UNIVERSITY OF CAMBRIDGE

Faculty of Economics

Institute for New Economic Thinking

Cambridge-INET Institute

Cambridge-INET Working Paper Series No: 2015/13

Cambridge Working Paper in Economics: 1556

LARGE FIRM DYNAMICS AND THE BUSINESS CYCLE

Vasco M. Carvalho (University of Cambridge)

Basile Grassi (University of Oxford)

Do large firm dynamics drive the business cycle? We answer this question by developing a quantitative theory of aggregate fluctuations caused by firm-level disturbances alone. We show that a standard heterogeneous firm dynamics setup already contains in it a theory of the business cycle, without appealing to aggregate shocks. We offer a complete analytical characterization of the law of motion of the aggregate state in this class of models – the firm size distribution – and show that the resulting closed form solutions for aggregate output and productivity dynamics display: (i) persistence, (ii) volatility and (iii) time-varying second moments. We explore the key role of moments of the firm size distribution – and, in particular, the role of large firm dynamics – in shaping aggregate fluctuations, theoretically, quantitatively and in the data.

Large Firm Dynamics and the Business Cycle*

Vasco M. Carvalho† Basile Grassi‡

April, 7, 2015

[Link to the latest version]

Abstract

Do large firm dynamics drive the business cycle? We answer this question by developing a quantitative theory of aggregate fluctuations caused by firm-level disturbances alone. We show that a standard heterogeneous firm dynamics setup already contains in it a theory of the business cycle, without appealing to aggregate shocks. We offer a complete analytical characterization of the law of motion of the aggregate state in this class of models – the firm size distribution – and show that the resulting closed form solutions for aggregate output and productivity dynamics display: (i) persistence, (ii) volatility and (iii) time-varying second moments. We explore the key role of moments of the firm size distribution – and, in particular, the role of large firm dynamics – in shaping aggregate fluctuations, theoretically, quantitatively and in the data.

Keywords: Large Firm Dynamics; Firm Size Distribution; Random Growth; Aggregate Fluctuations

*The authors would like to acknowledge helpful comments from Árpád Ábrahám, Axelle Arquié, Florin Bilbiie, David Backus, Paul Beaudry, Ariel Burstein, Jeff Campbell, Francesco Caselli, Gian Luca Clementi, Julian di Giovanni, Jan Eeckhout, Xavier Gabaix, Wouter den Haan, Hugo Hopenhayn, Jean Imbs, Oleg Itskhoki, Boyan Jovanovic, Andrei Levchenko, Isabelle Mejean, Ezra Oberfield, Michael Peters, Morten Ravn, Esteban Rossi-Hansberg, Edouard Schaal, Rob Shimer and Gianluca Violante. The authors thank seminar participants at Banque de France, CREi, CREST-INSEE, EUI, Essex, LSE, NYU, Princeton, NY Fed, Oxford, PSE, UCL and the Conference in Honor of Lars Hansen at the University of Chicago. We thank Margit Reischer for excellent research assistance. Part of this work was carried out during Grassi's stay at the NYU Economics department and he is grateful for their hospitality. Carvalho acknowledges the financial support of the European Research Council grant #337054, the Cambridge-INET Institute and the Keynes Fellowship.

†University of Cambridge and CREi, Universitat Pompeu Fabra, Barcelona GSE and CEPR. vcarvalho@crei.cat

‡Nuffield College, Department of Economics, University of Oxford. basile.grassi@economics.ox.ac.uk

1 Introduction

Aggregate prices and quantities exhibit persistent dynamics and time-varying volatility. Business cycle theories have typically resorted to exogenous aggregate shocks in order to generate such features of aggregate fluctuations. A recent literature has instead proposed that the origins of business cycles may be traced back to micro-level disturbances.[1] Intuitively, the prominence of a small number of firms leaves open the possibility that aggregate outcomes may be affected by the dynamics of large firms.[2] And yet, we lack a framework that enables a systematic evaluation of the link between the micro-level decisions driving firm growth, decline and churning and the persistence and volatility of macro-level outcomes.

This paper seeks to evaluate the impact of large firm dynamics on aggregate fluctuations. Building on a standard firm dynamics setup, we develop a quantitative theory of aggregate fluctuations arising from firm-level shocks alone. We derive a complete analytical characterization of the law of motion of the firm size distribution – the aggregate state variable in this class of models – and show that the resulting aggregate output and productivity dynamics are *endogenously* (*i*) persistent, (*ii*) volatile and (*iii*) exhibit time-varying second moments. We explore the key role of moments of the firm size distribution – and, in particular, the role of large firm dynamics – in shaping aggregate fluctuations, theoretically, quantitatively and in the data. Our results imply that large firm dynamics induce sizeable movements in aggregates and account for one quarter of aggregate fluctuations.

Our setup follows Hopenhayn's (1992) industry dynamics framework closely. Firms differ in their idiosyncratic productivity level, which is assumed to follow a discrete Markovian process. Incumbents have access to a decreasing returns to scale technology using labor as the only input. They produce a unique good in a perfectly competitive market. They face an operating fixed cost in each period which, in turn, generates

[1]See Gabaix (2011), Acemoglu et *al.* (2012), di Giovanni and Levchenko (2012), Carvalho and Gabaix (2013) and di Giovanni, Levchenko and Mejean (2014)

[2]For example, in the fall of 2012, JP Morgan predicted that the upcoming "*release of the iPhone 5 could potentially add between 1/4 to 1/2%-point to fourth quarter annualized GDP growth*" (JP Morgan, 2012). Apple's prominence in the US economy is comparable to that of a small number of very large firms. For example, Walmart's 2014 US sales amounted to 1.9% of US GDP. Taken together, according to Business Dynamics Statistics (BDS) data, the largest 0.02% of US firms account for about 20% of all employment.

endogenous exit. As previous incumbents exit the market, they are replaced by new entrants.

The crucial difference relative to Hopenhayn (1992) – and much of the large literature that follows from it – is that we do not rely on the traditional "continuum of firms" assumption in order to characterize the law of motion for the firm size distribution. Instead, we characterize the law of motion for any finite number of firms. Our first theoretical result shows that, generically, the firm size distribution is time-varying in a stochastic fashion. As is well known, this distribution is the aggregate state variable in this class of models. An immediate implication of our findings is therefore that aggregate productivity, aggregate output and factor prices are themselves stochastic. In a nutshell, we show that the standard workhorse model in the firm dynamics literature – once the assumption regarding a continuum of firms is dropped – already features aggregate fluctuations.

We then specialize our model to the case of random growth dynamics at the firm level. Given our focus on large firm dynamics, the evidence put forth by Hall (1987) in favor of Gibrat's law for large firms makes this a natural baseline to consider.[3] With this assumption in place, our second main theoretical contribution is to solve analytically for the law of motion of the aggregate state in our model. This closed form characterization is key to our analysis and enables us to provide a sharp characterization of the equilibrium firm size distribution and the dynamics of aggregates.

Our third theoretical result is to show that the steady state firm size distribution is Pareto distributed. We discuss the role of random growth, entry and exit and decreasing returns to scale in generating this result. The upshot of this is that our model can endogenously deliver a first-order distributional feature of the data: the co-existence of a large number of small firms and a small, but non-negligible, number of very large firms, orders of magnitude larger than the average firm in the economy.

Our final set of theoretical results sheds light on the micro origins of aggregate persistence, volatility and time-varying uncertainty. Leveraging on our characterization of the law of motion of the aggregate state we are able to show, *analytically*, that: (i) persistence in aggregate output is increasing with firm level productivity persistence and with the share of economic activity accounted by large firms; (ii) aggregate volatility decays only slowly with the number of firms in the economy, and that this rate of decay

[3] See also Evans (1987) and the discussion in Luttmer (2010).

is generically a function of the size distributions of incumbents and entrants, as well as the degree of decreasing returns to scale and (*iii*) aggregate volatility dynamics are endogenously driven by the evolution of the cross-sectional dispersion of firm sizes.

We then explore the quantitative implications of our setup. Due to our characterization of aggregate state dynamics, our numerical strategy is substantially less computational intensive than that traditionally used when solving for heterogeneous agents' models. This allows us to solve the model featuring a very large number of firms and thus match the firm size distribution accurately.

Our first set of quantitative results shows that the standard model of firm dynamics with no aggregate shocks is able to generate sizeable fluctuations in aggregates: aggregate output (aggregate productivity) fluctuations amount to 26% (17%, respectively) of that observed in the data. These fluctuations have their origins in large firm dynamics. In particular, we show how fluctuations at the upper end of the firm size distribution – induced by shocks to very large firms – lead to movements in aggregates. We supplement this analysis by showing that the same correlation holds true empirically: aggregate output and productivity fluctuations in the data coincide with movements in the tail of the firm size distribution.

We then focus on the origins of time-varying aggregate volatility. Consistently with our analytical characterization, our quantitative results show that the evolution of aggregate volatility is determined by the evolution of the cross-sectional dispersion in the firm size distribution. Unlike the extant literature, the latter is the endogenous outcome of firm-level idiosyncratic shocks and not the result of exogenous aggregate second moment shocks. Again, we compare these results against the data and find consistent patterns: aggregate volatility is high whenever cross-sectional dispersion high.

The paper relates to two distinct literatures: an emerging literature on the micro-origins of aggregate fluctuations and the more established firm dynamics literature. Gabaix's (2011) seminal work introduces the "Granular Hypothesis": whenever the firm size distribution is fat tailed, idiosyncratic shocks average out at a slow enough rate that it is possible for these to translate into aggregate fluctuations.[4] Relative to Gabaix (2011), our main contribution is to ground the granular hypothesis in a well specified firm dynamics setup: in our setting, firms' entry, exit and size decisions reflect optimal

[4] Other contributions in this literature include Acemoglu et *al.* (2012), di Giovanni and Levchenko (2012), Carvalho and Gabaix (2013) and di Giovanni, Levchenko and Mejean (2014).

forward-looking choices, given firm-specific productivity processes and (aggregate) factor prices. Further, the firm size distribution is an equilibrium object of our model.[5] This allows us to both generalize the existent theoretical results and to quantify their importance. The recent contribution of di Giovanni, Levchenko and Mejean (2014) provides a valuable empirical benchmark to this literature and, in particular, to our quantification exercise discussed above. Working with census data for France, their variance decomposition exercise finds that large firm dynamics account for just under 20% of aggregate volatility. Our quantitative results show that the magnitude of aggregate fluctuations implied by our firm dynamics environment is of the same order of magnitude.

This paper is also related to the firm dynamics literature that follows from the seminal contribution of Hopenhayn (1992).[6] Some papers in this literature have explicitly studied aggregate fluctuations in a firm dynamics framework (Campbell and Fisher (2004), Lee and Mukoyama (2008), Clementi and Palazzo (2015) and Bilbiie et *al.* (2012)). A more recent strand of this literature has focused on the time-varying nature of aggregate volatility and its link with the cross-sectional distribution of firms (e.g. Bloom et al, 2014). Invariably, in this literature, business cycle analysis is restricted to the case of common, aggregate shocks which are superimposed on firm-level disturbances. Relative to this literature, we show that its standard workhorse model – once the assumption regarding a continuum of firms is dropped and the firm size distribution is fat tailed – already contains in it a theory of aggregate fluctuations and time-varying aggregate volatility. We show this both theoretically and quantitatively in an otherwise transparent and well understood setup. We eschew the myriad of frictions - capital adjustment costs, labor market frictions, credit constraints or limited substitution possibilities across goods - that Hopenhayn's (1992) framework has been able to support. We do this because our focus is on large firm dynamics which are arguably less encumbered by such frictions.

The paper is organized as follows. Section 2 presents the basic model setup. Sections 3 and 4 develop our theoretical results. Section 5 describes the calibration of the model,

[5]The interplay between the micro-level decisions of firms and the equilibrium size distribution is also the object of analysis in Luttmer (2007, 2010 and 2012). Relative to this body of work, our contribution is to focus on the implications of firm dynamics on aggregate fluctuations rather than long-run growth paths.

[6]See, for example, Campbell (1998), Veracierto (2002), Clementi and Hopenhayn (2006), Rossi-Hansberg and Wright (2007), Khan and Thomas (2008) and Acemoglu and Jensen (2015).

our quantitative results and our empirical exercises. Finally, Section 6 concludes.

2 Model

We analyze a standard firm dynamics setup (Hopenhayn, 1992) with a finite but possibly large number of firms. We show how to solve for and characterize the evolution of the firm size distribution without relying on the usual law of large numbers assumption. We prove that, in this setting, the firm size distribution does not converge to a stationary distribution, but instead fluctuates stochastically around it. As a result, we show that aggregate prices and quantities are not constant over time as the continuum assumption in Hopenhayn (1992) - repeatedly invoked by the subsequent literature - does not apply. To do this, we start by describing the economic environment. As is standard in this class of models, this involves specifying a firm-level productivity process, the incumbents' problem and the entrants' problem.

2.1 Model Setup

The setup follows Hopenhayn (1992) closely. Firms differ in their productivity level, which is assumed to follow a discrete Markovian process. Incumbents have access to a decreasing returns to scale technology using labor as the only input. They produce a unique good in a perfectly competitive market. They face an operating cost at each period, which in turn generates endogenous exit. There is also a large (but finite) number of potential entrants that differ in their productivity. To operate next period, potential entrants have to pay an entry cost. The economy is closed in a partial equilibrium fashion by specifying a labor supply function that increases with the wage.

Productivity Process

We assume a finite but potentially large number of idiosyncratic productivity levels. The productivity space is thus described by a S-tuple $\mathit{\Phi} := \{\varphi^1, \ldots, \varphi^S\}$ with $\varphi > 1$ such that $\varphi^1 < \ldots < \varphi^S$. The idiosyncratic state-space is evenly distributed in logs, where φ is the log step between two productivity levels: $\frac{\varphi^{s+1}}{\varphi^s} = \varphi$. A firm is in state (or productivity state) s when its idiosyncratic productivity is equal to φ^s. Each firm's productivity level is assumed to follow a Markov chain with a transition matrix P.

We denote $F(.|\varphi^s)$ as the conditional distribution of the next period's idiosyncratic productivity $\varphi^{s'}$ given the current period's idiosyncratic productivity φ^s.[7]

Incumbents' Problem

The only aggregate state variable of this model is the distribution of firms on the set Φ. We denote this distribution by a $(S \times 1)$ vector μ_t giving the number of firms at each productivity level s at time t. For the current setup description, we abstract from explicit time t notation, but will return to it when we characterize the law of motion of the aggregate state. Given an aggregate state μ, and an idiosyncratic productivity level φ^s, the incumbent solves the following static profit maximization problem:

$$\pi^*(\mu, \varphi^s) = \underset{n}{\text{Max}} \{\varphi^s n^\alpha - w(\mu)n - c_f\}$$

where n is the labor input, $w(\mu)$ is the wage for a given aggregate state μ, and c_f is the operating cost to be paid every period. It is easy to show that π^* is increasing in φ^s and decreasing in w for a given aggregate state μ. The output of a firm is then $y(\mu, \varphi^s) = (\varphi^s)^{\frac{1}{1-\alpha}} \left(\frac{\alpha}{w(\mu)}\right)^{\frac{\alpha}{1-\alpha}}$. In what follows, the size of a firm will refer to its output level if not otherwise specified.

The timing of decisions for incumbents is standard and described as follows. The incumbent first draws its idiosyncratic productivity φ^s at the beginning of the period, pays the operating cost c_f and then hires labor to produce. It then decides whether to exit at the end of the period or to continue as an incumbent the next period. We denote the present discounted value of being an incumbent for a given aggregate state μ and idiosyncratic productivity level φ^s by $V(\mu, \varphi^s)$, defined by the following Bellman Equation:

$$V(\mu, \varphi^s) = \pi^*(\mu, \varphi^s) + \max \left\{ 0, \beta \int_{\mu' \in \Lambda} \sum_{\varphi^{s'} \in \Phi} V(\mu', \varphi^{s'}) F(\varphi^{s'}|\varphi^s) \Gamma(d\mu'|\mu) \right\}$$

where β is the discount factor, $\Gamma(.|\mu)$ is the conditional distribution of μ', tomorrow's aggregate state and $F(.|\varphi^s)$ is the conditional distribution of tomorrow's idiosyncratic productivity for a given today's idiosyncratic productivity of the incumbent.

[7] That is, for a given productivity level φ^s, the distribution $F(.|\varphi^s)$ is given by the s^{th}-row vector of the matrix P.

The second term on the right hand side of the value function above encodes an endogenous exit decision. As is standard in this framework, this decision is defined by a threshold level of idiosyncratic productivity given an aggregate state. Formally, since the instantaneous profit is increasing in the idiosyncratic productivity level, there is a unique index $s^*(\mu)$ for each aggregate state μ, such that: (i) for $\varphi^s \geq \varphi^{s^*(\mu)}$ the incumbent firm continues to operate next period and, conversely (ii) for $\varphi^s \leq \varphi^{s^*(\mu)-1}$ firm decides to exit next period.

After studying the incumbents' problem, we now turn to the problem of potential entrants.

Entrants' Problem

There is an exogenously given, constant and finite number of prospective entrants M. Each potential entrant has access to a signal about their potential productivity next period, should they decide to enter today. To do so, they have to pay a sunk entry cost which, in turn, leads to an endogenous entry decision which is again characterized by a threshold level of initial signals.

Formally, the entrants' signals are distributed according $G = (G_q)_{q\in[1\ldots S]}$, a discrete distribution over $\varPhi$. There is a total of M potential entrants every period, so that the MG_q gives the number of potential entrants for each signal level φ^q. If a potential entrant decides to pay the entry cost c_e, then she will produce next period with a productivity level drawn from $F(.|\varphi^q)$. Given this, we can define the value of a potential entrant with signal φ^q for a given the aggregate state μ as $V^e(\mu, \varphi^q)$:

$$V^e(\mu, \varphi^q) = \beta \int_{\mu'\in\Lambda} \sum_{\varphi^{q'}\in\varPhi} V(\mu', \varphi^{q'}) F(\varphi^{q'}|\varphi^q)\Gamma(d\mu'|\mu)$$

Prospective entrants pay the entry cost and produce next period if the above value is greater or equal to the entry cost c_e. As in the incumbent's exit decision, this now induces a threshold level of signal, $e^*(\mu)$, for a given aggregate state μ such that (i) for $\varphi^q \geq \varphi^{e^*(\mu)}$ the potential entrant starts operating next period and, conversely (ii) for $\varphi^q \leq \varphi^{e^*(\mu)-1}$ the potential entrant decides not to do so.

For simplicity, henceforth we assume that the entry cost is normalized to zero: $c_e = 0$ which in turn implies that $\varphi^{e^*(\mu)} = \varphi^{s^*(\mu)}$.

Labor Market and Aggregation

We assume that the supply of labor at a given wage w is given by $L^s(w) = Mw^\gamma$ with $\gamma > 0$. We assume that, for a given wage level, the labor supply function is a linear function of M, the number of potential entrants. This assumption is necessary because in what follows we will be interested in characterizing the behavior of aggregate quantities and prices as we let M increase. Note that if total labor supply were to be kept fixed, increasing M would lead to an increase in aggregate demand for labor. Therefore, the wage would increase mechanically. We therefore make this assumption to abstract from this mechanical effect of increasing M on the equilibrium wage.[8]

To find equilibrium wages, we derive aggregate labor demand in this economy. To do this, note that if Y_t is aggregate output, i.e. the sum of all individual incumbents' output, then $Y_t = A_t(L_t^d)^\alpha$ where L_t^d is the aggregate labor demand, the sum of all incumbents' labor demand in period t. A_t, aggregate total factor productivity, is then given by:

$$A_t = \left(\sum_{i=1}^{N_t} \left(\varphi^{s_{i,t}}\right)^{\frac{1}{1-\alpha}} \right)^{1-\alpha}$$

where $\varphi^{s_{i,t}}$ is the productivity level at date t of the i^{th} firm among the N_t operating firms at date t. This can be rewritten by aggregating over all firms that have the same productivity level:

$$A_t = \left(\sum_{s=1}^{S} \mu_{s,t} \left(\varphi^{s}\right)^{\frac{1}{1-\alpha}} \right)^{1-\alpha} = \left(B'\mu_t\right)^{1-\alpha} \tag{1}$$

where B is the $(S \times 1)$ vector of parameters $\left((\varphi^1)^{\frac{1}{1-\alpha}}, \ldots, (\varphi^S)^{\frac{1}{1-\alpha}}\right)$. As discussed above, the distribution of firms μ_t across the discrete state space $\Phi = \{\varphi^1, \ldots, \varphi^S\}$ is a $(S \times 1)$ vector equal to $(\mu_{1,t}, \ldots, \mu_{S,t})$ such that $\mu_{s,t}$ is equal to the number of operating firms in state s at date t. By the same argument, it is easy to show that aggregate labor demand is given by $L^d(w_t) = \left(\frac{\alpha A_t}{w_t}\right)^{\frac{1}{1-\alpha}}$. Note that the model behaves as a one factor model with aggregate TFP A_t.

[8] More generally, any increasing function of M will be possible. For simplicity, we assume a linear function. This assumption ensures that the equilibrium wage is independent of the number of potential entrants. To see this, note that given the labor market equilibrium condition (Equation 2), and under this assumption, the equilibrium wage is now a function of $\hat{\mu}_t := \frac{\mu_t}{M}$, the normalized productivity distribution across productivity levels. Given that M is a parameter of the model, we can therefore use μ_t or $\hat{\mu}_t$ interchangeably as the aggregate state variable.

The market clearing condition then equates labor supply and labor demand, i.e. $L^s(w_t) = L^d(w_t)$. Given date t productivity distribution μ_t, we can then solve for the equilibrium wage to get:

$$w_t = \left(\alpha^{\frac{1}{1-\alpha}} \frac{B'\mu_t}{M} \right)^{\frac{1-\alpha}{\gamma(1-\alpha)+1}} \tag{2}$$

This last Equation leads to the following expression for aggregate output:

$$Y_t = A_t L_t^{\alpha} \tag{3}$$

From these expressions, note that the wage and aggregate output is fully pinned down by the distribution μ_t. Given a current-period distribution of firms across productivity levels we can solve for all equilibrium quantities and prices. We are left to understand how this distribution evolves over time, i.e. how to solve for for $\Gamma(.|\mu_t)$, which is addressed in the following section.

3 Aggregate State Dynamics and Uncertainty: General Results

In this section, we first show how to characterize the law of motion for the productivity distribution, the aggregate state in this economy. We will prove that, generically, the distribution of firms across productivity levels is time-varying in a stochastic fashion. An immediate implication of this result is that aggregate productivity A_t and aggregate prices are themselves stochastic as they are simply a function of this distribution. Additionally, we then show that the characterization of the stationary firm productivity distribution offered in Hopenhayn (1992) is nested in our model when we take uncertainty to zero.

Law of Motion of the Productivity Distribution

In a setting with a continuum of firms, Hopenhayn (1992) shows that by appealing to a law of large numbers, the law of motion for the productivity distribution is in fact deterministic. In the current setting, with a finite number of incumbents, a similar

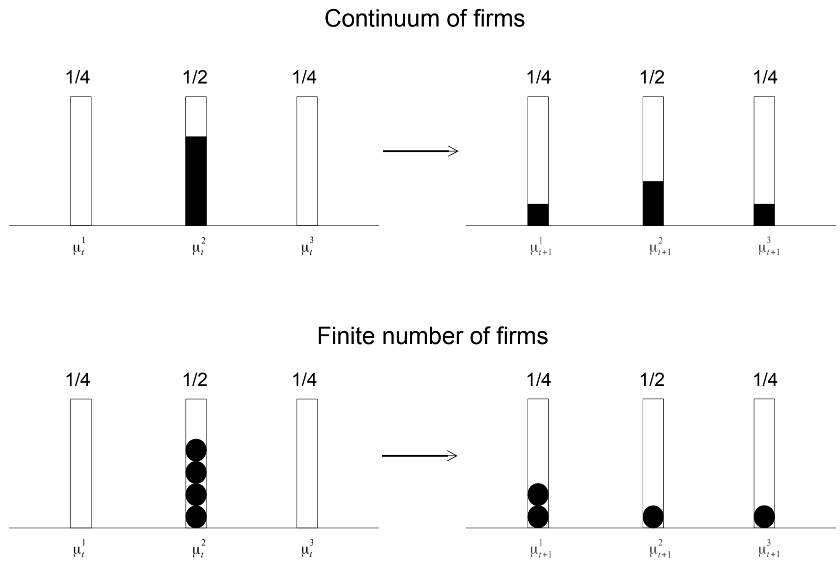

Figure 1: Why the vector μ_{t+1} follows a multinomial distribution.

NOTE: Top panel: continuum of firms case, unique possible outcome. Bottom panel: finite number of firms, one (of many) possible outcome.

argument cannot be made. We now show how to characterize the law of motion for the productivity distribution when we move away from the continuum case.

In order to build intuition for our general result below, we start by exploring a simple example where, for simplicity, we ignore entry and exit of firms. Assume there are only three levels of productivity ($S = 3$) and four firms. At time period t these firms are distributed according to the bottom-left panel of figure 1, i.e. all four firms produce with the intermediate level of productivity. Further assume that these firms have an equal probability of 1/4 of going up or down in the productivity ladder and that the probability of staying at the same intermediate level is 1/2. That is, the second row of the transition matrix P in this simple example is given by $(1/4, 1/2, 1/4)$. First note that, if instead of four firms we had assumed a continuum of firms, the law of large numbers would hold such that at $t + 1$ there would be exactly 1/4 of the (mass of) firms at the highest level of productivity, 1/2 would remain at the intermediate level and 1/4 would transit to the lowest level of productivity (top panel of figure 1). This is not the case here, since the number of firms is finite. For instance, a distribution of firms such as the one presented in the bottom-right panel of figure 1 is possible with a positive probability. Of course, many other arrangements would also be possible outcomes. Thus, in this example, the number of firms in each productivity bin at $t + 1$

follows a multinomial distribution with a number of trials of 4 and an event probability vector $(1/4, 1/2, 1/4)'$.

In this simple example, all firms are assumed to have the same productivity level at time t. It is easy however to extend this example to any initial arrangement of firms over productivity bins. This is because, for any initial number of firms *at a given* productivity level, the distribution of *these* firms across productivity levels next period follows a multinomial. Therefore, the *total* number of firms in each productivity level next period, is simply a sum of multinomials, i.e. the result of transitions from *all* initial productivity bins.

More generally, for S productivity levels, and an (endogenous) finite number of incumbents, N_t, making optimal employment and production decisions and accounting for entry and exit decisions, the following Theorem holds.

Theorem 1 *The number of firms at each productivity level at $t+1$, given by the $(S \times 1)$ vector μ_{t+1}, conditional on the current vector μ_t, follows a sum of multinomial distributions and can be expressed as:*

$$\mu_{t+1} = m(\mu_t) + \epsilon_{t+1} \tag{4}$$

where ϵ_{t+1} is a random vector with mean zero and a variance-covariance matrix $\Sigma(\mu_t)$ and

$$m(\mu_t) = (P_t^*)'(\mu_t + MG)$$

$$\Sigma(\mu_t) = \sum_{s=s^*(\mu_t)}^{S} (MG_s + \mu_{s,t})W_s$$

where P_t^* is the transition matrix P with the first $(s^*(\mu_t) - 1)$ rows replaced by zeros. $W_s = diag(P_{s,.}) - P_{s,.}^{'} P_{s,.}$ where $P_{s,.}$ denotes the s-row of the transition matrix P.

Proof. The distribution of firms μ_t across the discrete state space $\Phi = \{\varphi^1, \dots, \varphi^S\}$ is a $(S \times 1)$ vector equal to $(\mu_{1,t}, \dots, \mu_{S,t})$ such that $\mu_{s,t}$ is equal to the number of operating firms in state s at date t. The next period's distribution of firms across the (discrete) state space $\Phi = \{\varphi^1, \dots, \varphi^S\}$ is given by the dynamics of both incumbents and successful entrants.

In what follows, we define two conditional distributions. First, the distribution of incumbent firms at date $t+1$ conditional on the fact that incumbents were in state s at date t is denoted as $f_{t+1}^{\cdot,s}$. This $(S \times 1)$ vector is such that for each state k in $\{1, \ldots, S\}$, $f_{t+1}^{k,s}$, the k^{th} element of $f_{t+1}^{\cdot,s}$ gives the number of incumbents in state k at $t+1$ which were in state φ^s at t.

Similarly, let us define $g_{t+1}^{\cdot,s}$ the distribution of successful entrants at date $t+1$ given that they received the signal φ^s at date t. This $(S \times 1)$ vector is such that for each state k in $\{1, \ldots, S\}$, $g_{t+1}^{k,s}$, the k^{th} element of $g_{t+1}^{\cdot,s}$ gives the number of entrants in state k at $t+1$ which received a signal φ^s at t.

Period $t+1$ distribution is the sum of all these conditional distributions and thus the vector μ_{t+1} satisfies:

$$\mu_{t+1} = \sum_{s=s^*(\mu_t)}^{S} f_{t+1}^{\cdot,s} + \sum_{s=s^*(\mu_t)}^{S} g_{t+1}^{\cdot,s} \tag{5}$$

Note that $f_{t+1}^{\cdot,s}$ and $g_{t+1}^{\cdot,s}$ are now multivariate random vectors implying that μ_{t+1} also is a random vector.

At date $t+1$ for $s \geq s^*(\mu_t)$, $f_{t+1}^{\cdot,s}$ follows a multinomial distribution with two parameters: the integer $\mu_{s,t}$ and the $(S \times 1)$ vector $P'_{s,\cdot}$ where $P_{s,\cdot}$ is the s^{th} row vector of the matrix P. Similarly, at date $t+1$ for $s \geq s^*(\mu_t)$, $g_{t+1}^{\cdot,s}$ follows a multinomial distribution with two parameters: the integer MG_q and the $(S \times 1)$ vector $P'_{q,\cdot}$.

Recall that the mean and variance-covariance matrix of a multinomial distribution, $\mathcal{M}(m,h)$, is respectively the $(S \times 1)$ vector mh and the $(S \times S)$ matrix $H = diag(h) - hh'$. So let us define $W_s = diag(P_{s,\cdot}) - P'_{s,\cdot} P_{s,\cdot}$. From the right hand side of Equation 5, using the fact that the $f_{t+1}^{\cdot,s}$ and $g_{t+1}^{\cdot,s}$ follow multinomials, μ_{t+1} has a mean $m(\mu_t)$ and a variance-covariance matrix $\Sigma(\mu_t)$ where

$$\begin{aligned} m(\mu_t) &:= \sum_{s=s^*(\mu_t)}^{S} \left[\mu_t^s P'_{s,\cdot} + MG_s P'_{s,\cdot} \right] = (P_t^*)'(\mu_t + MG) \\ \Sigma(\mu_t) &:= \sum_{s=s^*(\mu_t)}^{S} (MG_s + \mu_t^s) W_s \end{aligned}$$

where P_t^* is the transition matrix P with the first $(s^*(\mu_t) - 1)$ rows replaced by zeros.

Equation 5 can be rewritten in a simple way as the sum of its mean and a zero-mean shock:

$$\mu_{t+1} = m(\mu_t) + \epsilon_{t+1}$$

where

$$\epsilon_{t+1} = \sum_{s=s^*(\mu_t)}^{S} \left[f_{t+1}^{\cdot,s} - \mu_t^s P_{s,\cdot}^{'} \right] + \sum_{s=s^*(\mu_t)}^{S} \left[g_{t+1}^{\cdot,s} - MG_s P_{s,\cdot}^{'} \right]$$

i.e ϵ_{t+1} is the demeaned version of μ_{t+1}. This gives us the result stated in the Theorem.

□

After taking into account the dynamics of incumbent firms and entry/exit decisions, the law of motion (Equation 4) of the aggregate state – the distribution of firms over productivity levels – is remarkably simple: tomorrow's distribution is an affine function of today's distribution up to a stochastic term, ϵ_{t+1}, that reshuffles firms across productivity levels.

It is easy to understand this characterization by recalling our simple example economy above without entry and exit. In this simple example, given the state transition probabilities, we should for example observe that *on average* the number of firms remaining at the intermediate level of productivity is twice that of those transiting to the highest level of productivity. This is precisely what the affine part of Equation 4 captures: the term $m(\mu_t)$ reflects these typical transitions, which are a function of matrix P alone. However, with a finite number of firms, in any given period there will be stochastic deviations from these typical transitions as we discuss above. In the Theorem, this is reflected in the "reshuffling shock" term, ϵ_{t+1}, that enters in the law of motion given by Equation 4. How important this reshuffling shock is for the evolution of the firm distribution is dictated by the variance-covariance matrix $\Sigma(\mu_t)$ which, in turn, is a function of the transition matrix P, the current firm distribution μ_t, and, in the general case with entry and exit, the signal distribution available to potential entrants.

Dynamics Under No Aggregate Uncertainty

The above characterization - in particular, Equation 4 - is instructive of the differences of the current setup relative to a standard Hopenhayn economy. The latter corresponds to the case where all of the relevant firm dynamics are encapsulated by the affine term $m(\mu_t)$. In order to understand this important benchmark case in the context of our discussion, we now briefly define the stationary distribution that obtains when the variance-covariance matrix is set to zero. Clearly, in this case, the aggregate state μ_t becomes non-stochastic and our "reshuffling shock", ϵ_{t+1}, operating on the cross-sectional distribution of firms would be absent.

The following corollary to Theorem 1 shows that the dynamics of the productivity distribution under no aggregate uncertainty are similar to the one in Hopenhayn (1992) framework.[9]

Corollary 1 *Let us define* $\hat{\mu}_t := \frac{\mu_t}{M}$ *for any* t. *When aggregate uncertainty is absent,* $\epsilon_{t+1} = 0$*:*

$$\hat{\mu}_{t+1} = (\bar{P}_t)'(\hat{\mu}_t + G) \tag{6}$$

where $\bar{P}_t$ *is the transition matrix* P *where the first* $\bar{s}(\mu_t) - 1$ *rows are replaced by zeros, and where* $\bar{s}(\mu_t)$ *is the threshold of the exit/entry rule when the variance-covariance of the* ϵ_{t+1} *is zero.*

Proof. This follows from Theorem 1 by taking $\mathbb{V}ar[\epsilon_{t+1}] = 0$ and dividing both side by M. $\square$

Under this special case, the law of motion for the distribution of firms across productivity levels is deterministic and its evolution is given by Equation 6. An immediate consequence of this is that under appropriate conditions on the transition matrix P, a stationary distribution exists and is given by:

$$\hat{\mu} = (I - \bar{P}')^{-1}\bar{P}'G \tag{7}$$

where $\bar{P}$ is the transition matrix P where the first $(\bar{s}(\hat{\mu}) - 1)$ rows are now replaced by zeros to account for equilibrium entry and exit dynamics. In the current setting, this is the analogue to Hopenhayn's (1992) stationary distribution. Henceforth, we call this object the stationary distribution, which can be interpreted as the deterministic steady-state of our model.

Taking stock, we have derived a law of motion for any finite number of firms and shown that, generically, the distribution of firms across productivity levels is time-varying in a stochastic fashion. An immediate implication of this is that aggregate productivity

[9] Note that, as is clear from the example in Figure 1 and the proof of Theorem 1, assuming that there is no aggregate uncertainty is equivalent to assuming that there is a continuum of firms.

A_t is itself stochastic. Corollary 1 implies that, in the continuum case, the distribution converges to a stationary object and, as a result, there are no aggregate fluctuations.

4 Aggregate State Dynamics under Gibrat's Law

In this section, we analyze a special case of the Markovian process driving firm level productivity: random growth dynamics. With this assumption in place, we then solve for the law of motion of a sufficient statistic with respect to aggregate productivity. By solving for this law of motion, we are then able to characterize how aggregate fluctuations, aggregate persistence and time-varying aggregate volatility arise as an endogenous feature of equilibrium firm dynamics.

4.1 Gibrat's law implies power law in the steady state

We now specialize the general Markovian process driving the evolution of firm-level productivity to the case of random growth. After exploring the firm-level implications of this assumption, we revisit the steady-state results described in the previous section.

Assumption 1 *Firm-level productivity evolves as a Markov Chain on the state space* $\Phi = \{\varphi^s\}_{s=1..S}$ *with transition matrix*

$$P = \begin{pmatrix} a+b & c & 0 & \cdots & \cdots & 0 & 0 \\ a & b & c & \cdots & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & a & b & c \\ 0 & 0 & 0 & \cdots & 0 & a & b+c \end{pmatrix}$$

This is a restriction on the general Markov process P in section 2. It provides a parsimonious parametrization for the evolution of firm-level productivity by only considering, for each productivity level, the probability of improving, c, the probability of declining, a, and their complement, $b = 1 - a - c$, the probability of remaining at the same productivity level. This process also embeds the assumption that there are reflecting barriers in productivity, both at the top and at the bottom, inducing a well-defined maximum and minimum level for firm-level productivity. This simple parametrization will be key in obtaining the closed-form results below.

The Markovian process defined in Assumption 1 has been first introduced by Champernowne (1953) and Simon (1955) and studied extensively in Córdoba (2008). For completeness, we now summarize the properties proved in the latter.

Properties 1 *For a given firm i at time t with productivity level $\varphi^{s_{i,t}}$ following the Markovian process in Assumption 1, we have the following:*

1. *The conditional expected growth rate and conditional variance of firm-level productivity are given by*

$$\mathbb{E}\left[\frac{\varphi^{s_{i,t+1}}-\varphi^{s_{i,t}}}{\varphi^{s_{i,t}}}|\varphi^{s_{i,t}}\right]=a(\varphi^{-1}-1)+c(\varphi-1)$$
$$\mathbb{V}ar\left[\frac{\varphi^{s_{i,t+1}}-\varphi^{s_{i,t}}}{\varphi^{s_{i,t}}}|\varphi^{s_{i,t}}\right]=\sigma_e^2$$

   where σ_e^2 is a constant. Both the conditional expected growth rate and the conditional variance are independent of i's productivity level, $\varphi^{s_{i,t}}$.

2. *As $t\to\infty$, the probability of firm i having productivity level φ^s is*

$$\mathbb{P}\left(\varphi^{s_{i,t}}=\varphi^s\right)\underset{t\to\infty}{\longrightarrow}K\left(\varphi^s\right)^{-\delta}$$

   where $\delta=\frac{\log(a/c)}{\log\varphi}$ and K is a normalization constant. Therefore, the stationary distribution of the Markovian Process in Assumption 1 is Pareto with tail index $\delta=\frac{\log(a/c)}{\log\varphi}$.

In short, Córdoba (2008) shows that the Markov process in Assumption 1 is a convenient way to obtain Gibrat's law on a discrete state space. In particular, Córdoba (2008) shows that whenever firm-level productivity follows this process, its conditional expected growth rate and its conditional variance are independent of the current level (part 1 of the properties above). Importantly, Córdoba (2008) additionally shows that the stationary distribution associated with this Markovian process is a power law distribution with tail index $\delta=\frac{\log(a/c)}{\log\varphi}$ (part 2 of the properties above).

The above assumption yields a tractable way of handling firm dynamics over time. At several points of the analysis below we will also be interested in understanding how the economy behaves with an ever larger number of firms. This raises the question of whether the maximum possible level of firm-level productivity should be kept fixed. If this was the case, and given the tight link between size and productivity implied by our model, increasing the number of firms would imply a declining share of economic activity commanded by the largest firms. As di Giovanni and Levchenko (2012) show this is counter-factual: in cross-country data, whenever the number of firms is larger, the share of the top firms in the economy increases. To accord with this evidence, in the following assumption we allow the maximum productivity-level to increase with the number of firms.

Assumption 2 *Assume that* $\varphi^S = ZN^{1/\delta}$

This assumption restricts the rate at which the maximum-level of productivity scales with the number of firms. To understand why this is a natural restriction to impose, first note that the stationary distribution of the Markovian process in Assumption 1 discussed above is also the cross-sectional distribution of a sample of size N of firms. Since the former is power-law distributed so is the latter. Second, from Newman (2005), the expectation of the maximum value of a sample N of random variables drawn from a power law distribution with tail index δ is proportional to $N^{1/\delta}$. Thus, under Assumption 2, for any sample of size N following the Markovian process in Assumption 1 the stationary distribution of this sample is Pareto distributed with a *constant* tail index δ.

With these two assumptions in place, we are now ready to revisit the main results in the previous section. We start by characterizing further the stationary distribution in Corollary 1, which we are now able to solve in closed-form. In particular, in Corollary 2 below we study the limiting case when the number of firms goes to infinity under Assumptions 1 and 2.[10]

Corollary 2 *Assume 1 and 2. If the potential entrants' productivity distribution is Pareto (i.e* $G_s = K_e (\varphi^s)^{-\delta_e}$*) then, as* $N \to \infty$*, the stationary productivity distribution converges point-wise to:*

$$\hat{\mu}_s = K_1 \left(\frac{\varphi^s}{\varphi^{s^*}}\right)^{-\delta} + K_2 \left(\frac{\varphi^s}{\varphi^{s^*}}\right)^{-\delta_e} \qquad \textit{for } s \geq s^*$$

where $\delta = \frac{\log(a/c)}{\log(\varphi)}$ *and* K_1 *and* K_2 *are constants, independent of* s *and* N*.*

[10] The proof of this Corollary is in two steps: (i) we first solve closed form for the stationary distribution given a maximum level of productivity φ^S under Assumption 1; (ii) we then take the limit of this distribution when the number of firms goes to infinity under the Assumption 2 stating how the maximum level of productivity scales with the number of firms.

Proof: See appendix A.1 $\square$.

Thus, the stationary productivity distribution for surviving firms (i.e. for $s \geq s^*$), is a mixture of two Pareto distributions: (i) the stationary distribution of the Markovian process assumed in 1 with tail index δ and (ii) the potential entrant distribution with tail index δ_e.

The first of these distributions is a consequence of Gibrat's law and a lower bound on the size distribution. This works in a similar way to the existent random growth literature. In the context of our model, this lower bound friction results from optimal entry and exit decisions by firms. Every period there is a number of firms whose productivity draws are low enough to induce to exit. These are replaced by low-productivity entrants inducing bunching around the exit/entry threshold, s^*, as in Luttmer (2007, 2010, 2012) . Unlike Luttmer however, our entrants can enter at every productivity level, according to a Pareto distribution. This leads to the second term in the productivity distribution above.

While the Corollary above characterizes the stationary firm-level productivity distribution it is immediate to apply these results to the firm size distribution. This is because the firm size distribution maps one to one to μ_t. To see this, recall that the output of a firm with productivity level φ^s is given by: $y_s = (\varphi^s)^{\frac{1}{1-\alpha}} \left(\frac{\alpha}{w}\right)^{\frac{\alpha}{1-\alpha}}$. Therefore, in the steady state, the number of firms of size y_s is given by μ_s.

Our Corollary 2 therefore implies that, for sufficiently large firms, the tail of the firm size distribution is Pareto distributed with tail index given by $\min\{\delta(1-\alpha), \delta_e(1-\alpha)\}$.[11] Note that the discrepancy between the firm (output) size distribution and the firm productivity distribution is governed by the degree of returns to scale α. The higher the degree of returns to scale, the lower the ratio between the productivity distribution tail, δ, and the firm size distribution tail $\delta(1-\alpha)$.

[11] To see this, note that for high productivity levels (i.e. for large s) the tail of the productivity distribution is given by the smaller tail index, i.e. the fattest-tail distribution among the two.

4.2 Aggregate Dynamics: A Complete Characterization

With the above assumptions in place, we now offer a complete analytical characterization of the evolution of equilibrium aggregate output in our model. Recall from Theorem 1 that we have already derived the law of motion of the aggregate state, i.e. the productivity distribution, for any Markov process governing the evolution of firm-level productivity. Under assumptions 1 and 2 we can specialize this law of motion to the particular case of random growth dynamics at the firm-level.

In particular, we derive the law of motion of

$$T_t := B'\mu_t = \sum_{s=1}^{S} (\varphi^s)^{\frac{1}{1-\alpha}} \mu_{t,s}$$

From the expressions for aggregate productivity (Equation 1) and the equilibrium wage (Equation 2), it is immediate that T_t is a sufficient statistic to solve for the relative price of labor in our model. By deriving the law of motion for T_t we are therefore able to characterize the law of motion for aggregate prices, productivity and output.

Theorem 2 *Assume 1 and 2. If the potential entrants' productivity distribution is Pareto (i.e $G_s = K_e (\varphi^s)^{-\delta_e}$) then,*

$$T_{t+1} = \rho T_t + \rho E_t(\varphi) + O_t^T + \sigma_t \varepsilon_{t+1} \tag{8}$$

$$\sigma_t^2 = \varrho D_t + \varrho E_t(\varphi^2) + O_t^\sigma \tag{9}$$

where $\mathbb{E}[\varepsilon_{t+1}] = 0$ and $\mathbb{V}ar[\varepsilon_{t+1}] = 1$. The persistence of the aggregate state is $\rho = a\varphi^{\frac{-1}{1-\alpha}} + b + c\varphi^{\frac{1}{1-\alpha}}$. The net entry term is the difference between the entry and exit contributions: $E_t(x) = \left(M \sum_{s=s_t}^{S} G_s (x^s)^{\frac{1}{1-\alpha}}\right) - \left((x^{s_t-1})^{\frac{1}{1-\alpha}} \mu_{s_t-1,t}\right)$. The term D_t is given by $D_t := \sum_{s=s_t-1}^{S} \left((\varphi^s)^{\frac{1}{1-\alpha}}\right)^2 \mu_{s,t}$ and $\varrho = a\varphi^{\frac{-2}{1-\alpha}} + b + c\varphi^{\frac{2}{1-\alpha}} - \rho^2$. The terms O_t^T and O_t^σ are a correction for the upper reflecting barrier in the idiosyncratic state space.

Proof: See appendix A.3 □.

Theorem 2 provides a full description of aggregate state dynamics in our model. It can be understood intuitively in terms of the evolution of aggregate productivity by noting that $T_t^{(1-\alpha)} = A_t$, i.e. T_t is simply a convex function of aggregate productivity.

Thus, up to this transformation, the Theorem states that the aggregate productivity of incumbents tomorrow is the sum of *(i)* ρT_t, the expected aggregate productivity of today's incumbents, conditional on their survival, *(ii)* $\rho E_t(\varphi)$, the expected aggregate productivity of today's net entrants conditional on their survival, and *(iii)* $\sigma_t \varepsilon_{t+1}$ a mean zero aggregate productivity shock. The term O_t^T is a correction term, arising from having imposed bounds on the state-space. This term vanishes as the state-space bounds increase; we relegate its precise functional form and further discussion of this term to the appendix.

Given the law of motion for the aggregate state, it is straightforward to characterize the law of motion for equilibrium output in our model. Corollary 3 does this by describing the dynamics of $\widehat{Y}_{t+1}$, the percentage deviation of output from its steady-state value.

Corollary 3 *Assume 1 and 2. If the potential entrants' productivity distribution is Pareto (i.e $G_s = K_e\left(\varphi^s\right)^{-\delta_e}$) then, aggregate output (in percentage deviation from steady state) has the following law of motion:*

$$\widehat{Y}_{t+1} = \rho \widehat{Y}_t + \rho \kappa_1 \widehat{E}_t(\varphi) + \kappa_2 \widehat{O_t^T} + \frac{\sigma_t}{T}\epsilon_{t+1} \tag{10}$$

where σ_t is given by Equation 9, $\widehat{E}_t(\varphi)$ is the percentage deviation from steady-state of $E_t(\varphi)$, $\widehat{O_t^T}$ is the percentage deviation from steady-state of O_t^T, κ_1, κ_2 are constants defined in the appendix and T is the steady-state value of the aggregate state variable T_t.

Proof: See appendix A.4 □.

The law of motion for T_t thus implies that aggregate output is persistent, as parametrized by ρ, and displays time-varying volatility, given by σ_t. Again, it is worth noting that there are no aggregate shocks or aggregate sources of persistence in our setting. Rather, these two properties emerge from the aggregation of firm-level dynamics alone. To better understand these aggregation results, and building on the expressions in Theorem 2 and Corollary 3, we now detail how persistence in aggregates depend on micro-level parameters. We then turn our attention to the (firm-level) origins of time-varying volatility in aggregates.

Aggregate Persistence

The following Proposition characterizes how the persistence of aggregate output, ρ, depends on parameters governing firm-level dynamics.

Proposition 1 *Let $\delta = \frac{\log \frac{a}{c}}{\log \varphi}$ be the tail index of the stationary productivity distribution as in Corollary 2. If $\delta \geq \frac{1}{1-\alpha}$ then the persistence of the aggregate state, ρ, satisfies the following properties:*

i) Holding δ constant, aggregate persistence is increasing in firm-level persistence: $\frac{\partial \rho}{\partial b} \geq 0$

ii) Holding b constant, aggregate persistence is decreasing in the tail index of the stationary productivity distribution: $\frac{\partial \rho}{\partial \delta} \leq 0$

iii) If the productivity distribution is Zipf, aggregate state dynamics contain a unit root: if $\delta = \frac{1}{1-\alpha}$, $\rho = 1$

Proof: See appendix A.5 □.

To interpret the condition under which the Proposition is valid, recall that $\delta(1-\alpha)$ gives the tail index of the stationary firm size distribution. Hence, the Proposition applies to Pareto distributions that are (weakly) thinner than Zipf. According to the Proposition, (i) the persistence of the aggregate state (and hence aggregate productivity, wages and output) is increasing in the probability, b, that firms do not change their productivity from one time period to the other. Intuitively, the higher is firm-level productivity persistence, the more persistent are aggregates.[12]

Further, according to (ii) in the Proposition, aggregate persistence will decrease with the tail index of the stationary firm-level productivity distribution. To understand this, note that this tail index is given by $\frac{a}{c}$. The thinner the tail, the larger this ratio is and thus, the larger is the relative probability of a firm having a lower productivity tomorrow. This therefore induces stronger mean reversion in productivity (and size) at the firm level which, in turn, leads to lower aggregate persistence. Thus, a fatter tail in the size distribution implies heightened aggregate persistence. In the limiting case

[12]Note that we are holding the tail index of the stationary productivity distribution, δ, constant. In terms of model primitives, we are keeping fixed the ratio $\frac{a}{c}$ while maintaining the adding-up constraint $a + b + c = 1$.

where the stationary size distribution is given by Zipf's law ($\delta(1-\alpha)=1$ in case (iii)), aggregate persistence is equal to 1. That is, Zipf's law implies unit-root type dynamics in aggregates.

(Time-Varying) Aggregate Volatility

We are now interested in understanding how aggregate volatility - and its evolution - depend on the parameters driving the micro-dynamics. To do this, we find it convenient to first rewrite the expression for the conditional volatility of $\widehat{Y}_{t+1}$ as:

$$\mathbb{V}ar_t\left[\widehat{Y}_{t+1}\right]=\frac{\sigma_t^2}{T^2}=\varrho\frac{D}{T^2}\frac{D_t}{D}+\varrho\frac{E(\varphi^2)}{T^2}\frac{E_t(\varphi^2)}{E(\varphi^2)}+\frac{O^\sigma}{T^2}\frac{O_t^\sigma}{O^\sigma} \tag{11}$$

where T, E and D are the the steady-state counterparts of T_t, E_t and D_t. To interpret these objects, first recall that $D_t=\sum_{s=s_t-1}^{S}\left((\varphi^s)^{\frac{1}{1-\alpha}}\right)^2\mu_{s,t}$ is proportional to the second moment of the firm size distribution at time t, a well defined measure of dispersion.[13] D is therefore proportional to the steady state dispersion in firm size. By the same argument, $E_t(\varphi^2)=\left(M\sum_{s=s_t}^{S}\left((\varphi^s)^{\frac{1}{1-\alpha}}\right)^2G_s\right)-\left(\left((\varphi^{s_t-1})^{\frac{1}{1-\alpha}}\right)^2\mu_{s_t-1,t}\right)$, is proportional to dispersion of firm size among successful entrants [14] and E is the corresponding object at the steady state.

Note also that the expression above implies that the *unconditional* expectation of conditional variance can be written as:

$$\mathbb{E}\frac{\sigma_t^2}{T^2}=\varrho\frac{D}{T^2}+\varrho\frac{E(\varphi^2)}{T^2}+\frac{O^\sigma}{T^2} \tag{12}$$

With these two objects in place, the following Proposition characterizes how aggregate volatility and its dynamics depend on the primitives of the model.

[13]To see this, recall that the size at time t of a firm with productivity level φ^s is given by $y_{s,t}=(\varphi^s)^{\frac{1}{1-\alpha}}(\alpha/w_t)^{\frac{\alpha}{1-\alpha}}$. The second moment of the firm size distribution is then $\sum_{s=1}^{S}y_{s,t}^2\mu_{s,t}=\sum_{s=1}^{S}(\varphi^s)^{\frac{2}{1-\alpha}}(\alpha/w_t)^{\frac{2\alpha}{1-\alpha}}\mu_{s,t}=(\alpha/w_t)^{\frac{2\alpha}{1-\alpha}}D_t$. In other words, D_t is proportional to the second moment of the firm size distribution at time t.

[14]Where the last term in the expression corrects for exit, and is proportional to the dispersion in the size of exiters.

Proposition 2 *Let $\delta = \frac{\log \frac{a}{c}}{\log \varphi}$ be the tail index of the stationary productivity distribution as in Corollary 2. Let δ_e be the tail index associated with the productivity distribution of potential entrants. Then*

i) Under assumption 2 and if $1 < \delta(1-\alpha) < 2$ and $1 < \delta_e(1-\alpha) < 2$, the unconditional expectation of aggregate variance satisfies:

$$\mathbb{E}\left[\frac{\sigma_t^2}{T^2}\right] \underset{N\to\infty}{\sim} \frac{\varrho D_1}{N^{2-\frac{2}{\delta(1-\alpha)}}} + \frac{\varrho D_2}{N^{1+\frac{\delta_e}{\delta}-\frac{2}{\delta(1-\alpha)}}} \tag{13}$$

where D_1 and D_2 are functions of model parameters but independent of N and M.

ii) The dynamics of conditional aggregate volatility depend on the dispersion of firm size:

$$\frac{\partial \mathbb{V}ar_t\left[\widehat{Y}_{t+1}\right]}{\partial D_t} = \frac{\partial \mathbb{V}ar_t\left[\widehat{Y}_{t+1}\right]}{\partial E_t} = \frac{\varrho}{T^2} \geq 0$$

Proof: See appendix A.6 $\square$.

Part (i) of Proposition 2 characterizes the average level of volatility of aggregate output growth in our model. It builds on our result that, under random growth dynamics for firm-level productivity, the stationary incumbent size distribution is Pareto distributed. It assumes further that the size distribution of potential entrants is also power law distributed. The assumptions on $\delta(1-\alpha)$ and $\delta_e(1-\alpha)$ ensure that these distributions are sufficiently fat tailed.

The $\underset{N\to\infty}{\sim}$ notation means that, in expectation, the conditional variance of the aggregate growth rate scales with N, the number of incumbent firms at the steady state, at a rate that is equal to the rate of the expression on the right hand side. The latter reflects the separate contributions of (i) surviving time t incumbents (i.e. the first term in the expression) and that of (ii) time t entrants that start producing at $t+1$ (i.e. the second term in the expression).

The key conclusion of the first part of Proposition 2 is therefore that, for $1 < \delta(1-\alpha) < 2$ and $1 < \delta_e(1-\alpha) < 2$, the variance of aggregate growth scales at a slower rate than $1/N$. Recall that the latter would be the rate of decay implied by a shock-diversification argument relying on standard central limit Theorems. This is not the case when the firm size distribution is fat tailed, as it is here. Rather, as the Proposition makes clear, the rate of decay of aggregate volatility depends on the tail indexes of the size distributions of entrants and incumbents. Whenever the size distribution of incumbents has a lower tail index than the size distribution of entrants - i.e. whenever $\delta < \delta_e$ - the rate of decay of aggregate volatility is a function of the tail index of incumbents alone. Conversely, whenever the size distribution of entrants has a fatter tail, the rate of decay is a function of the tail behavior of both incumbents and entrants. For either case, the closer are these distributions to Zipf's law, the slower is the rate of the decay.

This Proposition thus generalizes the main result in Gabaix (2011) to an environment where: (i) firm dynamics are the result of optimal intratemporal (i.e size) and intertemporal firm-level decisions (i.e. exit and entry), given the idiosyncratic productivity process characterized by the Markovian process in Assumption 1 and (ii) the Pareto distribution of firm sizes is an equilibrium outcome consistent with optimal firm decisions. More generally, the first part of Proposition implies that the voluminous literature that builds on the framework of Hopenhayn (1992) has overlooked the potentially non-negligible aggregate dynamics implied by the model, even when the number of firms entertained is large.

Part (ii) of Proposition 2 shows that the evolution of aggregate volatility over time – i.e. the *conditional* variance of aggregate output – mirrors that of D_t. As discussed above, D_t is proportional to the second moment of the firm size distribution at time t. Thus, whenever the firm size distribution at time t is more dispersed than the stationary distribution ($D_t > D$), aggregate volatility is higher.

The second part of the Proposition is therefore related to a literature looking at the connection between micro and macro uncertainty (see Bloom et al, 2014 and Kehrig, 2014). Consistent with the results of this literature, the Proposition yields a direct, positive, link between the two levels of uncertainty. Unlike this literature however, this link between the cross-sectional dispersion of micro-units and conditional aggregate volatility is endogenous and emerges without resorting to any exogenous aggregate shocks influencing the first and second moments of firms' growth.

5 Quantitative Results

In this section we present the quantitative implications of our model. We solve the model under the particular case of firm-level random productivity growth, which we have discussed in the previous section. We first calibrate the stationary steady-state solution of the model to match firm-level moments.

Based on this calibration, we then use the law of motion of the aggregate state to solve numerically for the firms' policy function. Thanks to our analytical characterization, our numerical strategy is substantially less computational intensive than that traditionally used when solving for heterogeneous agents models under aggregate uncertainty. This allows us to solve the model under a large-dimensional state-space. We discuss this in detail below. Using this numerical solution we quantitatively assess the performance of the model with respect to standard business cycle statistics and inspect the mechanism rendering firm-level idiosyncratic shocks into aggregate fluctuations. We then quantitatively explore the role of large firms in shaping the business cycle. Throughout we show empirical evidence that is consistent with our mechanism.

5.1 Steady-state calibration

We choose to calibrate the production units in our model to firm level data. To calibrate the model to the US economy, we first set the value of deep parameters. The span of control parameter α is set at 0.8. This value is chosen to be on the lower end of estimates, such as Basu and Fernald (1997) and Lee (2005). The discount factor β is set at 0.95 so that the implied annual gross interest rate is 4%, a value in line with the business cycle literature. The labor supply elasticity parameter, γ, is chosen to be 2 following Rogerson and Wallenius' (2009) argument linking micro and macro elasticities of labor supply. Finally, we set the fixed cost of production, c_f, to 1 every period.[15]

Statistic	Model	Data	References
Entry Rate	0.109	0.109	BDS firm data
Idiosyncratic Vol. σ_e	0.08	$0.1 - 0.2$	Castro et al. (forthcoming)
Tail index of Estab size dist.	1.097	1.097	BDS firm data
Tail index of Entrant Estab size dist.	1.570	1.570	BDS firm data
Share of Employment of the largest firm	0.2%	1%	Share of Wall-Mart
Number of establishments	4.5×10^6	4.5×10^6	BDS firm data

Table 1: Targets for the calibration of parameters

[15]In model simulations we found that the quantitative performance of the model is not affected by the value of fixed costs.

We then assume random productivity growth at the firm level, i.e. we follow Assumption 1 in the previous section. This implies that the stationary productivity distribution is Pareto-distributed with tail index δ. We additionally assume that the productivity of potential entrants is Pareto distributed with tail index δ_e. With these assumptions in place, the behavior of the model is exactly characterized by Theorem 2. This will also imply a Pareto distribution for firm size in the stationary steady state.

To obtain data counterparts for these and further moments discussed below, we use publicly available tabulations of firm size and firm size by age from the Business Dynamics Statistics (BDS) data between 1977 and 2012. These are in turn computed from the Longitudinal Business Database of the US census and ensure a near full coverage of the population of US firms. For a full description of this dataset and our computations below, please refer to the Data Appendix B.

According to our model, we can read off the tail of the productivity distribution of incumbents from its empirical counterpart by using the relation $\delta(1-\alpha) = 1.097$ and our assumed value for α. According to Corollary 2, this fixes the ratio between the parameters a and c in the firm-level productivity process, up to the state space parameter φ. Similarly, we fix the tail index for entrant distribution such that $\delta(1 - \alpha) = 1.570$. To obtain these numbers, we estimate the tail index from the BDS data at the US census. The data counterpart to the stationary size distribution in our model is given by the average (across years) of the size distribution of all firms. The corresponding object for entrants is given by the average (across years) size distribution of age 0 firms in the BDS data. We obtain tail estimates by using the estimator proposed in Virkar and Clauset (2014). According to our estimates the size distribution of incumbents is more fat tailed than that of the corresponding distribution of entrants; this is intuitive as the probability of observing very large entrants should indeed be smaller. While we are not aware of any such estimation for entrants, our tail index estimates for incumbents compare well with published estimates by Axtell (2001), Gabaix (2011) and Luttmer (2007).

We are left with two parameters to calibrate: the state space parameter φ and the parameter governing persistence in firm-level productivity, b. We calibrate these parameters jointly to match the volatility of firm-level growth rates, $\sigma_e = 8\%$ and the stationary entry rate, i.e. the ratio of entrants to incumbents, to be equal to 10.9%. Our choice for firm-level volatility is an intentionally conservative choice relative to the typical values reported in the literature. Working with Compustat data, Comin and Phillipon (2006) and Davis et al (2007) report sales growth volatility estimates for publicly listed firms between 0.1 and 0.2. Davis et al (2007) report even higher values for employment volatility at privately held firms, based on the Longitudinal Database of Businesses. Working with establishment level data, Foster et al (2008) and Castro et al (forthcoming) report an average value for annual productivity (TFPR) volatility of about 20% . By choosing a value for volatility at the lower end of these estimates, we acknowledge that large firms are typically less volatile than the average firm. Our choice for the entry rate target comes from computing the average entry rate from the BDS data. This is consistent with, for example, the values reported by Dunne et al (1988).[16]

Finally, we need to choose values for S and M. By choosing S we are fixing the largest possible productivity of a firm. Our choice of S implies that the largest firm accounts for 0.2% of total employment. This is again a conservative choice: Walmart for example is reported to have 1.4 million employees based in the US, about 1% of the labor force in the US. Recall that M is a free scale parameter as discussed in the setup of the model. We calibrate M such that the total number of firms is about 4.5 million, the mean total number of establishments reported in the BDS data. In table 1, we report the establishment moments that we match for the calibration. In table 2, the implied parameters by our targets.

We are interested in accurately matching the characteristics of large firms. Recall that our calibration procedure is intended to match well the *tail* of the firm size distribution. The left panel of Figure 2, plots the *entire* firm size distribution (in terms of employees) as implied by our model against that in the data. The right panel plots the corresponding distribution for entrants. These are plots of the counter-cumulative (CCDF) distribution of firm size giving, in the x-axis, the employment size category of a given firm and, in the y-axis, the empirical probability of finding a firm larger than the corresponding x-axis employment size category. The solid line reports the stationary size distribution in the model.

Filled (black) circles give the size distribution derived from the Business Dynamics Statistics (BDS) from the US Census which we have used to estimate the tail index. Note that the largest bin in the BDS data only pins down the *minimum* size of the

[16] We discuss in detail our data sources and computations in the Data Appendix.

Parameters	Value	Description
a	0.6129	Pr. of moving down
c	0.3870	Pr. of moving up
S	36	Number of productivity levels
φ	1.0874	Step in pdty bins
Φ	$\{\varphi^s\}_{s=1..S}$	Productivity grid
γ	2	Labor Elasticity
α	0.8	Production function
c_f	1.0	Operating cost
c_e	0	Entry cost
β	0.95	Discount rate
M	$4.8581 * 10^7$	Number of potential entrants
G	$\{MK_e(\varphi^s)^{-\delta_e}\}_{s=1..S}$	Entrant's distr. of the signal
K_e	0.9313	Tail parameter of the distr. G
$\delta_e(1-\alpha)$	1.570	Scale parameter of the distr. G

Table 2: Baseline calibration

largest firms in the data, corresponding to those with more than 10000 employees. In order to go beyond this data limitation, we supplement the BDS tabulations with Compustat data. Specifically, the hollow (red) circles in Figure 2 are computed from tabulating frequencies for Compustat firms *above* 10000 employees. Our assumption is that, for firms above 10000 employees, the distribution of firms in Compustat is similar to the one for all firms in the U.S. economy.

The model does well in matching the firm size distribution: it captures well the mass of small firms in the BDS data and the mass of large firms in the Compustat data. These latter moments are not targets of our calibration strategy (only the tail estimated on BDS data alone is). The same general pattern holds for the entrant distribution. The model slightly under-predicts the probability of finding very large firms. For instance, in our model the probability of finding a firm with more than 10000 employees is just under 0.0001 while the corresponding probability in the data is 0.0003. This is again consistent with our conservative calibration strategy and ensures that our results below are not driven by firms that are too large with respect to the data.

Turning to heterogeneity in productivity, and in particular, to how productive large firms are in our model, our calibration implies that the interquartile ratio in firm-level productivity is 1.29. Looking further at highly productive firms in the model, the ratio in total factor productivity between a firm at the 95th percentile and the 5th percentile

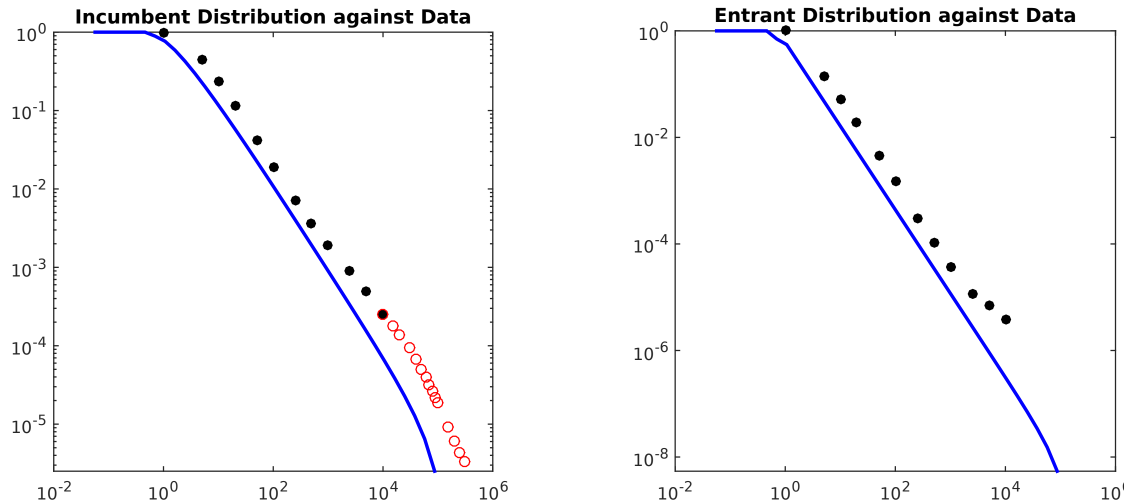

Figure 2: Counter Cumulative Distribution Functions (CCDF) of the firms size distribution of incumbents (left) and entrants (right) in the model (blue solid line) against data (circles).

NOTE: Black filled circle report the CCDF of firm size distribution for less than 10000 employees in the BDS. The red circle are tabulation from Compustat for firms with more than 10000 employees assuming that for this range the distribution of firms in Compustat is similar to the one of firms in the whole economy.

is 1.80. While we are not aware of any such computations with actual firm level data, these numbers are smaller but comparable to the *establishment* level moments, reported by Syverson (2004) which finds an interquartile ratio of 1.34 and a 95th-5th quantile ratio of 2.55.

5.2 Business cycle implications

We now solve the model outside the stationary steady state and provide a quantification of its performance as a theory of the business cycle. Due to the fact that we are able to solve for the aggregate state dynamics analytically, our numerical strategy is related to but less computational intensive than other commonly employed strategies. We start by describing this briefly. Based on our numerical solution, we then compute aggregate business cycle statistics and compare them against the data. We then inspect the mechanism in our model by performing a simple impulse response exercise, as traditional in the business cycle literature. The key difference is that here we track the aggregate response to an idiosyncratic shock which endogenously translates into an aggregate perturbation.

5.2.1 Numerical Strategy

The analytical solution to the law of motion of the aggregate state in Theorem 2 is key to our numerical strategy. Recall that in the model firms make optimal intertemporal decisions (entry and exit) by forming expectations of future aggregate conditions which are summarized by the state variable T_t. As Theorem 2 renders clear, the dynamics of T_t in turn depend on D_t, a term that is proportional to the second moment of the firm size distribution. Since the firm size distribution is a stochastic and time-varying object so is D_t. In order to solve the model numerically we will make the assumption that D_t is perceived by firms to be fixed at its steady-state value. Given this assumption, the firms' problem can be solved by standard value function iteration methods. We discuss this numerical algorithm in detail in Appendix C.

This is similar in spirit to the Krusell-Smith approach in that agents only take into account a reduced set of moments of the underlying high-dimensional state variable. Unlike Krusell and Smith (1998) however, we have a closed-form solution for the law of motion of the first moment of this distribution and hence, in our numerical solution, agents know the law of motion of T_t, up to second moments. In Krusell and Smith (1998) this law of motion has to be solved for, which imposes a simulation step with a high computational cost. Our procedure is also similar to that of Den Haan and Rendahl (2010) in that we exploit recurrence Equations linking different moments of the state distribution and then assume that agents' expectations do not depend on higher order moments of the distribution. Unlike them we do not need to solve for the law of motion of these moments. The consequence of a lower computational cost relative to the literature is that it allows us to solve for a large state space and therefore to better capture firm-level heterogeneity in productivity.[17]

5.2.2 Business cycle statistics

Using the calibration in 2 and the numerical algorithm describes in Appendix C, we compute the business cycle statistics. We simulate time series for output, hours and

[17] This solution method is arguably applicable to a large class of heterogeneous agents models whenever these incorporate random growth processes.

	Model			Data		
	$\sigma(x)$	$\frac{\sigma(x)}{\sigma(y)}$	$\rho(x,y)$	$\sigma(x)$	$\frac{\sigma(x)}{\sigma(y)}$	$\rho(x,y)$
Output	0.47	1.0	1.0	1.83	1.0	1.0
Hours	0.31	0.66	1.0	1.78	0.98	0.90
Agg. Productivity	0.21	0.46	1.0	1.04	0.57	0.66

Table 3: Business Cycle Statistics

NOTE: The model statistics are computed for the baseline calibration (cf. Table 2) for an economy simulated for 20,000 periods. The data statistics are computed from annual data in deviations from an HP trend. The source of the data is Fernald (2014). For further details refer to Appendix B.

aggregate TFP using the law of motion (4) of the productivity distribution. These statistics are presented in Table 3.

The standard deviation of aggregate output in the model is 0.47%, 26% of the annual volatility of HP-filtered real GDP in the data. The standard deviation of hours is 0.31%, about 17% of the annual volatility of total hours worked (in deviations from trend). As a result, the volatility of hours relative to output is 0.66, about two-thirds of relative volatility of hours to GDP in the data. The dampened behavior of hours relative to output in our model is not unlike that of a baseline RBC model. As in the latter class of models, aggregate dynamics in our model follow from the dynamics of aggregate TFP. In our baseline calibration, the standard deviation of aggregate productivity is 0.21%, about one fifth of the volatility of the (aggregate) Solow residual in data.

Crucially, in our model, these aggregate TFP dynamics are *not* the result of an exogenous "aggregate" shock. Rather they are the *endogenous* outcome of (*i*) the evolution of micro-level productivity and (*ii*) optimal decisions by firms regarding size, entry and exit. Thus, idiosyncratic firm dynamics account for one fifth of aggregate TFP variability in the data.

There are two benchmarks against which to compare this number. The first reflects the maintained assumption in the firm dynamics literature: in a model featuring 4.5 million firms - as our model does - the law of large numbers should hold exactly and therefore we should obtain *zero* aggregate volatility. The fact that we do not is the result of acknowledging the role of the firm size distribution in rendering idiosyncratic disturbances into aggregate ones; as first emphasized by Gabaix (2011) and generalized by our Proposition 2.[18]

The second benchmark is given by the recent empirical work of di Giovanni, Levchenko, Mejean (2014). Using a database covering the universe of French firms they conclude that firm-level idiosyncratic shocks account for about half of aggregate fluctuations and that a quarter of this can in turn be attributed to the direct impact of large firm dynamics on aggregates (the other three quarters being attributed to linkages as in Acemoglu et al 2012). Our quantification gives a structural interpretation to these numbers and the implied magnitudes in rough agreement with the (reduced-form) empirical estimates of di Giovanni, Levchenko, Mejean (2014) .

5.2.3 Inspecting the mechanism

As Proposition 2 renders clear, large firm dynamics are at the heart of the aggregate dynamics summarized above. Intuitively, the endogenous Pareto distribution of firm size implies that a relatively small group of very large firms have a probability mass that is non-negligible: individually, each firm accounts for a sizeable share of aggregate activity. Further, the number of very large firms is small enough that idiosyncratic shocks may not average out: if a large firm suffers a negative productivity shock, it is unlikely that a comparable sized firm suffers a positive shock that exactly compensates for the former.

By the same token, it is important to note that no individual firm drives the economy. Rather, the number of large firms is small enough that idiosyncratic shocks hitting these firms might actually appear (to the econometrician) like a correlated disturbance. For example, if there are only ten very large firms, the probability of eight of them suffering a negative shock is non-negligible; thus the probability of the economy entering a recession is also non-negligible even though there are no aggregate shocks. To put it simply, in our model business cycles have at a "small sample" origin.

To render this intuition concrete, we now compute an impulse response function giving the dynamic impact on aggregates of a negative one standard deviation shock to the

[18]This has important implications for a rich literature analyzing the interaction between firm dynamics and aggregate fluctuations; one where, invariably, aggregate shocks play a first order role. See, for example, Bloom et al (2014), Campbell (1998), Clementi and Palazzo (2015), Khan and Thomas (2008) and Veracierto (2002).

productivity growth rate of the largest firm. Figure 3 shows the response of aggregate output, hours, productivity along with the average marginal productivity of labor across firms.[19]

The top panel of Figure 3 shows the responses of aggregate output and aggregate hours to this large-firm shock. The dynamics of both variables closely mirror that of aggregate productivity, as displayed in the bottom left panel. Thus, after a one standard deviation negative shock to the largest firm, aggregate productivity decreases by 0.016%. In turn, this decline in aggregate productivity has a non-negligible effect on aggregate output, which declines by about 0.035% on impact. As a consequence, aggregate labor demand declines and hours worked fall by 0.023%. If these magnitudes seem small, recall that this is the result of a shock to a *single* firm out of 4.5 million firms.

The qualitative dynamics implied by the aggregate responses are consistent with that of a representative firm RBC model. However, in our model these aggregate responses to an idiosyncratic pulse also reflect the adjustment of all other firms in the economy. The bottom right panel in Figure 3 displays the response of the marginal productivity of the largest firm's competitors.

The intuition is as follows. Since the largest firm becomes less productive, it will shrink optimally and cut its labor demand. This in turn induces a decline in aggregate labor demand and thus in the equilibrium wage. Competitors of this firm, producing the same good and not having changed their productivity, now face a lower wage and optimally increase their size. In short, as a result of the shock, production is reallocated to the less productive competitors of the largest firm in the economy.

Note however that due to decreasing returns to scale, these firms' marginal productivity of labor also decreases. Therefore, though this process of reallocation towards competitors mitigates the aggregate response to the idiosyncratic shock, it is not strong enough to undo the initial effect of the shock. Were we to shut down this reallocation effect - by keeping the wage fixed - aggregate output would decrease by 0.081%, more than twice the effect in our baseline calibration.

[19] From the structure of the model, computing the impulse response is straightforward. From Equation 4, note that the transition of the firm size distribution between date t and date $t+1$ is a linear operator. Therefore, after computing the initial shock ϵ_t, we do not need to simulate a large number of paths and to take the average. Instead, we assume ϵ_t to be zero for $t \geq 1$ and thanks to the linearity of transition described by Equation 4 the result is exactly the same.

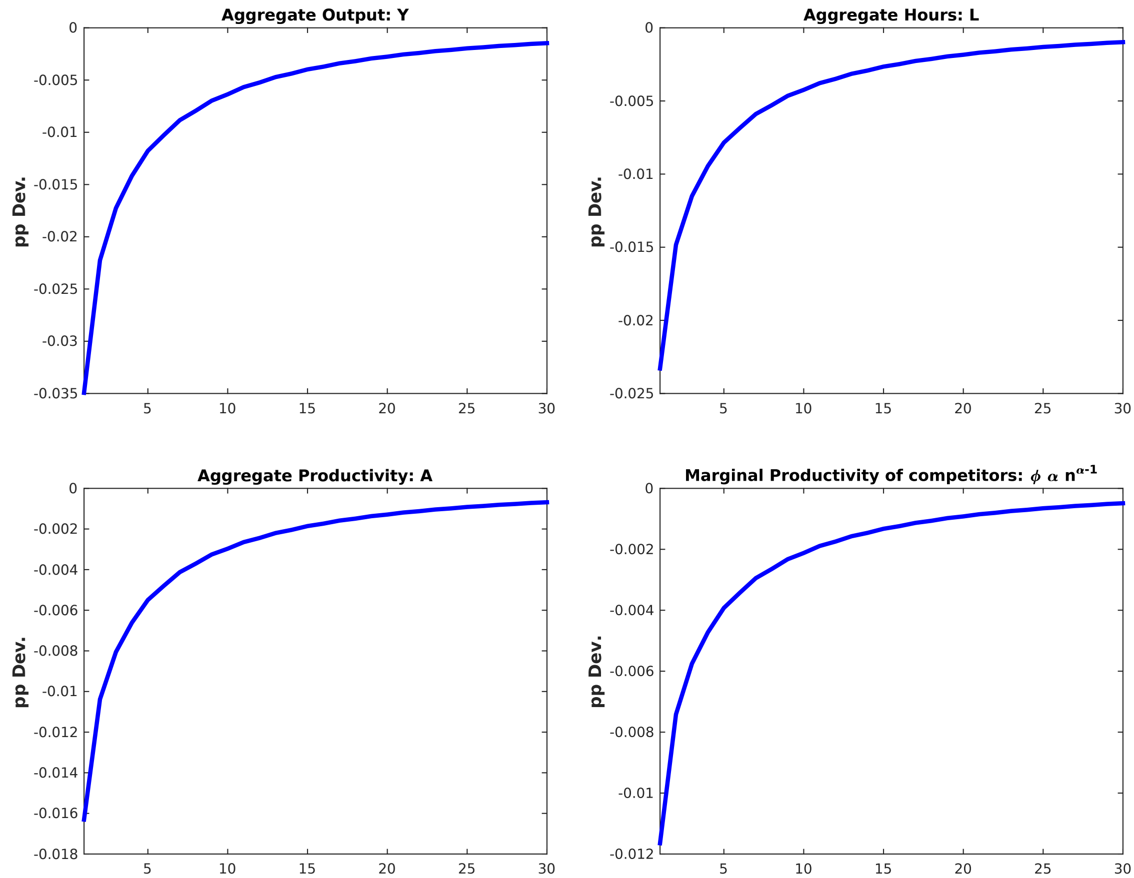

Figure 3: Impulse response to a one standard deviation negative productivity shock on the largest firm.

5.3 Large Firm Dynamics over the Business Cycle

Our model delivers two first-order implications for understanding aggregate fluctuations and the evolution of aggregate volatility. First, large firm dynamics drive aggregate growth and, hence, the business cycle. Second, cross-sectional dispersion in firm size drives aggregate volatility. In this section, we explore quantitatively these two implications of our model and show empirical evidence that is consistent with our mechanism.

To understand the impact of large firm dynamics on aggregate first moments we focus on the dynamic relationship between the tail of the firm size distribution and aggregate growth. To understand why we choose time variation in the tail index as a summary statistic for large firm dynamics, notice the following. It is clear that (both in the model

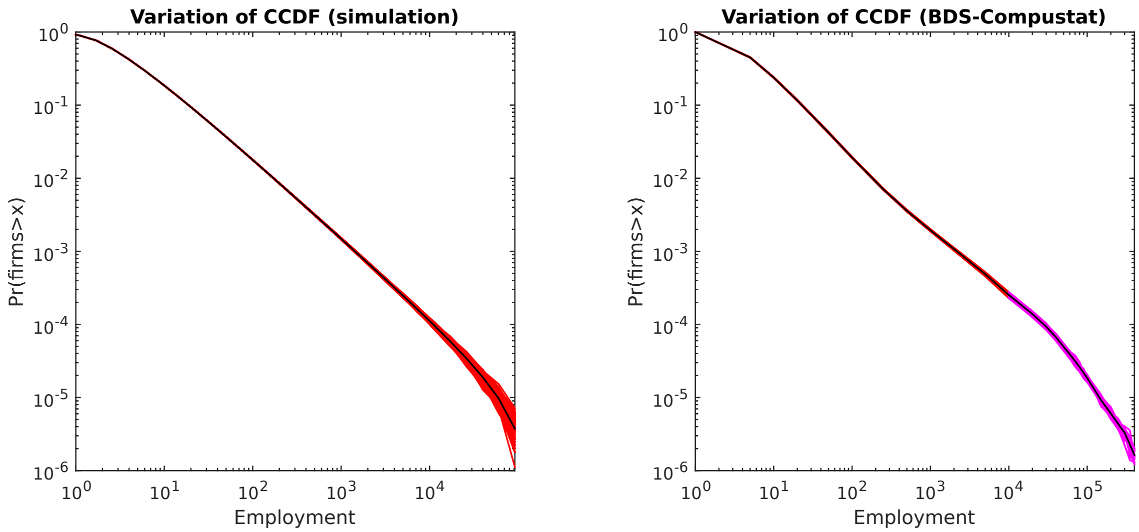

Figure 4: Variation of the Counter Cumulative Distribution Function (CCDF) in simulated data (left) and in the BDS data (right).

NOTE: The simulated data are the results of a 25000 periods sample (where the first 5000 are dropped). For the BDS data, we compute the CCDF for each year on the sample 1977-2008. The dashed black line is the mean of each sample.

and in the data) large firms comove with the business cycle.[20] By our argument in the previous subsection, since the number of large firms is relatively small, fluctuations within this group of firms will not cancel out. Note that this is not the case for the typical firm in the economy: precisely because there are many small firms in the economy, idiosyncratic fluctuations should cancel out. The upshot of this observation is that (*i*) fluctuations in large firms should show up as movements in the right tail of firm size distribution while (*ii*) the rest of the firm size distribution should be relatively stable over time.

Figure 4 plots counter-cumulative distributions (CCDF) of firm size *over time*, both in the model and in the binned BDS and Compustat data. The left panel of the Figure overlays 20000 CCDF's for firm size (measured by the number of employees), one for each sample period along a long simulation run of the model. The right panel displays all CCDF's associated with the BDS and Compustat data, running from 1977 to 2008.

[20]We do not address the literature debating whether large firms are more or less cyclical than small firms (see Moscarini and Postel-Vinay, 2012, Chari, Christiano and Kehoe, 2013, and Fort et al, 2013). The focus of this paper is on large firm dynamics over the business cycle; in order to keep this analysis as simple and transparent as possible we do not introduce frictions that are arguably important in capturing small firm dynamics.

Sample	Firms with more than	10k	15k	20k
Model	Correlation in level	−0.64	−0.57	−0.48
		(0.000)	(0.000)	(0.000)
	Correlation in growth rate	−0.41	−0.42	−0.44
		(0.000)	(0.000)	(0.000)
Data	Correlation in (HP filtered) level	−0.34	−0.51	−0.46
		(0.008)	(0.000)	(0.000)
	Correlation in growth rate	−0.33	−0.43	−0.38
		(0.011)	(0.001)	(0.004)

Table 4: Correlation of tail estimate with aggregate output.

NOTE: The tail in the model are estimated for simulated data for the baseline calibration (cf. Table 2) for an economy simulated during 20,000 periods. The tail are estimated on Compustat data over the period 1958-2008. The aggregate output data comes form the St-Louis Fed.

Consistent with our intuition above, variation (over time) in the firm size distribution is larger in the upper tail, both in model simulations and in the data.[21]

The question is now whether fluctuations in the right tail of the firm size distribution correlate with aggregate fluctuations. To perform this exercise in the model, we estimate the tail index generated by the model's firm size distribution. We do this for every period over a 20000 period simulation of the model's dynamics, under our baseline calibration. We then correlate this with the level of aggregate output in our model.

For the empirical counterpart to this correlation we use Compustat data only. While Compustat is not an accurate description of the population of U.S. firms, it contains detailed firm-level data that is particularly informative for large firms as these are more likely to be publicly listed. The number of large firm observations helps us to more accurately identify large firm dynamics and to better capture tail movements over time. From this data we obtain tail index estimates for each year between 1958 and 2008. We then correlate this with a measure of aggregate output growth for the corresponding year. Table 4 summarizes the results.

[21] As is clear from the Figure, the model's simulations imply a wider range of tail variation relative to the data. This is because the model's sample is much larger when compared to the relatively small number of bins in the data; hence sampling variation from extreme outcomes is more likely. Additionally, as a result of our conservative calibration strategy the number of very large firms is smaller than in the data and hence variability in the upper tail more pronounced. Recall that our aim in this section is not to match the amount of variability at the tail, rather to inspect the correlation of this variability with aggregate output.

For our baseline case, we choose to estimate tail indexes based on information for firms with more than 10000 employees, both in the model and in the data.[22]. The correlation between the tail estimate and the level of aggregate output in the model is negative (-0.64) and significant. The corresponding exercise in the data correlates the HP filtered aggregate output series and correlates this with the tail indexes estimated from Compustat. This correlation is again negative (-0.34) and significant. Intuitively, in periods when large firms suffer negative shocks, the tail index estimate is larger, i.e. the firm size distribution is less fat tailed. Both in the model and in the data, these periods coincide with below-trend performance at the aggregate level.

The remaining columns in the Table 4 present different robustness checks. First, we assess whether our results are sensitive to the cutoff choice when estimating tails. To do this, we re-estimate the empirical tail series in Compustat using larger scale cutoffs. The results are, if anything, stronger: the more we focus on the behavior of very large firms the stronger is the correlation in the data.[23] Our second robustness check, assesses whether this correlation survives a growth rate specification. The correlation between tail indices and aggregate output growth is again negative and does not depend on the particular cutoff chosen for the tail estimation.

Our model also has implications for the evolution of aggregate volatility over time. According to our discussion following Proposition 2 in the previous section, periods of high cross-sectional dispersion in firm size are periods of high volatility in aggregate output and aggregate TFP.

Our measure of cross-section dispersion in data is again sourced from Compustat. For each year, we compute the variance of (the logarithm of) real sales of manufacturing firms. We deflate the original nominal sales values in Compustat by the corresponding industry 4-digit price deflator from the NBER-CES Manufacturing database. For a measure of aggregate volatility, we follow Bloom et al (2014) and take the conditional volatility estimates from a GARCH(1,1) specification on aggregate TFP growth and aggregate GDP growth. Since both series contain low frequency movements we HP filter each series.

To obtain the model counterpart of the conditional volatility in aggregate output we make use of our analytical solution in Equation 11. We then correlate this series with

[22]For the model, we implement the estimator by Virkar and Clauset (2014) with a fixed cutoff. For the Compustat data, we follow Gabaix and Ibragimov (2011)

[23]In the Data Appendix we also show that this is robust to considering smaller cutoffs.

Sample	Aggregate Volatility	Dispersion of Real Sales	Dispersion of Employment
Model	Aggregate Volatility	0.9968 (0.000)	0.9983 (0.000)
Data	Aggr. Vol. in TFP growth	0.3461 (0.016)	0.2690 (0.065)
	Aggr. Vol. in GDP growth	0.2966 (0.041)	0.1782 (0.226)

Table 5: Correlation of Dispersion and Aggregate Volatility.

NOTE: In the model: aggregate volatility are computed using the Corollary 3. The model statistics are computed on a simulated sample of 20000 periods. In the data: we measure dispersion by computing cross-sectional variance. Variance of employment and real sales are computed using compustat data from 1960 to 2008 for manufacturing firms. Price are deflated using the NBER-CES Manufacturing Industry Database 4-digits price index. Aggregate volatility is measured by fitted values of an estimated GARCH on growth rate of TFP and Output that comes from Fernald (2014).

the cross-sectional variance of firm-level output and employment as given by the solution of the model. Both series are computed over a 20000 simulation of the model under our baseline calibration.

Table 5 summarizes the results. Both in the model and in the data, the cross-sectional variance of sales is positively correlated with aggregate volatility of TFP and GDP as our Proposition implies. Using the cross-sectional variance of employment also implies a positive (but weaker) correlation. In the Data Appendix we show that these findings are robust to considering other measures of cross-sectional dispersion, both at the firm level – using the interquartile range of real sales in Compustat – and at the establishment level – using either the productivity dispersion across establishments producing durable goods (from Kehrig, 2015, based on Census data) or the interquartile range of establishment level growth (from Bloom et al, 2014, based on Census data).

Taken together, the evidence in this section is consistent with the main predictions of our model: aggregate dynamics are driven by the dynamics of large firms and aggregate volatility follows movements in the dispersion of the cross-section of firms. Unlike the existing literature, our model delivers these predictions without resorting to aggregate shocks to first or second moments. Rather, aggregate output and aggregate volatility dynamics are the equilibrium outcome of micro-level dynamics.

5.4 Distributional Dynamics and the Business Cycle

The core of our argument is that the firm size distribution is a "sufficient statistic" for understanding fluctuations in aggregates. The aggregate state in our model *is* the firm size distribution. The evolution of this object over time determines aggregate output and its volatility. In the quantitative exercises above we have shown that certain moments of this distribution – its maximum in the impulse response analysis exercise and its tail and cross-sectional second moment in the previous subsection – do influence aggregates and that their quantitative impact is non-negligible.

In this section, we take this "sufficient statistic" argument one step further. Suppose we had access to a single time series object – the evolution of the firm size distribution over time as given by the BDS data. Using our calibrated model as an aggregation device we ask: what would be the implied history of aggregate fluctuations and volatility, based on this data alone?

To do this we use the expressions that aggregate the information in the firm size distribution and deliver aggregate productivity (Equation 1), aggregate output (Equation 3) and aggregate volatility (Equation 11) in the model.[24] As a result of this exercise, we obtain from our model three aggregate time series - for aggregate output, aggregate productivity and the volatility of aggregate output - whose time variation reflects movements in the size distribution over time *alone*. Figure 5 plots these three series (solid lines) against their HP filtered data counterparts (dashed lines).

As Figure 5 renders clear, when interpreted through the lenses of our model, the empirical evolution of the firm size distribution implies aggregate fluctuations that track well the historical fluctuations in aggregate data. In all three cases, the correlation between the implied and actual aggregates is positive and statistically significant. The estimates implied by the firm size distribution data track the evolution of aggregate

[24]To feed these expressions with data on the firm size distribution we must additionally resolve one issue: the model takes as a primitives the productivity grid Φ. Instead, the data on the firm size distribution by the BDS takes as primitives the bins of the firm size, which are fixed over time. Thus, given a size distribution at time t, we need to solve for the productivity bins in our model that are consistent with the observed size distribution. This can be easily obtained by using the optimality condition with respect to employment and the labor market clearing condition. To be precise, denote the information in the data by n_s^{BDS}, the observed number of firms with employment in bin s. We then solve for the productivity grid Φ_t^{BDS} and wage rate w_t^{BDS} such that $(i)\forall s, \quad \varphi_{s,t}^{BDS} = \frac{w_t^{BDS}}{\alpha}(n_s^{BDS})^{1-\alpha}$ and $(ii) w_t^{BDS} = \left(\alpha^{\frac{1}{1-\alpha}} \frac{\sum_{s=1}^{S^{BDS}} \left(\varphi_{s,t}^{BDS}\right)^{\frac{1}{1-\alpha}} \mu_{s,t}^{BDS}}{M} \right)^{\frac{1-\alpha}{\gamma(1-\alpha)+1}}$.

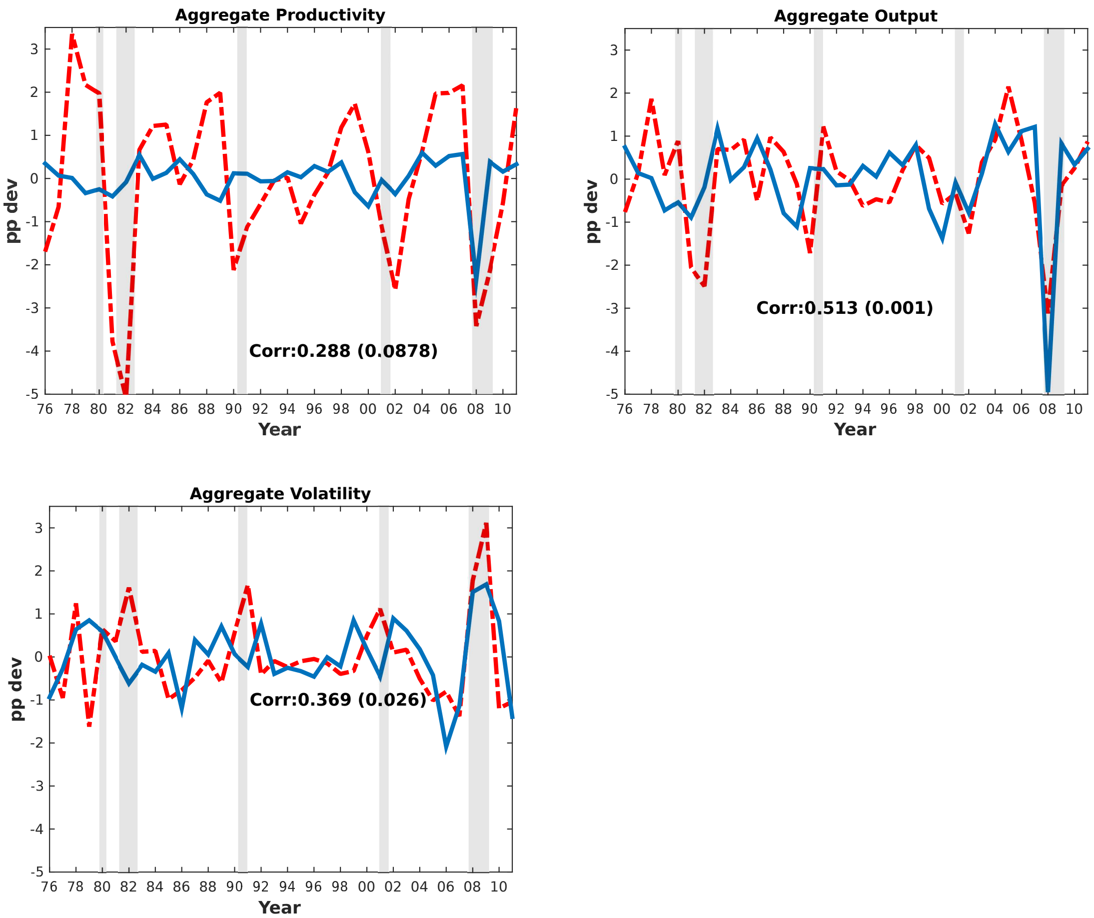

Figure 5: Dynamics of aggregate variable from firm size distribution data (BDS).
NOTE: The blue solid lines give the evolution of the aggregate series of our model when we use the BDS distribution of firm size and Equations 1, 3 and 29 to compute aggregate productivity, output and aggregate volatility. The red dashed line are the aggregate series in the data: aggregate TFP and output are taken from Fernald 2014. For aggregate volatility dynamics, we estimate a GARCH(1,1) on GDP growth following Bloom et al. (2014).

output (in deviations from trend) particularly well with a highly significant correlation of 0.533. For aggregate productivity and the conditional volatility of aggregate TFP, this correlation is about 1/3 and noisier, which mostly reflects a weaker correlation in the earlier part of the sample. Conversely, the data on the firm size distribution implies remarkably accurate aggregate dynamics in the period leading up to, during and after the Great Recession.

In summary, coupling the information contained in the dynamics of the firm size distribution with our quantitative model delivers a history of aggregate fluctuations that is not unlike what we observe in the aggregate data.

6 Conclusion

A small number of firms accounts for a substantial share of aggregate economic activity. This opens the possibility of doing away with aggregate shocks, instead tracing back the origins of aggregate fluctuations to large firm dynamics. We build a quantitative firm dynamics model in which we cast this hypothesis.

The first part of our analysis characterizes, analytically, the law of motion of the firm size distribution and shows that the implied aggregate output and productivity dynamics are persistent, volatile and exhibit time-varying second moments.

In the second part of the paper, we explore quantitatively and in the data, the role of the firm size distribution – and, in particular, that of large firm dynamics – in shaping aggregate fluctuations. Taken together, our results imply that a large fraction of aggregate dynamics can be rationalized by large firm dynamics.

The results in this paper suggest at least two fruitful ways of extending our analysis. First, while we have intentionally focused our attention on large firm dynamics as drivers of aggregate fluctuations, a complete analysis of firm dynamics over the business cycle should also match those of small firms. This surely implies moving away from the frictionless environment presented here and understanding how small firm dynamics are distorted by adjustment costs, credit constraints and other frictions. The fact that Hopenhayn's (1992) framework is tractable enough to handle such frictions, render our analysis easily extensible to such environments.

Second, in our framework, a firm does not internalize its own effect on aggregate prices and factor costs. In other models of firms dynamics, the assumption that the law of large numbers holds, justifies thinking about firms as infinitesimal price takers. However, we have shown that, in a standard firm dynamics framework, large firms have a non-negligible effect on aggregates. Thus, the price taker assumption should be taken carefully. This points to generalizations of our setup that do away with this assumption and take on market structure and market power as further determinants of aggregate dynamics.

References

[1] Daron Acemoglu, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016, 09 2012.

[2] Daron Acemoglu and Martin K. Jensen. Robust comparative statics in large dynamic economies. *Journal of Political Economy*, Forthcoming, 2015.

[3] Robert L. Axtell. Zipf distribution of u.s. firm sizes,. *Science 7*, 293(5536):1818–1820, September 2001.

[4] Susanto Basu and John G Fernald. Returns to scale in u.s. production: Estimates and implications. *Journal of Political Economy*, 105(2):249–83, April 1997.

[5] Florin O. Bilbiie, Fabio Ghironi, and Marc J. Melitz. Endogenous entry, product variety, and business cycles. *Journal of Political Economy*, 120(2):304 – 345, 2012.

[6] Nicholas Bloom, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J. Terry. Really Uncertain Business Cycles. Working Papers 14-18, Center for Economic Studies, U.S. Census Bureau, March 2014.

[7] Jeffrey R. Campbell. Entry, exit, embodied technology, and business cycles. *Review of Economic Dynamics*, 1(2):371–408, April 1998.

[8] Jeffrey R. Campbell and Jonas D.M. Fisher. Idiosyncratic risk and aggregate employment dynamics. *Review of Economic Dynamics*, 7(2):331 – 353, 2004.

[9] Vasco Carvalho and Xavier Gabaix. The Great Diversification and Its Undoing. *American Economic Review*, 103(5):1697–1727, August 2013.

[10] Rui Castro, Gian Luca Clementi, and Yoonsoo Lee. Cross-sectoral variation in firm-level idiosyncratic risk. *Journal of Industrial Economics*, Forthcoming.

[11] David G. Champernowne. A model of income distribution. *Economic Journal*, 63(250):318–351, June 1953.

[12] Varadarajan V. Chari, Lawrence J. Christiano, and Patrick J. Kehoe. The gertler-gilchrist evidence on small and large firm sales. mimeo, July 2013.

[13] Gian Luca Clementi and Berardino Palazzo. Entry, exit, firm dynamics, and aggregate fluctuations. Working Paper 19217, National Bureau of Economic Research, July 2015.

[14] Gina Luca Clementi and Hugo A Hopenhayn. A Theory of Financing Constraints and Firm Dynamics. *The Quarterly Journal of Economics*, 121(1):229–265, 02 2006.

[15] Diego A. Comin and Thomas Philippon. The rise in firm-level volatility: Causes and consequences. In *NBER Macroeconomics Annual 2005, Volume 20*, NBER Chapters, pages 167–228. National Bureau of Economic Research, Inc, October 2006.

[16] Juan Carlos Córdoba. A generalized gibrat's law. *International Economic Review*, 49(4):1463–1468, November 2008.

[17] Steven J. Davis, John Haltiwanger, Ron Jarmin, and Javier Miranda. Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms. In *NBER Macroeconomics Annual 2006, Volume 21*, NBER Chapters, pages 107–180. National Bureau of Economic Research, Inc, October 2007.

[18] Wouter J. Den Haan and Pontus Rendahl. Solving the incomplete markets model with aggregate uncertainty using explicit aggregation. *Journal of Economic Dynamics and Control*, 34(1):69–78, January 2010.

[19] Julian di Giovanni and Andrei A. Levchenko. Country size, international trade, and aggregate fluctuations in granular economies. *Journal of Political Economy*, 120(6):1083 – 1132, 2012.

[20] Julian di Giovanni, Andrei A. Levchenko, and Isabelle Mejean. Firms, Destinations, and Aggregate Fluctuations. *Econometrica*, 82(4):1303–1340, 07 2014.

[21] Julian di Giovanni, Andrei A. Levchenko, and Isabelle Mejean. Granular comovment. Technical report, December 2014.

[22] Timothy Dunne, Mark J. Roberts, and Larry Samuelson. Patterns of firm entry and exit in u.s. manufacturing industries. *The RAND Journal of Economics*, 19(4):pp. 495–515, 1988.

[23] David S Evans. The Relationship between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries. *Journal of Industrial Economics*, 35(4):567–81, June 1987.

[24] John Fernald. A quarterly, utilization-adjusted series on total factor productivity. Working Paper Series 2012-19, Federal Reserve Bank of San Francisco, 2014.

[25] Teresa C Fort, John Haltiwanger, Ron S Jarmin, and Javier Miranda. How Firms Respond to Business Cycles: The Role of Firm Age and Firm Size. *IMF Economic Review*, 61(3):520–559, August 2013.

[26] Lucia Foster, John Haltiwanger, and Chad Syverson. Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review*, 98(1):394–425, March 2008.

[27] Xavier Gabaix. The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772, 05 2011.

[28] Xavier Gabaix and Rustam Ibragimov. Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents. *Journal of Business & Economic Statistics*, 29(1):24–39, 2011.

[29] Bronwyn H Hall. The Relationship between Firm Size and Firm Growth in the U.S. Manufacturing Sector. *Journal of Industrial Economics*, 35(4):583–606, June 1987.

[30] Hugo A Hopenhayn. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, 60(5):1127–50, September 1992.

[31] JPMorgan. Can one little phone impact gdp? Michael Feroli, North America Economic Research, `https://mm.jpmorgan.com/EmailPubServlet?doc=GPS-938711-0.html&h=-825pgod`, 2012.

[32] Matthias Kehrig. The cyclical nature of the productivity distribution. January 2015.

[33] Aubhik Khan and Julia K. Thomas. Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica*, 76(2):395–436, 03 2008.

[34] Per Krusell and Anthony A. Smith. Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896, October 1998.

[35] Yoonsoo Lee. The importance of reallocations in cyclical productivity and returns to scale: evidence from plant-level data. Working Paper 0509, Federal Reserve Bank of Cleveland, 2005.

[36] Yoonsoo Lee and Toshihiko Mukoyama. Entry, exit and plant-level dynamics over the business cycle. Working Paper 0718, Federal Reserve Bank of Cleveland, 2008.

[37] Erzo G. J. Luttmer. Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics*, 122(3):1103–1144, 08 2007.

[38] Erzo G.J. Luttmer. Models of growth and firm heterogeneity. *Annual Review of Economics*, 2(1):547–576, 09 2010.

[39] Erzo G.J. Luttmer. Technology diffusion and growth. *Journal of Economic Theory*, 147(2):602–622, 2012.

[40] Mario J. Miranda and Paul L. Fackler. *Applied Computational Economics and Finance*, volume 1 of *MIT Press Books*. The MIT Press, June 2004.

[41] Giuseppe Moscarini and Fabien Postel-Vinay. The contribution of large and small employers to job creation in times of high and low unemployment. *American Economic Review*, 102(6):2509–39, October 2012.

[42] Mark E. J. Newman. Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46:323–351, 2005.

[43] Richard Rogerson and Johanna Wallenius. Micro and macro elasticities in a life cycle model with taxes. *Journal of Economic Theory*, 144(6):2277–2292, November 2009.

[44] Esteban Rossi-Hansberg and Mark L. J. Wright. Establishment Size Dynamics in the Aggregate Economy. *American Economic Review*, 97(5):1639–1666, December 2007.

[45] Herbert A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3-4):425–440, 1955.

[46] Chad Syverson. Product Substitutability and Productivity Dispersion. *The Review of Economics and Statistics*, 86(2):534–550, May 2004.

[47] Marcelo L. Veracierto. Plant-level irreversible investment and equilibrium business cycles. *American Economic Review*, 92(1):181–197, March 2002.

[48] Yogesh Virkar and Aaron Clauset. Power-law distributions in binned empirical data. *Annals of Applied Statistics*, 8(1):89–119, 2014.

Online Appendix to "Large Firm Dynamics and the Business Cycle"

Vasco M. Carvalho and Basile Grassi

Not for Publication. March 2015

A Proof Appendix

In this appendix, we first solve for the stationary distribution under random growth described in Corollary 2 (Appendix A.1). We then find the asymptotic value of the ratio between the number of incumbents and the number of potential entrants, when the former goes to infinity (Appendix A.2). This intermediate result will be used in the rest of this appendix. In Appendix A.3, we prove Theorem 2 describing the dynamics of the aggregate state T_t. In Appendix A.4, we solve for the dynamics of aggregate output up to a first order approximation as in Corollary 3. Finally, in Appendix A.5 and A.6, we prove Proposition 1 and 2 containing our comparative statics' results for aggregate persistence and volatility respectively.

A.1 Proof of Corollary 2

In this appendix, we prove that the productivity stationary distribution is a mixture of two distributions: (i) the stationary distribution associated with the Markovian firm-level productivity process and (ii) the distribution of entrants. These are respectively weighted by the constants K_1 and K_2. Formally, we show that $K_1 = -\frac{c}{a}\frac{(\varphi^{\delta_e}-1)(\varphi^{-\delta_e})^{s^*}}{(1-\varphi^{-\delta_e})(a\varphi^{-\delta_e}-c)}$ and $K_2 = \frac{(\varphi^{\delta_e}-1)(a(\varphi^{-\delta_e})^2+b\varphi^{-\delta_e}+c)}{a(\varphi^{-\delta_e})^2-\varphi^{-\delta_e}(a+c)+c}(\varphi^{s^*})^{-\delta_e}$. In the Corollary in the main text, we only reported the value of the productivity stationary distribution for productivity levels above the exit threshold. In this appendix, for completeness, we describe this distribution over the full idiosyncratic state-space. We then show that:

$$\hat{\mu}_s = \begin{cases} -\frac{c}{a}\frac{\left(\varphi^{\delta_e}-1\right)(\varphi^{-\delta_e})^{s^*}}{(1-\varphi^{-\delta_e})(a\varphi^{-\delta_e}-c)}\left(\frac{\varphi^s}{\varphi^{s^*}}\right)^{-\delta} + \frac{\left(\varphi^{\delta_e}-1\right)\left(a(\varphi^{-\delta_e})^2+b\varphi^{-\delta_e}+c\right)}{a(\varphi^{-\delta_e})^2-\varphi^{-\delta_e}(a+c)+c}(\varphi^s)^{-\delta_e} & \text{if } s \geq s^* \\ a\left(\varphi^{\delta_e}-1\right)\left(\frac{-c/a}{(1-\varphi^{-\delta_e})(a\varphi^{-\delta_e}-c)} + \frac{a(\varphi^{-\delta_e})^2+b\varphi^{-\delta_e}+c}{a(\varphi^{-\delta_e})^2-(a+c)\varphi^{-\delta_e}+c}+1\right)(\varphi^{s^*})^{-\delta_e} & \text{if } i = s^*-1 \\ 0 & \text{if } s < s^*-1 \end{cases}$$

with $\delta = \frac{\log(a/c)}{\log(\varphi)}$.

The proof of this Corollary follows two steps: (i) we first solve closed-form for the stationary distribution given a maximum level of productivity φ^S under Assumption 1; (ii) we then take the limit of this distribution when the number of firms goes to infinity under the Assumption 2.

Step 1: for a given S and a given number of firms

This step consists of solving for the stationary distribution given a fixed maximum level of productivity S. Formally, in this step, we will show that the following Lemma holds true.

Lemma 1 *For a given S, if (i) the entrant distribution is Pareto (i.e $G_s = K_e\left(\varphi^s\right)^{-\delta_e}$) and (ii) the productivity process follows Gibrat's law (Assumption 1) with parameters a and c on the grid defined by φ, then the stationary distribution (i.e when $\mathbb{V}ar_t\epsilon_{t+1} = 0$) is:*
For $s^* \le s \le S$:

$$\mu_s = \mathbb{P}\{\varphi = \varphi^s\} = MK_eC_1\left(\frac{\varphi^s}{\varphi^{s^*}}\right)^{-\delta} + MK_eC_2\left(\varphi^s\right)^{-\delta_e} + MK_eC_3$$

and $\mu_{s^*-1} = a\left(\mu_{s^*} + MK_e\left(\varphi^{s^*}\right)^{-\delta_e}\right)$ and $\mu_s = 0$ for $s < s^* - 1$.

Where $\delta = \frac{\log(a/c)}{\log(\varphi)}$ and C_1, C_2, C_3 are constants, independent of s, and where $C_1 = \frac{c\left(a(\varphi^{-\delta_e})^{S+2} - a(\varphi^{-\delta_e})^{s^*} - c(\varphi^{-\delta_e})^{S+3} + c(\varphi^{-\delta_e})^{s^*}\right)}{a(1-\varphi^{-\delta_e})(a-c)(a\varphi^{-\delta_e}-c)}$, $C_2 = \frac{(a(\varphi^{-\delta_e})^2 + b\varphi^{-\delta_e} + c)}{(a(\varphi^{-\delta_e})^2 - \varphi^{-\delta_e}(a+c) + c)}$ and $C_3 = \frac{-(\varphi^{-\delta_e})^{S+1}}{(1-\varphi^{-\delta_e})(a-c)}$.

Proof: To find the stationary distribution of the Markovian process described by the transition matrix P, we need to solve for μ in $\mu = (P_t^*)'(\mu + MG)$ where P is given by assumption 1 and where μ is the $(S \times 1)$ vector $(\mu_1, \ldots, \mu_S)'$. For simplicity, we assume $M = 1$.

The matrix Equation $\mu = (P_t^*)'(\mu + MG)$ can be equivalently written as the following system of Equations:

For $s < s^* - 1$:

$$\mu_s = 0 \tag{14}$$

For $s = s^* - 1$:

$$\mu_{s^*-1} = a(\mu_{s^*} + G_{s^*}) \tag{15}$$

For $s = s^*$:

$$\mu_{s^*} = b(\mu_{s^*} + G_{s^*}) + a(\mu_{s^*+1} + G_{s^*+1}) \tag{16}$$

For $s = S$:

$$\mu_S = c(\mu_{S-1} + G_{S-1}) + (b + c)(\mu_S + G_S) \tag{17}$$

For $s^* + 1 \le s \le S - 1$:

$$\mu_s = c(\mu_{s-1} + G_{s-1}) + b(\mu_s + G_s) + a(\mu_{s+1} + G_{s+1}) \tag{18}$$

The system of Equations 16, 17 and 18 gives a linear second difference Equation with two boundary conditions. The system has a exogenous term given by the distribution of entrants G. For this system, we define the associated homogeneous system by the same Equations with $G_s = 0, \forall s$. To solve for a linear second order difference Equation, we follow four steps: (i) Solve for the general solution of the homogeneous system; these solutions are parametrized by two constants (ii) Find one particular solution for the full system (iii) The general solution of the full system is then given by the sum of the general solution of the homogeneous system and the particular solution we have found (iv) Solve for the undetermined coefficient using the boundary conditions.

The recurrence Equation of the homogeneous system is equivalent to $c\mu_{s-1} - (a + c)\mu_s + a\mu_{s+1} = 0$ since $b = 1 - a - c$. To find the general solution of this Equation, let us solve for the root of the polynomial $aX^2 - (a+c)X + c$. This polynomial is equal to $a(X - c/a)(X - 1)$ and thus its roots are $r_1 = c/a$ and 1. The general solution of the homogeneous system associated to Equation 18 is then $\mu_s = A(c/a)^s + B$ where A and B are constants.

Using the form of the entrant distribution $G_s = K_e(\varphi^{-\delta_e})^s$, and assuming that $\varphi^{-\delta_e} \neq \frac{a}{c}$, a particular solution is $K_e\frac{a(\varphi^{-\delta_e})^2 + b\varphi^{-\delta_e} + c}{a(\varphi^{-\delta_e})^2 - (a+c)\varphi^{-\delta_e} + c}(\varphi^{-\delta_e})^s$.

The general solution of the second order linear difference Equation is then $A(c/a)^s + B + K_e\frac{a(\varphi^{-\delta_e})^2+b\varphi^{-\delta_e}+c}{a(\varphi^{-\delta_e})^2-(a+c)\varphi^{-\delta_e}+c}(\varphi^{-\delta_e})^s$. By substituting this general solution in the boundary condition 16 and 17, we find $A = K_e\left(\frac{c}{a}\right)^{-s^*}\frac{c\left(a(\varphi^{-\delta_e})^{S+2}-a(\varphi^{-\delta_e})^{s^*}-c(\varphi^{-\delta_e})^{S+3}+c(\varphi^{-\delta_e})^{s^*}\right)}{a(1-\varphi^{-\delta_e})(a-c)(a\varphi^{-\delta_e}-c)}$ and $B = K_e\frac{-(\varphi^{-\delta_e})^{S+1}}{(1-\varphi^{-\delta_e})(a-c)}$

Since the s^{th} productivity level is φ^s, then $s = \frac{\log\varphi^s}{\log\varphi}$ and thus $\left(\frac{c}{a}\right)^s = (\varphi^s)^{-\frac{\log a/c}{\log\varphi}}$. Let us define $\delta = \frac{\log a/c}{\log\varphi}$. The expression of the stationary distribution is then:

$$\mu_s = K_eC_1\left(\frac{\varphi^s}{\varphi^{s^*}}\right)^{-\delta} + K_eC_2\left(\varphi^s\right)^{-\delta_e} + K_eC_3 \tag{19}$$

for $s^* \leq s \leq S$. The value of μ_{s^*-1} is given by 15 and $\forall s < s^*-1, \mu_s = 0$. $\square$

Step 2: Taking the number of firms to infinity

The second step of the proof consists of finding the limit of constants K_e, C_1, C_2 and C_3 as the number of firms N goes to infinity. After finding these limits, we take the limit of Equation 19 in the previous Lemma.

Let us first describe the asymptotic behavior of K_e. Recall that the entrant distribution sums to one.[25]

$$1 = \sum_{s=1}^{S} G_s = K_e\sum_{s=1}^{S}\left(\varphi^s\right)^{-\delta_e} = K_e\sum_{s=1}^{S}\left(\varphi^{-\delta_e}\right)^s = K_e\frac{\varphi^{-\delta_e} - \left(\varphi^{-\delta_e}\right)^{S+1}}{1-\varphi^{-\delta_e}}$$

Rearranging terms, it follows that

$$K_e = \frac{1-\varphi^{-\delta_e}}{\varphi^{-\delta_e} - \left(\varphi^{-\delta_e}\right)^{S+1}}$$

Under assumption 2 and since $\delta_e, \delta > 0$ we have

$$\left(\varphi^{-\delta_e}\right)^S = \left(\varphi^S\right)^{-\delta_e} = \left(ZN^{1/\delta}\right)^{-\delta_e} = Z^{-\delta_e}N^{-\delta_e/\delta} \underset{N\to\infty}{\longrightarrow} 0$$

by applying these results to the expression for K_e, it follows that $K_e \underset{N\to\infty}{\longrightarrow} \varphi^{\delta_e} - 1$.

Let us now focus on the asymptotic behavior of C_3, C_2 and C_1. From Corollary 1, we have

$$C_3 = \frac{-(\varphi^{-\delta_e})^{S+1}}{(1-\varphi^{-\delta_e})(a-c)} = \frac{-\varphi^{-\delta_e}(\varphi^S)^{-\delta_e}}{(1-\varphi^{-\delta_e})(a-c)} = \frac{-\varphi^{-\delta_e}Z^{-\delta_e}}{(1-\varphi^{-\delta_e})(a-c)}N^{-\delta_e/\delta} \underset{N\to\infty}{\longrightarrow} 0$$

We also have that

$$C_2 := \frac{\left(a(\varphi^{-\delta_e})^2 + b\varphi^{-\delta_e} + c\right)}{(a(\varphi^{-\delta_e})^2 - \varphi^{-\delta_e}(a+c) + c)}$$

which is independent of S and thus of N.

[25]The way we define the model, we assume that G sums to one. We also assume that the number of potential entrants in bin s is MG_s, so that the total number of potential entrants is M.

Finally, we have

$$\begin{aligned}C_1 =&\frac{c\left(a(\varphi^{-\delta_e})^{S+2} - a(\varphi^{-\delta_e})^{s^*} - c(\varphi^{-\delta_e})^{S+3} + c(\varphi^{-\delta_e})^{s^*}\right)}{a(1-\varphi^{-\delta_e})(a-c)(a\varphi^{-\delta_e}-c)}\\ \underset{N\to\infty}{\longrightarrow}&\frac{c\left(-a(\varphi^{-\delta_e})^{s^*} + c(\varphi^{-\delta_e})^{s^*}\right)}{a(1-\varphi^{-\delta_e})(a-c)(a\varphi^{-\delta_e}-c)} = \frac{c}{a}\frac{-(a-c)(\varphi^{-\delta_e})^{s^*}}{(1-\varphi^{-\delta_e})(a-c)(a\varphi^{-\delta_e}-c)}\end{aligned}$$

and therefore

$$C_1 \underset{N\to\infty}{\longrightarrow} C_1^\infty := \frac{c}{a}\frac{(\varphi^{-\delta_e})^{s^*}}{(1-\varphi^{-\delta_e})(c-a\varphi^{-\delta_e})}$$

We have just found the limit of K_e, C_1, C_2 and C_3 when N goes to infinity. We then apply these results to the stationary distribution by taking N to infinity. According to Lemma 1, we have for $s^* \leq s \leq S$:

$$\frac{\mu_s}{M} = K_eC_1\left(\frac{\varphi^s}{\varphi^{s^*}}\right)^{-\delta} + K_eC_2(\varphi^s)^{-\delta_e} + K_eC_3$$

Under assumption 2, we have just shown that when the number of firms, N, goes to infinity, the stationary distribution is given by:

$$\frac{\mu_s}{M} = \left(\varphi^{\delta_e}-1\right)\frac{c}{a}\frac{(\varphi^{-\delta_e})^{s^*}}{(1-\varphi^{-\delta_e})(c-a\varphi^{-\delta_e})}\left(\frac{\varphi^s}{\varphi^{s^*}}\right)^{-\delta} + \left(\varphi^{\delta_e}-1\right)\frac{\left(a(\varphi^{-\delta_e})^2 + b\varphi^{-\delta_e}+c\right)}{(a(\varphi^{-\delta_e})^2 - \varphi^{-\delta_e}(a+c)+c)}(\varphi^s)^{-\delta_e}$$

□

A.2 Intermediate result: the link between the number of incumbents N and the number of potential entrants M

In this appendix, we are interested in the relationship between the number of incumbents N, the number of potentials entrants M, and the value of their ratio when N goes to infinity. We show that as N goes to infinity, the ratio M/N goes to a constant. This means that taking the endogenous variable N or the exogenous parameter M to infinity is strictly equivalent.

The number of firms is simply the sum of the number of firms in each bin:

$$\begin{aligned}N =&\sum_{s=1}^{S}\mu_s = \mu_{s^*-1} + \sum_{s=s^*}^{S}\mu_s\\ =&a\left(MK_eC_1 + MK_eC_2(\varphi^{s^*})^{-\delta_e} + MK_eC_3 + MK_e(\varphi^{s^*})^{-\delta_e}\right) + MK_eC_3\sum_{s=s^*}^{S}1\\ &+ MK_eC_1(\varphi^{s^*})^{\delta}\sum_{s=s^*}^{S}(\varphi^s)^{-\delta} + MK_eC_2\sum_{s=s^*}^{S}(\varphi^s)^{-\delta_e}\\ =&a\left(MK_eC_1 + MK_eC_2(\varphi^{s^*})^{-\delta_e} + MK_eC_3 + MK_e(\varphi^{s^*})^{-\delta_e}\right) + MK_eC_3(S-s^*+1)\\ &+ MK_eC_1(\varphi^{s^*})^{\delta}\frac{\left(\varphi^{-\delta}\right)^{s^*} - \left(\varphi^{-\delta}\right)^{S}}{1-\varphi^{-\delta}} + MK_eC_2\frac{\left(\varphi^{-\delta_e}\right)^{s^*} - \left(\varphi^{-\delta_e}\right)^{S}}{1-\varphi^{-\delta_e}}\end{aligned}$$

thus, by dividing both side by M, we have

$$\frac{N}{M} = a\left(K_eC_1 + K_e(C_2+1)(\varphi^{s^*})^{-\delta_e} + K_eC_3\right) + K_eC_3(S-s^*+1) + K_eC_1(\varphi^{s^*})^{\delta}\frac{\left(\varphi^{-\delta}\right)^{s^*} - \left(\varphi^{-\delta}\right)^{S}}{1-\varphi^{-\delta}} + K_eC_2\frac{\left(\varphi^{-\delta_e}\right)^{s^*} - \left(\varphi^{-\delta_e}\right)^{S}}{1-\varphi^{-\delta_e}}$$

Let us note that under assumption 2

$$\left(\varphi^{-\delta}\right)^S = \left(\varphi^S\right)^{-\delta} = \left(ZN^{1/\delta}\right)^{-\delta} = Z^{-\delta}N^{-1} \underset{N\to\infty}{\longrightarrow} 0$$

and that, since $S = \frac{1}{\log\varphi}(\log Z + \frac{1}{\delta}\log N)$, we have

$$SC_3 = \frac{1}{\log\varphi}(\log Z + \frac{1}{\delta}\log N)\frac{-\varphi^{-\delta_e}Z^{-\delta_e}}{(1-\varphi^{-\delta_e})(a-c)}N^{-\delta_e/\delta} \underset{N\to\infty}{\longrightarrow} 0$$

Thus, we have that

$$\begin{aligned}\frac{N}{M} \underset{N\to\infty}{\longrightarrow} (E^\infty)^{-1} :=& a\left((\varphi^{\delta_e}-1)C_1^\infty + (\varphi^{\delta_e}-1)(C_2+1)(\varphi^{s^*})^{-\delta_e}\right)\\ &+ (\varphi^{\delta_e}-1)C_1^\infty\frac{1}{1-\varphi^{-\delta}} + (\varphi^{\delta_e}-1)C_2\frac{\left(\varphi^{-\delta_e}\right)^{s^*}}{1-\varphi^{-\delta_e}}\end{aligned}$$

where E^∞ is the ratio of the number of potential entrants M and the number of incumbents, when there is an infinite number of incumbents. The last Equation shows that M and N are equivalent when the number of incumbents is large. Thus, taking N to infinity is the same as taking M to infinity i.e $M \underset{N\to\infty}{\sim} E^\infty N$.

A.3 Proof of Theorem 2

In this appendix, we prove Theorem 2. To do so, we use Theorem 1, where we have $\mu_{t+1} = (P_t^*)'(\mu_t + MG) + \epsilon_{t+1}$. We are interested in the law of motion of $T_t = B'\mu_t$ where $B = \{b_i\}_i$ is a vector with $b_i = (\varphi^{\frac{1}{1-\alpha}})^i$. This law of motion is derived directly from the law of motion of the productivity distribution described in Theorem 1. To prove Theorem 2, we then proceed in two steps. First, we show that $\mathbb{E}_t T_{t+1}$ is a function of T_t and the primitives of the model. Second, we solve for the conditional variance of T_{t+1}.

Formally, we assume that idiosyncratic productivities follow Gibrat's law (Assumption 1). We also assume that G, the distribution of potential entrants, is Pareto distributed with tail index δ_e, i.e we have $G_s = K_e(\varphi^{-\delta_e})^s$. In the first step, we show the following Equation for the law of motion of T_t:

$$\begin{aligned} T_{t+1} =& \rho T_t && \text{main dynamics}\\ &+ \rho M K_e\frac{(\varphi^{-\delta_e}\varphi^{\frac{1}{1-\alpha}})^{S+1} - (\varphi^{-\delta_e}\varphi^{\frac{1}{1-\alpha}})^{s_t}}{(\varphi^{-\delta_e}\varphi^{\frac{1}{1-\alpha}})-1} && \text{contribution of entrant}\\ &- \rho(\varphi^{\frac{1}{1-\alpha}})^{s_t-1}\mu_{s_t-1,t} && \text{contribution of exit}\\ &- c(\varphi^{\frac{1}{1-\alpha}}-1)(\varphi^{\frac{1}{1-\alpha}})^S\mu_{S,t} - c(\varphi^{\frac{1}{1-\alpha}}-1)M(\varphi^{-\delta_e}\varphi^{\frac{1}{1-\alpha}})^S && \text{upper reflecting barrier}\\ &+ \sigma_t\varepsilon_{t+1} && \text{aggregate uncertainty} \end{aligned} \tag{20}$$

where $\mathbb{E}[\varepsilon_{t+1}] = 0$, $\mathbb{V}ar[\varepsilon_{t+1}] = 1$ and $\rho = a\varphi^{-1/(1-\alpha)} + b + c\varphi^{1/(1-\alpha)}$.

The second step consists in finding an expression for the conditional variance of T_{t+1}. We show that the time-varying volatility σ_t^2 is:

$$
\begin{aligned}
\sigma_t^2 = \mathbb{V}ar_t[B'\epsilon_{t+1}] = & \varrho D_t \\
& + \varrho M K_e \frac{(\varphi^{-\delta_e}\varphi^{\frac{2}{1-\alpha}})^{S+1} - (\varphi^{-\delta_e}\varphi^{\frac{2}{1-\alpha}})^{s_t}}{(\varphi^{-\delta_e}\varphi^{\frac{2}{1-\alpha}}) - 1} \\
& - \varrho(\varphi^{\frac{2}{1-\alpha}})^{s_t-1}\mu_{s_t-1,t} \\
& - (\varrho - \varrho')MK_e(\varphi^{-\delta_e}(\varphi^{\frac{1}{1-\alpha}})^2)^S - (\varrho - \varrho')(\varphi^{\frac{1}{1-\alpha}})^{2S}\mu_{S,t}
\end{aligned}
\tag{21}
$$

where $D_t = \sum_{s=s_t-1}^{S}\left((\varphi^s)^{\frac{1}{1-\alpha}}\right)^2 \mu_{s,t}$, and $\varrho = a\varphi^{-2/(1-\alpha)} + b + c\varphi^{2/(1-\alpha)} - \rho^2$.

Step 1: Proof of Equation 20

Here we prove Equation 20. To do so, we first consider the term $B'm(\mu_t) = B'(P_t^*)'(\mu_t + MG)$ and express it as a function of T_t and the primitives of the model. The transition P matrix is

$$
P = \begin{pmatrix}
a+b & c & 0 & 0 & \cdots & \cdots & 0 & 0 \\
a & b & c & 0 & \cdots & \cdots & 0 & 0 \\
0 & a & b & c & \cdots & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & 0 & \cdots & a & b & c \\
0 & 0 & 0 & 0 & \cdots & 0 & a & b+c
\end{pmatrix}
$$

because of entry/exit we have

$$
P_t^* = \begin{pmatrix}
0 & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
0 & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
0 & \cdots & a & b & c & \cdots & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & a & b & c \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & a & b+c
\end{pmatrix}
$$

where the first non-zero row is the row indexed by s_t. Let us take the transpose of this last matrix.

$$
(P_t^*)' = \begin{pmatrix}
0 & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
0 & \cdots & 0 & a & 0 & \cdots & \cdots & 0 & 0 \\
0 & \cdots & 0 & b & a & \cdots & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & c & b & a \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & c & b+c
\end{pmatrix}
$$

Now the first non-zero row is the one index by $s_t - 1$. Let us look at an element of the $(S \times 1)$ vector $(P_t^*)'\mu_t$.

$$
[(P_t^*)'\mu_t]_i = \sum_{k=1}^{S}[(P_t^*)']_{i,k}\mu_{k,t} = \begin{cases}
0 & \text{if } i < s_t - 1 \\
a\mu_{s_t,t} & \text{if } i = s_t - 1 \\
b\mu_{s_t,t} + a\mu_{s_t+1,t} & \text{if } i = s_t \\
c\mu_{i-1,t} + b\mu_{i,t} + a\mu_{i+1,t} & \text{if } s_t < i < S \\
c\mu_{S-1,t} + (b+c)\mu_{S,t} & \text{if } i = S
\end{cases}
$$

Now let us pre-multiply this $(S \times 1)$ vector by the $(1 \times S)$ vector B':

$$
\begin{aligned}
B'(P_t^*)'\mu_t =& \sum_{i=1}^{S} b_i[(P_t^*)'\mu_t]_i \\
=& b_{s_t-1}a\mu_{s_t,t} + b_{s_t}(b\mu_{s_t,t} + a\mu_{s_t+1,t}) \\
&+ \sum_{i=s_t+1}^{S-1} b_i(c\mu_{i-1,t} + b\mu_{i,t} + a\mu_{i+1,t}) \\
&+ b_S(c\mu_{S-1,t} + (b+c)\mu_{S,t})
\end{aligned}
$$

Focusing on the third term of the second equality:

$$
\begin{aligned}
\sum_{i=s_t+1}^{S-1} b_i(c\mu_{i-1,t} + b\mu_{i,t} + a\mu_{i+1,t}) &= c\sum_{i=s_t+1}^{S-1} b_i\mu_{i-1,t} + b\sum_{i=s_t+1}^{S-1} b_i\mu_{i,t} + a\sum_{i=s_t+1}^{S-1} b_i\mu_{i+1,t} \\
&= c\sum_{i=s_t}^{S-2} b_{i+1}\mu_{i,t} + b\sum_{i=s_t+1}^{S-1} b_i\mu_{i,t} + a\sum_{i=s_t+2}^{S} b_{i-1}\mu_{i,t}
\end{aligned}
$$

Since $b_i = (\varphi^{1/(1-\alpha)})^i$ then $b_i = \varphi^{1/(1-\alpha)}b_{i-1}$, we have

$$
\begin{aligned}
\sum_{i=s_t+1}^{S-1} b_i(c\mu_{i-1,t} + b\mu_{i,t} + a\mu_{i+1,t}) =& c\varphi^{1/(1-\alpha)}\sum_{i=s_t}^{S-2} b_i\mu_{i,t} + b\sum_{i=s_t+1}^{S-1} b_i\mu_{i,t} + a(\varphi^{1/(1-\alpha)})^{-1}\sum_{i=s_t+2}^{S} b_i\mu_{i,t} \\
=& \left(c(\varphi^{1/(1-\alpha)}) + b + a(\varphi^{1/(1-\alpha)})^{-1}\right)\sum_{i=s_t+2}^{S-2} b_i\mu_{i,t} \\
&+ c\varphi^{1/(1-\alpha)}\left(b_{s_t}\mu_{s_t,t} + b_{s_t+1}\mu_{s_t+1,t}\right) \\
&+ b\left(b_{s_t+1}\mu_{s_t+1,t} + b_{S-1}\mu_{S-1,t}\right) \\
&+ a(\varphi^{1/(1-\alpha)})^{-1}\left(b_{S-1}\mu_{S-1,t} + b_S\mu_{S,t}\right)
\end{aligned}
$$

Introducing this term in the Equation for $B'(P_t^*)'\mu_t$ yields:

$$
\begin{aligned}
B'(P_t^*)'\mu_t =& \left(c\varphi^{1/(1-\alpha)} + b + a(\varphi^{1/(1-\alpha)})^{-1}\right)\sum_{i=s_t+2}^{S-2} b_i\mu_{i,t} \\
&+ c\varphi^{1/(1-\alpha)}\left(b_{s_t}\mu_{s_t,t} + b_{s_t+1}\mu_{s_t+1,t}\right) \\
&+ b\left(b_{s_t+1}\mu_{s_t+1,t} + b_{S-1}\mu_{S-1,t}\right) \\
&+ a(\varphi^{1/(1-\alpha)})^{-1}\left(b_{S-1}\mu_{S-1,t} + b_S\mu_{S,t}\right) \\
&+ b_{s_t-1}a\mu_{s_t,t} + b_{s_t}(b\mu_{s_t,t} + a\mu_{s_t+1,t}) \\
&+ b_S(c\mu_{S-1,t} + (b+c)\mu_{S,t})
\end{aligned}
$$

Rearranging terms, we have

$$\begin{aligned}
B'(P_t^*)'\mu_t = &\left(c\varphi^{1/(1-\alpha)} + b + a(\varphi^{1/(1-\alpha)})^{-1}\right)\sum_{i=s_t+2}^{S-2} b_i\mu_{i,t} \\
&+ b_{s_t}\mu_{s_t,t}\left(c\varphi^{1/(1-\alpha)} + b + a(\varphi^{1/(1-\alpha)})^{-1}\right) \\
&+ b_{s_t+1}\mu_{s_t+1,t}\left(c\varphi^{1/(1-\alpha)} + b + a(\varphi^{1/(1-\alpha)})^{-1}\right) \\
&+ b_{S-1}\mu_{S-1,t}\left(c\varphi^{1/(1-\alpha)} + b + a(\varphi^{1/(1-\alpha)})^{-1}\right) \\
&+ b_S\mu_{S,t}\left((b+c) + a(\varphi^{1/(1-\alpha)})^{-1}\right)
\end{aligned}$$

and plugging them back in the sum

$$\begin{aligned}
B'(P_t^*)'\mu_t = &\left(c\varphi^{1/(1-\alpha)} + b + a(\varphi^{1/(1-\alpha)})^{-1}\right)\sum_{i=s_t}^{S-1} b_i\mu_{i,t} \\
&+ b_S\mu_{S,t}\left(\varphi^{1/(1-\alpha)}c - (\varphi^{1/(1-\alpha)} - 1)c + b + a(\varphi^{1/(1-\alpha)})^{-1}\right)
\end{aligned}$$

which finally yields

$$B'(P_t^*)'\mu_t = \left(c\varphi^{1/(1-\alpha)} + b + a(\varphi^{1/(1-\alpha)})^{-1}\right)\sum_{i=s_t}^{S} b_i\mu_{i,t} - c(\varphi^{1/(1-\alpha)} - 1)b_S\mu_{S,t}$$

We have $B'\mu_t = \sum_{i=1}^{S} b_i\mu_{i,t}$ but note that $\mu_{i,t} = 0$ for $i < s_t - 1$. Thus $B'\mu_t = b_{s_t-1}\mu_{s_t-1,t} + \sum_{i=s_t}^{S} b_i\mu_{i,t}$. Let us define $\rho = c\varphi^{1/(1-\alpha)} + b + a(\varphi^{1/(1-\alpha)})^{-1}$. It follows that

$$\begin{aligned}
B'(P_t^*)'\mu_t =&\rho\left(T_t - b_{s_t-1}\mu_{s_t-1,t}\right) - c(\varphi^{1/(1-\alpha)} - 1)b_S\mu_{S,t} \\
=&\rho T_t - \rho(\varphi^{1/(1-\alpha)})^{s_t-1}\mu_{s_t-1,t} - c(\varphi^{1/(1-\alpha)} - 1)(\varphi^{1/(1-\alpha)})^S\mu_{S,t}
\end{aligned}$$

Similarly, for the term $B'(P_t^*)'G$ we have

$$B'(P_t^*)'G = \rho\sum_{i=s_t}^{S} b_iG_i - c(\varphi^{1/(1-\alpha)} - 1)(\varphi^{1/(1-\alpha)})^S G_S$$

Assuming that G is Pareto distributed, we have $G_i = K_e(\varphi^{-\delta_e})^i$.

$$B'(P_t^*)'G =\rho K_e\frac{(\varphi^{1/(1-\alpha)-\delta_e})^{S+1} - (\varphi^{1/(1-\alpha)-\delta_e})^{s_t}}{(\varphi^{1/(1-\alpha)-\delta_e}) - 1} - c(\varphi^{1/(1-\alpha)} - 1)(\varphi^{1/(1-\alpha)-\delta_e})^S$$

Collecting terms yields:

$$\begin{aligned}
B'm(\mu_t) =&\rho T_t && (22)\\
&+ \rho MK_e\frac{(\varphi^{1/(1-\alpha)-\delta_e})^{S+1} - (\varphi^{1/(1-\alpha)-\delta_e})^{s_t}}{(\varphi^{1/(1-\alpha)-\delta_e}) - 1} - \rho(\varphi^{1/(1-\alpha)})^{s_t-1}\mu_{s_t-1,t} && (23)\\
&- c(\varphi^{1/(1-\alpha)} - 1)(\varphi^{1/(1-\alpha)})^S\mu_{S,t} - c(\varphi^{1/(1-\alpha)} - 1)M(\varphi^{1/(1-\alpha)-\delta_e})^S && (24)
\end{aligned}$$

This completes the proof of Equation 20.□

Step 2: Proof of Equation 21

Here we prove Equation 21 describing the conditional variance of T_{t+1}, $\mathbb{V}ar_t\left[B'\epsilon_{t+1}\right]$ as a function of model primitives and of a moment of the productivity distribution at time t. To do so, we first recall some results on multinomial distributions. Then we go back to the definition of the random vector ϵ_{t+1} in the proof of Theorem 1.

To show this formula it is first useful to recall some basic results on multinomial distributions. Let assume that $X \rightsquigarrow \mathcal{M}ulti(n,p)$ with n being the number of trials and p the $(S\times 1)$ vector of event probabilities p_i. Let us define $X=[x_i]_{1\leq i\leq S}$. We have $\mathbb{E}[x_i]=np_i$ and that the variance-covariance matrix is $\mathbb{V}ar[X]=diag(p)-pp'$. We thus have $\mathbb{V}ar[x_i]=\mathbb{C}ov(x_i,x_i)=np_i(1-p_i)$ and $\mathbb{C}ov(x_i,x_j)=-np_ip_j$ for $i\neq j$.

For any $(S\times 1)$ vector B, we have

$$\begin{aligned}
\mathbb{V}ar[B'X] =& \mathbb{V}ar\left[\sum_{i=1}^{S} b_i x_i\right]\\
=&\sum_{i=1}^{S}\mathbb{V}ar[b_ix_i]+2\sum\sum_{1\leq i<j\leq S}\mathbb{C}ov(b_ix_i,b_jx_j)\\
=&\sum_{i=1}^{S} b_i^2np_i(1-p_i)+2\sum\sum_{1\leq i<j\leq S} b_ib_j\left(-np_ip_j\right)\\
=&n\sum_{i=1}^{S} b_i^2p_i(1-p_i)-2n\sum\sum_{1\leq i<j\leq S} b_ib_jp_ip_j
\end{aligned}\tag{25}$$

With this result in mind let us go back to the proof of Equation 21. By the definition of the $(S\times 1)$ random vector ϵ_{t+1} in the proof of Theorem 1 we have

$$\epsilon_{t+1}=\sum_{s=s_t}^{S}\left[f_{t+1}^{.,s}-\mu_{s,t}.P_{s,.}^{'}\right]+\sum_{s=s_t}^{S}\left[g_{t+1}^{.,s}-MG_s.P_{s,.}^{'}\right]$$

Let us pre-multiply by B', which yields

$$B'\epsilon_{t+1}=\sum_{s=s_t}^{S}\left[B'f_{t+1}^{.,s}-\mu_{s,t}.B'P_{s,.}^{'}\right]+\sum_{s=s_t}^{S}\left[B'g_{t+1}^{.,s}-MG_s.B'P_{s,.}^{'}\right]$$

Note that, since $\mathbb{E}\left[\epsilon_{t+1}\right]=0$ we have $\mathbb{E}\left[B'\epsilon_{t+1}\right]=0$. Our aim is to compute the variance of the above univariate random variable $B'\epsilon_{t+1}$. Taking the variance conditional on time t information and using the fact that, at time t, both $s_t,\mu_{s,t}$ are known and that the $f_{t+1}^{.,s}$ and $g_{t+1}^{.,s}$ are independent random vectors, yields:

$$\mathbb{V}ar_t[B'\epsilon_{t+1}]=\sum_{s=s_t}^{S}\mathbb{V}ar\left[B'f_{t+1}^{.,s}\right]+\sum_{s=s_t}^{S}\mathbb{V}ar\left[B'g_{t+1}^{.,s}\right]\tag{26}$$

Let us now focus on the term $\mathbb{V}ar\left[B'f_{t+1}^{.,s}\right]$ for $s_t\leq s<S$. Recall that the $(S\times 1)$ random vector $f_{t+1}^{.,s}$ follows a multinomial distribution with number of trials $\mu_{s,t}$ and probability vector $P_{s,.}'=(0,\cdots,0,a,b,c,0\cdots,0)'$ where the term b is at row s. Let us apply the results

on multinomial distributed random vector that we show above (Equation 25) to the random vector $f^{\cdot,s}_{t+1}$. This leads to

$$\begin{aligned}\mathbb{V}ar\left[B'f^{\cdot,s}_{t+1}\right] =&\mu_{s,t}\left[a(1-a)b^2_{s-1}+b(1-b)b^2_s+c(1-c)b^2_{s+1}-2\left(abb_{s-1}b_s+acb_{s-1}b_{s+1}+bcb_sb_{s+1}\right)\right]\\ =&\mu_{s,t}b^2_s\left[a(1-a)(\varphi^{1/(1-\alpha)})^{-2}+b(1-b)+c(1-c)(\varphi^{1/(1-\alpha)})^2-2ab(\varphi^{1/(1-\alpha)})^{-1}-2ac-2bc(\varphi^{1/(1-\alpha)})\right]\end{aligned}$$

where we have used (in the second equality) the fact that $b_s = b_{s-1}\varphi^{1/(1-\alpha)}$. Let us define $\varrho = \left[a(1-a)(\varphi^{1/(1-\alpha)})^{-2}+b(1-b)+c(1-c)(\varphi^{1/(1-\alpha)})^2-2ab(\varphi^{1/(1-\alpha)})^{-1}-2ac-2bc(\varphi^{1/(1-\alpha)})\right]$. Note that it can be shown that $\varrho = a\varphi^{-2/(1-\alpha)}+b+c\varphi^{2/(1-\alpha)}-\rho^2$. We then have

$$\mathbb{V}ar\left[B'f^{\cdot,s}_{t+1}\right] = \varrho\mu_{s,t}b^2_s \quad \text{for} \quad s_t \le s < S$$

Let us now compute $\mathbb{V}ar\left[B'f^{\cdot,S}_{t+1}\right]$. The random vector $f^{\cdot,S}_{t+1}$ follows a multinomial distribution with number of trials $\mu_{S,t}$ and probability vector $P'_{S,\cdot} = (0,\cdots,0,a,b+c)'$. Using Equation 25, we have

$$\begin{aligned}\mathbb{V}ar\left[B'f^{\cdot,S}_{t+1}\right] =&\mu_{S,t}\left[a(1-a)b^2_{S-1}+(b+c)(1-b-c)b^2_S-2b_{S-1}b_Sa(b+c)\right]\\ =&\mu_{S,t}\left[a(1-a)b^2_{S-1}+(b+c)(1-b-c)b^2_S-2b_{S-1}b_Sa(b+c)\right]\\ =&\mu_{S,t}\left[a(1-a)(\varphi^{1/(1-\alpha)})^{-2}b^2_S+(b+c)(1-b-c)b^2_S-2(\varphi^{1/(1-\alpha)})^{-1}b^2_Sa(b+c)\right]\\ =&\mu_{S,t}b^2_S\left[a(1-a)(\varphi^{1/(1-\alpha)})^{-2}+(b+c)(1-b-c)-2a(b+c)(\varphi^{1/(1-\alpha)})^{-1}\right]\end{aligned}$$

Let us define $\varrho' = a(1-a)(\varphi^{1/(1-\alpha)})^{-2}+(b+c)(1-b-c)-2a(b+c)(\varphi^{1/(1-\alpha)})^{-1}$, we then have $\mathbb{V}ar\left[B'f^{\cdot,S}_{t+1}\right] = \varrho'\mu_{S,t}b^2_S$.

Taking this together, we have thus

$$\sum_{s=s_t}^{S}\mathbb{V}ar\left[B'f^{\cdot,s}_{t+1}\right] =\varrho\sum_{s=s_t}^{S}b^2_s\mu_{s,t}-(\varrho-\varrho')b^2_S\mu_{S,t}$$

Similarly we have that

$$\sum_{s=s_t}^{S}\mathbb{V}ar\left[B'g^{\cdot,s}\right] =\varrho\sum_{s=s_t}^{S}b^2_sg_s-(\varrho-\varrho')b^2_Sg_S$$

Note that $b^2_sg_s = K_e((\varphi^{1/(1-\alpha)})^2\varphi^{-\delta_e})^s$ and let us define D_t, a measure of firm-level productivity dispersion by

$$D_t := (B\circ B)'\mu_t = b^2_{s_t-1}\mu_{s_t-1,t}+\sum_{s=s_t}^{S}b^2_s\mu_{s,t} = \sum_{s=s_t-1}^{S}\mu_{s,t}(\varphi^{1/(1-\alpha)})^2$$

where $.\circ.$ is the Hadamard product of matrices (element-wise product).

Taking the results above yields the result:

$$\begin{aligned}\mathbb{V}ar[B'\epsilon_{t+1}] =&\varrho D_t+\varrho K_e\frac{((\varphi^{1/(1-\alpha)})^2\varphi^{-\delta_e})^{S+1}-((\varphi^{1/(1-\alpha)})^2\varphi^{-\delta_e})^{s_t}}{(\varphi^{1/(1-\alpha)})^2\varphi^{-\delta_e}-1}-\varrho(\varphi^{1/(1-\alpha)})^{2(s_t-1)}\mu_{s_t-1,t}\\ &-(\varrho-\varrho')(\varphi^{1/(1-\alpha)})^{2S}\mu_{S,t}-K_e(\varrho-\varrho')((\varphi^{1/(1-\alpha)})^2\varphi^{-\delta_e})^S\end{aligned}$$

$\square$.

A.4 Proof of Corollary 3

In this appendix we prove Corollary 3 which describes the dynamics of aggregate output. This result is a direct implication of Theorem 2 regarding the dynamics of the univariate state variable T_t.

To prove this Corollary, we first solve for aggregate output, Y_t, as a function of the univariate state variable T_t analytically. We then study their first order relationship. The next step is then to take the first-order approximation of the Equation describing the dynamics of the univariate state variable T_t. Finally, we find the implied first-order dynamics of Y_t.

Let us first compute aggregate output Y_t as a function of T_t only:

$$Y_t = \sum_{i=1}^{N_t} y_t^i = \sum_{s=1}^{S} \mu_{s,t} (\varphi^s)^{\frac{1}{1-\alpha}} \left(\frac{\alpha}{w_t}\right)^{\frac{\alpha}{1-\alpha}} = \left(\frac{\alpha}{w_t}\right)^{\frac{\alpha}{1-\alpha}} T_t$$

Recall that $w_t = \left(\alpha^{\frac{1}{1-\alpha}} \frac{T_t}{LM}\right)^{\frac{1-\alpha}{\gamma(1-\alpha)+1}}$. Substituting the expression of the wage in the latter Equation yields $Y_t = \alpha^{\frac{\alpha\gamma}{\gamma(1-\alpha)+1}} \left(\frac{1}{L(M)}\right)^{\frac{-\alpha}{\gamma(1-\alpha)+1}} (T_t)^{1-\frac{\alpha}{\gamma(1-\alpha)+1}}$. This last equality taken at the first order implied that:

$$\widehat{Y}_t = \left(1 - \frac{\alpha}{\gamma(1-\alpha)+1}\right) \widehat{T}_t \tag{27}$$

where $\widehat{X}_t$ of a variable X_t is the percentage deviation from its steady-state value X: $\widehat{X}_t = (X_t - X)/X$.

We then take the percentage deviation from steady-state of Equation 8:

$$\begin{aligned}
T_{t+1} =& \rho T_t + \rho E_t - O_t^T + \sigma_t \varepsilon_{t+1} \\
T =& \rho T + \rho E - O^T \\
T_{t+1} - T =& \rho(T_t - T) + \rho(E_t - E) - (O_t^T - O) + \sigma_t \varepsilon_{t+1} \\
\frac{T_{t+1} - T}{T} =& \rho \frac{T_t - T}{T} + \rho \frac{E}{T} \frac{E_t - E}{E} - \frac{O^T}{T} \frac{O_t^T - O}{O} + \frac{\sigma_t}{T} \varepsilon_{t+1}
\end{aligned}$$

$$\begin{aligned}
\widehat{T_{t+1}} =& \rho \widehat{T}_t + \rho \frac{E}{T} \widehat{E_t} - \frac{O^T}{T} \widehat{O_t^T} + \frac{\sigma_t}{T} \varepsilon_{t+1} \\
\widehat{Y}_{t+1} =& \rho \widehat{Y}_t + \left(1 - \frac{\alpha}{\gamma(1-\alpha)+1}\right) \rho \frac{E}{T} \widehat{E}_t - \left(1 - \frac{\alpha}{\gamma(1-\alpha)+1}\right) \frac{O^T}{T} \widehat{O_t^T} + \frac{\sigma_t}{T} \varepsilon_{t+1}
\end{aligned}$$

where the second line is Equation 8 at the steady-state; in the third line we subtract the second to the first line; in the fourth line we divide both sides by the steady-state value of T and in the last line we use Equation 27. $\square$

A.5 Proof of Proposition 1: Aggregate Persistence

In this appendix, we prove Proposition 1 regarding the comparative statics results for aggregate persistence, ρ. We first express ρ as a function of b, a measure of micro-level persistence, and of δ, the tail of the productivity stationary distribution.

First, note that from definition $\delta = \frac{\log(a/c)}{\log \varphi}$, it follows that $c = a\varphi^{-\delta}$. Secondly, from the fact that $b = 1 - a - c = 1 - a(1 + \varphi^{-\delta})$ we have that $a = \frac{1-b}{1+\varphi^{-\delta}}$. From Theorem 2, aggregate persistence is $\rho = a\varphi^{\frac{-1}{1-\alpha}} + b + c\varphi^{\frac{1}{1-\alpha}}$. In this last Equation, let us substitute c and a using $c = a\varphi^{-\delta}$ and $a = \frac{1-b}{1+\varphi^{-\delta}}$:

$$\rho = \frac{1-b}{1+\varphi^{-\delta}}\varphi^{\frac{-1}{1-\alpha}} + b + \varphi^{-\delta}\varphi^{\frac{1}{1-\alpha}}\frac{1-b}{1+\varphi^{-\delta}}$$

$$\rho = \frac{1-b}{1+\varphi^{-\delta}}\left[\varphi^{\frac{-1}{1-\alpha}} - \varphi^{-\delta} + \varphi^{-\delta}\varphi^{\frac{1}{1-\alpha}} - 1\right] + 1$$

First, it is clear that if $\delta = \frac{1}{1-\alpha}$, then it follows that $\rho = 1$. This is exactly (iii) of the Proposition 1.

Second, from the expression of ρ, it is clear that $\frac{d\rho}{db} > 0$ if and only if $g(\delta) = \varphi^{\frac{-1}{1-\alpha}} - \varphi^{-\delta} + \varphi^{-\delta}\varphi^{\frac{1}{1-\alpha}} - 1 < 0$. Note that $g(\frac{1}{1-\alpha}) = 0$ and $g(\delta) \underset{\delta\to\infty}{\longrightarrow} \varphi^{\frac{-1}{1-\alpha}} - 1 < 0$ since $\varphi > 1$. The derivative of g is $g'(\delta) = -(-\log\varphi)\varphi^{-\delta} + (-\log\varphi)\varphi^{-\delta+\frac{1}{1-\alpha}} < 0$. It follows that for $\delta > \frac{1}{1-\alpha}$, then $g(\delta) < 0$ and thus $\frac{d\rho}{db} > 0$. We have just shown (i).

Finally to show (ii), let us rewrite $\rho = -\frac{(b-1)g(\delta)}{1+\varphi^{-\delta}} + 1$. We have shown that for $g(\delta)$ is decreasing in δ, since $b < 1$ it is clear that $(b-1)g(\delta)$ is increasing in δ. Note that $\frac{1}{1+\varphi^{-\delta}}$ is also increasing in δ. It follows that $\frac{(b-1)g(\delta)}{1+\varphi^{-\delta}}$ is increasing in δ which then implies that ρ is decreasing in δ, which is the statement in (ii).

□

A.6 Proof of Proposition 2: Aggregate Volatility

In this appendix, we prove Proposition 2 describing how aggregate volatility decays with the number of firms N.

To prove this Proposition, we study the asymptotic behavior of T, D and deduce the one for D/T^2, again when the number of firms N goes to infinity. We complete the proof by studying the behavior of the remaining terms $E(\varphi^2)$ and O^σ and $E(\varphi^2)/T^2$ and O^σ/T^2.

Step 1: How T evolves when the number of incumbents converges to infinity

For a given number of firms, let us look at the expression of T:

$$\begin{aligned}
T =& \sum_{s=1}^{S} (\varphi^s)^{\frac{1}{1-\alpha}} \mu_s \\
=& (\varphi^{s^*-1})^{\frac{1}{1-\alpha}} \mu_{s^*-1} + \sum_{s=s^*}^{S} (\varphi^s)^{\frac{1}{1-\alpha}} \mu_s \\
=& (\varphi^{s^*-1})^{\frac{1}{1-\alpha}} a \left(MK_eC_1 + MK_eC_2 \left(\varphi^{s^*}\right)^{-\delta_e} + MK_eC_3 + MK_e(\varphi^{s^*})^{-\delta_e} \right) \\
&+ \sum_{s=s^*}^{S} (\varphi^s)^{\frac{1}{1-\alpha}} \left(MK_eC_1 \left(\frac{\varphi^s}{\varphi^{s^*}}\right)^{-\delta} + MK_eC_2 \left(\varphi^s\right)^{-\delta_e} + MK_eC_3 \right)
\end{aligned}$$

By dividing both sides by M, we get

$$\begin{aligned}
\frac{T}{M} =& a(\varphi^{s^*-1})^{\frac{1}{1-\alpha}} \left(K_eC_1 + K_eC_2 \left(\varphi^{s^*}\right)^{-\delta_e} + K_eC_3 + K_e(\varphi^{s^*})^{-\delta_e} \right) \\
&+ K_eC_1 \left(\varphi^{s^*}\right)^{\delta} \sum_{s=s^*}^{S} \left(\varphi^{-\delta+\frac{1}{1-\alpha}}\right)^s + K_eC_2 \sum_{s=s^*}^{S} \left(\varphi^{-\delta_e+\frac{1}{1-\alpha}}\right)^s + K_eC_3 \sum_{s=s^*}^{S} (\varphi^{\frac{1}{1-\alpha}})^s \\
=& a(\varphi^{s^*-1})^{\frac{1}{1-\alpha}} \left(K_eC_1 + K_eC_2 \left(\varphi^{s^*}\right)^{-\delta_e} + K_eC_3 + K_e(\varphi^{s^*})^{-\delta_e} \right) \\
&+ K_eC_1 \left(\varphi^{s^*}\right)^{\delta} \frac{\left(\varphi^{-\delta+\frac{1}{1-\alpha}}\right)^{s^*} - \left(\varphi^{-\delta+\frac{1}{1-\alpha}}\right)^{S+1}}{1-\varphi^{-\delta+\frac{1}{1-\alpha}}} + K_eC_2 \frac{\left(\varphi^{-\delta_e+\frac{1}{1-\alpha}}\right)^{s^*} - \left(\varphi^{-\delta_e+\frac{1}{1-\alpha}}\right)^{S+1}}{1-\varphi^{-\delta_e+\frac{1}{1-\alpha}}} \\
&+ K_eC_3 \frac{(\varphi^{\frac{1}{1-\alpha}})^{s^*} - (\varphi^{\frac{1}{1-\alpha}})^{S+1}}{1-\varphi^{\frac{1}{1-\alpha}}}
\end{aligned}$$

Recall that under assumption 2, we have

$$\begin{aligned}
(\varphi^{\frac{1}{1-\alpha}})^S &= (\varphi^S)^{\frac{1}{1-\alpha}} = (ZN^{1/\delta})^{\frac{1}{1-\alpha}} = Z^{\frac{1}{1-\alpha}} N^{\frac{1}{\delta(1-\alpha)}} \\
\left(\varphi^{-\delta+\frac{1}{1-\alpha}}\right)^S &= \left(\varphi^S\right)^{-\delta+\frac{1}{1-\alpha}} = \left(ZN^{1/\delta}\right)^{-\delta+\frac{1}{1-\alpha}} = Z^{-\delta+\frac{1}{1-\alpha}} N^{-1+\frac{1}{\delta(1-\alpha)}} \\
\left(\varphi^{-\delta_e+\frac{1}{1-\alpha}}\right)^S &= \left(\varphi^S\right)^{-\delta_e+\frac{1}{1-\alpha}} = \left(ZN^{1/\delta}\right)^{-\delta_e+\frac{1}{1-\alpha}} = Z^{-\delta_e+\frac{1}{1-\alpha}} N^{-\frac{\delta_e}{\delta}+\frac{1}{\delta(1-\alpha)}}
\end{aligned}$$

Since we assume that $\delta(1-\alpha) > 1$ and $\delta_e(1-\alpha) > 1$ we have both $-\frac{\delta_e}{\delta} + \frac{1}{\delta(1-\alpha)} < 0$ and $-1 + \frac{1}{\delta(1-\alpha)} < 0$ and thus both $\left(\varphi^{-\delta+\frac{1}{1-\alpha}}\right)^S$ and $\left(\varphi^{-\delta_e+\frac{1}{1-\alpha}}\right)^S$ converge to zero when N goes to infinity. We also have that

$$C_3(\varphi^{\frac{1}{1-\alpha}})^S = \frac{-\varphi^{-\delta_e} Z^{-\delta_e}}{(1-\varphi^{-\delta_e})(a-c)} Z^{\frac{1}{1-\alpha}} N^{-\delta_e/\delta + \frac{1}{\delta(1-\alpha)}} \underset{N\to\infty}{\longrightarrow} 0$$

Putting these results together yields

$$\frac{T}{M} \underset{N\to\infty}{\longrightarrow} T^{\infty} := a(\varphi^{s^*-1})^{\frac{1}{1-\alpha}} \left((\varphi^{\delta_e}-1)C_1^{\infty} + (\varphi^{\delta_e}-1)(C_2+1)\left(\varphi^{s^*}\right)^{-\delta_e} \right)$$
$$+ (\varphi^{\delta_e}-1)C_1^{\infty} \frac{\left(\varphi^{\frac{1}{1-\alpha}}\right)^{s^*}}{1-\varphi^{-\delta+\frac{1}{1-\alpha}}} + (\varphi^{\delta_e}-1)C_2 \frac{\left(\varphi^{-\delta_e+\frac{1}{1-\alpha}}\right)^{s^*}}{1-\varphi^{-\delta_e+\frac{1}{1-\alpha}}}$$

In other words, under assumption 2 and if $\delta(1-\alpha) > 1$ and $\delta_e(1-\alpha) > 1$ then $T \underset{M\to\infty}{\sim} T^{\infty}M$ or

$$T \underset{N\to\infty}{\sim} E^{\infty}T^{\infty}N \tag{28}$$

Step 2: How D evolves when the number of incumbents converges to infinity

For a given number of firms, the steady-state value of D:

$$\begin{aligned}
D =& \sum_{s=1}^{S} \left((\varphi^{s})^{\frac{1}{1-\alpha}} \right)^2 \mu_s \\
=& \left((\varphi^{s^*-1})^{\frac{1}{1-\alpha}} \right)^2 \mu_{s^*-1} + \sum_{s=s^*}^{S} \left((\varphi^{s})^{\frac{1}{1-\alpha}} \right)^2 \mu_s \\
\frac{D}{M} =& (\varphi^{s^*-1})^{\frac{2}{1-\alpha}} \hat{\mu}_{s^*-1} + K_e C_1 (\varphi^{s^*})^{\delta} \sum_{s=s^*}^{S} (\varphi^{s})^{\frac{2}{1-\alpha}-\delta} + K_e C_2 \sum_{s=s^*}^{S} (\varphi^{s})^{\frac{2}{1-\alpha}-\delta_e} + K_e C_3 \sum_{s=s^*}^{S} (\varphi^{s})^{\frac{2}{1-\alpha}} \\
=& a(\varphi^{s^*-1})^{\frac{2}{1-\alpha}} \left(K_e C_1 + K_e (C_2+1)\left(\varphi^{s^*}\right)^{-\delta_e} + K_e C_3 \right) \\
&+ K_e C_1 (\varphi^{s^*})^{\delta} \frac{(\varphi^{\frac{2}{1-\alpha}-\delta})^{s^*} - (\varphi^{\frac{2}{1-\alpha}-\delta})^{S+1}}{1-\varphi^{\frac{2}{1-\alpha}-\delta}} \\
&+ K_e C_2 \frac{(\varphi^{\frac{2}{1-\alpha}-\delta_e})^{s^*} - (\varphi^{\frac{2}{1-\alpha}-\delta_e})^{S+1}}{1-\varphi^{\frac{2}{1-\alpha}-\delta_e}} \\
&+ K_e C_3 \frac{(\varphi^{\frac{2}{1-\alpha}})^{s^*} - (\varphi^{\frac{2}{1-\alpha}})^{S+1}}{1-\varphi^{\frac{2}{1-\alpha}}}
\end{aligned}$$

Under assumption 2, we have

$$(\varphi^{\frac{2}{1-\alpha}-\delta})^{S} = (\varphi^{S})^{\frac{2}{1-\alpha}-\delta} = (ZN^{1/\delta})^{\frac{2}{1-\alpha}-\delta} = Z^{\frac{2}{1-\alpha}-\delta} N^{\frac{2}{\delta(1-\alpha)}-1}$$
$$(\varphi^{\frac{2}{1-\alpha}-\delta_e})^{S} = (\varphi^{S})^{\frac{2}{1-\alpha}-\delta_e} = (ZN^{1/\delta})^{\frac{2}{1-\alpha}-\delta_e} = Z^{\frac{2}{1-\alpha}-\delta_e} N^{\frac{2}{\delta(1-\alpha)}-\frac{\delta_e}{\delta}}$$
$$C_3(\varphi^{\frac{2}{1-\alpha}})^{S} = C_3(\varphi^{S})^{\frac{2}{1-\alpha}} = \frac{-\varphi^{-\delta_e}Z^{-\delta_e}}{(1-\varphi^{-\delta_e})(a-c)} (Z)^{\frac{2}{1-\alpha}} N^{\frac{2}{\delta(1-\alpha)}-\frac{\delta_e}{\delta}}$$

Under the assumption that $\delta(1-\alpha)<2$ and $\delta_e(1-\alpha)<2$, these terms diverge when N goes to infinity. Thus we are able to look at the asymptotic equivalent of D/M,

$$\begin{aligned}\frac{D}{M} \underset{N\to\infty}{\sim} & a(\varphi^{s^*-1})^{\frac{2}{1-\alpha}}\left((\varphi_e^\delta-1)C_1^\infty+(\varphi_e^\delta-1)(C_2+1)\left(\varphi^{s^*}\right)^{-\delta_e}\right)\\ &+(\varphi_e^\delta-1)C_1^\infty(\varphi^{s^*})^\delta\frac{-\varphi^{\frac{2}{1-\alpha}-\delta}}{1-\varphi^{\frac{2}{1-\alpha}-\delta}}Z^{\frac{2}{1-\alpha}-\delta}N^{\frac{2}{\delta(1-\alpha)}-1}\\ &+\left((\varphi_e^\delta-1)C_2\frac{-\varphi^{\frac{2}{1-\alpha}-\delta_e}}{1-\varphi^{\frac{2}{1-\alpha}-\delta_e}}Z^{\frac{2}{1-\alpha}-\delta_e}+(\varphi_e^\delta-1)\frac{-\varphi^{\frac{2}{1-\alpha}}}{1-\varphi^{\frac{2}{1-\alpha}}}\frac{-\varphi^{-\delta_e}Z^{-\delta_e}}{(1-\varphi^{-\delta_e})(a-c)}(Z)^{\frac{2}{1-\alpha}}\right)N^{\frac{2}{\delta(1-\alpha)}-\frac{\delta_e}{\delta}}\end{aligned}$$

By using the intermediate result above on the link between N and M, we have

$$\begin{aligned}D \underset{N\to\infty}{\sim} & a(\varphi^{s^*-1})^{\frac{2}{1-\alpha}}\left((\varphi_e^\delta-1)C_1^\infty+(\varphi_e^\delta-1)(C_2+1)\left(\varphi^{s^*}\right)^{-\delta_e}\right)E^\infty N\\ &+(\varphi_e^\delta-1)C_1^\infty(\varphi^{s^*})^\delta\frac{-\varphi^{\frac{2}{1-\alpha}-\delta}}{1-\varphi^{\frac{2}{1-\alpha}-\delta}}Z^{\frac{2}{1-\alpha}-\delta}E^\infty N^{\frac{2}{\delta(1-\alpha)}}\\ &+\left((\varphi_e^\delta-1)C_2\frac{-\varphi^{\frac{2}{1-\alpha}-\delta_e}}{1-\varphi^{\frac{2}{1-\alpha}-\delta_e}}Z^{\frac{2}{1-\alpha}-\delta_e}+(\varphi_e^\delta-1)\frac{-\varphi^{\frac{2}{1-\alpha}}}{1-\varphi^{\frac{2}{1-\alpha}}}\frac{-\varphi^{-\delta_e}Z^{-\delta_e}}{(1-\varphi^{-\delta_e})(a-c)}(Z)^{\frac{2}{1-\alpha}}\right)E^\infty N^{\frac{2}{\delta(1-\alpha)}-\frac{\delta_e}{\delta}+1}\end{aligned}$$

By introducing the appropriate constant D_1^∞, D_2^∞ and D_3^∞ such that we have, under assumption 2

$$D \underset{N\to\infty}{\sim} D_1^\infty N + D_2^\infty N^{\frac{2}{\delta(1-\alpha)}} + D_3^\infty N^{\frac{2}{\delta(1-\alpha)}-\frac{\delta_e}{\delta}+1} \tag{29}$$

Step 3: How D/T^2 evolves with N:

The first term of aggregate volatility described by Equation 12 is $\frac{D}{T^2}$. Let us look at its equivalent when N goes to infinity by combining Equations 28 and 29

$$\frac{D}{T^2} \underset{N\to\infty}{\sim} \frac{\frac{D_1^\infty}{(E^\infty T^\infty)}}{N} + \frac{\frac{D_2^\infty}{(E^\infty T^\infty)}}{N^{2-\frac{2}{\delta(1-\alpha)}}} + \frac{\frac{D_3^\infty}{(E^\infty T^\infty)}}{N^{1+\frac{\delta_e}{\delta}-\frac{2}{\delta(1-\alpha)}}}$$

Under the assumptions that $\delta(1-\alpha)<2$ and $\delta_e(1-\alpha)<2$, then $2-\frac{2}{\delta(1-\alpha)}<1$ and $1+\frac{\delta_e}{\delta}-\frac{2}{\delta(1-\alpha)}<1$. In other words, the last two terms dominate the first term and thus:

$$\frac{D}{T^2} \underset{N\to\infty}{\sim} \frac{\frac{D_2^\infty}{(E^\infty T^\infty)}}{N^{2-\frac{2}{\delta(1-\alpha)}}} + \frac{\frac{D_3^\infty}{(E^\infty T^\infty)}}{N^{1+\frac{\delta_e}{\delta}-\frac{2}{\delta(1-\alpha)}}} \tag{30}$$

□

Step 4: How $E(\varphi^2)$ and O^σ evolve with N

Here we prove a similar result for the remaining terms in Equation 12, i.e. $E(\varphi^2)/T^2$ and O^σ/T^2. Let us first find an equivalent for $\frac{E(\varphi^2)}{M}$ and then for $\frac{O^\sigma}{M}$ when $N\to\infty$. The steady-state expression of $E(\varphi^2)$ is

$$\begin{aligned}
E(\varphi^2) &= \left(M\sum_{s=s^*}^{S} G_s\left(\varphi^{2s}\right)^{\frac{1}{1-\alpha}}\right) - \left(\left(\varphi^{2(s^*-1)}\right)^{\frac{1}{1-\alpha}}\mu_{s^*-1,t}\right)\\
&= \left(MK_e\sum_{s=s^*}^{S}(\varphi^s)^{-\delta_e}\left(\varphi^{2s}\right)^{\frac{1}{1-\alpha}}\right) - \left(\left(\varphi^{2(s^*-1)}\right)^{\frac{1}{1-\alpha}}\mu_{s^*-1,t}\right)\\
&= MK_e\frac{\left(\varphi^{-\delta_e+\frac{2}{1-\alpha}}\right)^{S+1} - \left(\varphi^{-\delta_e+\frac{2}{1-\alpha}}\right)^{s^*}}{\left(\varphi^{-\delta_e+\frac{2}{1-\alpha}}\right)-1} - \left(\left(\varphi^{\frac{2}{1-\alpha}}\right)^{(s^*-1)}\mu_{s^*-1,t}\right)
\end{aligned}$$

Under assumption 1, we still have

$$(\varphi^{\frac{2}{1-\alpha}-\delta_e})^S = (\varphi^S)^{\frac{2}{1-\alpha}-\delta_e} = (ZN^{1/\delta})^{\frac{2}{1-\alpha}-\delta_e} = Z^{\frac{2}{1-\alpha}-\delta_e}N^{\frac{2}{\delta(1-\alpha)}-\frac{\delta_e}{\delta}}$$

Thus, it follows

$$\begin{aligned}
\frac{E(\varphi^2)}{M} =& K_e\frac{\left(Z^{\frac{2}{1-\alpha}-\delta_e}N^{\frac{2}{\delta(1-\alpha)}-\frac{\delta_e}{\delta}}\varphi^{-\delta_e+\frac{2}{1-\alpha}}\right) - \left(\varphi^{-\delta_e+\frac{2}{1-\alpha}}\right)^{s^*}}{\left(\varphi^{-\delta_e+\frac{2}{1-\alpha}}\right)-1}\\
&- \left(\left(\varphi^{\frac{2}{1-\alpha}}\right)^{(s^*-1)}\left(K_eC_1 + K_e(C_2+1)\varphi^{s^*} + K_eC_3\right)\right)\\
=& K_eN^{\frac{2}{\delta(1-\alpha)}-\frac{\delta_e}{\delta}}\frac{\left(Z^{\frac{2}{1-\alpha}-\delta_e}\varphi^{-\delta_e+\frac{2}{1-\alpha}}\right) - N^{\frac{-2}{\delta(1-\alpha)}+\frac{\delta_e}{\delta}}\left(\varphi^{-\delta_e+\frac{2}{1-\alpha}}\right)^{s^*}}{\left(\varphi^{-\delta_e+\frac{2}{1-\alpha}}\right)-1}\\
&- \left(\left(\varphi^{\frac{2}{1-\alpha}}\right)^{(s^*-1)}\left(K_eC_1 + K_e(C_2+1)\varphi^{s^*} + K_eC_3\right)\right)
\end{aligned}$$

Under the assumption that $\delta_e(1-\alpha) < 2$, we have

$$\frac{E(\varphi^2)}{M} \sim K_eN^{\frac{2}{\delta(1-\alpha)}-\frac{\delta_e}{\delta}}\frac{\left(Z^{\frac{2}{1-\alpha}-\delta_e}\varphi^{-\delta_e+\frac{2}{1-\alpha}}\right)}{\left(\varphi^{-\delta_e+\frac{2}{1-\alpha}}\right)-1} - \left(\left(\varphi^{\frac{2}{1-\alpha}}\right)^{(s^*-1)}\left(\varphi^{\delta_e}-1\right)\left(C_1^\infty + (C_2+1)\varphi^{s^*}\right)\right)$$

Recall that $M \sim E^\infty N$, then for some constant E_1^∞ and E_2^∞, we have $E(\varphi^2) \sim E_1^\infty N^{1-\frac{\delta_e}{\delta}+\frac{2}{\delta(1-\alpha)}} + E_2^\infty N$. Using the fact that $T^2 \underset{N\to\infty}{\sim} E^\infty T^\infty N$ and the above Equation, we get for some other constant $\mathcal{E}_1^\infty$ and $\mathcal{E}_2^\infty$:

$$\frac{E(\varphi^2)}{T^2} \sim \frac{\mathcal{E}_1^\infty}{N^{1+\frac{\delta_e}{\delta}-\frac{2}{\delta(1-\alpha)}}} + \frac{\mathcal{E}_2^\infty}{N} \sim \frac{\mathcal{E}_1^\infty}{N^{1+\frac{\delta_e}{\delta}-\frac{2}{\delta(1-\alpha)}}} \tag{31}$$

where the last equivalence comes from the fact that $\delta(1-\alpha) > 2$ and $\delta_e(1-\alpha) > 2$ and thus $1 > 1 + \frac{\delta_e}{\delta} - \frac{2}{\delta(1-\alpha)}$.

The steady-state expression for O^σ is:

$$\begin{aligned}\frac{O^\sigma}{M} &= -K_e(\varrho-\varrho')(\varphi^{-\delta_e}(\varphi^{\frac{1}{1-\alpha}})^2)^S - (\varrho-\varrho')(\varphi^{\frac{1}{1-\alpha}})^{2S}\hat{\mu}_S \\ &= -K_e(\varrho-\varrho')(\varphi^{-\delta_e}(\varphi^{\frac{1}{1-\alpha}})^2)^S - (\varrho-\varrho')(\varphi^{\frac{1}{1-\alpha}})^{2S}\left(K_eC_1(\varphi^{s^*})^{-\delta}(\varphi^S)^{-\delta} + K_eC_2(\varphi^S)^{-\delta_e} + K_eC_3\right) \\ &= -K_e(\varrho-\varrho')(\varphi^{-\delta_e+\frac{2}{1-\alpha}})^S - (\varrho-\varrho')\left(K_eC_1(\varphi^{s^*})^{-\delta}(\varphi^{-\delta+\frac{2}{1-\alpha}})^S + K_eC_2(\varphi^{-\delta_e+\frac{2}{1-\alpha}})^S + K_eC_3(\varphi^{\frac{2}{1-\alpha}})^S\right)\end{aligned}$$

Recall that under assumption 2,

$$\begin{aligned}(\varphi^{\frac{2}{1-\alpha}-\delta})^S &= (\varphi^S)^{\frac{2}{1-\alpha}-\delta} = (ZN^{1/\delta})^{\frac{2}{1-\alpha}-\delta} = Z^{\frac{2}{1-\alpha}-\delta}N^{\frac{2}{\delta(1-\alpha)}-1} \\ (\varphi^{\frac{2}{1-\alpha}-\delta_e})^S &= (\varphi^S)^{\frac{2}{1-\alpha}-\delta_e} = (ZN^{1/\delta})^{\frac{2}{1-\alpha}-\delta_e} = Z^{\frac{2}{1-\alpha}-\delta_e}N^{\frac{2}{\delta(1-\alpha)}-\frac{\delta_e}{\delta}} \\ C_3(\varphi^{\frac{2}{1-\alpha}})^S &= C_3(\varphi^S)^{\frac{2}{1-\alpha}} = \frac{-\varphi^{-\delta_e}Z^{-\delta_e}}{(1-\varphi^{-\delta_e})(a-c)}(Z)^{\frac{2}{1-\alpha}}N^{\frac{2}{\delta(1-\alpha)}-\frac{\delta_e}{\delta}}\end{aligned}$$

Using the above relations, we then have, for some constants O_1^∞ and O_2^∞,

$$O^\sigma \sim O_1^\infty N^{1-\frac{\delta_e}{\delta}+\frac{2}{\delta(1-\alpha)}} + O_2^\infty N^{\frac{2}{\delta(1-\alpha)}}$$

from which it follows that for some other constant $\mathcal{O}_1^\infty$ and $\mathcal{O}_2^\infty$

$$\frac{O^\sigma}{T^2} \sim \frac{\mathcal{O}_1^\infty}{N^{1+\frac{\delta_e}{\delta}-\frac{2}{\delta(1-\alpha)}}} + \frac{\mathcal{O}_2^\infty}{N^{2-\frac{2}{\delta(1-\alpha)}}} \tag{32}$$

Putting Equations 30, 31 and 32 together yields the results in Equation 13. □

B Data Appendix

In this appendix, we describe the different data sources used in the paper. The first data source is the Business Dynamics Statistics (BDS), giving firm counts by size and age on the universe of firms in the US economy. Compustat data contains information on publicly traded firms. Finally, we use publicly available aggregate time series.

B.1 BDS data

According to the US Census Bureau, the Business Dynamics Statistics (BDS) provides annual measures of firms' dynamics covering the entire economy. It is aggregated into bins by firm characteristics such as size and size by age. The BDS is created from the Longitudinal Business Database (LBD), a US firm-level census. The BDS database gives us the number of firms by employment size categories (1-5, 5-10, 10-20,20-50,50-100,100-250,500-1000,1000-2500,2500-5000,5000-10000) for the period ranging from 1977 to 2012. Note that the number of firms in each bin is the number of firms on March 12 of each year. We also source from the BDS the number of firms of age zero by employment size. We call the latter entrants.

We compute the empirical counterpart of the steady-state stationary distribution in our model based on this data, by taking the average of each bin over years. We do this for the entrant and incumbent distributions. We then estimate the tail of these distribution following Virkar and Clauset (2014). We find that the tail estimate for the (average) incumbent size distribution is 1.0977 with a standard-deviation of 0.0016. For entrants, this estimate is 1.5708 with standard deviation of 0.0166. To compute the entry rate, we divide the average number of entrants over the period 1977-2012 by the average number of incumbents. Over this period there are 48,8140 entrant firms and 4,477,300 incumbent firms; the entry rate is then 10.9%.

To perform the exercise described in section 5.4, we need to compute the model counterpart of the time t firm size distribution. According to Theorem 1, these are deviations of the firm size distribution around the (deterministic) stationary firm size distribution. However, in the BDS data, the trend of each bin is different. We thus HP-filter each bin of the BDS data with a smoothing parameter $\lambda = 6.25$. Each bins is thus decomposed $\mu_{s,t}^{BDS} = \mu_{s,t}^{BDS-trend} + \mu_{s,t}^{BDS-dev}$ where $\mu_{s,t}^{BDS}$ is the original bin value, $\mu_{s,t}^{BDS-trend}$ is its HP-trend and $\mu_{s,t}^{BDS-dev}$ is the HP-deviation from trend. The empirical counterpart of time t firm size distribution in our model is thus $\mu_{s}^{BDS-average} + \mu_{s,t}^{BDS-dev}$ where $\mu_{s}^{BDS-average}$ is the average of bins s over the period 1977-2012. We then use Equations 1, 3 and 11 to compute the time series for A_t, Y_t and $\frac{\sigma_t^2}{T^2}$ which we plot in Figure 5 along with data aggregate time series describe below.

B.2 Compustat

The Compustat database is compiled from mandatory public disclosure documents by publicly listed firm in the US. It is a firm-level yearly (unbalanced) panel with balance sheet information. Apart from firm-level identifiers, year and sector (4 digit SIC) information, we use two variables from Compustat: employment and sales. We use data from the year 1958 to 2009. Sales is a nominal variable. We deflate it using the price deflator given by the NBER-CES Manufacturing Industry Database for shipments (PISHIP) in the corresponding SIC industry.

Using this dataset, we estimate tail indexes following Gabaix and Ibragimov (2011), performing a log rank-log size regression on the cross-section of firms each year. Our measure of size is given by the number of employees. We compute tail estimates for firms above 1k, 5k, 10k, 15k and 20k employees. We then HP-filter the resulting time-series of tail estimates (with a smoothing parameter of 6.25).

For each year, we also compute the cross-sectional variance of real sales and then HP-filter the time series using a smoothing parameter of 6.25.

B.3 Aggregate Data

The aggregate data comes from two sources. We take quarterly time series of aggregate TFP and Output from Fernald (2014). For the exercise in section 5.4, since the BDS data are computed on March 12 of each year, we compute the average over 4 quarters up to, and including, March. For example, for the year 1985 we compute the average of 1984Q2, 1984Q3, 1984Q4 and 1985Q1. We do this for TFP and Output before HP-filtering the resulting time series with a smoothing parameter of 6.25. The other source for annual time series on aggregate output is taken from the St-Louis FED. We use this series for the correlations reported in Table 4, either after HP-filtering with smoothing parameter 6.25 or by computing its growth rate.

For the results in Table 5, we estimate a GARCH(1,1) on the de-meaned growth rate of both aggregate TFP and output, both at a quarterly frequency. The source for this data is Fernald (2014). We take the square of 4 quarter-average of the conditional standard deviation vector resulting from the estimated GARCH. We then HP-filter these series with a smoothing parameter of 6.25.

B.4 Robustness Check

Sample	Firms with more than	1k	5k	10k	15k	20k
Model	Correlation in level	−0.50	−0.71	−0.64	−0.57	−0.48
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	Correlation in growth rate	−0.11	−0.35	−0.41	−0.42	−0.44
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Data	Correlation in (HP filtered) level	−0.36	−0.17	−0.34	−0.51	−0.46
		(0.005)	(0.20)	(0.008)	(0.000)	(0.000)
	Correlation in growth rate	−0.29	−0.21	−0.33	−0.43	−0.38
		(0.030)	(0.114)	(0.011)	(0.001)	(0.004)

Table 6: Correlation of tail estimate with aggregate output.

Note: The tail in the model is estimated for simulated data based on our baseline calibration (cf. Table 2) for an economy simulated during 20,000 periods. The tail in the data is estimated on Compustat data over the period 1958-2008. The aggregate output data is from the St-Louis Fed.

	(1) IQR of Real Sales (Compustat)	(2) STD of Pdy (Durables) (Kehrig 2015)	(3) IQR of real sales (Bloom et al. 2014)
Aggregate Volatility in TFP growth	0.2532	0.3636	0.3583
	(0.0825)	(0.0269)	(0.030)
Aggregate Volatility in GDP growth	0.1911	0.2923	0.3504
	(0.1932)	(0.079)	(0.034)

Table 7: Correlation of Dispersion and Aggregate Volatility

Note: In this table, we display the correlation of various measures of micro-level dispersion with two measures of aggregate volatility. Aggregate volatility is measured by the fitted values of an estimated GARCH on growth rates of TFP and output. Both are sourced from Fernald (2014) (see description above). In column (1) the Inter Quartile Range (IQR) of real sales is computed using Compustat data from 1960 to 2008 for manufacturing firms. Nominal values are deflated using the NBER-CES Manufacturing Industry Database 4-digits price index. In column (2) we take the establishment-level median standard deviation of productivity (levels) from Kherig (2015) who, in turn, computes it from Census data. In column (3) we take the establishment-level IQR of sales growth from Bloom at al. (2014).

C Numerical Solution Appendix

In this appendix, we describe the numerical algorithm used to solve the model described in the paper. Recall that given the Equation 2, T_t is a sufficient statistic to describe the wage. Using Equation 8, it is clear that the law of motion of T_t is a function of past values of T_t and D_t. As described in the main text, we are assuming that firms do not take into account the time-varying volatility of T_t and form their expectations by assuming that D_t is constant and equal to its steady-state value D. From the perspective of the firms, T_t does only depend on its past values.

It follows that the value of being a incumbent only depends on the aggregate state T. To solve the model we simply have to solve for the following Bellman Equation:

$$V(T, \varphi^s) = \pi^*(T, \varphi^s) + \max\left\{0, \beta \int_T \sum_{\varphi^{s'} \in \Phi} V(T', \varphi^{s'}) F(\varphi^{s'}|\varphi^s) \Upsilon(dT'|T)\right\}$$

where $\Upsilon(.|T)$ is the conditional distribution of next period's state T', given the current period state T. This distribution is given by Equation 8 with $D_t = D$. We also assume that the shock ε_{t+1} in this last Equation follows a standard normal distribution.

To solve for the above Bellman Equation we are using a standard value function iteration algorithm implemented in Matlab with the Compecon toolbox developed by Miranda and Fackler (2004). To do so, we define a grid for T (in logs) along with productivity grid of the idiosyncratic state space Φ described in the paper. We then form a guess on the value function as a function of $\log(T)$ and $\log(\varphi^s)$, and plug it to the right hand side of the above Bellman Equation. This is repeated until convergence. This algorithm converges to the solution of the above Bellman Equation and allows us to compute the policy function $s(T)$.