Cross-Sectional Heterogeneity and the Persistence of Aggregate Fluctuations

Claudio Michelacci
CEMFI

Working Paper No. 9906
Junio 1999

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.
(E-mail address: c.michelacci@cemfi.es).

CEMFI, Casado del Alisal 5, 28014 Madrid, Spain.
Tel: 34 91 4290551, fax: 34 91 4291056, www.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 transmission 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, 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 behaviour 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 continuously". 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 suggested that the low frequency behaviour of aggregate time series might be

[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.

the 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 addresses, incidentally, the question 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 behaviour 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 5 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. Section 4 shows that the model can replicate the observed degree of aggregate persistence. Section 5 discusses the roles of each assumption. Section 6 relates the model developed here to Granger (1980). Section 7 concludes. The appendix contains the derivation of most of the results contained in 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 X_t, $t \geq 0$ (if it exists) reads like

$$X_t = X_0 + \gamma\, t + \sum_{n=0}^{t} \phi_n\, \epsilon_{t-n}, \tag{1}$$

where X_0 and $\gamma \geq 0$ capture respectively initial conditions and a deterministic trend, the quantities ϕ_n's are the Wold coefficients while the shocks

[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).

ϵ_t (usually assumed to be white noise) are called 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.

A very general way to model the rate of decay of the Wold coefficients consists of assuming that

$$\phi_n = \tilde{\phi}_n + d\, n^{d-1} + o\left(\tilde{\phi}_n\right), \tag{2}$$

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, while $o\left(\tilde{\phi}_n\right)$ indicates a quantity of lower order than $\tilde{\phi}_n$, that is $\lim_{n\to\infty} \frac{o\left(\tilde{\phi}_n\right)}{\tilde{\phi}_n} = 0$. 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 representation (2) nests standard time series model in an environment that maintains a great degree of continuity. For example, a standard trend stationary process with $ARMA$ disturbance exhibits Wold coefficients ϕ_n's decaying no more slowly than an exponential rate. This implies that the persistence is low and that the parameter of fractional integration d is equal to zero. 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 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. In fact, in the $ARIMA$ processes shocks can either propagate at constant (or increasing) rates or have permanent effects, like in the case of the unit root. For example, an exponential rate of absorption, $\phi_n \sim \rho^n$[10], 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 $n+1$. Hence, $ARIMA$ 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

[10]'$\sim$' denotes asymptotic equivalence for $n \uparrow \infty$, that is that the ratio of the left- and right-hand side tends to a bounded quantity bounded away from zero. Henceforth all asymptotic equivalences, if not otherwise specified, are taken with respect to the index n.

constant (or increasing) rates or have permanent effects. That is why it is useful to allow for orders of integration different from zero and one.

An order of integration different from zero ($d \neq 0$) allows for the possibility of decreasing rates of absorption. In fact when the rate of absorption is hyperbolic, the Wold coefficients ϕ_n behave like $n^{d-1} \sim (1 - \frac{1-d}{1})(1 - \frac{1-d}{2}) \cdots (1 - \frac{1-d}{n})$[11]. 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 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 the absolute lack of memory.

There are different ways to estimate the order of integration of the time series. The most famous is based on the log-periodogram regression proposed by Geweke and Porter Hudak (1983)[12]. Diebold and Rudebusch (1989) and Michelacci and Zaffaroni (1998) ran log-periodogram regressions for the GDP per capita for a set of different OECD economies and show that the parameter d, the order of integration of the time series, is between zero and one[13]. 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* [14].

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 1).

[11] See for example equation (34) in the appendix for a derivation of the result.

[12] 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.

[13] The log-periodogram regression is a semi-parametric estimator. It implies that the econometrician must choose the bandwidth, that is the number of Fourier frequencies used to approximate the behaviour of the periodogram around zero frequency. For the case of US aggregate GDP, the results are very robust with respect to the choice of the bandwidth.

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

This implies that the new information delivered by the Wold innovations ϵ_t is irrelevant for forecasting on very long horizons and is incompatible with a unit root in output[15].

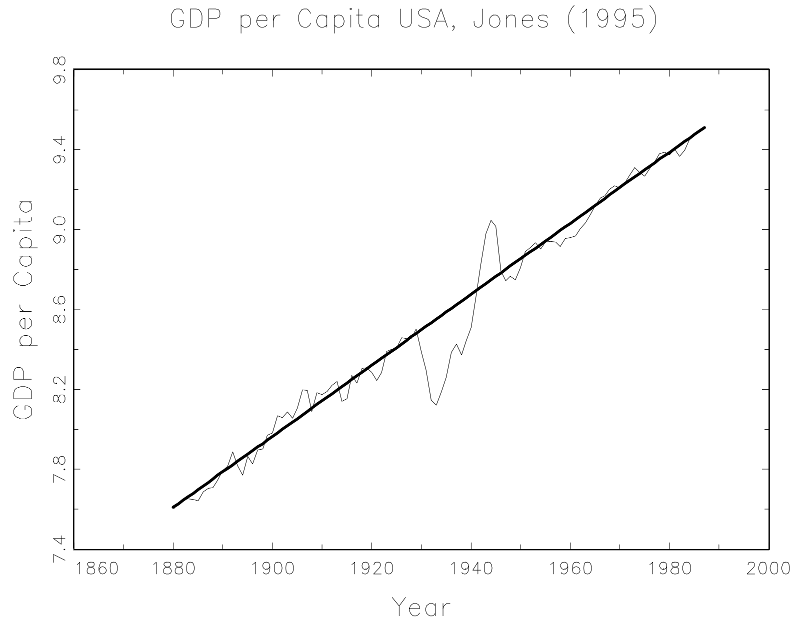

Figure 1: **Per Capita GDP in the United States, 1880-1987** (Natural Logarithm). The Data are from Maddison (1982,1989) as used in Jones (1996). The solid bold line represents the time trend calculated using data only from 1880 to 1929.

The empirical evidence suggests that the underlying stochastic process for aggregate GDP exhibits some form of long memory even if the degree of memory might remain uncertain, i.e. if the particular case of the unit root can be ruled out. In this paper we draw on the observed empirical relation between the expected growth of a firm and its size to provide a microfoundation for the observed degree of aggregate persistence, an order

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

of integration (weakly) between zero and one. 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 might be a key element in explaining aggregate persistence.

3 A vintage 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 response to the shock. The model has two key ingredients. Firstly, technological adoption is *costly:* to capture the productivity benefits of technical changes, older capital vintages must be replaced with more recent equipment. 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[16].

3.1 The Model

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

The rate of technological progress is exogenous at rate γ.

The number of firms in the economy is fixed with Lebesgue measure equal to one. Firms are infinitely lived, risk-neutral and maximizes expected returns in output units discounted at rate $r > 0$.

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 $t - i$. Firms using different technologies are able to produce different quantities of

[16] See for example, Solow (1960), Aghion and Howitt (1994), Caballero and Hammour (1994, 1996) and Jovanovic and Nyarko (1996). The model could be easily embedded in a search theoretic framework with fixed amount of resources, where each operating firm requires a given amount of resources and non operating firms must wait for these resources to be freed before using them. See for example Pissarides (1990). The assumption that the number of firms is fixed (see below) would then be the result of a *free entry condition* equating the benefits of creating a new firm with the cost of creation. 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.

goods, more exactly a firm in state i at time t produces a quantity of goods equal to $\gamma(t-i)$[17].

We indicate with π_t the vector of countably infinite dimension collecting the measure of firms in each state. The ith element of the vector π_t measures the number of firms using technology $t-i+1$ at time t. π_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 switching to a better technology. Technological adoption implies some costs which are assumed to be fixed and independent of the technology adopted[18]. Therefore, if adopting a new technology, the firm always invests in the leading technology in the economy and it will be in state zero in the next period. We 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 across units and over time with common distribution F over the support (possibly unbounded) $\Lambda \subseteq \Re$ and zero expected value. λ 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

[17] It is possible to think that all variables are denominated in logs, implying that differences indicate growth rates while arithmetic averages indicate the logarithm of geometric ones. If so, 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.

[18] This follows, for example, Caballero and Engel (1991, 1993 and 1998), Caballero, Engel and Haltiwanger (1997) and Mortensen and Pissarides (1998).

be read indifferently as either a technological or a demand shock[19]. Risk neutrality of firms implies that 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

$$\begin{aligned} V(i,\,t,\,\lambda) &= \max_{s\in\{0,1\}} \gamma(t-i) - s\,(c_i+\lambda) + \\ &\quad +\beta\,(1-s)\,V^e\,(i+1,\,t+1) + \beta\,s\,V^e\,(0,\,t+1). \end{aligned} \tag{3}$$

$0<\beta=\frac{1}{1+r}<1$ 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 λ. It follows from dynamic programming arguments that the problem is well defined[20]. In particular the value function $V(i,\,t,\,\lambda)$ is linear in t, weakly decreasing in λ 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 λ is such that

$$\beta\,[V^e\,(0,\,t+1) - V^e\,(i+1,\,t+1)] \geq c_i+\lambda. \tag{4}$$

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 (4) occurs[21], 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

$$1-p_i = F\left(\beta\,[V^e\,(0,\,t+1) - V^e\,(i+1,\,t+1)] - c_i\right), \;\; \forall i. \tag{5}$$

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

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

[20] 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 (3) and the corresponding sequential problem.

[21] 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.

$$P=\begin{bmatrix} 1-p_0 & p_0 & 0 & 0 & 0 & 0 & \cdots \\ 1-p_1 & 0 & p_1 & 0 & 0 & 0 & \cdots \\ 1-p_2 & 0 & 0 & p_2 & 0 & 0 & \cdots \\ 1-p_3 & 0 & 0 & 0 & p_3 & 0 & \cdots \\ 1-p_4 & 0 & 0 & 0 & 0 & p_4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{6}$$

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 π_t at time t into the cross-sectional distribution π_{t+1} at time $t+1$.

The next sub-section focus on the steady state properties of the system, while the following 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 sections 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 (weakly) 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 γ and β, 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 (3) and (5). Lemma 1 guarantees the validity of this 'semi-structural' approach: any assumption on the transmission mechanism P is the result of a corresponding assumption on the structural parameters of the model. In section 6 we will do an exercise of reverse engineering and we will analyze the structural interpretation of a set of assumptions on the transmission mechanism P.

Lemma 1 (Validity of the 'semi-structural' approach) *Given a distribution function $F(\cdot)$, the parameters γ and β, and an arbitrary sequence of probabilities $\{p_i, i \geq 0\}$, there does exist a sequence of adjustment costs $\{c_i, i \geq 0\}$, whose solution to the firm problem, defined by equations (3) and (5), 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 (5). 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[22]. Models with switching costs and human capital specificity suggest, however, why the probabilities $1- p_i$ may be decreasing in i[23]. The cost of adopting the leading technology c_i is in general positively related to i[24] and firms using new technologies are more likely to end up on the technological frontier rather than firms far away

[22] For example in Aghion and Howitt (1994), Caballero and Hammour (1994b) and Mortensen and Pissarides (1996), both the probability distribution of the idiosyncratic shocks λ 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.

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

[24] 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 (4). Discounting, $\beta < 1$, implies however that induced changes on the left hand side are always smaller than those on impact on the right hand side of equation (4). As a result an increase in c_i always reduces the value of $1 - p_i$. For a formal proof see equation (19) in the appendix.

from it. This reflects the fact that the higher the technology gap the more 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. Therefore, we impose that, whatever its current state, a firm sooner or later will adjust with probability one.

Assumption 1 Indicate with β_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' of the economy.

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. To analyze under which conditions a steady state distribution exists it is useful to 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 guarantees that the transmission mechanism P is reducible.

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 and 6. In the paper we consider a technical modification of Assumption 2 and we call it Assumption 2'. It ensures that, once entered the recurrent class, units do not jump deterministically from one state to the other[25]. Neither Assumption 2 or 2' will be maintained assumptions throughout the analysis, their role is limited and in section 6 we will remove them.

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

Assumption 2' Assume that Assumption 2 holds and that if $i^* > 1$, it does exist $1 \square i < i^*$ such that $\beta_i \neq 1$.

Lemma 3 (Existence of a Steady-State distribution) *Assumption 1 is given. The transmission mechanism P is reducible if and only if Assumption 2 holds. 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 ϵ_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, ϵ_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 λ becomes equal to $c_i + \lambda + \epsilon_t$. Equations (3) and (5) 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 λ is such that

$$\beta \left[V^e (0, t+1) - V^e (i+1, t+1)\right] \geq c_i + \lambda + \epsilon_t. \qquad (4')$$

We indicate with $1-\ p_i(\epsilon_t)$ the probability that the event (4') occurs, that is

$$1 - p_i(\epsilon_t) = F\left(\beta \left[V^e (0, t+1) - V^e (i+1, t+1)\right] - c_i - \epsilon_t\right), \ \ \forall i, \qquad (5')$$

is the probability that, when the aggregate shock is equal to ϵ_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 (6) collecting the probabilities $p_i(\epsilon_t)$. The dynamics of the cross-sectional distribution of vintages currently in use, π_t, are therefore described by the equation

$$\pi_t = \delta_t + P' \pi_{t-1} \tag{7}$$

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 π_{t-1} at time $t-1$ into the cross-sectional distribution π_t at time t. The infinite dimensional column vector δ_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 δ_t measures the size and structure of the reallocation and has two general properties. Firstly, the sum by column of its entries, δ_t^i, $i \geq 1$, is exactly equal to 0, that is

$$1' \delta_t = 1' \left(\bar{P}(\epsilon_t) - P\right)' \pi_{t-1} = 0, \tag{8}$$

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

$$I\left(\delta_t^1 > 0\right) = I\left(-\delta_t^j > 0\right) = I\left(-\epsilon_t > 0\right), \ \forall \left|\delta_t^j\right| \neq 0, \ j \neq 1. \tag{9}$$

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

We assume that the reallocation structure δ_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 δ_t where only a finite number of coefficients δ_t^i are strictly positive.

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

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 π_{t-1} at time $t-1$, the level of output at time $t+n$ is equal to[26]

$$Y_{t+n} = \gamma (t+n) - \gamma \pi_{t-1}{}' P^n O - \gamma \delta_t{}' P^n O \tag{10}$$

while it would have been equal to

$$Y_{t+n} = \gamma (t+n) - \gamma \pi_{t-1}{}' P^n O \tag{11}$$

in the absence of the shock. The difference between (10) and (11) gauges the dynamic response of the aggregate economy to the shock ϵ_t. In other words the quantities

$$\phi_n = -\gamma \delta_t' P^n O, \ \forall n \tag{12}$$

are analogous to the Wold coefficients analysed in the previous section: ϕ_n gauges the effect in the economic system of the shock ϵ_t, n periods ahead. Therefore as in equation (1) the rate of decay of ϕ_n measures the persistence of the shock in the model.

4 The Persistence of Aggregate Fluctuations

We assume that Assumptions 1 and 3 hold throughout the analysis. We measure the degree of persistence of the shock ϵ_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

[26] Lemma 4 in the appendix shows that the infinite dimensional matrix products $\delta_t{}' P^n O$ are bounded and well defined for all n.

coefficients in (1), the rate of decay of the pseudo Wold coefficients ϕ_n in (12) gauges the persistence of the shock in the model. In particular we say that the transmission mechanism P, together with the reallocation structure δ_t, exhibits an order of integration d if, for $n \uparrow \infty$, $\phi_n \sim n^{d-1}$ as in (1).

Firstly we show under which conditions the model cannot generate a order of integration different from zero. 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 the model can not generate an order of integration in the aggregate different from zero.

Proposition 1 (Increasing hazard functions) *Assumption 3 holds. Suppose that there does exist a state* $\square \geq i^*$ *such that* $p_{\square} < 1$ *and that the probabilities* p_i *are weakly decreasing for any* $i \geq \square$*. Then aggregate output always behaves like a trend stationary process with ARMA disturbances, that is the quantities* ϕ_n *in (12) decay at least exponentially, i.e.* $|\phi_n| \,\square\, K\rho^n$, $0 \,\square\, \rho < 1$, *where* K *is a positive bounded quantity.*

Proof: See appendix.

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[27]. This reflects the fact that the bigger the technological gap the more difficult is technological adoption.

[27]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.

We now draw on the relation between the expected growth of a firm and its size to show the model may generate a positive order of integration 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[28]. We classify as 'big' the firms that are adopting vintages close to the technological frontier, while all other firms are 'small'. We assume, temporarily, that Assumption 2 holds and we say that 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 γ as the technological frontier. Small firms might grow at rates different from γ. 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[29]. Proposition 2 shows that if all small firms are growing in the same way, i.e. $g_i = h\gamma$ independent of i (see Assumption A5), the model is capable at generating orders of integration different from zero. 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 $h = 1$.

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

[29] Under the conditions analyzed in footnote (17), g_i is the expected growth rate of a firm in state i just as γ is the growth rate of the aggregate economy.

Proposition 2 (Robust transmission mechanism) *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* ϕ_n *in (12) are such that, as* $n \uparrow \infty$, $\phi_n \sim n^{d-1}$, *where* $d = 2 - h$.

Proof: See appendix.

The additional assumption ($A4$) requires firstly that some units end up in the non recurrent set ($\delta_t^{s+i+1} \neq 0$ for some i). Secondly, it puts some boundaries (in addition to Assumption 3) on the amount of reallocation that is driven by the aggregate shock. For example ($A4$) is satisfied in the case of 'finite' reallocations, that is for reallocation structures δ_t where only a finite number of coefficients δ_t^{s+i+1} are strictly different from zero. Its role is technical and we will discuss it further in the next section. We will then show that ($A4$) bounds the persistence generated by the model: once we relax it, the degree of persistence generated by the model increases.

In the next section we will argue that ($A5$) is the crucial assumption to replicate the observed degree of aggregate persistence: all small firms are growing in the same way, i.e. $g_i = h\gamma$ independent of i. Indeed, 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 ϵ_t hits the system, firms are reallocated across sizes according to the reallocation structure δ_t. Whether the shock will be absorbed or not depends then on the relative growth of small versus big firms. 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 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 degree of persistence observed in the time series of aggregate output.

Assumption ($A5$) allows one to summarize Proposition 2 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[30].

(ii) If $h < 1$, big firms grow faster than small ones and the first difference of aggregate output exhibits long memory[31]. 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 exactly 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, $d = 1$.

We conclude that the transmission mechanism P is *robust* because it generates an order of integration different from zero independently of the relative growth of small versus big firms.

5 Interpretation of the Assumptions

Proposition 2 is particularly dependent on Assumptions 2', ($A4$) and ($A5$). We now discuss in more detail their roles and we analyze the consequences of removing both assumption 2 and ($A4$), but keeping ($A5$) as maintained assumption. We will see that, with some qualifications, the model still replicates the observed degree of aggregate persistence. That is why we argue that ($A5$) drives the results and that the empirical relation between expected

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

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

growth and firm size might provide a microfoundation for aggregate persistence. The section concludes with deriving properties of the sequence of adjustments costs $\{c_i, i \geq 0\}$ that implies $(A5)$.

Removing Assumption 2

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 and we maintain the assumption that a steady state distribution exists, $h > 1$ (see Lemma 3)[32], the impact effect of the shock cannot be amplified without limit and the model is unable to generate orders of integration greater than one (see Proposition 3). In other words Proposition 3 shows that Assumption 2' plays the role of increasing the *robustness* of the transmission mechanism.

Proposition 3 (Persistence in the irreducible case) *Suppose that the transmission mechanism P is irreducible, that Assumptions 3 and $(A4)$ hold for some s. Assume also that the expected growth of a unit in state i, g_i, is such that, for some arbitrary s^*,*

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

Then the order of integration of aggregate output d is equal to $2 - h$, that is the quantities ϕ_n in (12) are such that, as $n \uparrow \infty$, $\phi_n \sim n^{d-1}$, where $d = 2 - h$.

Proof: See appendix.

[32]In particular, if the transmission mechanism P is irreducible, h □ 1 implies that, for $i \uparrow \infty$, $\beta_i \sim i^{-h}$ and no steady state distribution exists by Lemma 3. In this case the shock has permanent and bounded effect on the level of output. This would be equivalent to a unit root in output, $d = 1$. In general, if the transmission mechanism P is irreducible and a steady state distribution does exist, the order of integration of aggregate output is always strictly smaller than one. I analyze this case in my Phd dissertation. See Michelacci (1998). See also (26) in Lemma 6 for an application of this result.

Removing (A4)

Assumption $(A4)$ can fail either because $\sum_{i=0}^{\infty} \frac{\left|\delta_t^{s+i+1}\right|}{\beta_i^s} = \infty$ or because $\sum_{i=0}^{\infty} \frac{\left|\delta_t^{s+i+1}\right|}{\beta_i^s} = 0$. It is important to distinguish one case from the other. In the first case $(A4)$ limits the size of the reallocation driven by the aggregate shock and it bounds the persistence generated by the model. We consider now a counter-example where $(A4)$ fails and $\sum_{i=0}^{\infty} \frac{\left|\delta_t^{s+i+1}\right|}{\beta_i^s} = \infty$. We assume for example that all firms are initially in the tail of the distribution of vintages, so that $\pi_{t-1}^i \sim \beta_i^{\square}$ as $i \uparrow \infty$ for some $\square$. For example, this would be the case, if the system is in steady state and the transmission mechanism is irreducible. We consider then the effect of a negative aggregate shock $\epsilon_t < 0$, so large in absolute value that $\delta_t^{s+i} = -a\,\beta_i^s$, $a > 0$. If $F(\cdot)$ is uniform, a would then be equal to the absolute value of the aggregate shock, $a = |\epsilon_t|$. Proposition 4 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 δ_t.

Proposition 4 (Small versus large shocks) *Assume that*

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

for some finite $\square$ *and that the expected growth of a unit in state* i, g_i, *is such that*

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

Then the order of integration of aggregate output d *is equal to* $3 - h$, *that is the quantities* ϕ_n *in (12) are such that, as* $n \uparrow \infty$, $\phi_n \sim n^{d-1}$, *where* $d = 3 - h$.

Proof: See appendix.

Corollary (The reallocation structure matters) *Assumptions 1 and 3 hold. For a given transmission mechanism* P, *different reallocation structures* δ_t *can generate different rate of absorption of the shocks as measured by the quantities* ϕ_n *in (12).*

Proposition 4 is interesting for three reasons. Firstly, it shows that the model can generate asymmetric responses to shocks. In fact $\delta_t = \left(\bar{P}(\epsilon_t) - P\right)' \pi_{t-1}$, and if $p_i = 1 - \frac{\gamma h}{1+1}$, only a negative shock can generate a difference between $p_i(\epsilon_t)$ and p_i equal to a constant, which is why a in ($A4$') may only be negative. Secondly, Proposition 4 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 4 delivers an alternative and suggestive characterization of aggregate persistence. According to this view most shocks generate low persistence[33]; 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, it may be well represented by a standard weak memory process[34]. 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.

Alternatively, assumption ($A4$) can fail if $\sum_{i=0}^{\infty} \frac{\left|\delta_t^{s+i+1}\right|}{\beta_i^s} = 0$. That is, given the assumption that P is reducible, ($A4$) 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 and the transmission mechanism P is reducible, assumption ($A4$) implies that the shock ϵ_t is positive. 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. However, when ($A4$) fails in this direction, aggregate output always behaves like a trend stationary process with $ARMA$ disturbances, see Proposition 5.

[33] For example $h > 2$ does not generate a positive order of integration under the conditions assumed in Propositions 2 and 3.

[34] See for example Perron (1989).

Proposition 5 (The role of cross-sectional heterogeneity) *Suppose that Assumptions 2' and 3 hold. Assume however that*

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

where $s = \max\{i : i \geq 0, \, p_i = 0\} + 1$. *Then aggregate output always behaves like a trend stationary process with ARMA disturbances, that is the quantities* ϕ_n *in (12) 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.

Proposition 5 is interesting. In fact, $\sum_{i=0}^{\infty} \frac{\left|\delta_t^{s+i+1}\right|}{\beta_i^s} = 0$ is equivalent to assuming that all units always remain in a finite dimensional set close to the technological frontier or, alternatively, that the number of (feasible) states in the system is finite and equal to s. If so, the order of integration of output is always equal to zero. Given that a variable with a finite number of states is bounded, it is not surprising that the aggregate can never be a non stationary process ($d \geq \frac{1}{2}$)[35]. This theorem however claims more, as it states that orders of integration different from zero ($d \neq 0$) can not be generated if cross-sectional heterogeneity is uniformly bounded. In this respect, Proposition 5 allows one to generalize results by Bertola and Caballero (1990, 1994), Caballero and Hengel (1990) in the sS literature. Despite the large non linearities implied in their framework, the rate at which the expected value of the cross-sectional distribution converges to its long run value is always exponential[36].

The proposition also illustrates the role of ongoing growth in producing the amount of cross-sectional heterogeneity required to generate a positive order of integration in the aggregate. When there is positive growth in technological progress and technological adoption is costly, a firm currently using

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

[36] The result is also consistent with the observation by Bertola and Caballero (1994), that an increase in the variance of the idiosyncratic uncertainty reduces the speed of convergence of the aggregate variable even if, in their set-up, the rate of convergence still remains exponential.

a given vintage can wait for an arbitrary number of periods before adopting a new technology. That is, technological progress makes cross-sectional heterogeneity to be not uniformly bounded.

A structural interpretation for ($A5$)

These considerations lead us to conclude that ($A5$) is the crucial assumption to generate a positive order of integration in aggregate output and to replicate the observed degree of aggregate persistence (see Proposition 2, 3 and 4)[37]. ($A5$) requires that all small firms grow in the same way, i.e. $g_i = h\gamma$ independent of i. It implies that the adjustment cost c_i must grow sufficiently fast when the technology gap i increases. An interesting question is how fast the sequence of adjustment cost $\{c_i, i \geq 0\}$ must grow to imply ($A5$). To analyze this question, some assumptions on the distribution function of the idiosyncratic shock, $F(\cdot)$, are required (see Lemma 1). For simplicity sake we rely on a uniform distribution.

Proposition 6 (Adjustment costs and the growth-size relation) *Assume that ($A5$) holds and that the distribution function of the idiosyncratic shock, $F(\cdot)$, is uniform with support $[0, 1]$, then, it does exist c_0 such that, the sequence of adjustment costs $\{c_i, i \geq 0\}$ satisfies*

$$c_i = a + \frac{\gamma}{\beta^{-1} - 1}(i+1) + f(i), \tag{13}$$

where a is a constant while , as $i \uparrow \infty$, $0 > f(i) \sim \frac{1}{i}$.

Proof: See appendix.

The values of the probability $1 - p_i$ are the outcome of two contrasting forces. On the one hand, the bigger the technological gap i, the greater is the gain to adopt new technologies. On the other hand, the bigger the technological gap the costlier is technological adoption. It can be proved that a

[37]There are some technical qualifications here. Proposition 2 allowed for any $h > 0$, Proposition 3 required $h > 1$ to ensure the existence of a steady state distribution, whereas Proposition 4 required $h > 2$ in order to have Assumption 3 satisfied.

constant hazard function, $p_i = p \ \forall i \geq 0$, implies that the sequence of adjustment cost $\{c_i, i \geq 0\}$ grows linearly[38]. That is, the two forces mentioned above, balance exactly and the present discounted value of the gains to adopt new technologies, equal to $\frac{\gamma}{\beta^{-1}-1}(i+1)$, grows at the same rate as the cost of technological adoption c_i. (13) implies that for $(A5)$ to hold the second force must dominate: as the technological gap, i, raises, the difference between the benefits of technological adoption and the associated cost falls as $\frac{1}{i}$. If so, the hazard function relative to adopting new technologies is decreasing in the size of the technological gap and the transmission mechanism P replicates the observed degree of aggregate persistence.

The empirical relation between expected growth and firm size might provide a justification for $(A5)$[39]. That is why we conclude with a note on the literature. In fact, conditional on survival small firms grow faster but also tend to die more often. The question to be addressed then is what growth we should attribute to a dead firm. In general equilibrium dead firms free resources that potential investors can now exploit, and therefore death might be associated with high growth. Once this sample selection problem is taken into account, the empirical evidence tends to conclude that small firms grow faster than big ones[40]. That is why we feel that our reading of the literature is the closest to the macro approach we are pursuing here and that micro and macro evidence matches quite closely.

6 Propositions 2, 3 and 4 and Granger (1980)

Granger (1980) showed that the aggregation of $AR(1)$ processes can lead to positive orders of integration in the aggregate if some suitable conditions on the cross-sectional distribution of the first order correlation of the process

[38] Following the same steps as in the proof of Proposition 6, it can be shown that, $p_i = p$ $\forall i \geq 0$ implies

$$c_{i+1} = \frac{\gamma}{\beta^{-1}-1} - (1-p) + \frac{\gamma(i+1)}{\beta^{-1}-1}, \qquad \forall i \geq 0.$$

[39] See also footnote (17) and (29).

[40] 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) correct for the sample selection problem and conclude that small firms grow faster than big ones.

are satisfied. In this section we relate propositions 2, 3 and 4 to Granger (1980) and we show which elements are required to generate a positive order of integration in the aggregate variable.

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

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

where i denotes the individual agent, n is the number of agents, ϵ_t is an aggregate shock, whereas ϵ_t^i is an idiosyncratic shock both assumed to be *iid* over time. He further assumes that the coefficients ρ_i are independent drawings from an absolute continuous function with density $g(\cdot)$ such that for $\rho \to 1^-$

$$g(\rho) \sim (1-\rho)^{h-2} \tag{14}$$

where h is a real parameter $\in (1, \infty)$ by the integrability constraint[41]. Granger shows that when n is sufficiently large, the aggregate variable $X_t = \frac{1}{n}\sum_{i=1}^{n} x_t^i$ behaves like a process with order of integration $d = 2-h$. We now relate this result with propositions 2, 3 and 4. There are indeed important similarities.

Consider for the example the dynamics of the state of a firm implied by (6). 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

$$x_t^i = \mu_i + \rho_i x_{t-1}^i + \epsilon_t^i \tag{15}$$

where $\rho_i = p_i$, $\mu_i = p_i$ while ϵ_t^i has zero mean and conditional distribution given by

$$\epsilon_t^i \mid x_{t-1}^i = \begin{cases} 1-p_i & p_i \\ -p_i\,(i+1) & (1-p_i)\,(i+1) \end{cases}. \tag{16}$$

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 3 the number of units $g(\rho)$ with persistence $\rho_i = p_i = \rho$ for $\rho \to 1^-$ is such that (in steady state)

$$g(\rho) \sim (1-\rho)^{h-2},$$

[41] Strictly speaking, Granger consider the case where the distribution $g(\cdot)$ is Beta(p, q). Robinson (1978) and Goncalves and Gourieroux (1988) show however, that the low frequency behaviour of the aggregate is determined only by the shape of the cross sectional distribution $g(\cdot)$ around one.

where $h > 1$. Proposition 3 seems surprisingly similar to Granger (1980). There are, however, important differences which explain proposition 2 and 4. In Granger, the degree of persistence of a unit i is fixed within the entire life of the firm. In the vintage model developed here the degree of persistence ρ_i changes as the firm changes state. This explain why, differently from Granger, the model can generate orders of integration greater than one, thus generating a unit root as a particular case (see proposition 2). Moreover, the model deals with aggregate shocks that are not *iid*. The model can then show the different behavior of the aggregate system in response to large versus small shocks (proposition 4).

The main message of this section is that generating an order of integration different from zero in output requires three necessary elements. It requires firstly 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 $g(\rho)$, and, finally, a sufficient number of units with very large persistence as implied by (14). In Granger (1980) the three elements are assumed one independently of the other. In the paper, we have shown that the first element might generate endogenously the other two, thus generating interesting implications for aggregate dynamics. The decreasing hazard function (see proposition 1) implied by $(A5)$ generates endogenously the required number of units with a sufficient amount of persistence which are required to replicate the observed degree of aggregate persistence[42].

7 Conclusions

We have analyzed how an aggregate shocks propagates in the economic system in the context of a standard vintage model. We have used the model to provide some microfoundations for the slow adjustment of output in response to shocks. In particular, we have argued that the empirical relation between growth and firm size might be able to reconcile the macro with the micro evidence. If Gybrat´s law holds, all firms grow at the same rate indepen-

[42]This analysis also suggests that the binary nature of the adjustment process implied by (3) does not seem crucial to generate the three elements together.

dently of their size and a unit root characterizes 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 the aggregate time series exhibit a fractional order of integration. Whether or not aggregate time series exhibit a positive order of integration has been much debated and certainly it will still be in the future. Formal investigation, however, has suggested that long memory might well represent the low frequency behaviour of time series. If so, micro and macro evidence matches quite closely and the process of ongoing churning and catching up that takes place in the economy might be a key factor in explaining aggregate persistence.

The tools introduced in the paper could potentially be used to analyze aggregate dynamics in other vintage models proposed in the literature. 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? Can aggregate fluctuations arise from the aggregation of purely idiosyncratic shocks? What is the role of technological progress (here γ) and the size of the market (here the number of firms) in replicating the observed degree of aggregate persistence? The ultimate task is to provide microfoundations for the many features which characterize aggregate dynamics. These issues will be addressed by further research.

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.

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.

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.

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.

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

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. (1980), "Long Memory Relationships and the Aggregation of Dynamic Models", *Journal of Econometrics* 14, 227-238.

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

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), "Growth with Cross-Sectional Heterogeneity", chapter 4, *Phd Dissertaion, London School of Economics.*

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.

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

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.

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.

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.

8 Appendix

8.1 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

$$1 - p_i = F\left(\lambda_R^i\right), \quad \forall i,$$

where $\tilde{V}^e(j)$ indicates the expected value of $\tilde{V}(j,\, \lambda)$ taken with respect to the random variable λ while λ_R^i is the solution to the equation

$$\beta\,[\tilde{V}^e(0) - \tilde{V}^e(i+1)] = c_i + \lambda_R^i, \quad \forall i. \tag{17}$$

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

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

Using (17) and taking expectations we obtain

$$\tilde{V}^e(i) = -\gamma i - p_i \lambda_R^i - c_i - \int_{-\infty}^{\lambda_R^i} s\; dF(s) + \beta \tilde{V}^e(0), \quad \forall i. \tag{18}$$

(17) together with (18) yields

$$\begin{aligned} c_{i+1} &= p_0 \lambda_R^0 + c_0 + \left(c_i + \lambda_R^i\right) \frac{1}{\beta} + \int_{-\infty}^{\lambda_R^0} s\; dF(s) - \gamma(i+1) + \\ &\quad -p_{i+1} \lambda_R^{i+1} - \int_{-\infty}^{\lambda_R^{i+1}} s\; dF(s), \quad \forall i. \end{aligned} \tag{19}$$

For any given sequence of probabilities $\{p_i, i \geq 0\}$, equation (19) 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 (17) and (18) a sequence of adjustment costs $\{c_i, i \geq 0\}$ that solves (19) delivers the given sequence of probabilities $\{p_i, i \geq 0\}$ as a solution of the firm problem. *Q.E.D.*

Proof of Lemma 2 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 communicate or state i is transient (see for example Karlin and Taylor 1975, 1981). 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 See Billingsley 1986, theorem 8.8. and example 8.13. *Q.E.D.*

8.2 Proofs of results in section 4

The next three Lemmas are used throughout the proofs.

Lemma 4 (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 of Lemma 4 $\delta_t = \delta^+ - \delta^-$ 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 bounded $\forall n$. This follows from Assumption 3, (9) and the fact that each element of $P^n O$ has increments bounded above from one. *Q.E.D.*

Lemma 5 (Reducible case) $\square = \max\{i : p_i = 0, i \geq 0\} + 1$ *while*

$$s = \begin{cases} \square \;\; if \;\; \square < \infty \\ i^* \;\; if \;\; \square = \infty \end{cases} \tag{20}$$

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,

$$\delta_t' P^n O = A + B + C - D, \tag{21}$$

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

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

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

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

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

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

Proof of Lemma 5 By definition (20) 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 Lemmas 2 and 3 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 P^n, the nth iterate of P, as follows

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

where r_j^n, $j \geq 0$, 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 β_n^{s+i}, $i \geq 0$, are in row $s+i+1$ and column $s+i+1+n$. It can be proved by recursion that for $i \geq s$

$$r_i^n = \sum_{j=0}^{n-1} \beta_j^i \left(1 - p_{i+j}\right) r_0^{n-1-j}, \tag{22}$$

where $\beta_0^i = 1$, while $r_0^0 = e_1'$, e_1 indicating a vector of dimension $s \times 1$ whose first element is equal to one and zero otherwise.

To prove (21) we proceed as follows. Firstly, we show that r_i^n, $0 \square i < s$ converges at least exponentially to π, the vector of dimension $s \times 1$ corresponding to the steady state distribution of $\tilde{P}$, that is

$$|r_i^n - \pi'| \square H \rho^n 1', \quad \forall i < s, \quad \forall n > 0 \tag{23}$$

where $1 \geq H \geq 0$, $0 \square \rho < 1$ 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

$$r_i^n = \left(1 - \beta_{n-1}^i\right) \pi' + \beta_{n-1}^i \left(1 - p_{i+n-1}\right) e_1' + a_i^n, \tag{24}$$

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. (23) together with (24), Assumption 3 and the properties (8) and (9) of δ_t will then imply that $\delta_t' P^n O$ is equal to (21) completing the proof.

We now prove (23). 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 π 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 (23) 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 $s - i^*$. This proves (23).

We now prove (24). From (22) and (23) 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 (24). *Q.E.D.*

Lemma 6 (Irreducible case) *Indicate with $R_0^{(n)}$ the expected position after n iterations of a unit starting in state 0. $Y_n = R_0^{(n+1)} - R_0^{(n)}$ satisfies the recursive relation given by*

$$Y_n = -\sum_{i=0}^{n-1} \beta_{i+1} Y_{n-i-1} + \beta_{n+1} \left(n+1 \right), \ \forall n > 0, \tag{25}$$

where $Y_0 = 0$.

Moreover, if the transmission mechanism is irreducible and a steady state distribution does exist

$$\lim_{n \to \infty} Y_n = 0, \tag{26}$$

$$\delta_t' P^n O = \sum_{j=0}^{\infty} Y_{n+j} \sum_{i=j}^{\infty} \delta_t^{i+1} \frac{\beta_{i-j}}{\beta_{i+1}}, \tag{27}$$

where $\beta_i = \beta_i^0 = \prod_{k=0}^{i-1} p_k > 0$ and $\beta_0 = 1$.

Proof of Lemma 6 By the law of iterated expectations, it follows that

$$R_0^{(n)} = \sum_{i=0}^{\infty} \beta_i (1 - p_i) R_0^{(n-1-i)} + \beta_n\, n, \ n \geq 0,$$

where $R_0^{(n)}=0$ if $n \leq 0$,while $\beta_i=\prod_{k=0}^{i-1} p_k$. It follows from the definition of Y_n that

$$Y_n = R_0^{(n+1)} - R_0^{(n)} = \sum_{i=0}^{\infty} \beta_i (1-p_i) Y_{n-i-1} + \beta_{n+1}\,(n+1) - \beta_n\, n, \; n \geq 0, \quad (28)$$

where $Y_n = 0$ if $n \leq 0$. Given (28), (25) can be proved by recursion.

To prove (26), we first note that a limit for Y_n always exists, with $\lim_{n\to\infty} Y_n = a, \quad 0 \leq a \leq 1$. This follows from recurrency, the basic limit theorem of Markov chains (see Karlin and Taylor 1975, theorem 1.2.) and the fact Y_n is uniformly bounded by one. We now show that if a steady state distribution does exist, $\lim_{n\to\infty} Y_n = 0$. Let us argue by contradiction and suppose that $\lim_{n\to\infty} Y_n = a > 0$. If so, $\forall \epsilon$, it does exist N^*, such that $\forall n > N^*$

$$|Y_n - a| = \left| - \sum_{i=0}^{\infty} \beta_{i+1} (Y_{n-i-1} - a) - a \sum_{i=0}^{\infty} \beta_i + \beta_{n+1}\,(n+1) \right| < \frac{\epsilon}{\sum_{i=0}^{\infty} \beta_{i+1}}. \quad (29)$$

Given any two quantities A and B,

$$|A| - |B| \leq |A+B| \leq |A| + |B|\,.$$

We apply this result to equation (29) with

$$\begin{aligned} A &= -a \sum_{i=0}^{n+1-N^*} \beta_i, \\ B &= -\sum_{i=0}^{n-N^*} \beta_{i+1} [Y_{n-i-1} - a] - \sum_{i=n-N^*+1}^{n} \beta_{i+1} Y_{n-i-1} + \beta_n\,(n+1)\,, \end{aligned}$$

so that $|Y_n - a| = |A + B|$. The triangle inequality implies

$$\begin{aligned} |B| &\leq \left| - \sum_{i=0}^{n-N^*} \beta_{i+1} (Y_{n-i-1} - a) \right| + \left| \sum_{i=n-N^*+1}^{n} \beta_{i+1} Y_{n-i-1} \right| + \beta_n\,(n+1) \leq \\ &\leq \epsilon + (N^*)^2 \sup_{n-N^* \leq i \leq n} \beta_i + \beta_n\,(n+1)\,, \end{aligned}$$

as Y_{n-i}, for $i \geq 0$ can have a jump of size at most equal to n. As $\lim_{n\to\infty} \beta_n n = 0$, by Lemma 3, $|B|$ is arbitrarily small, for large n, so that equation (29) might be satisfied only if $|A| = 0$ that is $a = 0$. This is a contradiction and proves (26).

We now prove (27). The assumption that P is irreducible and Lemma 3 imply that, $\forall i$, $\beta_i = \beta_i^0 = \prod_{k=0}^{i-1} p_k > 0$. Indicate with $R_{j-1}^{(n)}$ the generic element in place $j \geq 1$ of the vector $R^{(n)} = P^n O$. For each $i \geq 0$, $R_i^{(n)}$ follows the recursion

$$R_i^{(n)} = (1 - p_i) R_0^{(n-1)} + p_i \, R_{i+1}^{(n-1)}, \tag{30}$$

where $R^{(1)} = PO$. If we solve for $R_{i+1}^{(n-1)}$ in equation (30) and we substitute backwards for $R_i^{(n)}$, we obtain that, $\forall i > 0$,

$$R_i^{(n)} = \frac{\beta_0}{\beta_i} Y_{n+i-1} + \frac{\beta_1}{\beta_i} Y_{n+i-2} + \frac{\beta_2}{\beta_i} Y_{n+i-3} + \cdots + \frac{\beta_{i-1}}{\beta_i} Y_n + R_0^{(n)}, \ \ \forall i > 0, \tag{31}$$

where $Y_{n+i-1} = R_0^{(n+i)} - R_0^{(n+i-1)}$. (31) and (8) lead to (27). *Q.E.D.*

Proof of Proposition 1 (Increasing hazard functions) Assume first that Assumption 2' holds. If so, we proceed as follows. We show that Assumptions 1, 2' and 3 hold and we apply Lemma 5, so that $\delta_t' P^n O$ is equal to (21). 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 ρ^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 for this case.

If the probabilities p_i are decreasing in i, for $i \geq \square$, β_j^i is such that, for large j, $\forall i$,

$$\beta_j^i = \beta_\square^i \beta_{j-\square}^\square \ \square \ \beta_\square^i (p_\square)^{j-\square} \ \square \ (p_\square)^{j-\square} \tag{32}$$

so that Assumption 1 holds.

As assumptions 1, 2' and 3 hold, Lemma 5 guarantees that $\delta_t' P^n O$ is equal to (21). We know that A is always $O(\rho^n)$, $0 \ \square \ \rho < 1$. Assumption 3 together with (32) guarantee that also B, C and D are $O(\rho^n)$, $0 \ \square \ \rho < 1$.

Assume now that Assumption 2' does not hold and P is irreducible. We proceed as follows. We rely on Lemma 6 and we show that (25) implies that $Y_n = O(\rho^n)$, $0 \ \square \ \rho < 1$. We will then use Lemma 4 to show that (27) implies $\delta_t' P^n O = O(\rho^n)$, $0 \ \square \ \rho < 1$, concluding the proof.

Without loss of generality and to simplify notation we assume that $\square = 0$. If the probability p_i are decreasing in i, $\beta_i \ \square \ p_\square^i$, so that by Lemma 3 a steady state distribution does exist. Lemma 6 implies that $\lim_{n \to \infty} Y_n = 0$. We want to show that there does exist a number $\frac{1}{\rho} > 1$ such that $Y_n = O(\rho^n)$. We choose ρ such that $\frac{p_0}{\rho} < 1$ that is well defined as by assumption, $0 < p_0 < 1$. Given equation (25) it follows that

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

where $\tilde{Y}_n = \frac{Y_n}{\rho^n}$, $\tilde{\beta}_{i+1} = \frac{\beta_{i+1}}{\rho^{i+1}} \square \left(\frac{p_0}{\rho}\right)^{i+1}$. As the series $\sum_{i=0}^{\infty} \tilde{\beta}_{i+1}$ converges, we can show that $\lim_{n\to\infty} \tilde{Y}_n = 0$ by arguments similar to those used in the proof of Lemma 6. The fact that $Y_n = O(\rho^n)$ and irreducibility imply that $\forall \epsilon$, it does exist N^*, such that $\forall n > N^*$, $0 < \frac{Y_n}{\rho^n} < \epsilon$, which implies that $Y_{n+j} < \rho^{n-N^*} Y_{N^*+j}, \quad \forall j > 0$. This together with (27) imply that

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

where K is a positive bounded quantity by Lemma 4 and (9).$Q.E.D.$

Proof of Proposition 2 (Robust transmission mechanism) To prove that, as $n \uparrow \infty$, $\phi_n = \gamma \delta_t' P^n O \sim n^{1-h} = n^{d-1}$, $d = 2 - h$, we proceed as follows. Firstly, we show that Assumptions 1, 2' and 3 hold. We then apply Lemma 5, so that $\delta_t' P^n O$ is equal to (21). We know that $A = O(\rho^n)$, $0 \square \rho < 1$. We then show that $B \sim n^{1-h}$, $C = O(n^{1-h})$ and $D \sim n^{-h}$. 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, as $n \uparrow \infty$, $\beta_n^{s^*} \sim \beta_n^s \sim n^{-h}$, 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)

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha), \tag{33}$$

and $(A5)$ follow that

$$\beta_n^{s^*} = \frac{\Gamma(s^* + 1)}{\Gamma(s^* + n + 1)} \frac{\Gamma(s^* + n + 1 - h)}{\Gamma(s^* + 1 - h)} \sim \beta_n^s \sim n^{-h} \; as \; n \uparrow \infty, \tag{34}$$

where the last asymptotic equivalence used the fact that

$$\lim_{n\to\infty} n^{b-a} \frac{\Gamma(n + a)}{\Gamma(n + b)} = 1. \tag{35}$$

See Abramowitz and Stegun (1972), formula 6.1.46.

As Assumptions 1, 2' and 3 hold, by Lemma 5, $\delta_t' P^n O$ is equal to (21).

We now show that $B \sim n \beta_n^s \sim n^{1-h}$. 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 $(A4)$ is bounded for all n. β_i^{s+n} is a strictly positive quantity bounded above by one and below by zero such that, for $n \geq s^* - s$,

$$\beta_i^{s+n} = \frac{\Gamma(n+s+1)}{\Gamma(n+s+1-h)} \frac{\Gamma(n+s+i+1-h)}{\Gamma(n+s+1+i)}. \tag{36}$$

Condition $(A4)$, the fact that by (36) and (35) $\lim_{n\to\infty} \beta_i^{s+n} = 1, \forall i$, together with the Lebesgue dominated convergence theorem guarantees that

$$\lim_{n\to\infty} \sum_{i=0}^{\infty} \frac{\left|\delta_t^{s+i+1}\right|}{\beta_i^s} \beta_i^{s+n} = \sum_{i=0}^{\infty} \frac{\left|\delta_t^{s+i+1}\right|}{\beta_i^s} < \infty. \tag{37}$$

Given (9) and (34), it follows that $B \sim n\,\beta_n^s \sim n^{1-h}$.

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

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

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

$$\sum_{i=1}^{\infty} \frac{\Gamma(b-1-a+i)}{\Gamma(b+i)} = \frac{\Gamma(b-a)}{a\Gamma(b)}, \ \forall b > a > 0, \tag{38}$$

for any integer b (see Gradshteyn and Ryzhik , 1997, formula 0.247 and Granger and Joyeux, 1980, p. 18).

Consider first the case in which $h > 1$. From assumption $(A4)$ together with $(C$'$)$, (36), (38) and (35) it follows that

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

Consider now the case h □ 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 is bounded for all n. We now show that if h □ 1, the quantity $\frac{\beta_i^{s+n}}{(s+n)}$ is always decreasing in n. (9), 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 □ 1, $C = o\,(n\,\beta_n^s)$ thus concluding that, for any general h, $C = O(n\,\beta_n^s) = O(n^{1-h})$ by (34).

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), we obtain that

$$\frac{d\left(\ln \frac{\beta_i^{s+n}}{s+n}\right)}{dn} = [\psi(n+s) - \psi(n+s+1-h) + \psi(n+s+i+1-h) - \psi(n+s+i+1)]. \tag{39}$$

h □ 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 (39) is negative.

Finally, we show that $D \sim \beta_n^s$ so that $D \sim n^{-h}$ by (34). 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 is bounded for all n. β_i^{s+n-1} is a strictly positive quantity bounded above by one and below from zero. Condition $(A4)$, (9), the fact that by (36) and (35), $\lim_{n\to\infty} \beta_i^{s+n} = 1$, $\forall i$, together with the Lebesgue dominated convergence theorem imply that $D \sim \beta_n^s$. *Q.E.D.*

8.3 Proofs of results in section 5

Proof of Proposition 3 (Persistence in the irreducible case) To prove the assertion we proceed as follows. Firstly, we use (25) to show that, as $n \uparrow \infty$, $Y_n \sim n^{1-h}$. Secondly, we use this result and (27) to show that $\delta_t' P^n O \sim n^{1-h}$, concluding the proof.

We now show that $Y_n \sim n^{1-h}$. Let us argue by contradiction and suppose that $\lim_{n\to\infty}(n+b)^{h-1}Y_n = 0$ where b is an arbitrary positive quantity (similar reasoning would apply to the contradiction $\lim_{n\to\infty}(n+b)^{h-1}Y_n = \infty$). It follows from equation (25) that

$$(n+b)^{h-1}Y_n = \tilde{Y}_n = -\sum_{i=0}^{n-1} \beta_{i+1}\, h(n,i)\, \tilde{Y}_{n-i-1} + (n+b)^{h-1}\beta_{n+1}\,(n+1)\,,\ \forall n > 0,$$

where $h(n,i) = \frac{(n+b)^{h-1}}{(n-i-1)^{h-1}}$. It implies that

$$(n+b)^{h-1}\beta_{n+1}\,(n+1) - \left| -\sum_{i=0}^{\infty} \beta_{i+1}\, h(n,i)\, \tilde{Y}_{n-i-1} \right| \ □ \ \left|\tilde{Y}_n\right| \ □$$

$$\Box\ (n+b)^{h-1}\beta_{n+1}\ (n+1) + \left| -\sum_{i=0}^{\infty} \beta_{i+1}\, h(n,i)\ \tilde{Y}_{n-i-1} \right| . \tag{40}$$

We now show that the positive quantity

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} \beta_{i+1} h(n,i) \tag{41}$$

is bounded. Infact, the same reasoning that led to (34) together with $(A5)$ imply that, for $i \uparrow \infty$, $\beta_i \sim i^{-h}$. It follows that

$$\sum_{i=0}^{n-1} \beta_{i+1} h(n,i) \sim \sum_{i=0}^{\frac{n}{2}} \frac{(n+b)^{h-1}}{(n-i-1+b)^{h-1}(i+1)^h} + \sum_{i=\frac{n}{2}+1}^{n-1} \frac{(n+b)^{h-1}}{(n-i-1+b)^{h-1}(i+1)^h}$$

$$\Box\ (\frac{1}{2}+\frac{b}{n})^{1-h} \sum_{i=0}^{\frac{n}{2}} \frac{1}{(i+1)^h} + (\frac{n}{2})^{-h} \sum_{i=\frac{n}{2}+1}^{n-1} \left(1-\frac{i}{n}+\frac{b}{n}\right)^{1-h},$$

where the first term is bounded as $h > 1$, while the second is equal to

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

and it goes to zero when $n \uparrow \infty$. We can then follow the same procedure as in the proof of Lemma 6 to show that under the assumption that $\lim_{n\to\infty} \tilde{Y}_n = 0$, $\forall \epsilon$, it does exist N^* such that $\forall n > N^*$

$$(n+b)^{h-1}\beta_n\, n - \epsilon\ \Box\ \left|\tilde{Y}_n\right|\ \Box\ (n+b)^{h-1}\beta_n\, n + \epsilon,$$

that is a contradiction since, as $n \uparrow \infty$, $\beta_n\, n \sim n^{1-h}$.

We now show that $\delta_t' P^n O \sim n^{1-h}$. In fact, $Y_n \sim n^{1-h}$ implies that $\lim_{n\to\infty} \frac{Y_{n+i}}{Y_n} = 0$, $\forall i$. As a result it does exist N^* such that $\forall n > N^*$, $\forall j$

$$0 < Y_{n+i} < A Y_n \tag{42}$$

where A is a positive bounded quantity. Moreover, $\beta_n \sim n^{-h}$, $h > 1$ together with $(A4)$ implies that

$$\sum_{j=0}^{\infty} \sum_{i=j}^{\infty} \delta_t^{i+1} \frac{\beta_{i-j}}{\beta_{i+1}} = \sum_{i=1}^{\infty} \frac{\delta_t^i}{\beta_i} \sum_{j=0}^{i-1} \beta_j < \infty.$$

From this together with (42),(27) and the Lebesgue dominated convergence theorem it follows that $\delta_t' P^n O \sim Y_n \sim n^{1-h}$. *Q.E.D.*

Proof of Proposition 4 (Small versus large shocks) We wish to show that, as $n \uparrow \infty$, $\phi_n = \gamma \delta_t' P^n O \sim n^{d-1}$, where $d = 3 - h$. We assume first that Assumption 2' holds and we proceed as follows. We show that Assumptions 1, 2' and 3 hold. We then apply Lemma 5, so that $\delta_t' P^n O$ is equal to (21). We know that $A = O(\rho^n)$ $0 \square \rho < 1$. We then show that $B \sim n^{2-h}$, $C \sim n^{2-h}$ and $D \sim n^{1-h}$. Consequently $\phi_n = \gamma \delta_t' P^n O \sim n^{2-h}$, and completing the proof for this case.

We now show that Assumptions 1 and 3 hold (Assumption 2' holds by hypothesis). ($A5$) implies that, $\forall j, i > 0$,

$$\beta_j^{s+i} = \frac{\Gamma(i+s+1)}{\Gamma(i+s+1-h)} \frac{\Gamma(i+s+j+1-h)}{\Gamma(i+s+1+j)}. \tag{43}$$

As $h > 2$, (35) together with (43) implies that Assumptions 1 and 3 are satisfied. Hence, by Lemma 5, $\delta_t' P^n O$ is equal to (21).

To show that $B \sim n^{2-h}$, we use the fact that (38), (35) 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 $C \sim n^{2-h}$. (38) implies that, for any integer b,

$$\sum_{i=1}^{\infty} i \frac{\Gamma(b-2-a+i)}{\Gamma(b+i)} = \frac{\Gamma(b-1-a)}{a(1+a)\Gamma(b-1)}, \ \forall a > 0. \tag{44}$$

(35), (38) and (C) together with (44) imply that

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

To show that $D \sim n^{1-h}$, we use the fact that (38), (35) and (D) imply that

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

We now assume that Assumption 2' fails and P is irreducible. To simplify notation and without loss of generality we assume that $\square = 0$. From similar arguments as those used in the proof Proposition 3 it follows that $Y_n \sim n^{1-h}$. This together with (27), (43) and ($A4$') imply that

$$\delta_t' P^n O = -a \sum_{j=0}^{\infty} Y_{n+j} \sum_{i=0}^{\infty} \beta_i \sim n^{2-h}.$$

Q.E.D.

Proof of Proposition 5 (The role of cross-sectional heterogeneity) Since Assumptions 1, 2' and 3 hold, Lemma 5 applies. Since $\delta_t^i = 0, \forall i > s$,by assumption, (21) produces $\delta_t' P^n O = A = O(\rho^n)$, $0 \square \rho < 1$. *Q.E.D.*

Proof of Proposition 6 (Adjustment costs and the growth-size relation) As $F \sim U[0,1]$, we have

$$\begin{aligned} 1 - p_i &= \lambda_R^i, \\ \int_{-\infty}^{\lambda_R^{i+1}} s\, dF(s) &= \frac{1}{2}(1 - p_i)^2 . \end{aligned}$$

As a result, (19) is equal to

$$c_{i+1} = A + \beta^{-1} c_i + \beta^{-1}(1 - p_i) + \frac{1}{2} p_{i+1}^2 - \gamma(i+1), \quad i \geq 0$$

where $A = c_0 - \frac{1}{2} p_0^2$. Substituting backwards we obtain that, $\forall i \geq 0$,

$$\begin{aligned} c_{i+1} &= A \sum_{j=0}^{i} \beta^{-j} + \beta^{-(i+1)} c_0 + \beta^{-1} \sum_{j=0}^{i} \beta^{-j} (1 - p_{i-j}) + \\ & + \frac{1}{2} \sum_{j=0}^{i} \beta^{-j} (p_{i+1-j})^2 - \gamma \sum_{j=0}^{i} \beta^{-j} (i + 1 - j) . \end{aligned} \tag{45}$$

To prevent the adjusting costs from exploding exponentially at rate β^{-1},and given Lemma 1, we set c_0 as equal to

$$c_0 = \frac{1}{2} p_0^2 - 1 + \frac{\gamma}{\beta^{-1} - 1} + (1 - \beta) \sum_{j=0}^{\infty} \beta^j \left(p_j - \frac{1}{2} p_j^2 \right) .$$

It follows that (45) is equal to

$$c_{i+1} = K + \frac{\gamma(i+1)}{\beta^{-1} - 1} + \sum_{j=0}^{\infty} \beta^j p_{i+j+1} - \frac{1}{2} \sum_{j=1}^{\infty} \beta^j (p_{i+j+1})^2, \quad i \geq 0, \tag{46}$$

where $K = \frac{c_0 - \frac{1}{2} p_0^2}{1 - \beta^{-1}} + \frac{\gamma \beta^{-1}}{(1 - \beta^{-1})^2}$. Given ($A5$), $\forall i \geq s^*$, (46) produces

$$c_{i+1} = K + \frac{1}{1 - \beta} - \frac{1}{2} \frac{\beta}{1 - \beta} + \frac{\gamma(i+1)}{\beta^{-1} - 1} - \frac{h\gamma}{2 + i} - \frac{1}{2} \sum_{j=1}^{\infty} \beta^j \frac{(h\gamma)^2}{(2 + i + j)^2}$$

which, together with the Lebesgue dominated convergence theorem, yields (13). *Q.E.D.*