger, C.W.J. (1966)**, "The Typical Spectral Shape of an Economic Variable", *Econometrica* 34-1, 150-161.

**Granger, C.W.J. and Joyeux, R. (1980)**, "An Introduction to Long-Memory Time Series Models and Fractionally Differencing", *Journal of Time Series Analysis*, 1-1, 15-29.

**Grossman, G. and Helpman E. (1991)**, *Innovation and Growth in the Global Economy,* Cambridge MA, MIT Press.

**Hall, B. (1987)**, "The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector", *The Journal of Industrial Economics,* 35-4, June, 583-606

**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,* 16-1, 17-41.

**Jones,C. (1995)**, "Time Series Tests of Endogenous Growth Models", *Quarterly Journal of Economics*, 110, 495-525.

**Jones, R. and Newman, G. (1995)**, "Adaptive Capital, Information Depreciation and Schumpeterian Growth", *Economic Journal 105,* July, 897-915.

**Jovanovic, B. and Nyarko, Y. (1996)**, "Learning by Doing and the Choice of Technology", *Econometrica*, Vol. 64, No. 6, November, 1299-1310.

**Kalecki, M. (1945)**, "On the Gibrat Distribution", *Econometrica*, April, 13-2, 161-170.

**Karlin, S. and Taylor, H.M. (1975)**, *A First Course in Stochastic Processes*, Academic Press, Inc. (London) LTD.

**Karlin, S. and Taylor, H.M. (1981)**, *A Second Course in Stochastic Processes*, Academic Press, Inc. (London) LTD.

**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.

**Mansfield, E. (1962)**, "Entry, Gibrat's Law, Innovation, and the Growth of Firms", *American Economic Review,* 52-5, December, 1023-1051.

**Michelacci, C. (1998)**, "The Macroeconometrics of Cross-Sectional Heterogeneity", *Mimeo, CEMFI,* Madrid.

**Michelacci, C. and Zaffaroni, P. (1998)**, "(Fractional) Beta Convergence", *Journal of Monetary Economics, forthcoming.*

**Mortensen, D.T. and Pissarides, C. (1998)**, "Technological Progress, Job Creation, and Job Destruction", *Review of Economic Dynamics*, 1-4, October, 733-753.

**Nelson, C. and Plosser, C. (1982)**, "Trends and Random Walks in Macroeconomic Time Series: some Evidence and Implications", *Journal of Monetary Economics*, 10, 139-162.

**Perron, P. (1989)**, "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis", *Econometrica, 57, 1361-1401.*

**Pissarides, C. A. (1990)**, *Equilibrium Unemployment Theory*, Oxford: Basil Blackwell 1990.

**Priestley, M. (1965)**, "Evolutionary Spectra and Non-Stationary Processes", *Journal of the Royal Statistical Association,* 27, 204-237.

**Quah, D. (1993)**, "Galton's fallacy and tests of the convergence hypothesis", *Scandinavian Journal of Economics,* 95-4, December, 427-443.

**Robinson, P. M. (1994)**, "Time Series with Strong Dependence", in *Advances in Econometrics, Sixth World Congress*, ed. by C.A. Sims, Vol. 1, 47- 95, Cambridge University Press.

**Robinson, P.M. (1995)**, "Log-Periodogram Regression of Time Series with Long Range Dependence", *Annals of Statistics* , 23-3, 1048-1072.

**Robinson, P.M. and Marinucci, D. (1997)**, "Semiparametric Frequency-Domain Analysis of Fractional Cointegration", *Mimeo, London School of Economics.*

**Rotemberg, J. and Woodford, M. (1996)**, "Real-Business-Cycle Models and the Forecastable Movements in Output, Hours, and Consumption", *American Economic Review,* 86-1, March, 71-89.

**Saint-Paul, G. (1993)**, "Productivity Growth and the Structure of the Business Cycle", *European Economic Review* 37, 861-890.

**Seneta, E. (1976)**, *Regularly Varying Functions,* Springer, Berlin.

**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.

**Stokey, N. Lucas, R.E. and Prescotts, E. (1988)**, *Recursive Methods in Economic Dynamics*, Cambridge MA, Harvard University Press.

**Sutton, J. (1995)**, "The Size Distribution of Business Part I: a Benchmark Case", *The Economics of Industry Group, Discussion Paper Series EI/9,* December.

**Sutton, J. (1996)**, "Gibrat's Legacy", *The Economics of Industry Group, Discussion Paper Series EI/14,* October.

**Velasco, C. (1996)**, "Non-Stationary Log-Periodogram Regression", *Mimeo, London School of Economics.*