DISCUSSION PAPER SERIES

No. 4302

CROSS-SECTIONAL HETEROGENEITY AND THE PERSISTENCE OF AGGREGATE FLUCTUATIONS

Claudio Michelacci

LABOUR ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP4302.asp

ISSN 0265-8003

CROSS-SECTIONAL HETEROGENEITY AND THE PERSISTENCE OF AGGREGATE FLUCTUATIONS

Claudio Michelacci, Centre for Monetary and Financial Studies (CEMFI) and CEPR

Discussion Paper No. 4302
March 2004

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 4302

March 2004

ABSTRACT

Cross-Sectional Heterogeneity and the Persistence of Aggregate Fluctuations*

The micro evidence indicates that small firms grow faster than big firms. I argue that this relationship between the expected growth rate of a firm and its size may provide a micro foundation for the well-known high degree of persistence of shocks to aggregate output. The logic goes as follows. Almost any shock tends to temporarily alter firms' incentives to invest in growth thereby leading to a reallocation of firms across size categories. If small firms grow faster than big ones, the impact effect of the shock on aggregate output is gradually absorbed. But, as fast growing small firms become big and start to grow at the lower rate of big firms, the rate at which the shock is absorbed decreases over the adjustment path. As a result, shocks are absorbed, yet at a very low decreasing rate that induces long memory in aggregate output. I argue that this transmission mechanism may reconcile 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.

JEL Classification: C43, E10, E32 and L11
Keywords: Gibrat's law, long memory and vintage models

Claudio Michelacci
CEMFI
Casado del Alisal 5
28014 Madrid
SPAIN
Tel: (34 91) 4290 551
Fax: (34 91) 4291 056
Email: c.michelacci@cemfi.es

*I would like to thank Orazio Attanasio, Cristopher Bliss, Ricardo Caballero, Liudas Giraitis, Marco Lippi, Domenico Marinucci, Chris Pissarides, Peter M Robinson, Javier Suarez and Paolo Zaffaroni for very useful conversations and suggestions. I am also particularly grateful to a referee for his many detailed comments. The usual disclaimer applies.

Submitted 12 February 2004

NON-TECHNICAL SUMMARY

1 Introduction

It is well known from time series analysis that shocks to (detrended) aggregate output have very persistent effects. Since Nelson and Plosser (1982) have argued that GDP exhibits a unit root —and therefore that temporary shocks have permanent effects on the level of output—, many studies have debated about the exact degree of aggregate persistence. But what type of firm behavior lies behind the persistence of shocks observed in the data? In this paper I argue that the empirical relationship between the expected growth rate of a firm and its size may provide a microfoundation for such aggregate persistence.

Gibrat (1931) first investigated the relationship between expected growth 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. Albeit not conclusive, more recent studies question Gibrat's law and argue that small firms tend to grow faster than big firms.[1]

To see the implications of these findings for aggregate persistence, suppose at first that Gibrat's law holds. When so, any aggregate shock that reallocates firms across sizes has a permanent effect on the level of output, inducing a unit root in its time series formulation.[2] Indeed, once the shock hits the system, firms are reallocated across sizes. But then, and given Gibrat's law, firms keep growing at the same rate, thereby perpetuating forever the impact effect of the shock on aggregate output. On the contrary, in a world where Gibrat's law fails and small firms grow faster than big ones, the same shock is absorbed, yet at very low decreasing rates. Indeed, as small firms grow faster than big ones, the initial effect of the shock on output is gradually absorbed. But, as fast growing small firms become big and grow at the lower rate of big firms, the rate at which the shock is absorbed decreases over the adjustment path. Thus, the empirical relationship between firm growth and firm size suggests that the persistence of aggregate fluctuations is very high, that shocks are absorbed

[1] Sutton (1997) surveys the empirical debate on Gibrat's law. For more direct empirical evidence see, among others, Evans (1987) and Dunne et al. (1989).

[2] The idea that Gibrat's law and a unit root in output are closely related was implicitly contained in Kalecki (1945). Indeed 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 continuously". As a matter of fact a distinctive feature of a process with a unit root is that its variance is a linear function of time.

and that the rate of absorption is decreasing over the adjustment process.

To give further economic content to the claim, I consider a version of the standard Solow (1960) vintage model where older capital vintages must be replaced with more recent ones in order to reap the productivity gains of technological change. In the model firms using vintages far away from (close to) the technological frontier are *small* (*big*) since they produce less (greater) output. At each point in time, a firm weighs the benefits of switching to a better technology with the opportunity cost of investing part of its own resources in technological adoption. These costs vary across firms as well as over time. Importantly, I assume that they tend to be larger the more obsolete is the technology currently operated by the firm.[3] Thus the probability of moving into the technological lead decreases as the firm falls behind in the technological ladder.

I focus on whether temporary common shocks to firms may translate into persistent changes in the level of aggregate output. I consider aggregate shocks to the opportunity cost of technological adoption that cause a reallocation of firms across technological vintages. 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 present in the model, i.e. the different responses of the existing firms and the way they add up into aggregate output.

To gauge the degree of persistence of a shock I borrow the notions of *long memory* and *order of integration* of a stochastic process from time series econometrics.[4] Empirical investigation suggests that the low frequency behavior of aggregate output is well characterized by a long memory process, including the unit root as a particular case.[5] A distinctive feature of a long memory process is that a shock propagates at decreasing rates which means that the rate of absorption of the shock at each stage n of the adjustment process is a decreasing function of n.

In the model, this type of dynamics arises naturally as a result of aggregating the processes of initial churning and subsequent catching up that a shock originates at the firms' level. Indeed, once the shock hits the system, firms are reallocated

[3] See Jones and Newman (1995), and Jovanovic and Nyarko (1996) for examples of models where firms using more obsolete technologies face greater costs in adopting new technologies.

[4] See Robinson (1994) for a survey on the properties of long memory processes.

[5] See Diebold and Rudebusch (1989), Gil-Alana and Robinson (1997) and Michelacci and Zaffaroni (2000) for empirical evidence supporting the claim that aggregate GDP is well approximated by a long memory process with an order of integration between zero and one.

across technological vintages. But, if the probability of moving into the technological lead decreases as the firm falls behind in the technological ladder, a firm that does not adopt a new technology today will be less likely to do it so tomorrow when, due to technological progress, its technology will be even more obsolete. Hence the probability that firms down in the technological ladder catch up decreases over the adjustment path, the shock propagates at decreasing rates and aggregate output exhibits long memory.

The model can generate any order of integration in output strictly below two and thus a unit root as a particular case. The order of integration in fact depends on the exact linkage between a firm's expected growth and the technology it uses. If firms down in the technological ladder tend to grow faster than the whole economy, the shock will be eventually absorbed, otherwise its effect can either persist forever or even get amplified without limit. An interesting particular case arises when 'Gibrat's law' holds so that all firms grow at the same rate independently of the technology currently in operation. In this case output exhibits a unit root. Indeed, after the initial churning up produced by the shock, firms keep growing at the same rate, thereby perpetuating forever the impact effect of the shock on output.

To relate my paper to some previous results in time series econometrics, I trace the logic of my findings back to Granger (1980) famous result on how the aggregation of AR(1) processes can generate a positive order of integration in the aggregate. I show that some features of my steady state distribution of firms' technologies resemble the conditions needed for Granger's result to hold.

The remainder of the paper is divided into 6 sections. Section 2 introduces my metrics for aggregate persistence. Section 3 describes the model. Section 4 gives conditions to generate long memory. Section 5 discusses how to extend the results to more general set-ups. Section 6 relates my findings to Granger (1980). Section 7 concludes. The appendix contains the derivation of all the results of the paper.

2 Measuring aggregate persistence

Standard measures of persistence are based on the related notions of impulse response and Wold representation. More formally, the Wold representation of a time series

y_t, $t \geq 0$, (if it exists) reads like

$$y_t = y_0 + \nu t + \sum_{n=0}^{t} \phi_n \, \epsilon_{t-n}, \tag{1}$$

where y_0 and $\nu \geq 0$ capture initial conditions and a deterministic trend, respectively. The quantities ϕ_n's are the Wold coefficients while the shocks ϵ_t are the Wold innovations. The Wold coefficient ϕ_n gauges the fraction of the shock ϵ_{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.

Let ρ denote a quantity greater than zero but strictly smaller than one, $0 \leq \rho < 1$. Then a very general way to model the rate of decay of ϕ_n consists of assuming that

$$\phi_n = d \, n^{d-1} + O\left(\rho^n\right), \tag{2}$$

where d captures the possibility that shocks are absorbed at rates slower than the exponential, while $O\left(\rho^n\right)$ indicates a quantity at most of order ρ^n, that is $\lim_{n \to \infty} \frac{O(\rho^n)}{\rho^n} < \infty$. The parameter d measures the *order of integration* of the time series.[6] If it is greater than zero, the time series exhibits *long memory,* while a d equal to 0 implies *weak memory*.

The representation (2) nests standard time series model. For example, in a trend stationary process with $ARMA$ disturbance, the Wold coefficients ϕ_n's decay no more slowly than at an exponential rate, so that the parameter of fractional integration d is equal to zero. In a process with a unit root, instead, temporary shocks have permanent effects on the level of the time series. Thus the Wold coefficients approach a constant and d is equal to 1. $ARIMA$ processes are, however, restrictive in allowing only for specific rates of propagation of the shocks. Notice that, at each stage n of the adjustment process, the fraction of the still unabsorbed part of the shock which will be absorbed by stage $n+1$ is equal to $1 - \frac{\phi_{n+1}}{\phi_n}$. Thus in $ARIMA$ processes, such a fraction is non decreasing in n and bounded below by $1 - \rho$. Therefore, $ARIMA$ models show a solution of continuity in approximating, at the limit, the case of the unit root: in this environment, shocks are either absorbed at constant (or increasing) rates or have permanent effects.

[6] See for example Beran (1994) and Robinson (1994) for a description of the distinctive properties of long versus weak memory processes.

An order of integration different from zero, $d \neq 0$, allows for the possibility of decreasing rates of absorption. To see this, notice that long memory implies that the Wold coefficients in (2) behave like n^{d-1} which satisfies $n^{d-1} \sim (1-\frac{1-d}{1})(1-\frac{1-d}{2}) \cdots (1-\frac{1-d}{n})$.[7] Thus the fraction of shock absorbed at each stage n of the adjustment process is $\frac{1-d}{n+1}$, which is (in absolute value) strictly decreasing in n. Hence, long memory allows for a variety of intermediate cases, and smoothly bridges the gap between the degree of persistence associated with the unit root and the constant (or increasing) rates of absorption associated with $ARIMA$ processes.

The empirical evidence supports the claim that aggregate GDP is well approximated by a long memory process with order of integration strictly between zero and one. That means that (differently from a unit root process) the effect of shocks vanishes over time, but at rate (much) slower than that implied by an arbitrary $ARMA$ process. Diebold and Rudebusch (1989) and Michelacci and Zaffaroni (2000) consider log-periodogram regressions and show that the order of integration of the GDP per capita of several OECD countries, is between zero and one.[8] Jones (1995) and Diebold and Senhadji (1996) also argue that some form of mean reversion actually takes place in the data. They show how a time trend, calculated using only past information, forecasts extremely well the current level of US GDP. This implies that the new information delivered by Wold innovations is irrelevant for forecasting on very long horizons and is incompatible with a unit root in output.

I next draw on the observed empirical relationship between the expected growth of a firm and its size to provide some microfoundation for the observed degree of aggregate persistence: an order of integration (weakly) between zero and one —i.e. I include the specific case of a unit root as a possible alternative.

3 The model

This section first lays down the structure of a vintage model. It then introduces a (once-and-for-all) aggregate shock and characterizes the dynamics of the system in

[7] The symbol '$\sim$' denotes asymptotic equivalence with respect (unless otherwise specified) to n, that is that the ratio of the left- and right-hand side tends to a bounded quantity bounded away from zero as n goes to infinity.

[8] Based on a different econometric methodology, Gil-Alana and Robinson (1997) provide further evidence that US GDP is well characterized by a long memory process.

response to the shock. The model has two key ingredients. Firstly, technological adoption is *costly* since a firm must invest resources to reap the benefits of technical changes. Secondly, the costs may *vary* across firms and thus firms using the same vintage can end up adopting different technologies. Versions of the model have been extensively analyzed in the literature.[9]

3.1 The set-up

Time is discrete and the rate of technological progress is exogenous at rate ν. The economy is populated by a measure one of firms which are infinitely lived, risk-neutral and maximize expected returns in output units discounted with factor $0 < \beta < 1$.[10]

At time t, a firm distant $i \geq 0$ from the technological frontier produces an amount of goods equal to $\nu(t-i)$. I indifferently refer to i as the *state* or the *technological distance* of the firm. Analogously, $t-i$ is the firm's *technology*.

Let π_t denote the column vector of countably infinite dimension whose element in row $j \geq 1$ represents the measure of firms using technology $t-j+1$ at time t. Analogously, denote by Q an infinite dimension column vector with the property that its jth element is exactly equal to $j-1$. Hence the level of aggregate output at time t, y_t, is equal to

$$y_t = \nu t - \nu \pi_t' Q, \tag{3}$$

where "$'$" indicates the transpose operator on the given vector hereafter always taken to be a column vector.

At a given point in time t a firm in state i has two possibilities: either keeping using technology $t-i$ so that in the next period the firm will be in state $i+1$, or switching to a better technology. Technological adoption, however, involves some costs which are assumed to be fixed and independent of the technology adopted.[11] Therefore, whenever adopting a new technology, the firm always invests in the leading technology in the economy and it will be in state zero in the subsequent period. I

[9] Examples of vintage models similar to mine include Solow (1960), Caballero and Hammour (1994) and Jovanovic and Nyarko (1996).

[10] In Section 5 I discuss the model's properties when the number of firms is discrete rather than a continuum.

[11] This way of modelling adjustment costs follows, among others, Bertola and Caballero (1990), Caballero and Engel (1999), Caballero et al. (1997) and Mortensen and Pissarides (1998).

assume, very parsimoniously, that the cost of adopting a new technology consists of two components, c_i and λ, which enter additively. c_i is a deterministic component function of the state i of the firm. λ is a random variable identically independently distributed, *iid*, across units and over time with common distribution $F(\cdot)$ over the bounded positive support $\Lambda \subset \mathbb{R}^+$. λ gauges the firm-specific (opportunity) cost of investing part of its own (capital or labour) resources in technological improvements. Therefore it can be interpreted indifferently as technology or demand driven.[12]

Let s denote a binary variable which is equal to one if the firm adopts a new technology while it is zero otherwise. Then risk neutrality implies that the value of a firm $V(t, i, \lambda)$ in state i at time t, whose cost of adopting the leading technology is $c_i + \lambda$, follows the Bellman equation

$$V(t, i, \lambda) = \max_{s \in \{0,1\}} \nu(t\text{-}i) - s(c_i + \lambda) + s\beta V^e(t+1, 0) + (1\text{-}s)\,\beta V^e(t+1, i+1), \quad (4)$$

where $V^e(t, i) = \int_\Lambda V(t, i, x) dF(x)$ indicates the expected value of $V(t, i, \lambda)$ taken with respect to the random variable λ. It follows from dynamic programming arguments that the problem is well defined.[13] In particular, the value function $V(t, i, \lambda)$ is linear in t, weakly decreasing in λ and finally strictly decreasing in i if $\nu i + c_i$ is strictly increasing in i.

In general the firm decides to adopt a new technology whenever the realization of the idiosyncratic shock λ is such that

$$\beta[V^e(t+1, 0) - V^e(t+1, i+1)] \geq c_i + \lambda, \quad (5)$$

which means that the firm weights the benefits of technological adoption $\beta\,[V^e(t+1, 0) - V^e(t+1, i+1)]$ with the associated costs $\lambda + c_i$. Let $1 - p_i$ denote the probability that the event (5) occurs. Then the assumption that the idiosyncratic shocks are *iid* with distribution function $F(\cdot)$, implies that, $\forall i \geq 0$,

$$1 - p_i = F(R_i), \quad (6)$$

[12] See, for example, Aghion and Saint Paul (1998) for a model where demand shocks affect firms' incentives to adopt new technologies.

[13] 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 (4) and the corresponding sequential problem.

where

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

denotes the *reservation adjustment cost* such that a firm in state i is indifferent between adjusting or sticking to the currently used technology. In particular notice that, as the value function $V(t,i,\lambda)$ is linear in t, R_i is function of i only. Consequently, the dynamics of the state of a firm is fully described by the infinite dimensional Markov chain P given by

$$P = \begin{bmatrix} 1-p_0 & p_0 & 0 & 0 & 0 & 0 & \cdots \\ 1-p_1 & 0 & p_1 & 0 & 0 & 0 & \cdots \\ 1-p_2 & 0 & 0 & p_2 & 0 & 0 & \cdots \\ 1-p_3 & 0 & 0 & 0 & p_3 & 0 & \cdots \\ 1-p_4 & 0 & 0 & 0 & 0 & p_4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{8}$$

where the element in row j and column k represents the probability that a firm in state $j-1$ will be in state $k-1$ in the next period. P is the *transmission mechanism* in the economy as it maps the time t *cross sectional distribution* of firms technological states π_t into the distribution π_{t+1} at time $t+1$.

3.2 Structure of the Transmission Mechanism

To characterize both the dynamics and the steady state properties of the system I 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 ν and β, and the sequence of adjustment 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 to the firm problem, defined by equations (4) and (6). Proposition 1 guarantees the validity of this 'semi-structural' approach: any assumption on the transmission mechanism P is the result of a corresponding set of assumptions on the structural parameters of the model.

Proposition 1 (Validity of the 'semi-structural' approach) *Given a distribution function $F(\cdot)$, the parameters ν and β, and an arbitrary sequence of probabilities $\{p_i, i \geq 0\}$, there does exist a unique sequence of adjustment costs $\{c_i, i \geq 0\}$, whose solution to the firm problem, defined by equations* (6) *and* (7)*, is the given sequence of probabilities. For any $i \geq 0$, such adjustment costs c_i's can be obtained by solving*

$$c_i = \frac{\beta^2 \nu}{(1-\beta)^2} - \frac{\beta (c_0 + R_0)}{1-\beta} + \frac{\beta \nu (i+1)}{1-\beta} - R_i + \sum_{j=1}^{\infty} \beta^j \int_{R_{i+j}}^{R_0} F(x)\, dx. \tag{9}$$

Equation (9) allows to recover the structural interpretation of any set of assumptions on the transmission mechanism P. Interestingly, equation (9) states that the adjustment costs c_i's must grow at a rate faster than $\frac{\beta\nu(i+1)}{1-\beta}$for the reservation adjustment costs R_i's —and the associated adjustment probabilities $1 - p_i$'s— to be decreasing in technological distance i. Also the converse can be proved: whenever c_i grows faster than $\frac{\beta\nu(i+1)}{1-\beta}$, R_i is indeed decreasing in i. To understand this result notice that the quantities R_i's are set so as to equate the gain to the cost of technological adoption for a firm in state i. The gain from technological adoption is equal to the induced permanent increase in output, $\nu(i+1)$, discounted with factor β from the next period onwards. Thus such gain is worth $\frac{\beta\nu(i+1)}{1-\beta}$ which is increasing in i. But, as technological distance i increases, also the costs of technological adoption (might) rise. In particular if the adjustment costs c_i's grow at a rate faster than $\frac{\beta\nu(i+1)}{1-\beta}$, the costs of technological adoption tend to increase faster than the associated gains so that restoring equality between the two requires the R_i's to fall as i increases.[14]

Arguably, firms must sooner or later adopt new technologies if they do not want to be driven out of the market. Thus I impose throughout the analysis that, whichever its current state is, a firm will eventually adjust with probability one. More formally, let $\gamma_j^i = \prod_{k=0}^{j-1} p_{i+k}$ denote the probability that a firm currently in state i does not adjust for j consecutive periods. Then:

Assumption 1 Suppose that, $\forall i$, $\lim_{j\to\infty} \gamma_j^i = 0$.

[14] Notice that the last term in (9) is zero if R_i is constant with i, positive if R_i is decreasing with i and negative otherwise. This term is the analogous of the capital gain/loss in the arbitrage equations and mitigates the extent in which R_i should be increasing (decreasing) with i when $\frac{\beta\nu(i+1)}{1-\beta}$grows faster (slower) than c_i. I thank a referee for this observation.

The side effect of this assumption is that the transmission mechanism P exhibits one and only one *recurrent* (ergodic) class containing state zero. Specifically:

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

The existence of a unique recurrent class implies that, after a shock, the system always converges back to the original situation where all units are in the set of recurrent states. In turn, this implies that any persistence generated by the shock is not due to a shift in the 'equilibrium' of the economy.

Lemma 1 guarantees that if a steady state distribution exists, it is unique and stable. To analyze existence, one should distinguish the case where the recurrent class consists of an infinite number of states (*irreducible transmission mechanism)* from that where such number is finite (*reducible transmission mechanism).* The former case implies that each firm will visit infinitely often all the states in the economy. Given Lemma 1, the latter corresponds instead to a situation where firms end up with probability one into a finite dimensional set of states close to the technological frontier. This last will be the case if the following assumption holds:

Assumption 2 Let $\bar{\imath} = \min \left\{ j : \gamma_j = \gamma_j^0 = \prod_{k=0}^{j-1} p_k = 0,\ j \geq 1 \right\}$. Assume that $\bar{\imath} < \infty$.

Assumption 2 implies that a firm in a state sufficiently close to the technological frontier $i \leq \bar{\imath}$ adjusts in a finite number of periods so that it always remains at most at distance $\bar{\imath}$ from the technological frontier. In the paper I consider a technical modification of Assumption 2 and I refer to it as Assumption 2'. It ensures that, once entered the recurrent class, units do not jump deterministically from one state to the other.[15]

Assumption 2' Let Assumption 2 hold and suppose that if $\bar{\imath} > 1$, it does exist $1 \leq j < \bar{\imath}$ such that $\gamma_j \neq 1$.

[15]Positing Assumption 2 rather than 2' would not affect any results of the paper, except those concerning the existence of a steady state distribution.

The following lemma completely characterizes the conditions under which a steady state distribution exists.

Lemma 2 (Existence of a Steady-State distribution) *The transmission mechanism P is reducible if Assumption 2 holds, otherwise it is irreducible. A steady state distribution exists if and only if either Assumption 2' holds or the transmission mechanism is irreducible and the series $\sum_{k=1}^{\infty} \gamma_k$ converges, where $\gamma_j = \gamma_j^0 = \prod_{k=0}^{j-1} p_k$. Either way, the steady state probability of being in state $i \geq 0$, $\bar{\pi}_i$, is equal to*

$$\bar{\pi}_i = \frac{\gamma_i}{\sum_{k=0}^{\infty} \gamma_k}, \tag{10}$$

where $\gamma_0 = 1$.

Throughout the paper, I analyze in details the consequences of positing Assumption 2 and 2' by carefully distinguishing between the properties of a reducible and an irreducible transmission mechanism.

3.3 An aggregate shock

I now introduce an aggregate shock ϵ_t that hits the system at time t and I characterize the dynamic response of aggregate output to the shock. The shock, ϵ_t, modifies, in a similar way, the adjustment cost of all firms in the economy. Specifically, the cost of adopting the leading technology for a firm in state i with idiosyncratic component equal to λ becomes equal to $c_i + \lambda + \epsilon_t$. Thus, when $\epsilon_t > 0$ ($\epsilon_t < 0$) the cost of technological adoption rises (falls) and in the next period there will be fewer (more) firms using the leading technology relative to the number that would be using it in the absence of the shock, $\epsilon_t = 0$.

More formally, when $\epsilon_t \neq 0$, a firm in state i decides to adjust whenever the realization of the idiosyncratic component λ is such that

$$\beta[V^e(t+1, 0) - V^e(t+1, i+1)] \geq c_i + \lambda + \epsilon_t,$$

so that, at time t, the probability that a firm in state i adopts the leading technology in the economy becomes equal to

$$1 - p_i(\epsilon_t) = F\left(\beta\left[V^e(t+1, 0) - V^e(t+1, i+1)\right] - c_i - \epsilon_t\right).$$

Let $\bar{P}(\epsilon_t)$ denote the Markov chain analogous to (8) collecting the probabilities $p_i(\epsilon_t)$. Hence, the dynamics of the cross-sectional distribution of firms technological states, π_t, is described by the equation

$$\pi_t = \delta_t + P' \pi_{t-1} \tag{11}$$

where $\delta_t = \left(\bar{P}(\epsilon_t) - P\right)' \pi_{t-1}$. In the absence of the aggregate shock, $\bar{P}(\epsilon_t) = P$, δ_t is a vector of zeros and the transmission mechanism P maps π_{t-1} into π_t. The infinite dimensional column vector δ_t is simply an error term which is equal to the difference between the observed cross-sectional distribution given by $\pi_t = \bar{P}(\epsilon_t)' \pi_{t-1}$ and the distribution which would have emerged in the absence of the aggregate shock, equal to $P'\pi_{t-1}$.

The aggregate shock induces a reallocation of firms across technological vintages. The implied *reallocation structure* δ_t has two general properties. Firstly, the sum by column of its entries is exactly equal to 0, that is

$$\bar{1}'\delta_t = \bar{1}' \left(\bar{P}(\epsilon_t) - P\right)' \pi_{t-1} = 0, \tag{12}$$

where $\bar{1}$ denotes a vector of ones. Secondly, a negative (positive) aggregate shock fosters (harms) technological adoption. More formally, let $I(\cdot)$ denote the characteristic function. Then from the monotonicity of $p_i(\epsilon_t)$ with respect to ϵ_t, it follows that

$$I\left(\delta_t^1 > 0\right) = I\left(-\delta_t^j > 0\right) = I\left(-\epsilon_t > 0\right), \ \forall \left|\delta_t^j\right| \neq 0, \ \forall j > 1, \tag{13}$$

where δ_t^j denotes the element in row j of the vector δ_t.

Throughout the analysis I impose conditions such that the reallocation structure δ_t has, on impact, a bounded effect on the level of aggregate output. In section 5, I discuss examples where δ_t has always a finite number of entries strictly different from zero, so that the impact effect of the shock is naturally bounded. In general, the following assumption will always be taken to hold:

Assumption 3 Suppose that $\delta_t' Q < \infty$.

Let P^n denote the *nth* iterated of P. Then Lemma 3 in the appendix shows that if Assumption 3 holds, the infinite dimensional matrix products $\delta_t' P^n Q$ are bounded and well defined for all n. This simple result allows to characterize the response of output to the shock at any period n after its occurrence. Indeed, given an initial distribution π_{t-1} at time $t-1$, the level of output at time $t+n$ is equal to

$$y_{t+n} = \nu (t+n) - \nu \pi_{t-1}{}' P^n Q - \nu \delta_t{}' P^n Q \tag{14}$$

while it would have been equal to

$$y_{t+n} = \nu (t+n) - \nu \pi_{t-1}{}' P^n Q \tag{15}$$

in the absence of the shock. The difference between (14) and (15) gauges the dynamic response of output to the shock ϵ_t. In other words the quantities

$$\phi_n = -\nu \delta_t' P^n Q, \; \forall n \geq 0 \tag{16}$$

are analogous to the Wold coefficients in the moving average representation of a time series, as they gauge at each stage n of the adjustment process the fraction of the shock ϵ_t which has not yet been absorbed. Therefore, as in (2), the rate of decay of ϕ_n measures the persistence of the shock in the model. Accordingly, I will say that the transmission mechanism P, together with the reallocation structure δ_t, generates an order of integration $d \neq 0$ in aggregate output if the *Wold coefficients* ϕ_n's in (16) are such that $\phi_n \sim n^{d-1}$. Conversely d is equal to zero if $\phi_n = O(\rho^n)$.

4 Integration in aggregate output

I first show that, to generate long memory in aggregate output, the adjustment probabilities must fall as the firm's technology becomes more obsolete. I then relate the model to the growth firm size literature and I give formal conditions to generate long memory in output.

4.1 Weak memory

Let me refer to the sequence of adjustment probabilities $\{1-p_i, \, i \geq 0\}$ as the *hazard function* which associates each technological distance i with an adjustment probability $1-p_i$. For example, an increasing hazard function means that the probability of moving onto the technological frontier is the greater the more obsolete is the firm's technology. When so, the order of integration of output is always zero. Specifically:

Proposition 2 (Increasing hazard function) *Let Assumption* 3 *hold and suppose that there does exist a state* m *with* $p_m < 1$ *such that the probabilities* p_i*'s are weakly decreasing in* i *for any* $i \geq m$*. Then output exhibits weak memory.*

When the hazard function is increasing, a firm that does not adopt a new technology today will be more likely to do so tomorrow when, due to technological progress, its technology will be more obsolete. Thus, during the adjustment process, the probability that a firm down in the technological ladder catches up increases, the shock is absorbed at increasing rates and aggregate output always exhibits weak memory.

4.2 Long memory

Conversely, when the hazard function is decreasing in technological distance, shocks naturally propagates at decreasing rather than increasing rates and aggregate output tends to exhibit long memory. Indeed, once the shock hits the system, firms are reallocated across technological vintages. But, if the probability of moving into the technological lead decreases as the firm falls behind in the technological ladder, a firm that does not adopt a new technology today will be less likely to do it so tomorrow when, due to technological progress, its technology will be even more obsolete. Hence the probability that firms down in the technological ladder catch up decreases over the adjustment path, the shock propagates at decreasing rates and output naturally tends to exhibit long memory.

In the model, a decreasing hazard function arises when the adjustment costs c_i's grow at a rate faster than $\frac{\beta\nu(i+1)}{1-\beta}$, see Proposition 1. More generally the hazard function tends to be decreasing in models where the firm accumulates human and physical capital specific to a given technology, therefore becoming particularly resistant to adopt a new one. Some empirical evidence also suggests the existence of an hazard function decreasing in technological distance.[16]

To interpret the model results in terms of the relationship between expected

[16] For instance, Blundell et al. (1999) find a robust and positive effect of market share on observable headcounts of innovation. To the extent that market share and productivity are positively related, this evidence suggests the existence of a positive correlation between the probability of technological adoption and the current level of firm's productivity. In accordance with this interpretation, Baily et al. (1992) also 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 productivity.

growth and firm size, notice that $g_i = \nu(i+1)(1-p_i)$ is the firm's expected growth since a firm in state i either raises output by $\nu(i+1)$, which occurs with probability $1-p_i$, or with probability p_i it keeps output constant. Also notice that, in the model, firms currently operating obsolete technologies produce less output and are therefore smaller than firms using technology close to the technological frontier. Specifically, when it exists $h > 0$ and a state i^* such that

$$g_i = \nu(i+1)(1-p_i) = \nu h, \quad \forall i \geq i^*, \tag{A4}$$

small firms —i.e. firms which are currently producing less output— grow at rate νh. The parameter h measures the growth of small firms relative to the aggregate economy: small firms are growing faster (slower) than the remaining *big* firms in the economy, if $h > (<)1$. Analogously, all firms grow at the same rate if $h = 1$. I next show that $(A4)$ is the crucial assumption to generate long memory in output.

4.2.1 A further bound on the reallocation structure

I start by imposing further bounds on the reallocation structure δ_t. Specifically:

Assumption 5 Let $s = 1 + \max\{i : p_i = 0, i \geq 0\}$ if P is reducible while let $s = 0$ if P is irreducible. Assume that $0 < \sum_{k=0}^{\infty} \frac{|\delta_t^{k+s+1}|}{\gamma_k^s} < \infty$.[17]

Assumption 5 requires first that, if the transmission mechanism is reducible, some units enter the set of non recurrent states ($\delta_t^{k+s+1} \neq 0$ for some k). Secondly, it imposes some further bounds (in addition to Assumption 3) to the extent of the reallocation induced by the shock. Such assumption is naturally satisfied if, for example, only a finite number of entries of δ_t is different from zero. In the next sub-section I investigate the consequences of relaxing it while section 5 discusses examples of the basic model where this assumption is naturally satisfied. Proposition 3 shows that $(A4)$ is enough to generate long memory in an arbitrary, reducible transmission mechanism.

Proposition 3 (Integration in the reducible case) *Let Assumptions 2', 3 and 5 hold and suppose that* $(A4)$ *is satisfied for some* $h > 0$. *Then the order of integration of output is equal to* $2 - h$.

[17] Generally s is finite since Assumption 5 is posited jointly with $(A4)$.

Proposition 3 can intuitively 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.

(ii) If $h < 1$, big firms grow faster than small ones and the first difference of aggregate output exhibits long memory. In this case the initial effect of the shock tends to be amplified without limits. 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 output exhibits a unit root, $d = 1$ —i.e. the shock has a permanent and bounded effect on the level of output.

A reducible transmission mechanism P (Assumption 2' holds) implies that all firms end up using a technology close to the technological frontier. Conversely, if P is irreducible, all firms keep wandering across all possible states in the economy. When so, it is still true that firms down in the technological ladder can catch up with those ahead, but it also happens that firms operating technologies close to the frontier will eventually fall down again in the ladder. This is why, in an irreducible transmission mechanism, the impact effect of the shock cannot be amplified without limit and output never exhibits an order of integration strictly greater than one.[18] To see an application of this result assume that all firm's grow at the same rate so that ($A4$) holds with $h \leq 1$ for any $i \geq 0$, that is $g_i = \nu h$, $\forall i \geq 0$. But then $\delta_t{}' P^n Q = \delta_t{}' Q$ for any n, which implies that output has an order of integration equal to one independently of h since $\phi_n = \phi_0$, $\forall n$ is the distinctive property of a random walk.

Moreover, if the transmission mechanism P is irreducible, ($A4$) implies that, for large j, $\gamma_j \sim j^{-h}$ so that Lemma 2 implies that $h > 1$ is required for a steady state distribution to exist. Thus if the transmission mechanism is irreducible and a steady state distribution exists, the order of integration of aggregate output is always strictly smaller than one. Specifically:

[18] I analyze this case in more details in my Phd dissertation, see Michelacci (1998).

Proposition 4 (Integration in the irreducible case) *Assume that the transmission mechanism is irreducible and that Assumptions* 3 *and* 5 *hold. Then if* $(A4)$ *is satisfied for some* $h > 1$, *a steady state distribution exists and the order of integration of output is equal to* $2 - h$.

I next analyze the consequences of removing Assumption 5, but still maintaining $(A4)$. I find that aggregate output still exhibits long memory so that imposing $(A4)$ is the crucial assumption to generate long memory in output.

4.2.2 Alternative reallocation structures

Assumption 5 can fail because either $\sum_{k=0}^{\infty} \frac{\left|\delta_t^{k+s+1}\right|}{\gamma_k^s} = \infty$ or $\sum_{k=0}^{\infty} \frac{\left|\delta_t^{k+s+1}\right|}{\gamma_k^s} = 0$. Let consider a counter-example of the former type. Assume that a sufficient mass of firms is in the tail of the cross sectional distribution of states, so that, for some κ, $\pi_{t-1}^i \sim \gamma_i^\kappa$ as i goes to infinity. Consider then the effect of a negative aggregate shock $\epsilon_t < 0$, such that $\delta_t^{\kappa+i} = -\bar{\delta}\,\gamma_i^k$, $\bar{\delta} > 0$.[19] Lastly notice that if $(A4)$ also hold, one has that, for large i, $\delta_t^{\kappa+i} \sim i^{-h}$. Hence $h > 2$ is required to make Assumption 3 satisfied. If so, the degree of persistence of the shock is greater than that generated in Propositions 3 and 4, despite the transmission mechanism P being the same. Indeed:

Proposition 5 (Small versus large shocks) *Assume that* $\forall i \geq 0$

$$\delta_t^{\kappa+i} = -\bar{\delta}\,\gamma_i^\kappa \neq 0,\ 0 < \bar{\delta} < 1, \tag{A5'}$$

for some finite κ *and suppose that* $(A4)$ *holds with* $h > 2$. *Then the order of integration of aggregate output is equal to* $3 - h$.

Proposition 5 has some interesting implications. Firstly, it shows that the model can generate asymmetric responses to shocks. Indeed as $\delta_t = \left(\bar{P}\left(\epsilon_t\right) - P\right)' \pi_{t-1}$, and $p_i = 1 - \frac{h}{1+1}$, 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 $\bar{\delta}$ in $(A5')$ can only be positive. Secondly, Proposition 5 is an application of the 'folk wisdom' claiming a positive relationship between amplification and propagation mechanism. When Assumption 5 fails, shocks become

[19] For example, this reallocation structure would emerge, if the system is in steady state, the transmission mechanism is irreducible and $F(\cdot)$ is uniform. In this case, $\bar{\delta}$ would simply be equal to the absolute value of the aggregate shock, $\bar{\delta} = |\epsilon_t|$.

more persistent, which means that the degree of persistence of shocks is the greater the larger is the extent of the reallocation induced by the shock. Finally, Proposition 5 suggests an alternative characterization of aggregate persistence. According to this view most shocks generate low persistence —notice that $h > 2$ does not produce a positive order of integration under the conditions assumed in Propositions 3 and 4. Sometimes, however, large negative shocks hit the system and generate the large degree of persistence which characterizes aggregate output. Many researchers have noticed that, after allowing for some structural breaks, the time series of US GDP may be well represented by a standard weak memory process. What is suggestive is that the model might also explain why most structural breaks identified in the literature (the 'big recession', World War II, oil price shocks) are associated with large negative shocks.

Alternatively, Assumption 5 can fail if P is reducible, and no unit enters the set of non recurrent states in response to the shock. If so, output always exhibits weak memory, see Proposition 6.

Proposition 6 (Uniformly bounded cross-sectional heterogeneity) *Let Assumptions* 2' *and* 3 *hold and assume that*

$$\sum_{k=0}^{\infty} \frac{\left|\delta_t^{k+s+1}\right|}{\gamma_k^s} = 0, \qquad \gamma_0^s = 1, \tag{A5"}$$

where $s = 1 + \max\{i : p_i = 0,\ i \geq 0\}$. *Then output exhibits weak memory.*[20]

$(A5")$ is equivalent to assuming that all units always remain into a finite dimensional set of states close to the technological frontier. A time series with order of integration $d \geq \frac{1}{2}$, has infinite variance and it is non-stationary in second moments since $\sum_{n=0}^{\infty} (\phi_n)^2$ diverges. Hence it is not surprising that a variable with a finite number of states (and therefore bounded) never exhibits an order of integration greater than one half. More generally, this theorem shows that d is equal to zero whenever cross-sectional heterogeneity is uniformly bounded. In this respect, Proposition 5 generalizes results by Bertola and Caballero (1990), Caballero and Engel (1991) in the sS literature. In their framework, cross-sectional heterogeneity is always bounded, thus

[20] If s is infinite, it can be shown that output always exhibits weak memory regardless of the reallocation structure.

the rate at which the expected value of the cross-sectional distribution converges to its long run value is necessarily exponential and the aggregate always exhibits weak memory. The proposition then elucidates why the vintage structure of the model is important in generating long memory. Indeed, in such environment, it fails to exist a technological distance that bounds uniformly the set of relevant firm's technological states since, due to technological progress, a firm can move arbitrarily far away from the technological frontier in the absence of further technological adoption.

5 Further discussion

In this section, I briefly discuss the robustness of my results to environments where (i) the number of firms is discrete rather than a continuum, (ii) aggregate shocks get repeated over time, (iii) firms die and new firms enter, and (iv) growth is exponential rather than linear. I show that assumption $(A4)$ is generally the only assumption required to generate a positive order of integration in aggregate output.

5.1 A discrete number of firms

Heretofore, I have assumed that the economy is populated by a continuum of firms. I did so for analytical convenience only, since, in this case, the adjustment path of output is deterministic and the implied impulse response can be defined in an intuitive manner. I next move away from this theoretical benchmark and I posit that the economy is populated by a discrete number of firms, n. I show that the previous results remain unaffected. Importantly, if the number of firms is discrete, the reallocation structure δ_t has just a finite number of entries different from zero and Assumptions 3 and 5 are always satisfied. Thus $(A4)$ is the only assumption required to make Propositions 3 and 4 satisfied.

Table 1 reports the average over 1000 replications of the estimate of the order of integration d of output for two economies labelled 'long' and 'weak' memory, respectively. In the former the adjustment costs c_i's are such that the adjustment probabilities are equal to $1 - p_0 = 0.8$ and $1 - p_i = \frac{1.5}{i+1}$, $\forall i \geq 1$, so that $(A4)$ holds and the order of integration implied by Proposition 3 and 4 would be one half. In the latter the adjustment probabilities are $1 - p_i = 0.2$, $\forall i \geq 0$, so that the order of integration implied by Proposition 2 would be zero. The number of firms is discrete and equal to $n = 10$. The idiosyncratic shocks λ, which are the only source of aggregate variability, are a drawing from a uniform distribution with support on $[0, 1]$. The estimates of d are obtained by running a log-periodogram regression on the time series of aggregate output once detrended by a linear trend estimated by OLS.[21] The results contained in columns 2 and 4 of the table confirm the validity of the results previously derived in an environment with a continuum of firms: the average estimate of d is not significantly different from its theoretical value, that means $d = 0.5$ and $d = 0$ in the long and weak memory economy, respectively.

INSERT TABLE 1 APPROXIMATELY HERE

5.2 Ongoing aggregate uncertainty

In the basic model the aggregate shock was unexpected and once-and-for-all. Columns 3 and 5 of Table 1 report the average estimated d in the weak and long memory economy previously described when the random component of the firm's adjustment cost is the sum of an idiosyncratic shock drawn from a uniform distribution with support $[0, 0.5]$ and a white noise aggregate shock which assumes values zero or one half with probability one half. Thus, from the firm's point of view the random shock to the cost of technological adoption is simply a uniform distribution with support on $[0, 1]$. The table shows that the conclusions previously derived remain unchanged: the estimated d is not significantly different from one half (zero) in the long (weak) memory economy.

5.3 Firms' entry and exit

Arguably firms can die and free resources that potential new entrants can exploit. I next discuss why allowing firm to die and then to be replaced by new firms would affect the main results of the paper in an ambiguous way. For example, if all firms are equally

[21] The log-periodogramm regression was originally proposed by Geweke and Porter Hudak (1983). Under the assumptions of stationarity and gaussianity, Robinson (1995) proved consistency and asymptotic normality of the estimator while Giraitis et al. (1997) proved that its asymptotic rate of convergence is optimal. The properties of the log-periodogram regression were proved to be robust in non stationary environments (see Velasco 1999) as well as in processes with non Gaussian innovations (see Velasco 2000).

likely to die and new firms enter with a technology that is a drawing from the current distribution of firms' technologies, the paper's results would remain unchanged.[22] Such pattern of firms entry and exit is certainly disputable. But the net effect on aggregate persistence of departing from such benchmark is also uncertain. On the one hand, firms death probability is arguably increasing in technological obsolescence. On the other hand, newly created firms tend to enter at the lower tail of the distribution of firms productivity—see for example table 3 in Baily et al. (1992). While the former effect would tend to undo the results of the paper, the latter would even reinforce them, thereby making any conclusion be (a priori) ambiguous.

5.4 Exponential growth

To introduce exponential growth one could simply think that all variables of the basic model are denominated in logs, implying that differences indicate growth rates while arithmetic averages indicate the logarithm of geometric ones. If so, one would be 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 a Cobb-Douglas aggregate production function which uses as intermediate inputs the output produced by each firm.[23] The model could then be closed by assuming that the sector producing final goods is perfectly competitive, so that intermediate goods are sold at a price equal to their marginal product.

In a related effort I first solve and then simulate the associated economy. I show that, if something like $(A4)$ holds in steady state, the estimated d is not significantly different from that predicted by the basic model. The introduction of aggregate demand complementarities, however, implies that the adjustment probabilities increase (fall) when aggregate output is high (low) relative to the average so that the economy tends to adjust more quickly in a 'boom' than in a 'recession'. In turn, this mechanism implies the existence of non linearities in aggregate dynamics.

[22] This last assumption is often made in theoretical models, see for example Howitt (1999).

[23] If the number of firms is discrete (a continuum), output would be given by a Cobb-Douglas production function $\prod_{j=1}^{n}(y_j)^{\frac{1}{n}}$ $(\exp\int_0^1 y_j dj)$ where y_j is firm j output.

6 Relation to Granger (1980)

Granger (1980) showed that the aggregation of $AR(1)$ processes can lead to a positive order of integration in the aggregate if some suitable conditions on the cross-sectional distribution of the first order correlation of the process are satisfied. I now relate Propositions 3, 4 and 5 to Granger (1980). I show that some features of my steady state distribution of firms' technologies resemble the conditions needed for Granger's result to hold. This allows me to speculate about the conditions required to generate a positive order of integration in the aggregate variable.

Granger (1980) considers the aggregation of $AR(1)$ processes like

$$x_t^i = \theta_i x_{t-1}^i + \epsilon_t^i + \epsilon_t, \qquad i = 1, ..., n,$$

where i refers to the individual, n is the number of agents considered, while ϵ_t^i and ϵ_t are, respectively, an idiosyncratic and aggregate shock both assumed to be *iid* over time. He further assumes that the coefficients θ_i are independent drawings from an absolute continuous function with density $f(\cdot)$ such that for $\theta \to 1^-$

$$f(\theta) \sim (1-\theta)^{h-2} \tag{17}$$

where h is a real parameter belonging to the open interval $(1, \infty)$ by the integrability constraint.[24] Granger shows that, for n large enough, the aggregate variable $X_t = \frac{1}{n}\sum_{i=1}^{n} x_t^i$ behaves like a process with order of integration $d = 2 - h$. There are important similarities between this result and Propositions 3, 4 and 5.

To see this, consider the dynamics of the state of a firm implied by (8). When the state of the firm at time $t-1$ is equal to $x_{t-1}^i = i \geq 0$, its state x_t^i at time t could be written as equal to

$$x_t^i = \eta_i + \theta_i x_{t-1}^i + \epsilon_t^i \tag{18}$$

where $\theta_i = p_i$, $\eta_i = p_i$ while ϵ_t^i has zero mean and distribution conditional on x_{t-1}^i given by

$$\epsilon_t^i \mid x_{t-1}^i = \begin{cases} -(i+1)\,p_i & \text{with probability } 1-p_i, \\ (i+1)\,(1-p_i) & \text{with probability } p_i. \end{cases} \tag{19}$$

[24] Strictly speaking, Granger considers the case where the function $f(\cdot)$ is a Beta distribution. Robinson (1978), Goncalves and Gourieroux (1988) and Lippi and Zaffaroni (1999) show however, that the low frequency behaviour of the aggregate is determined only by the shape of the cross sectional distribution $f(\cdot)$ around one.

Hence, the dynamics of the state of the firm one period ahead can (formally) be written as an $AR(1)$ process with first order correlation equal to p_i. Moreover, under the assumptions underlying Proposition 4 and, given (10), the number of units $f(\theta)$ with coefficients $\theta_i = \theta$ for $\theta \to 1^-$ is such that (in steady state)

$$f(\theta) \sim (1-\theta)^{h-2},$$

where $h > 1$. Proposition 4 seems then surprisingly similar to the results obtained by Granger (1980). There are, however, important differences which explain the content of Propositions 3 and 5. In Granger, the value of the first order correlation of any unit i is fixed within its entire life. Instead, in my vintage model the coefficient θ_i changes as the firm changes state. This explains why, differently from Granger, the model can generate orders of integration greater than one, and a unit root as a particular case (see Proposition 3). Moreover, the aggregate shocks considered in section 4 are not iid, so that the model can analyze the different response of the aggregate system to large versus small shocks, thus leading to Proposition 5.

In general, one can conclude that three elements are required to generate an order of integration different from zero in output. It requires first a positive relation between current and future level of output at the micro level, as implied by the $AR(1)$ process, secondly, a certain degree of cross-sectional heterogeneity, as implied by the density function $f(\theta)$, and, finally, a sufficient number of units with very large persistence as implied by (17). In Granger (1980) the three elements are assumed one independently of the other. In my model the first element can generate endogenously the other two. When the adjustment costs c_i's grow sufficiently fast so as to make $(A4)$ hold, the hazard function is decreasing in technological distance and the model produces endogenously the number of units with a sufficient amount of persistence which are required to generate a positive order of integration in the aggregate. These considerations also suggest that, insofar as the hazard function decreases sufficiently fast to produce a number of units with enough persistence, the binary nature of the adjustment process implied by (4) is not crucial for generating a positive order of integration in aggregate output.

7 Conclusions

I have analyzed how an aggregate shock propagates in a vintage model where the probability of moving into the technological lead decreases as the firm falls behind in the technological ladder. I have used the model to provide some microfoundations for the slow adjustment of output in response to shocks. In particular, I have interpreted my findings in terms of the relationship between firm growth and firm size. If Gibrat's law holds, all firms grow at the same rate independently of their size and a unit root characterizes output dynamics at the micro as well as at the macro level. If Gibrat's law fails and small firms grow faster than big firms (as recent empirical evidence suggests), the rate at which the shock propagates in the economic system is decreasing over the adjustment process and aggregate output exhibits a fractional order of integration.

The order of integration of aggregate output has been much debated. Formal investigation, however, has suggested that long memory might well represent the low frequency behavior of aggregate output. If so, micro and macro evidence matches quite closely. Moreover, models which do not deal explicitly with cross-sectional heterogeneity generate a dynamics in which shocks either have permanent effects or vanish at the usual exponential rate. Thus these models can not replicate the long memory feature that seems to characterize aggregate output. That is why I think that the process of ongoing churning and catching-up that takes place in the economy may be a key factor in explaining the observed degree of aggregate persistence.

The tools introduced in the paper could potentially be used to analyze dynamics in other vintage models. The model is also well suited to analyze further interesting questions. For example: how do non-linearities at the micro level affect aggregate dynamics, once compared with a standard linear process? What is the role of technological progress as well as firms' entry and exit in replicating the observed degree of aggregate persistence? Addressing these issues would be particularly relevant to bring the model to the data. The ultimate task would be to provide microfoundations for the many features which characterize aggregate dynamics.

8 Appendix

8.1 Proofs of results in section 3

Proof of Proposition 1 Hereafter I keep the convention that $F(x) = 0$ for any x out of the support of the distribution of the idiosyncratic shock λ. As the value function $V(t,i,\lambda)$ is linear in t, one can write

$$V(t,i,\lambda) = \frac{\nu t}{1-\beta} + \tilde{V}(i,\lambda).$$

From the monotonicity of $\tilde{V}(i,\lambda)$ with respect to λ and (5) it follows that

$$\tilde{V}(i,\lambda) = \begin{cases} -\nu i - c_i - \lambda + \beta\tilde{V}^e(0), & if\ \lambda < R_i, \\ -\nu i + \beta\tilde{V}^e(i+1), & if\ \lambda \geq R_i, \end{cases} \tag{20}$$

where $\tilde{V}^e(i) = \int_\Lambda \tilde{V}(i,x)dF(x)$. Taking expectations in (20), using (7) to substitute for $\tilde{V}^e(i+1)$ and after an integration by parts yields

$$\tilde{V}^e(i) = -\nu i - c_i - R_i + \int_{-\infty}^{R_i} F(x)\,dx + \beta\tilde{V}^e(0),\ \ \forall i, \tag{21}$$

which after using (7) leads to

$$\tilde{V}^e(i) = -\nu i + \beta\tilde{V}^e(i+1) + \int_{-\infty}^{R_i} F(x)\,dx. \tag{22}$$

First notice that $\lim_{i\to\infty} \beta^i \tilde{V}^e(i) = 0$ since a firm in state i can achieve a value of at least $-\sum_{j=0}^{\infty} \beta^j (i+j)$ by never adjusting. Then evaluating (22) at $i = 0$, and repeatedly substituting forward for $\tilde{V}^e(i+1)$ yields

$$\tilde{V}^e(0) = -\frac{\beta\nu}{(1-\beta)^2} + \sum_{j=0}^{\infty} \beta^j \int_{-\infty}^{R_j} F(x)\,dx. \tag{23}$$

Furthermore evaluating (21) at $i = 0$ yields

$$\tilde{V}^e(0) = \frac{1}{1-\beta}\left[-c_0 - R_0 + \int_{-\infty}^{R_0} F(x)\,dx\right]$$

which substituted in (23) leads to

$$c_0 + R_0 = \frac{\beta\nu}{1-\beta} + (1-\beta)\sum_{j=1}^{\infty} \beta^j \int_{R_j}^{R_0} F(s)\,ds. \tag{24}$$

Moreover (21) evaluated at $i+1$ and after using (7) yields

$$c_{i+1} + R_{i+1} = c_0 + R_0 - \nu\,(i+1) - \int_{R_{i+1}}^{R_0} F\,(x)\,dx + \beta^{-1}\,(c_i + R_i)\,. \tag{25}$$

For any given sequence of probabilities $\{p_i, i \geq 0\}$, equation (25) defines a difference equation of the first order in c_i, whose solution is unique once c_0 is set to satisfy (24). Besides, equations (7) and (21) imply that a sequence of adjustment costs $\{c_i, i \geq 0\}$ that solves (24) and (25) yields the given sequence of probabilities $\{p_i, i \geq 0\}$ as a solution of the firm problem.

Substituting backward for $c_i + R_i$ in (25) one obtains that $\forall i \geq 0$

$$c_{i+1} = \sum_{j=0}^{i} \beta^{-j} \left[c_0 + R_0 - \nu\,(i+1-j) - \int_{R_{i+1-j}}^{R_0} F\,(x)\ dx \right] + \beta^{-(i+1)}\,(c_0 + R_0)\,. \tag{26}$$

Collecting $\beta^{-(i+1)}$ in the right hand side, using (24) and after some algebra yields (9) evaluated at $\forall i \geq 1$. To conclude the proof just notice that (9) evaluated at $i = 0$ is equivalent to (24). $\|$

Proof of Lemma 1 Under Assumption 1, a firm starting from any state i, returns to state zero with probability one. This implies firstly that state zero is recurrent and secondly that either state i and state zero communicate or state i is transient (see for example Karlin and Taylor 1975, 1981). As state zero is recurrent, at least one recurrent class does exist and given the previous considerations this is unique. $\|$

Proof of Lemma 2 See Billingsley 1986, theorem 8.8. and example 8.13. $\|$

8.2 Proofs of results in section 4

Unless otherwise specified '$\sim$' denotes asymptotic equivalence for n going to infinity while ρ indicates a quantity such that $0 \leq \rho < 1$. The next three lemmas are used throughout the proofs.

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

Proof of Lemma 3 Let $\delta_t = \delta^+ - \delta^-$ where $\delta^+ \geq 0$ and $\delta^- \geq 0$. 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). Consequently $\delta' P^n Q = (\delta^+ - \delta^-)'\,P^n Q$ is well defined provided that for each n, $(\delta^-)'\,P^n Q$ and $(\delta^+)'\,P^n Q$ are bounded. This follows from Assumption 3, (13) and the fact that each element of P^nQ has increments bounded above by one. $\|$

Lemma 4 *Let the transmission mechanism be reducible so that* $s = 1 + \max\{i : \, p_i = 0, \, i \geq 0\}$. *Then under Assumptions 1, 2' and 3,*

$$\delta_t' P^n Q = A + B + C - D, \tag{27}$$

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

$$A = \sum_{i=1}^{\infty} \delta_t^i c_{i-1}^n, \tag{A}$$

$$B = n \sum_{i=0}^{\infty} \delta_t^{i+s+1} \gamma_n^{i+s}, \tag{B}$$

$$C = \sum_{i=0}^{\infty} \delta_t^{i+s+1} \gamma_n^{i+s} (i+s), \tag{C}$$

$$D = S \sum_{i=0}^{\infty} \delta_t^{i+s+1} \gamma_{n-1}^{i+s}, \tag{D}$$

where $0 \leq S < \bar{\imath}$, *is the expected value of the steady state distribution while the quantities* c_i^n*'s are such that* $0 \leq c_i^n \leq \rho^n K$, *with* K *being a bounded quantity independent of* i *and* n.

Proof of Lemma 4 Consider the submatrix $\tilde{P}$ of P identified by its first s rows and columns. $\tilde{P}$ is stochastic matrix as it is a positive square matrix such that each row sums up to one. One can then partition P^n as follows:

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

where e_j^n, $j \geq 0$, indicates a row vector of dimension $1 \times s$ corresponding to the first s elements of the row $j+1$ of the matrix P^n, while the element γ_n^{s+i}, $i \geq 0$, are in row $i+s+1$ and column $n+i+s+1$. It can then be easily proved by recursion that for $i \geq s$

$$e_i^n = \sum_{j=0}^{n-1} \gamma_j^i (1 - p_{i+j}) \, e_0^{n-1-j}, \tag{28}$$

where $\gamma_0^i = 1$, while e_0^0 denotes a vector of dimension $s \times 1$ whose first element is equal to one while all the others are equal to zero.

Let $\hat{\pi}$ denote the vector of dimension $s \times 1$ corresponding to the steady state distribution of $\tilde{P}$. To prove (27) one can proceed as follows. Firstly, show that e_i^n, $0 \leq i < s$ converges at least exponentially to $\hat{\pi}$, that is

$$|e_i^n - \hat{\pi}'| \leq H \rho^n \bar{1}', \;\; \forall i < s, \;\; \forall n > 0 \tag{29}$$

where H is a positive bounded quantity while $\bar{1}'$ is a vector of dimension $1 \times s$ whose elements are all equal to one.

Secondly, show that $\forall i \geq s$, $\forall n$, e_i^n is equal to

$$e_i^n = \left(1 - \gamma_{n-1}^i\right) \hat{\pi}' + \gamma_{n-1}^i \left(1 - p_{i+n-1}\right) \left(e_0^0\right)' + a_i^n, \tag{30}$$

where the a_i^n's are such that $\forall n$, $\forall i$, $0 \leq a_i^n \leq \rho^n K \bar{1}'$ with K being a bounded quantity independent of i and n. But then (29), (30), Assumption 3 together with (12) and (13) immediately imply (27), thus completing the proof.

To prove (29) consider first the case $\bar{\imath} = s$. Given Assumption 2', there exists a state $0 \leq i < \bar{\imath} - 1$ such that $p_i \neq 1$ and the Markov chain $\tilde{P}$ is both irreducible and aperiodic. When so a steady state distribution exists and the rate of convergence is exponential and independent of the initial distribution (see e.g. Stokey and Lucas 1988, theorem 11-4) and (29) holds.

Consider now the case $\bar{\imath} < s$. If so the matrix $\tilde{P}$ is reducible and the first $\bar{\imath} < s$ states of the Markov chain $\tilde{P}$ are recurrent while all the others $s - \bar{\imath}$ are transient. The structure of $\tilde{P}$ implies that a unit starting from any transient state $\bar{\imath} < i \leq s$ enter the recurrent class after a number of periods less or equal than $s - \bar{\imath}$. Hence the previous reasoning yields (29).

To prove (30), notice that from (28) and (29) and after using the triangle inequality it follows that, $\forall i \geq s$, $\forall n$,

$$\left| e_i^n - \hat{\pi}' \left(1 - \gamma_{n-1}^i\right) - \gamma_{n-1}^i \left(1 - p_{i+n-1}\right) \left(e_0^0\right)' \right| \leq H \left(1 - \gamma_{n-1}^i\right) \rho^n \bar{1}' \leq H \rho^n \bar{1}'.$$

||

Lemma 5 *Let E_0^n denote the expected state after n iterations of a unit starting in state 0. Then $W_n = E_0^{n+1} - E_0^n$ satisfies the recursion formula*

$$W_n = -\sum_{i=0}^{n-1} \gamma_{i+1} W_{n-i-1} + \gamma_{n+1} \left(n + 1 \right), \; \forall n > 0, \tag{31}$$

where $W_0 = 0$.

Moreover, if the transmission mechanism is irreducible and a steady state distribution does exist, the following two results hold:

$$\lim_{n\to\infty} W_n = 0, \tag{32}$$

$$\delta_t' P^n Q = \sum_{j=0}^{\infty} W_{n+j} \sum_{i=j}^{\infty} \delta_t^{i+1} \frac{\gamma_{i-j}}{\gamma_{i+1}}. \tag{33}$$

Proof of Lemma 5 The law of iterated expectations yields

$$E_0^n = \sum_{i=0}^{\infty} \gamma_i (1-p_i) E_0^{n-i-1} + \gamma_n n, \; n \geq 0,$$

with $E_0^n = 0$ if $n \leq 0$. It then follows from the definition of W_n that

$$W_n = E_0^{n+1} - E_0^n = \sum_{i=0}^{\infty} \gamma_i (1-p_i) W_{n-i-1} + \gamma_{n+1}(n+1) - \gamma_n n, \; n \geq 0, \tag{34}$$

where $W_{n-i} = 0$ for $i \geq n$. But given (34), (31) can be proved by recursion.

To prove (32), first notice that a limit for W_n always exists, with $\lim_{n\to\infty} W_n = z$ where $0 \leq z \leq 1$. This follows from recurrency, the basic limit theorem of Markov chains (see Karlin and Taylor 1975, theorem 1.2.) and the fact that W_n is uniformly bounded by one. Then argue by contradiction and suppose that $\lim_{n\to\infty} W_n = z > 0$. If so and given (31), $\forall \epsilon$, it does exist N^*, such that $\forall n > N^*$

$$|W_n - z| = \left| -\sum_{i=0}^{n-1} \gamma_{i+1}(W_{n-i-1} - z) - z \sum_{i=0}^{n-1} \gamma_i + \gamma_{n+1}(n+1) \right| < \frac{\epsilon}{\sum_{i=0}^{n-1} \gamma_{i+1}}, \tag{35}$$

where the existence of a steady state distribution together with Lemma 2 guarantee that $\sum_{i=0}^{n-1} \gamma_i$ is finite. Given any two quantities K and H, one has

$$|K| - |H| \leq |K + H| \leq |K| + |H|,$$

which can be applied to (35) with

$$K = -z \sum_{i=0}^{n-N^*} \gamma_i,$$

$$H = -\sum_{i=0}^{n-N^*} \gamma_{i+1}[W_{n-i-1} - z] - \sum_{i=n-N^*+1}^{n-1} \gamma_{i+1} W_{n-i-1} + \gamma_{n+1}(n+1),$$

so that $|W_n - z| = |K + H|$. But then the triangle inequality implies

$$\begin{aligned} |H| &\leq \left| -\sum_{i=0}^{n-N^*} \gamma_{i+1}(W_{n-i-1} - z) \right| + \left| \sum_{i=n-N^*+1}^{n-1} \gamma_{i+1} W_{n-i-1} \right| + \gamma_{n+1}(n+1) \leq \\ &\leq \epsilon + (N^* - 1)^2 \sup_{n-N^*+2 \leq i \leq n} \gamma_i + \gamma_{n+1}(n+1), \end{aligned}$$

since W_{n-i}, for $i \geq j$ is always bounded by $n - j$. As $\lim_{n\to\infty} \gamma_n n = 0$, by Lemma 2, $|H|$ is arbitrarily small, for large n, so that equation (35) can be satisfied only if $|K| = 0$ which in turn implies $z = 0$. This is a contradiction and proves (32).

To prove (33) notice that the assumption that P is irreducible and Lemma 2 imply that, $\forall i$, $\gamma_i > 0$. Let E^n_{j-1} denote the generic element in place $j \geq 1$ of the vector $E^{(n)} = P^n Q$. Then for any $i \geq 0$, E^n_i follows the recursion

$$E^n_i = (1 - p_i)E^{n-1}_0 + p_i E^{n-1}_{i+1}. \tag{36}$$

Solving for E^{n-1}_{i+1} in equation (36) by substituting backwards for E^n_i, yields

$$E^n_i = \frac{\gamma_0}{\gamma_i} W_{n+i-1} + \frac{\gamma_1}{\gamma_i} W_{n+i-2} + \frac{\gamma_2}{\gamma_i} W_{n+i-3} + \cdots + \frac{\gamma_{i-1}}{\gamma_i} W_n + E^n_0. \tag{37}$$

But then (12) and (37) immediately yield (33). $\|$

Proof of Proposition 2 Assume first that Assumption 2' holds. Then notice that, if the probabilities p_i are decreasing in i, for $i \geq m$, γ^i_j is such that, for large j, and $\forall i$,

$$\gamma^i_j = \gamma^i_m \gamma^{i+m}_{j-m} \leq \gamma^i_m (p_m)^{j-m} \leq (p_m)^{j-m} \tag{38}$$

so that Assumption 1 holds. But then, as Assumptions 1 and 3 also hold, Lemma 4 applies and $\delta'_t P^n Q$ is equal to (27). We know that $A = O(\rho^n)$ where $O(\rho^n)$ indicates a quantity at most of order ρ^n, that is $\lim_{n\to\infty} \frac{O(\rho^n)}{\rho^n} < \infty$. But Assumption 3 together with (38) guarantee that also B, C and D are $O(\rho^n)$. Hence $\delta'_t P^n Q = O(\rho^n)$ and concluding the proof for this case.

Consider now the case where Assumption 2' fails and P is irreducible. Then rely on Lemma 5 to prove first that $W_n = O(\rho^n)$ and then that $\delta'_t P^n Q = O(\rho^n)$, thus concluding the proof.

I now show that $W_n = O(\rho^n)$ where ρ is such that $\frac{p_0}{\rho} < 1$ that is well defined because, by assumption, $0 < p_0 < 1$. Without loss of generality and to simplify notation I set $m = 0$. If the probability p_i are decreasing in i, $\gamma_i \leq (p_m)^i$, so that by Lemma 2 a steady state distribution exists. Hence from equation (31) in Lemma 5 it follows that

$$\frac{W_n}{\rho^n} = \tilde{W}_n = -\sum_{i=0}^{\infty} \tilde{\gamma}_{i+1} \tilde{W}_{n-i-1} + \rho \tilde{\gamma}_{n+1} (n+1), \ \forall n > 0,$$

where $\tilde{W}_n = \frac{W_n}{\rho^n}$ while $\tilde{\gamma}_{i+1} = \frac{\gamma_{i+1}}{\rho^{i+1}} \leq \left(\frac{p_0}{\rho}\right)^{i+1}$. As the series $\sum_{i=0}^{\infty} \tilde{\gamma}_{i+1}$ converges, one can use arguments similar to those used in the proof of Lemma 5 to prove that $\lim_{n\to\infty} \tilde{W}_n = 0$. But then $W_n = O(\rho^n)$ and irreducibility imply that $\forall \epsilon$, it does exist N^*, such that $\forall n > N^*$, $0 < \frac{W_n}{\rho^n} < \epsilon$. Hence one can write $W_{n+j} < \rho^{n-N^*} W_{N^*+j}$, $\forall j > 0$ which together with (33) imply that

$$|\delta_t' P^n Q| < \rho^{n-N^*} \left|\delta_t' P^{N^*} Q\right| < \rho^n K$$

where K is a positive bounded quantity by Lemma 3. $\|$

Proof of Proposition 3 To prove that $\phi_n = \nu \delta_t' P^n Q \sim n^{1-h} = n^{d-1}$, $d = 2-h$, proceed as follows. Firstly, show that Assumptions 1, 2' and 3 hold. Then apply Lemma 4, to write $\delta_t' P^n Q$ as equal to (27). Finally, after noticing that $A = O(\rho^n)$, prove that $B \sim n^{1-h}$, $C = O(n^{1-h})$ and $D \sim n^{-h}$, thus concluding the proof.

Assumptions 2' and 3 hold by hypothesis. To check that Assumption 1 is also satisfied, notice that $\gamma_n^{i^*} \sim \gamma_n^s \sim n^{-h}$. Indeed, from the recursion of the Gamma function (see Abramowitz and Stegun, 1972, formula 6.1.15)

$$\Gamma(\alpha+1) = \alpha \Gamma(\alpha), \tag{39}$$

and $(A4)$ it follows that

$$\gamma_n^{i^*} = \frac{\Gamma(i^*+1)}{\Gamma(i^*+1-h)} \cdot \frac{\Gamma(n+i^*+1-h)}{\Gamma(n+i^*+1)} \sim \gamma_n^s \sim n^{-h}, \tag{40}$$

where the last asymptotic equivalence used the fact that

$$\lim_{n\to\infty} n^{b-a} \frac{\Gamma(n+a)}{\Gamma(n+b)} = 1, \tag{41}$$

see Abramowitz and Stegun (1972), formula 6.1.46. Hence Lemma 4 applies and $\delta_t' P^n Q$ is equal to (27).

To show that $B \sim n\,\gamma_n^s \sim n^{1-h}$, notice that (B) and the identity $\frac{\gamma_{n+i}^s}{\gamma_i^s} = \gamma_n^{i+s}$ yield

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

which by Assumption 5 is bounded for all n. γ_i^{n+s} is a strictly positive quantity bounded above by one and below by zero such that, for $n \geq i^* - s$,

$$\gamma_i^{n+s} = \frac{\Gamma(n+s+1)}{\Gamma(n+s+1-h)} \cdot \frac{\Gamma(n+i+s+1-h)}{\Gamma(n+i+s+1)}. \tag{42}$$

Assumption 5, the fact that by (41) and (42) $\lim_{n\to\infty} \gamma_i^{n+s} = 1, \forall i$, together with the Lebesgue dominated convergence theorem imply that

$$\lim_{n\to\infty} \sum_{i=0}^{\infty} \frac{\left|\delta_t^{i+s+1}\right|}{\gamma_i^s} \gamma_i^{n+s} = \sum_{i=0}^{\infty} \frac{\left|\delta_t^{i+s+1}\right|}{\gamma_i^s} < \infty. \tag{43}$$

But then (13), (41) and (42) immediately yield $B \sim n\,\gamma_n^s \sim n^{1-h}$.

To prove that $C = O(n^{1-h})$ I use two preliminary results. Firstly, from (C) and the identity $\frac{\gamma_{n+i}^s}{\gamma_i^s} = \gamma_n^{i+s}$ I obtain that

$$C = \gamma_n^s \sum_{i=0}^{\infty} \frac{\delta_t^{i+s+1}}{\gamma_i^s} \gamma_i^{n+s}\ (i+s)\,. \tag{C'}$$

Secondly, I use the following result concerning sums of Gamma functions:

$$\sum_{i=1}^{\infty} \frac{\Gamma(i+b-1-a)}{\Gamma(i+b)} = \frac{\Gamma(b-a)}{a\Gamma(b)}, \tag{44}$$

where $b > a > 0$ and b is an integer (see Gradshteyn and Ryzhik , 1997, formula 0.247 and Granger and Joyeux, 1980, p. 18).

Consider first the case where $h > 1$. From Assumption 5 together with (C'), (41), (42) and (44) it follows that

$$|C| \sim\ \gamma_n^s \frac{\Gamma(n+s+1)}{\Gamma(n+s+1-h)} \cdot \frac{\Gamma(n+s+1-h)}{(h-1)\,\Gamma(n+s)} \sim n\gamma_n^s.$$

Consider then the case $h \leq 1$ and write

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

which by Assumption 3 and Lemma 3 is bounded for all n. If one proves that $\frac{\gamma_i^{n+s}}{n+s}$ is decreasing in n, (13), the fact that $\lim_{n\to\infty} \frac{\gamma_i^{s+n}}{n+s} = 0$, together with the Lebesgue dominated convergence theorem imply that $C = o(n\,\gamma_n^s)$. Hence, for a generic h, $C = O(n\,\gamma_n^s)$ so that $C = O(n^{1-h})$ by (40).

After defining the Psi-function $\psi(\cdot)$

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

and using the recursion formula $\psi(z+1) - \frac{1}{z} = \psi(z)$ (see Abramowitz and Stegun, 1972, formula 6.3.5), I obtain that

$$\frac{d\left(\ln\frac{\gamma_i^{n+s}}{n+s}\right)}{dn} = [\psi\,(n{+}s)\text{-}\psi\,(n{+}s{+}1\text{-}h)+\psi\,(n{+}i{+}s{+}1\text{-}h)\text{-}\psi\,(n{+}i{+}s{+}1)]. \tag{45}$$

Thus $h \leq 1$, together with the strictly increasing nature of $\psi\,(\cdot)$ over the positive real line (see Abramowitz and Stegun, 1972, formula 6.3.16), imply that (45) is negative and $\frac{\gamma_i^{n+s}}{n+s}$ is indeed decreasing in n

Finally to show that $D \sim n^{-h}$, use (D) to write

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

which by Assumption 3 and Lemma 3 is bounded for all n. Notice that $\gamma_i^{n+s-1}$is a strictly positive quantity bounded above by one and below by zero. Hence Assumption 5, (13), the fact that by (41) and (42) , $\lim_{n\to\infty}\gamma_i^{n+s} = 1$, $\forall i$, together with the Lebesgue dominated convergence theorem immediately imply that $D \sim \gamma_n^{s}$, so that $D \sim n^{-h}$ by (40). $\|$

Proof of Proposition 4 I prove first that $W_n \sim n^{1-h}$ and then that $\delta_t' P^n Q \sim n^{1-h}$, thus concluding the proof.

To show that $W_n \sim n^{1-h}$ argue by contradiction and suppose that $\lim_{n\to\infty}(n+b)^{h-1}W_n = 0$ where b is an arbitrary positive quantity (similar reasoning would apply to the contradiction $\lim_{n\to\infty}(n+b)^{h-1}W_n = \infty$). It follows from equation (31) that $\forall n > 0$

$$(n+b)^{h-1}W_n = \tilde{W}_n = -\sum_{i=0}^{n-1}\gamma_{i+1}\,h(n,i)\;\tilde{W}_{n-i-1} + (n+b)^{h-1}\gamma_{n+1}\;(n+1)\,,$$

where $\tilde{W}_n = (n+b)^{h-1}W_n$ while$h(n,i) = \frac{(n+b)^{h-1}}{(n-i-1)^{h-1}}$. Hence one can write

$$(n+b)^{h-1}\gamma_{n+1}\;(n+1) - \left|-\sum_{i=0}^{\infty}\gamma_{i+1}\,h(n,i)\;\tilde{W}_{n-i-1}\right| \leq \left|\tilde{W}_n\right| \leq$$

$$\leq (n+b)^{h-1}\gamma_{n+1}\;(n+1) + \left|-\sum_{i=0}^{\infty}\gamma_{i+1}\,h(n,i)\;\tilde{W}_{n-i-1}\right|. \tag{46}$$

I now show that the positive quantity

$$\lim_{n\to\infty}\sum_{i=0}^{n-1}\gamma_{i+1}h(n,i) \tag{47}$$

is bounded. Indeed, the same reasoning that led to (40) together with $(A4)$ imply that, as $i \uparrow \infty$, $\gamma_i \sim i^{-h}$. But from this it follows that

$$\begin{aligned}\sum_{i=0}^{n-1}\gamma_{i+1}h(n,i) &\sim \sum_{i=0}^{\frac{n}{2}}\frac{(n+b)^{h-1}}{(n\text{-}i\text{-}1\text{+}b)^{h-1}(i\text{+}1)^h}+\sum_{i=\frac{n}{2}+1}^{n-1}\frac{(n+b)^{h-1}}{(n\text{-}i\text{-}1\text{+}b)^{h-1}(i\text{+}1)^h}\\ &\leq \left(\frac{1}{2}+\frac{b}{n}\right)^{1\text{-}h}\sum_{i=0}^{\frac{n}{2}}\frac{1}{(i+1)^h}+\left(\frac{n}{2}\right)^{\text{-}h}\sum_{i=\frac{n}{2}+1}^{n-1}\left(1-\frac{i}{n}+\frac{b}{n}\right)^{1\text{-}h},\end{aligned}$$

where the first term is bounded since $h > 1$, while the second goes to zero when n goes to infinity since

$$\left(\frac{n}{2}\right)^{\text{-}h}\sum_{i=\frac{n}{2}+1}^{n-1}\left(1-\frac{i}{n}+\frac{b}{n}\right)^{1-h} \sim \left(\frac{n}{2}\right)^{\text{-}h}\left[\sum_{i=0}^{\frac{n}{2}}\left(\frac{i}{n}\right)^{1-h}+H\right] \sim \left(\frac{n}{2}\right)^{\text{-}h}\frac{n^{2-h}}{n^{1-h}},$$

with H being a positive bounded quantity. But then using (46) and after following the same reasoning as that in the proof of Lemma 5 one obtains that under the assumption that $\lim_{n\to\infty}\tilde{W}_n = 0$, $\forall\epsilon$, it does exist N^* such that $\forall n > N^*$

$$(n+b)^{h-1}\gamma_n\, n-\epsilon \leq \left|\tilde{W}_n\right| \leq (n+b)^{h-1}\gamma_n\, n+\epsilon,$$

that is a contradiction since $\gamma_n\, n \sim n^{1-h}$.

To prove that $\delta_t' P^n Q \sim n^{1-h}$ notice that irreducibility together with $W_n \sim n^{1-h}$ imply that for all i it does exist N^* such that $\forall n > N^*$,

$$0 < W_{n+i} < HW_n \tag{48}$$

where H is a positive bounded quantity. Moreover, $\gamma_n \sim n^{-h}$, $h > 1$ together with Assumption 5 imply that

$$\sum_{j=0}^{\infty}\sum_{i=j}^{\infty}\delta_t^{i+1}\frac{\gamma_{i-j}}{\gamma_{i+1}} = \sum_{i=1}^{\infty}\frac{\delta_t^i}{\gamma_i}\sum_{j=0}^{i-1}\gamma_j < \infty.$$

But this together with (33), (48) and the Lebesgue dominated convergence theorem yield $\delta_t' P^n Q \sim W_n \sim n^{1-h}$. $\|$

Proof of Proposition 5 (Small versus large shocks) One wishes to show that $\phi_n = \nu\delta_t' P^n Q \sim n^{d-1}$, where $d = 3-h$. Without loss of generality and to simplify notation assume that $i^* = s = \kappa$.

Consider first the case where Assumption 2' holds. Then show that Assumptions 1 and 3 also hold and apply Lemma 4 to write $\delta_t' P^n Q$ as given by (27). Finally, after noticing that $A = O(\rho^n)$, show that $B \sim n^{2-h}$, $C \sim n^{2-h}$ and $D \sim n^{1-h}$. Consequently $\phi_n = \nu \delta_t' P^n Q \sim n^{2-h}$, and completing the proof for this case.

To check that Assumptions 1 and 3 hold, notice that $(A4)$ implies that, $\forall j, i > 0$,

$$\gamma_j^{i+s} = \frac{\Gamma(i+s+1)}{\Gamma(i+s+1-h)} \cdot \frac{\Gamma(i+j+s+1-h)}{\Gamma(i+j+s+1)}. \tag{49}$$

As $h > 2$, (41) together with (49) imply that Assumptions 1 and 3 are satisfied and Lemma 4 applies.

To show that $B \sim n^{2-h}$, notice that (41), (44), (B) and $(A5')$ imply that

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

To prove that $C \sim n^{2-h}$ notice that (44) implies that, for any positive rational a and any integer b such that $b > a+1$,

$$\sum_{i=1}^{\infty} i \frac{\Gamma(i+b-2-a)}{\Gamma(i+b)} = \frac{\Gamma(b-1-a)}{a(1+a)\Gamma(b-1)}. \tag{50}$$

But then (41), (C) and $(A5')$ together with (50) yield

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

To show that $D \sim n^{1-h}$ simply make use of (41), (44), (D) and $(A5')$ to write

$$D = -\bar{\delta} S \frac{\Gamma(s+1)}{\Gamma(s+1-h)} \cdot \frac{\Gamma(n+s-h)}{(h-1)\Gamma(n+s-1)} \sim n^{1-h}.$$

Consider now the case where Assumption 2' fails and P is irreducible. From similar arguments as those used in the proof Proposition 4 it follows that $W_n \sim n^{1-h}$. This together with (33), (49) and $(A5')$ imply that

$$\delta_t' P^n Q = -\bar{\delta} \sum_{j=0}^{\infty} W_{n+j} \sum_{i=0}^{\infty} \gamma_i \sim n^{2-h},$$

which concludes the proof. $\|$

Proof of Proposition 6 As Assumptions 1, 2' and 3 hold, Lemma 4 applies. By assumption, $\delta_t^i = 0, \forall i > s$, so that (27) immediately yields $\delta_t' P^n Q = A = O(\rho^n)$. $\|$

References

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

Aghion, P., and Saint-Paul, G., 1998. On the Virtues of Bad Times: Interaction between Productivity Growth and Economic Fluctuations, Macroeconomic Dynamics 2, 322-344.

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

Beran, J., 1994. Statistics for Long-Memory Processes (Chapman & Hall, New York).

Bertola, G., and Caballero, R., 1990. Kinked Adjustment Costs and Aggregate Dynamics, in: Blanchard, O., and Fischer, S., eds., NBER Macroeconomics Annual 1990, (MIT Press, Cambridge MA).

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

Blundell, R., Griffith, R., and Van Reenen, J., 1999. Market Share, Market Value and Innovation in a panel of British manufacturing Firms, Review of Economic Studies 66, 529-554.

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

Caballero, R., and Engel, E., 1999. Explaining Investment Dynamics in U.S. Manufacturing: a Generalized (S,s) Approach, Econometrica 67, 783-826.

Caballero, R., and Hammour, M., 1994. The Cleansing Effect of Recessions, American Economic Review 84, 1350-1368.

Caballero, R., Engel, E., and Haltiwanger, J., 1997. Aggregate Employment Dynamics: Building from Microeconomic Evidence, American Economic Review 87, 115–137.

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, 671-698.

Evans, D.S., 1987. Tests of Alternative Theories of Firm Growth, Journal of Political Economy 87, 657-674.

Feller, W., 1971. An Introduction to Probability Theory and its Applications, Volume II (John Wiley & Sons, New York).

Geweke, J., and Porter Hudak, S., 1983. The Estimation and Application of Long Memory Time Series Models, Journal of Time Series Analysis 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 (Librarie du Recueil Sirey, Paris).

Gil-Alana, L., and Robinson, P. M., 1997. Testing Unit Root and other Non-Stationary Hypotheses in Macroeconomic Time Series, Journal of Econometrics 80, 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, 49-60.

Goncalves, E. and Gourieroux, C., 1988. Agregation de processus autoregressifs d'ordre 1, Annales d'Economie et Statistique 12, 127-149.

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

Granger, C., 1980. Long Memory Relationships and the Aggregation of Dynamic Models, Journal of Econometrics 14, 227-238.

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

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

Howitt, P., 1999. Steady Endogenous Growth with Population and R&D Inputs Growing, Journal of Political Economy 107, 715-730.

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, 897-915.

Jovanovic, B., and Nyarko, Y., 1996. Learning by Doing and the Choice of Technology, Econometrica 64, 1299-1310.

Kalecki, M., 1945. On the Gibrat Distribution, Econometrica 13, 161-170.

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

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

Kemeny, J., Snell, J., and Knapp, A., 1966. Denumerable Markov Chains (Van Nostrand Company, Princeton New Jersey).

Lippi, M., and Zaffaroni, P., 1999. Contemporaneous Aggregation of Linear Dynamic Models in Large economies, Mimeo, Servizio Studi, Banca d'Italia.

Michelacci, C., 1998. Growth with Cross-Sectional Heterogeneity, chapter 4, (Phd Dissertation, London School of Economics).

Michelacci, C., and Zaffaroni, P., 2000. (Fractional) Beta Convergence, Journal of Monetary Economics 45, 129-153.

Mortensen, D., and Pissarides, C., 1998. Technological Progress, Job Creation, and Job Destruction, Review of Economic Dynamics 1, 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.

Robinson, P.M., 1978. Statistical Inference for a Random Coefficient Autoregressive Model, Scandinavian Journal of Statistics 5, 163-168.

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

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

Solow, R., 1960. Investment and Technical Progress, in: Arrow, K.J., Karlin, S., and Suppes, P., eds., Mathematical Methods in the Social Sciences (Stanford University Press, Stanford California).

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

Sutton, J., 1997. Gibrat's Legacy, Journal of Economic Literature 35, 40-59.

Velasco, G., 1999. Non-stationary log-periodogram regression, Journal of Econometrics 91, 325-371.

Velasco, G., 2000. Non-Gaussian log-periodogram regression, Econometric Theory 16, 44-79.

Statistics:	Long-memory		Weak-memory	
	No aggregate shock	With aggregate shock	No aggregate shock	With aggregate shock
Estimated-d (s.e)				
Bandwidth = $T^{0.3}$ (Asym. s.e. = 0.23)	0.45 (0.45)	0.45 (0.45)	-0.05 (0.35)	-0.07 (0.36)
Bandwidth =$T^{0.525}$ (Asym. s.e.= 0.10)	0.67 (0.22)	0.66 (0.22)	0.08 (0.12)	0.08 (0.13)
AR coef. (s.e)	0.55 (0.08)	0.50 (0.13)	0.80 (0.03)	0.79 (0.03)

Table 1: **Estimates of d**. The table reports the average of the estimate of d and its corresponding standard error. Each economy is simulated for 1000 times over 1000 periods. In running the log-periodogram regression, the trimming coefficient is set equal to zero while the bandwidth is equal to T^{α} where $T = 1000$ is the sample size. The AR-coefficients are estimated by OLS after fractionally differencing detrended output with the theoretical $d = 0.5$ and $d = 0$ in the long and weak memory economy, respectively. The initial state of any firm j is an independent drawing from the distribution function given by Lemma 2. The number of firms in the economy is discrete and equal to $n = 10$. The distribution function $F(\cdot)$ is uniform with support $[0, 1]$. If no aggregate shock is present, λ is *iid* over time and across units, while in the model with aggregate shock λ is the sum of an idiosyncratic component drawn from a uniform distribution with support $[0, 1/2]$ and a white noise aggregate component which get values zero or 1/2 with probability 1/2.