Large Firm Dynamics and the Business Cycle *

Vasco M. Carvalho†   Basile Grassi‡

May 2, 2017

[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 an analytical characterization of the law of motion of the aggregate state in this class of models – the firm size distribution – and show that 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 the editor, Mark Aguiar, two anonymous referees and Árpád Ábrahám, Axelle Arquié, Florin Bilbiie, David Backus, Paul Beaudry, Ariel Burstein, Jeff Campbell, Francesco Caselli, Larry Christiano, Gian Luca Clementi, Julian di Giovanni, Jan Eeckhout, Xavier Gabaix, Wouter den Haan, Hugo Hopenhayn, Jean Imbs, Oleg Itskhoki, Boyan Jovanovic, Pat Kehoe, Andrei Levchenko, Isabelle Mejean, Ezra Oberfield, Alexei Onatski, Michael Peters, Franck Portier, Morten Ravn, Esteban Rossi-Hansberg, Edouard Schaal, Rob Shimer, Gianluca Violante and Iván Werning. The authors thank seminar participants at Banque de France, Columbia Business School, CREi, CREST-INSEE, EUI, Essex, LSE, NY Fed, NYU, Minneapolis Fed, Princeton, Oxford, PSE, UCL, the SITE workshop at Stanford, the SED at Warsaw, the NBER-SI (EFCE), the CEPR Granularity workshop in Leuven 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 CEPR. `vmcarvalho.web@gmail.com`

‡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 an 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 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

[1] See Gabaix (2011), Acemoglu et *al.* (2012), di Giovanni and Levchenko (2012), Carvalho and Gabaix (2013), di Giovanni, Levchenko and Mejean (2014, 2017), Baqaee (2016), Baqaee and Fahri (2017) and Grassi (2017).

[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.

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 dynamics of aggregate productivity, up to the contribution of entry and exit. This characterization is key to our analysis and enables us to provide a analytical results on 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 productivity 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 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.

Taken together, our theoretical results also deliver a simple economic intuition for why large firm dynamics may drive the business cycle. Answering this question requires answering why, following an idiosyncratic shock to a very large firm (say, G.M.), its competitors (say, Ford) do not increase their scale and gain market share. If this were to happen, production would merely be reallocated from G.M. to Ford, rather than reduced in the aggregate. After all, if the shock is purely idiosyncratic to G.M., demand for auto-mobiles would be unaffected while primary input prices (e.g. wages in Detroit) should decline as G.M. sheds workers. What our results imply is that this reallocation effect should be second order: under a Pareto firm size distribution, the productivity gap between G.M. and Ford is large enough such that: (i) the shock to G.M. does have aggregate consequences and auto-worker wages do decline but, (ii) Ford is unproductive enough relative to G.M. such that, under decreasing returns to scale, the amount of reallocation is limited.[4, 5]

[3] There is a large literature assessing the empirical validity of Gibrat's law in data. Across all firms, the evidence on the relationship between firm size and firm growth is mixed. Hall (1987) and, more recently, Haltiwanger et al (2013), show that rejections of Gibrat's law in the data are attributable to the dynamics of small entrants and that there is no systematic relation between firm size and firm growth among large firms. See also Evans (1987) and the discussion in Luttmer (2010).

[4] That is, with Ford's productivity unchanged, and facing lower input costs, Ford would indeed increase its scale. However, in doing so, Ford's marginal productivity would decline given decreasing returns to scale thus limiting the amount of reallocation.

[5] Clearly, one could think of additional real frictions - such as imperfect substitutability between the goods of different firms or adjustment costs - that could stand in the way of such reallocation dynamics. Here we eschew this possibility

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 is 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. 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 forward-looking choices, given firm-specific productivity processes and (aggregate) factor prices. Further, the firm size distribution is an equilibrium object of our model.[6] 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 above 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.[7]

and explore a frictionless, worst-case, environment - firms produce a single homogeneous good and can adjust their scale freely.

[6] 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.

[7] To understand how we arrive to the 20% figure, note the following. di Giovanni, Levchenko, Mejean's (2014) headline finding is that "the standard deviation of the firm-specific shocks' contribution to aggregate sales growth amounts to 80% of the standard deviation of aggregate sales growth in the whole economy" (di Giovanni, Levchenko, Mejean (2014, p. 1304).

This paper is also related to the firm dynamics literature that follows from the seminal contribution of Hopenhayn (1992).[8] Some papers in this literature have explicitly studied aggregate fluctuations in a firm dynamics framework (Campbell and Fisher (2004), Lee and Mukoyama (2008,2015) , Clementi and Palazzo (2016) 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, 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.

But they also go on to further decompose this firm-specific contribution into two additional components: what they dub the 'direct' effect (i.e. purely idiosyncratic shocks to large firms) vs. the 'contribution of firm linkages' (i.e. covariance terms motivated by unobserved firm-to-firm linkages). On this latter decomposition, they find that the 'direct' component accounts 26% of firm-specific variability, (see di Giovanni, Levchenko, Mejean (2014, p. 1326)). Putting the first and the second decompositions together, the contribution of purely idiosyncratic shocks to large firms should be $0.26 \times 0.80 = 0.208$.

[8] 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).

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 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 $\Phi := \{\varphi^1, \ldots, \varphi^S\}$ 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.[9]

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

[9] 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.

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.

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 Φ. 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 aggregate state μ as $V^e(\mu,\varphi^q)$:

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

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, unless otherwise stated, 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.[10,11]

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 $Y_t = A_t^{1-\alpha}(L_t^d)^\alpha$ where L_t^d is the aggregate labor demand, the sum of all incumbents' labor demand in period t. Henceforth it will be convenient to differentiate between aggregate TFP, $A_t^{1-\alpha}$, and the term A_t itself. With some abuse of language, we refer to the latter as aggregate productivity, which is given by:

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

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:

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

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 $L^d(w_t) = \left(\frac{\alpha A_t^{1-\alpha}}{w_t}\right)^{\frac{1}{1-\alpha}}$. Note that the model behaves as a one factor model with aggregate TFP $A_t^{1-\alpha}$.

The market clearing condition then equates labor supply and labor demand, i.e. $L^s(w_t) = L^d(w_t)$.

[10] 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.

[11] By assuming an exogenous function for labor supply, we restrict our analysis to partial equilibrium. Note however that we do not take wages as given. Rather, the wage w will adjust to clear the labor market. Formally our reduced form labor supply function is equivalent to a general equilibrium setup where household preferences are linear in consumption and concave in labor disutility, the instantaneous utility $U(C,L) = C - M^{\frac{-1}{\gamma}} L^{1+\frac{1}{\gamma}}/(1+\frac{1}{\gamma})$ delivers the assumed labor supply function. A version of this model with concavity in consumption is certainly feasible computationally, by following the methodology developed in Khan and Thomas (2008) and Bachmann, Caballero and Engel (2013).

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:

$$Y_t = A_t^{1-\alpha} L_t^{\alpha} \tag{3}$$

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.

Finally note that if, as we will show below, idiosyncratic shocks to large firms do lead to variation of μ_t over time, then this will induce variations in the equilibrium wage as instructed by Equation 2. This, in turn, will lead to reallocation of economic activity across firms. To see this, define the elasticity of output of a given firm i to an idiosyncratic productivity change in another firm j:

$$\frac{\partial \log y_i}{\partial \log \varphi^{s_{j,t}}} = \frac{\partial \log y_i}{\partial \log w_t} \frac{\partial \log w_t}{\partial \log \varphi^{s_{j,t}}} = -\frac{\alpha}{1-\alpha} \frac{\partial \log w_t}{\partial \log \varphi^{s_{j,t}}} \tag{4}$$

Intuitively, the steeper decreasing returns to scale are - i.e. the steeper the decline in the marginal productivity of firm i as it expands, following a decline in wages - the lower this elasticity is. Thus, for a given wage response elasticity to large firm shocks, the strength of this countervailing reallocation effect is lower the smaller α is.

We are left to understand the behavior of the second component of this elasticity, i.e. the response of wages to idiosyncratic shocks. This, in turn, implies understanding how the aggregate state, the firm-size distribution, evolves over time as a function of idiosyncratic shocks. We address this 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.

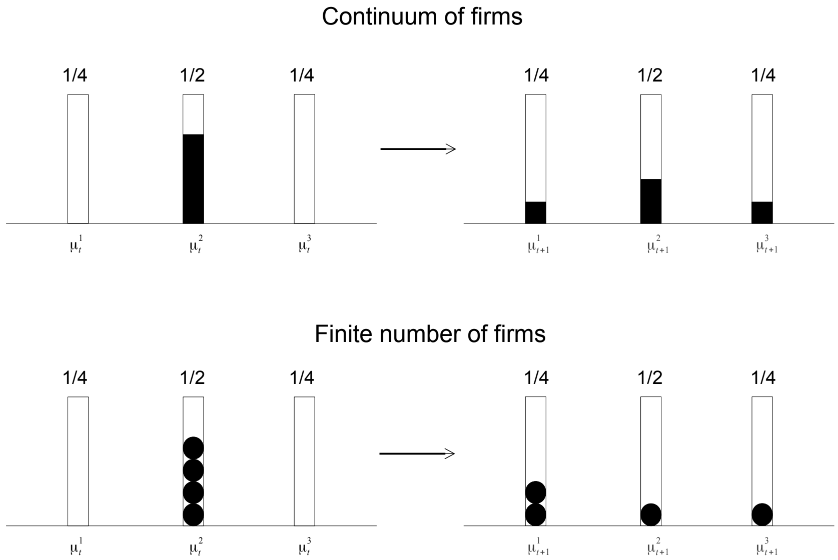

Figure 1: Why the vector μ_{t+1} follows a multinomial distribution.

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

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 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:*

$$\mu_{t+1} = m(\mu_t) + \epsilon_{t+1} \tag{5}$$

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 See Appendix A.1. □

After taking into account the dynamics of incumbent firms and entry/exit decisions, the law of motion (Equation 5) 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 5 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 5. 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.

Steady-State Equilibrium and the Stationary Distribution

The above characterization - in particular, Equation 5 - 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 particular, it is immediate to verify that, under a continuum of firms, the variance-covariance matrix in Theorem 1 is equal to zero and the aggregate state μ_t becomes non-stochastic. The following corollary to Theorem 1 shows that the deterministic dynamics of the productivity distribution under no aggregate uncertainty are similar to the one in Hopenhayn (1992) framework.

Corollary 1 *Define $\hat{\mu}_t := \frac{\mu_t}{M}$ for any t. With a continuum of firms, or equivalently, when aggregate uncertainty is absent, $\epsilon_{t+1} = 0$:*

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

where $\widetilde{P}_t$ is the transition matrix P where the first $\widetilde{s}(\mu_t) - 1$ rows are replaced by zeros, and where $\widetilde{s}(\mu_t)$ is the threshold of the entry and exit 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 sides by M. □

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, this law of motion converges to a self-reproducing distribution. This defines the deterministic steady-state equilibrium of our model where (i) the wage is constant and (ii) conditional on a value for firm-level productivity, the value and policy functions that solve the firms' problem are constant. Following Hopenhayn's (1992), we dub the distribution of firm-productivity levels which obtains at the steady-state as the stationary distribution. In our setting, this is defined as:

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

where $\widetilde{P}$ is the transition matrix P where the first $(s^*(\hat{\mu}) - 1)$ rows are now replaced by zeros to account for equilibrium entry and exit dynamics and where s^* is the steady-state value of the entry/exit thesholds. Note that this distribution does not depend on time t. Despite the presence of idiosyncratic shocks – implying firms transiting across productivity states and eventual exit – the mass of firms at each productivity level is constant.

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 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 aggregate productivity up to the policy function on firm entry and exit. 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.

We start by specializing 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 (Córdoba 2008) *For a given firm i at time t with productivity level $\varphi^{s_{i,t}}$ with $s_{i,t} \neq 1, S$ that follows 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.[12] 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 constant absolute size of the largest firms. As di Giovanni and Levchenko (2012) show this is counter-factual: in cross-country data, whenever the size of the economy increases, the absolute size of the top 10 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 $\varphi^S = ZN^{1/\delta}$, for some constant Z.*

[12]Note that, due to the fact that we have a bounded state space for productivity, Gibrat's law cannot apply for firms at the upper and lower boundaries of the productivity process. Thus, the conditional expected productivity growth rate is higher (resp. lower) at the lowest (resp. highest) productivity level than at any other productivity levels: $\mathbb{E}\left[\frac{\varphi^{s_{i,t+1}} - \varphi^{s_{i,t}}}{\varphi^{s_{i,t}}} | \varphi^{s_{i,t}} = \varphi^1\right] = c(\varphi - 1) > a(\varphi^{-1} - 1) + c(\varphi - 1) > a(\varphi^{-1} - 1) = \mathbb{E}\left[\frac{\varphi^{s_{i,t+1}} - \varphi^{s_{i,t}}}{\varphi^{s_{i,t}}} | \varphi^{s_{i,t}} = \varphi^S\right]$.

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 firms of size N. 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.

4.1 Steady State Equilibrium Characterization

With the above two assumptions in place, we now provide a detailed description of the steady-state equilibrium. We start by characterizing further the stationary distribution in Corollary 1, which we are now able to solve in closed-form. We then present a full solution of the firm's problem by deriving the policy and value functions in the steady-state.

First, in Corollary 2 below, we study the limiting case when the number of firms goes to infinity under Assumptions 1 and for $S \to \infty$.[13]

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

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

where $\overline{s^*}$ is the steady-state entry/exit thesholds for $S \to \infty$ and where $\delta = \frac{\log(a/c)}{\log(\varphi)}$ and K_1 and K_2 are constants, independent of s.

Proof: See Online Appendix B.3 □.

Thus, the stationary productivity distribution for surviving firms (i.e. for $s \geq \overline{s^*}$), is a mixture of two Pareto distributions: (i) the stationary distribution of the Markovian process given by Assumption 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, $\overline{s^*}$, as in Luttmer (2007, 2010, 2012) . Unlike Luttmer however, our entrants can enter at every productivity

[13] 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 productivity bins, S, goes to infinity.

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: $y_s = (\varphi^s)^{\frac{1}{1-\alpha}} \left(\frac{\alpha}{\overline{w}}\right)^{\frac{\alpha}{1-\alpha}}$ where $\overline{w}$ is the limit of the steady-state value of the wage when S goes to infinity. 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 $\min\{\delta(1-\alpha), \delta_e(1-\alpha)\}$.[14] 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 $\delta(1-\alpha)$. Note that $\delta(1-\alpha)$ is an observable quantity since it is the tail of the firm size distribution (for sufficiently large firms and as soon as $\delta_e < \delta$). We are using this feature in the calibration of the quantitative section of this paper.

The result above characterizes the firm productivity and firm size distribution in the steady-state, thus pinning down the aggregate state up to $\overline{s^*}$, the firm's policy function. In what follows, we now solve for this policy function along with the associated value function for incumbents, thus presenting a full characterization of the steady-state equilibrium.[15]

Proposition 1 *Under Assumption 1, when $S \to \infty$ and φ is small enough, then in the stationary equilibrium the value function of a firm facing productivity level φ^s and wage $\overline{w}$ is equal to*

$$V(\mu, \varphi^s) = \frac{-c_f}{1-\beta}\left[1 - \beta r_2^{[s-\overline{s^*}+1]^+}\right] + \frac{1-\alpha}{1-\rho\beta}\left(\frac{\alpha}{\overline{w}}\right)^{\frac{\alpha}{1-\alpha}}\left(\varphi^{\frac{1}{1-\alpha}}\right)^s\left[1 - \rho\beta\left(\frac{r_2}{\varphi^{\frac{1}{1-\alpha}}}\right)^{[s-\overline{s^*}+1]^+}\right]$$

where $[x]^+ = \text{Max}(x, 0)$ and r_2 is a constant defined in the appendix bounded above by one. The policy function is characterized by the threshold

$$\overline{s^*} = \left\lceil (1-\alpha)\log\left(\frac{c_f(1-r_2)(1-\rho\beta)}{\rho(1-\beta)(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}}(1-r_2\varphi^{\frac{-1}{1-\alpha}})}\right)(\log\varphi)^{-1} + \alpha(\log\overline{w})(\log\varphi)^{-1}\right\rceil$$

where $\lceil x \rceil$ is the ceiling function, i.e. the least succeeding integer of x.

Proof: See Online Appendix B.2 □.

For a sketch of the proof note that, under Assumption 1, the next period's value of (surviving) incumbents can only take one of three values, depending on their idiosyncratic productivity realization.

[14] 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.

[15] The steady-state equilibrium is given by the expressions in Proposition 1 and Equation 2, relating the equilibrium wage to aggregate productivity, itself a function of $\overline{s^*}$.

This implies that solving for the value function is equivalent to solving a second order difference equation. The constant r_2 in the proposition is the relevant solution of this equation.[16]

Intuitively, the value function of an incumbent is simply the present value of instantaneous profits adjusted by exit risk. To see this, note that the first term of the value function reflects the present discounted value of fixed operating costs. The second term reflects the present discounted value of variable profits. In turn, each of these terms are expressed as the product of (i) the present discounted value that would obtain in a world without exit and (ii) an adjustment for the exit risk, encoded in the square brackets terms. To understand the latter, note that the exit probability of an incumbent currently in a high idiosyncratic productivity bin is low. In this case, the term in square brackets is close to 1, i.e. the exit-risk adjustment of future profits is small. The converse holds for a low productivity incumbent.

The second result in the proposition gives the policy function for entry and exit decisions. A larger operating fixed cost, c_f, or a larger equilibrium wage, $\overline{w}$, increase the value of the entry/exit threshold, $\overline{s^*}$. Intuitively, either will reduce the value of being an incumbent and thus rendering it more likely that a firm exits or a potential entrant declines to enter.

As we will see below in Section 4.3 much of this intuition carries through in a special case where the aggregate state is time-varying.

4.2 A Complete Analytical Characterization for an Economy without Entry and Exit

Having described the steady-state equilibrium, we are now interested in characterizing the dynamics of aggregate productivity and output under aggregate uncertainty. In this section, we start by analysing the simpler case of an economy without entry and exit. This allows us to derive the law of motion of the aggregate state analytically and, as a byproduct, closed form expressions for aggregate output dynamics which exhibit persistence and time-varying volatility. As we shall see, our key results will generalize to the case with entry and exit.

Economies without entry and exit are a special case of the setup introduced in Section 2, when fixed operating costs, c_f, are zero and entry costs, c_e, are large enough. In this case, firms always have a positive present discounted value of profits, irrespective of their current idiosyncratic productivity draw and never choose to exit. For a large enough entry cost , no potential entrant will choose to start producing either, irrespective of the current aggregate state and idiosyncratic productivity signal. Therefore, the total number of firms is fixed at N, which is now a parameter of the model rather than an endogenous variable to be solved for.

[16]For the finite S case, this implies the existence of two polynomial roots, r_1 and r_2. In the case where $S \to \infty$, only the latter root is relevant when solving for the value function. We defer presenting the more general solution of the value function for a finite S case to the Online Appendix B.2 and focus on the $S \to \infty$ case in the main text.

Without entry and exit, we start by noting that the law of motion for the aggregate state, i.e. the productivity distribution, is a special case of Theorem 1 where Equation 5 is now:

$$\mu_{t+1} = P'\mu_t + \epsilon_{t+1} \tag{8}$$

Recall that P is the transition matrix for firm-level productivity and ϵ_{t+1} is a random vector with mean zero and variance-covariance matrix $\Sigma(\mu_t) = \sum_{s=1}^{S} \mu_{s,t} W_s$ and $W_s = diag(P_{s,.}) - P'_{s,.}P_{s,.}$ where $P_{s,.}$ denotes the s^{th}-row of the transition matrix P. Without endogenous exit, relative to Theorem 1, the transition matrix across productivity states is no longer time-varying and, under no entry, we no longer have to keep track of the contributions of entrants to the law of motion of the aggregate state. However, the variance of the ϵ_{t+1} is still time-varying as it remains a function of the lagged realization of the productivity distribution.

As Equation 8 shows, the law of motion of the aggregate state μ_t is stochastic, i.e. μ_t is a random vector. The following proposition describes the (unconditional) behavior of this random vector and thus the stochastic properties of the aggregate state variable of the model with no entry/exit.

Proposition 2 *For the no entry and exit case and under Assumption 1, the unconditional mean of μ_t is $\mu = (\mu_1, \ldots, \mu_s, \ldots, \mu_S)'$ and is given by*

$$\mathbb{E}\left[\mu_{s,t}\right] = \mu_s = N\frac{1-\varphi^{-\delta}}{\varphi^{-\delta}(1-(\varphi^S)^{-\delta})}\left(\varphi^s\right)^{-\delta}$$

where $\delta = \frac{\log(a/c)}{\log(\varphi)}$. Furthermore, the unconditional variance-covariance matrix of μ_t is

$$\mathbb{V}ar\left[\mu_t\right] = \sum_{k=0}^{\infty}(P')^k\left(\sum_{s=1}^{S}\mu_s W_s\right)P^k$$

where P is the transition matrix for firm-level productivity, and, $W_s = diag(P_{s,.}) - P'_{s,.}P_{s,.}$ where $P_{s,.}$ denotes the s^{th}-row of the transition matrix P in Assumption 1.

Proof: See Online Appendix B.4 □.

This proposition shows that the aggregate state, μ_t, fluctuates around its mean. In turn, this mean is the stationary distribution of firm productivity, μ, which is Pareto distributed with tail index δ. Note that the latter is simply a particular case of Corollary 2 for the case with no entry and exit. The second part of the proposition shows that deviations of μ_t from μ are governed by the variance-covariance matrix, $\mathbb{V}ar\left[\mu_t\right]$. The latter takes the conditional variance-covariance matrix of μ_t and adjusts it by the persistence in the law of motion of μ_t, as given by the transition matrix P.[17]

[17] To see why this is the case, it is useful to draw an analogy with the unconditional variance of the following univariate process: $x_{t+1} = (1-\rho_x)\bar{x} + \rho_x x_t + \sqrt{x_t}\sigma_u u_{t+1}$ where $\bar{x} = \mathbb{E}x_t$, $0 < \rho_x < 1$ and u_t is a white noise independent of x_t. It is easy to show that $\mathbb{V}ar\left[\mu_t\right] = \frac{\bar{x}\sigma_u^2}{1-\rho_x^2} = \sum_{k=0}^{\infty}\rho_x^{2k}\bar{x}\sigma_u^2$. The formula of the variance-covariance matrix, $\mathbb{V}ar\left[\mu_t\right]$, in Proposition 2 is a multivariate generalization of this simple univariate process, where W_s plays the role of σ_u^2, P the role of ρ_x and μ the role of $\bar{x}$.

The above discussion shows that the firm productivity distribution fluctuates stochastically around a Pareto distribution. We now build on this characterization to explore the dynamics of aggregate productivity and output. We start by deriving the law of motion of aggregate productivity:

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

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

Theorem 2 *Assume 1, then aggregate productivity dynamics is given by*

$$A_{t+1} = \rho A_t + O_t^A + \sigma_t \varepsilon_{t+1} \tag{9}$$

$$\sigma_t^2 = \varrho D_t + O_t^\sigma \tag{10}$$

Furthermore, aggregate output dynamics (in percentage deviation from its steady-state value, $\widehat{Y}_t$) is given by:

$$\widehat{Y}_{t+1} = \rho \widehat{Y}_t + \kappa \widehat{O_t^A} + \psi \frac{\sigma_t}{A} \varepsilon_{t+1} \tag{11}$$

where $\mathbb{E}[\varepsilon_{t+1}] = 0$ and $\mathbb{V}ar[\varepsilon_{t+1}] = 1$. The parameter $\rho = a\varphi^{\frac{-1}{1-\alpha}} + b + c\varphi^{\frac{1}{1-\alpha}}$ and $\varrho = a\varphi^{\frac{-2}{1-\alpha}} + b + c\varphi^{\frac{2}{1-\alpha}} - \rho^2$. The term D_t is given by $D_t := \sum_{s=1}^{S} \left((\varphi^s)^{\frac{1}{1-\alpha}} \right)^2 \mu_{s,t}$. The terms O_t^A and O_t^σ are a correction for the upper and lower reflecting barriers in the idiosyncratic state space. A is the steady-state value of the aggregate productivity A_t. κ is a constant defined in the Appendix, and, $\psi = \left(1 - \frac{\alpha}{\gamma(1-\alpha)+1}\right)$ is such that $\widehat{Y}_t = \psi \widehat{A}_t$ where $\widehat{A}_t$ (resp. $\widehat{O_t^A}$) is the percentage deviation from steady-state of A_t (resp. O_t^A).

Proof: See Appendix A.2 for a proof sketch and Online Appendix B.5 for a formal proof. □

Theorem 2 provides a full description of aggregate productivity and aggregate output dynamics in our model. The theorem states that the aggregate productivity tomorrow is the sum of (i) ρA_t, the expected aggregate productivity and (ii) $\sigma_t \varepsilon_{t+1}$ a mean zero aggregate productivity shock with time-varying volatility as instructed by the term D_t. The term O_t^A 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 productivity, it is straightforward to characterize the law of motion for equilibrium output, which is given by the second part of the theorem. The dynamic properties of A_t, persistence and time-varying volatility, carry through to the percentage deviation of aggregate output from its steady-state value, $\widehat{Y}_t$. This is not surprising, as in the aggregate our model behaves as a one factor RBC model.

Theorem 2 implies that aggregate productivity and aggregate output are persistent and exhibit time-varying volatility. Intuitively, since there are no aggregate shocks in our model, aggregate persistence simply reflects firm-level persistence in productivity. To understand the time-varying nature of aggregate volatility it is easiest to consider an extreme case with only two firms. In that case, aggregate productivity would simply reflect the weighted sum of the two firms' productivity. The volatility of the aggregate productivity would then depend on (the square of) the relative size of these two firms, which will be time-varying as they are subject to independent shocks. The term D_t in Theorem 2 gives the generalization of this intuitive argument for a large but finite number of firms.

To better understand this result, and building on the expressions in Theorem 2, 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.

To understand the persistence of aggregate output, ρ, note that, at the steady-state, ρ satisfies $\mathbb{E}_t\left[y_{i,t+1}|y_{i,t}\right] = \rho y_{i,t}$ for a given firm i. That is, aggregate persistence is nothing but firm-level persistence.[18] Building on this, the following proposition further characterizes how aggregate persistence depends on parameters governing firm-level dynamics.

Proposition 3 . *Let $\delta = \frac{\log \frac{a}{c}}{\log \varphi}$ be the tail index of the stationary productivity distribution as in Corollary 2. If $\delta(1-\alpha) \geq 1$ then the persistence of the aggregate output, 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 productivity dynamics contain a unit root: if $\delta = \frac{1}{1-\alpha}$, $\rho = 1$*

Proof: See Online Appendix B.6 $\square$.

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.[19]

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

[18] Note that from census firm-level data, this a potentially observable object.

[19] 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.

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:

$$\mathbb{V}ar_t\left[\widehat{Y}_{t+1}\right] = \psi\frac{\sigma_t^2}{A^2} = \psi\varrho\frac{D}{A^2}\frac{D_t}{D} + \psi\frac{O^\sigma}{A^2}\frac{O_t^\sigma}{O^\sigma} \tag{12}$$

where A and D are the the steady-state counterparts of A_t and D_t. To interpret these objects, first recall that $D_t = \sum_{s=1}^{S}\left((\varphi^s)^{\frac{1}{1-\alpha}}\right)^2\mu_{s,t}$ is proportional to the second moment of the firm size distribution at time t, a well defined measure of dispersion.[20] D is therefore proportional to the steady-state dispersion in firm size. Finally, note that at the steady-state, ϱ satisfies $\varrho = \mathbb{V}ar\left[\frac{y_{i,t+1}}{y_{i,t}}\right]$, that is, ϱ is given by the variance of firm-level output growth at the steady-state.[21] Thus, aggregate volatility is simply firm-level volatility scaled by the current level of dispersion in the economy. As the latter varies over time, so does aggregate volatility.

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

$$\mathbb{E}\frac{\sigma_t^2}{A^2} = \psi\varrho\frac{D}{A^2} + \psi\frac{O^\sigma}{A^2}$$

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

Proposition 4 *Let $\delta = \frac{\log\frac{a}{c}}{\log\varphi}$ be the tail index of the stationary productivity distribution as in Corollary 2, then*

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

$$\mathbb{E}\left[\frac{\sigma_t^2}{A^2}\right] \underset{N\to\infty}{\sim} \frac{\varrho G_1}{N^{2-\frac{2}{\delta(1-\alpha)}}} \tag{13}$$

where G_1 is a function of model parameters but independent of N.

$ii)$ *The dynamics of conditional variance of aggregate output depends on the dispersion of firm size:*

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

Proof: See Online Appendix B.8 □.

[20] 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.

[21] Note that as for ρ, this a potentially oberservable object.

Part (i) of Proposition 4 characterizes the average level of volatility of aggregate output 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. The assumption on $\delta(1-\alpha)$ ensures that this distribution is 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 firms, at a rate that is equal to the rate of the expression on the right hand side.

The key conclusion of the first part of Proposition 4 is therefore that, for $1 < \delta(1-\alpha) < 2$, the variance of aggregate output 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 arguments. 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 index of the size distribution of firms. The closer is this distribution 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 size decisions, 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.

Part (ii) of Proposition 4 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, 2015). 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 exogenous aggregate shocks influencing the first and second moments of firms' growth.

4.3 Aggregate Persistence and Volatility in Economies with Entry and Exit

We now extend the characterization of aggregate persistence and volatility to economies with endogenous, forward-looking, firm entry and exit decisions. Relative to the no entry and exit case discussed above, the law of motion for the firm productivity distribution now depends on the current realization of productivity signals across entrants and a potentially time-varying entry and exit threshold. The latter implies that, without further assumptions, we can no longer solve for the full solution of the model, which now includes a policy function describing how firms' entry and exit decisions depend on the aggregate state. Despite this, we are able to show that our analytical expressions for the persistence and volatility of aggregate output generalize to the case with entry and exit. In particular, whenever the entrant productivity signal distribution is thinner tailed than the productivity distribution of incumbents, we will show that the contribution of entry and exit to aggregate volatility is second-order.

To show this, we follow the same steps as in Section 4.2. Recalling the general law of motion for the aggregate state variable in Theorem 1,

$$\mu_{t+1} = (P_t^*)'(\mu_t + MG) + \epsilon_{t+1} \tag{14}$$

makes clear that the law of motion of the aggregate state depends on the distribution of productivity signals for potential entrants, G, and the current (endogenous) threshold for entry and exit decisions, $s^*(\mu_t)$. This is because P_t^*, the transition matrix of firms across productivity states, also encodes entry and exit transitions.

Specializing this setup to the Gibrat's law case (i.e. Assumption 1), it is then straightforward to show that this general law of motion of the aggregate state implies the following dynamics for aggregate productivity:

$$A_{t+1} = \rho A_t + \rho E_t(\varphi) + O_t^A + \sigma_t \varepsilon_{t+1} \tag{15}$$

$$\sigma_t^2 = \varrho D_t + \varrho E_t(\varphi^2) + O_t^\sigma \tag{16}$$

Relative to Theorem 2, the contribution of entry and exit appears in the net entry terms, E_t. These terms give the difference between the entry and exit contributions for both the level and the volatility of aggregate productivity with $E_t(x) = \left(M \sum_{s=s^*(\mu_t)}^{S} G_s \left(x^s\right)^{\frac{1}{1-\alpha}}\right) - \left(\left(x^{s^*(\mu_t)-1}\right)^{\frac{1}{1-\alpha}} \mu_{s^*(\mu_t)-1,t}\right)$, where x is either φ or φ^2.[22] Intuitively, the first term of this expression gives the contribution of today's entrants while the second term gives that of exiters. As a result, aggregate productivity of incumbents tomorrow, A_{t+1}, now depends on $\rho E_t(\varphi)$, the expected aggregate productivity of today's net entrants, conditional on their survival.

It is worth noting that entry and exit decisions do not alter the fact that, as in Section 4.2, aggregate productivity is persistent and exhibits time-varying volatility. Furthermore, given this law of motion for aggregate productivity, it is immediate to show that the law of motion for aggregate output inherits these same properties:

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

where κ_1 and κ_2 are constant defined in the appendix.[23] In particular, the parameter ρ in the above equation, is the same as that defined in Theorem 2, for the no entry and exit case. It follows that the persistence of aggregate productivity and output, depends on deep parameters governing firm-level dynamics in the same way as in Proposition 3.

[22] In particular note that the entry term affecting volatility, $E_t(\varphi^2) = M \sum_{s=s^*(\mu_t)}^{S} ((\varphi^s)^{\frac{1}{1-\alpha}})^2 G_s - ((\varphi^{s^*(\mu_t)-1})^{\frac{1}{1-\alpha}})^2 \mu_{s^*(\mu_t)-1,t}$, is proportional to dispersion of firm size among successful entrants, where the last term in the expression corrects for exit, and is proportional to the dispersion in the size of exiters.

[23] This equation is derived in the proof of Theorem 3 in Online Appendix B.5.

Turning to the volatility of aggregate output, we are now able to generalize Proposition 4 to the case of entry and exit as follows:

Proposition 5 *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:*

$$\mathbb{E}\left[\frac{\sigma_t^2}{A^2}\right] \underset{N\to\infty}{\sim} \frac{\varrho G_1}{N^{2-\frac{2}{\delta(1-\alpha)}}} + \frac{\varrho G_2}{N^{1+\frac{\delta_e}{\delta}-\frac{2}{\delta(1-\alpha)}}} \tag{17}$$

where G_1 and G_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:*

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

Proof: See Online Appendix B.8 □.

Relative to the simpler case without entry and exit, we now need to take a stand on the distribution of productivity signals across entrants. We follow our earlier approach in Corollary 2 and assume it to be Pareto distributed with tail parameter δ_e. With this assumption in place, the proposition implies that the characterization of aggregate volatility given for the no entry and exit case carries through to the current, more general setup.

Thus, part (ii) of the proposition remains unchanged, while part (i) of the proposition shows that the variance of aggregate output fluctuations still declines at a slower rate than $1/N$. However, relative to the no entry and exit case, this rate of decay now depends on the tail indexes of the size distributions of both incumbents ($\delta(1-\alpha)$) *and* entrants ($\delta_e(1-\alpha)$) . In particular, whenever the size distribution of incumbents has a lower tail index than the size distribution of entrants - i.e. whenever $\delta < \delta_e$ - the first term in the expression for aggregate volatility dominates. In this case, asymptotically, the rate of decay of aggregate volatility will be a function of the tail index of incumbents, at the first order, and we recover the result in Proposition 4.

Intuitively, whenever the size distribution of incumbents is close to Zipf's law (i.e. $\delta(1-\alpha)$ is close to 1) and the probability of observing very large entrants is small, aggregate volatility depends on large incumbent firms alone. This implies that, despite the fact that we cannot solve for entry dynamics explicitly, we can still describe the behaviour of aggregate volatility, as the contribution of entry and exit is second order and it suffices to track the dynamics of large incumbents. Further, as we will argue below, this is likely to be the empirically relevant regime, as both conditions (fat-tailed incumbents and relatively thinner tailed entrants) are met in the data.

Taken together, these results imply that the voluminous literature building 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.

4.3.1 Policy and Value Functions: A Special Case

As discussed above, with endogenous entry and exit decisions, and without further assumptions, it is not possible to make further headway analytically. This is because entry and exit decisions are forward looking and firms therefore need to forecast the future productivity distribution or, equivalently, future wages.

Despite the fact that, as shown in the previous section, our results regarding aggregate persistence and volatility do not depend on the extensive margin, it is nevertheless useful to understand entry and exit decisions on two counts. First, the behavior of the extensive margin following large firm shocks is of independent interest. Second, a quantitative evaluation of the general model requires a solution to these forward looking decisions.

Technically, the key challenge in solving for value and policy functions in our model is that these are nonlinear functions of the aggregate state variable μ_t. While not infinite dimensional (as in Hopenhayn, 1992), this is still a large dimensional object which we cannot solve for analytically.[24] Our proposed solution, following most of the literature on heterogeneous agents, is to reduce the dimensionality of the state-space.

Unlike most of the literature, we do know the form that the law of motion for A_t takes. This per se, does not solve our problem since E_t and σ_t are still functions of μ_t. Mathematically, this means that A_t is not a recursive map: A_t maps to past values of itself but also to other moments of the productivity distribution. However, it becomes a recursive map if we make the following assumption.

Assumption 3 *Assume that, when forming expectations about future wages, firms take $\frac{E_t}{A_t}$, $\frac{O_t^A}{A_t}$ and $\frac{\sigma_t}{A_t}$ to be constant and equal to $\frac{E}{A}$, $\frac{O^A}{A}$ and $\frac{\sigma}{A}$, respectively, i.e. their stationary equilibrium counterparts.*

In other words, when forecasting future wages, firms ignore (i) the time-varying contribution of net entry to aggregate productivity and (ii) the time-varying nature of aggregate volatility. These admittedly strong assumptions allow us to make progress analytically by rendering the wage forecasting

[24] In the baseline calibration of Section 5, the dimension of the aggregate state variable, μ_t, is pinned down by $S = 35$, large enough to render the problem computationally infeasible.

problem tractable.[25] This in turn, delivers closed form expressions for the value and policy functions, which we summarize in the following proposition.

Proposition 6 *Assume 1 and 3, when $S \to \infty$, the value of an incumbent firm under aggregate state μ_t and idiosyncratic productivity level $\varphi^s \geq \varphi^{s^*(\mu_t)}$ is:*

$$V(\mu_t, \varphi^s) = \frac{-c_f}{1-\beta}\left[1 - \frac{a\beta}{\widetilde{r_2}\left(\frac{1}{\beta} - \frac{1}{\widetilde{\beta}_2}\right) + a}\widetilde{r_2}^{\,s-s^*(\mu_t)+1}\right]$$
$$+ \frac{1-\alpha}{1-\rho\widetilde{\beta}_\alpha}\left(\frac{\alpha}{w_t}\right)^{\frac{\alpha}{1-\alpha}}\left(\varphi^{\frac{1}{1-\alpha}}\right)^s \left[1 + \frac{-\rho\beta a \varphi^{\frac{-1}{1-\alpha}}\widetilde{\beta}_\alpha + \rho\beta - \rho\widetilde{\beta}_\alpha}{\beta\left(\widetilde{r_2}\left(\frac{1}{\beta} - \frac{1}{\widetilde{\beta}_2}\right) + a\right)}\left(\varphi^{\frac{1}{1-\alpha}}\right)\left(\frac{\widetilde{r_2}}{\varphi^{\frac{1}{1-\alpha}}}\right)^{s-s^*(\mu_t)+1}\right]$$

with the corresponding policy function is given by by

$$s^*(\mu_t) = \left\lceil (1-\alpha)\log\left(\frac{c_f\left(1 - \frac{a}{\widetilde{r_2}\left(\frac{1}{\beta}-\frac{1}{\widetilde{\beta}_2}\right)+a}\frac{\beta}{\widetilde{\beta}_2}\widetilde{r_2}\right)(1-\rho\widetilde{\beta}_\alpha)}{\rho(1-\beta)(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}}\left(1 + \frac{-\beta a\varphi^{\frac{-1}{1-\alpha}}\widetilde{\beta}_\alpha - \widetilde{\beta}_\alpha + \beta}{\beta\left(\widetilde{r_2}\left(\frac{1}{\beta}-\frac{1}{\widetilde{\beta}_2}\right)+a\right)}\frac{\widetilde{r_2}}{\widetilde{\beta}_2}\right)}\right)(\log\varphi)^{-1} + \alpha(\log w_t)(\log\varphi)^{-1}\right\rceil$$

where $\lceil x \rceil$ is the ceiling function i.e. the least succeeding integer of x and $\widetilde{\beta}_\alpha$, $\widetilde{\beta}_2$ and $\widetilde{r_2}$ are constants defined in the Online Appendix B.9.

Proof: See Online Appendix B.9 □.

The proof of Proposition 6 follows a guess and verify strategy. It is easiest to understand this solution to the firms' problem by first noting the similarities with the steady-state solution given in Proposition 1. In particular, note that the value and policy functions assume the same form as before. The first term of the value function again reflects the present value of fixed operating costs, while the second term reflects the expected present discounted value of profits. Again, we see an adjustment for exit risk entering in the same form on both terms. Regarding the policy function, as before, the first term indicates the contribution of fixed cost considerations for the entry and exit decisions while the second term reflect variable profits. Indeed, it is easy to show that, if wages were fixed across time (i.e. if there was no aggregate uncertainty), the terms $\widetilde{\beta}_\alpha$=$\widetilde{\beta}_2$=β such that the value and policy functions in this proposition would exactly coincide with those given in the steady-state equilibrium.

However, the solution for both the value and policy functions now reflect the time-varying nature of aggregate risk. That is, even when holding firm-level productivity fixed, the value and entry and exit decisions of a firm are time-varying and depend on the current realization of wages. In the value function, aggregate risk appears in two intuitive ways. First, under aggregate uncertainty, the

[25] Nevertheless, note that in the data (and in our quantitative exercise below) the average year-on-year contribution of net entrants to aggregate productivity is small to begin with (see, for example, Foster et al. (2008) or Osotimehin (2016)). Relatedly, based on census data for French firms, Osotimehin (2016), concludes that time variation in entry and exit contributes little to the year-year variability of productivity.

threshold for entry and exit will now be time-varying, as instructed by the policy function. As such, the terms adjusting firm value for exit risk, now depend on the current value of this threshold (which, in turn, depends on the current realization of wages). Second, the expected present discounted value of profits is also time-varying and depends on the current realization of wages.

The policy function also reflects aggregate risk. Thus, for the same value of firm-level productivity, if the current realization of wages is high, the threshold for entry and exit increases. This is intuitive: as shown above aggregate productivity (and, hence, aggregate wages) is persistent and therefore, following a positive shock that increases the current value of wages, firms expect the cost of variable inputs to remain high in the future. This implies that the expected present value of profits is now lower which, in turn, implies that even relatively higher productivity firms might now find it optimal to exit (or not enter).

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 (Assumption 1), which we have discussed in the previous section. We first calibrate the 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 productivity and Assumption 3 to solve numerically for the firms' policy function. 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).[26] The discount factor β is set at 0.95 so that the implied annual gross interest rate is 5.3%, 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.[27]

[26] In the Online Appendix we explore an alternative calibration strategy where α is calibrated rather than fixed. We find that qualitatively all results shown below go through and, if anything, imply that large firm dynamics account for a larger share of aggregate volatility.

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

Statistic	Model	Data	References
Entry Rate	0.109	0.109	BDS firm data
Idiosyncratic Vol. σ_e	0.08	$0.1 - 0.2$	See main text
Tail index of Firm size dist.	1.097	1.097	BDS firm data
Tail index of Entrant Firm size dist.	1.570	1.570	BDS firm data
Share of Employment of the largest firm	0.14%	1%	Share of Wall-Mart
Number of Firms	4.5×10^6	4.5×10^6	BDS firm data

Table 1: Targets for the calibration of parameters

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. This will also imply a Pareto distribution for firm size in the steady-state equilibrium as shown in Corollary 2.

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 C.

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 $\delta_e(1 - \alpha) = 1.570$. 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 estimate for incumbents compares 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 the persistence of firm-level productivity, b. We calibrate these parameters jointly to match the volatility of firm-level productivity growth rates, $\sigma_e = 8\%$, and a steady-state entry rate, i.e. the ratio of entrants to incumbents, of 10.9%. Our choice for firm-level productivity volatility is at the low end of the values reported in the literature. Working with Compustat data, Gabaix (2011) reports a value for the volatility of labor productivity (measured as sales per employee) of 12%. Based on establishment-level census data, Foster et al (2008) and Castro et al (2015) report an average annual

Parameters	Value	Description
a	0.6129	Pr. of moving down
c	0.3870	Pr. of moving up
S	35	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	Scale parameter of the distr. G
$\delta_e(1-\alpha)$	1.570	Tail parameter of the distr. G

Table 2: Baseline calibration

productivity (TFPR) volatility of about 20%, a similar value to that used by Clementi and Palazzo (2016) in their firm-level calibration. In Section 5.2.2 below, we show that our quantitative results are largely unaffected by further lowering our (already conservative) target of $\sigma_e = 8\%$. Our choice for the entry rate target comes from directly computing the average entry rate from the BDS data and is consistent with, for example, the values reported by Dunne et al (1988).[28]

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 0.14% 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.

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

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 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 accurately reproduces both 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 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 steady-state equilibrium and provide a quantification of its performance as a theory of the business cycle. We start by briefly describing our numerical strategy. Based on our numerical solution, we 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.

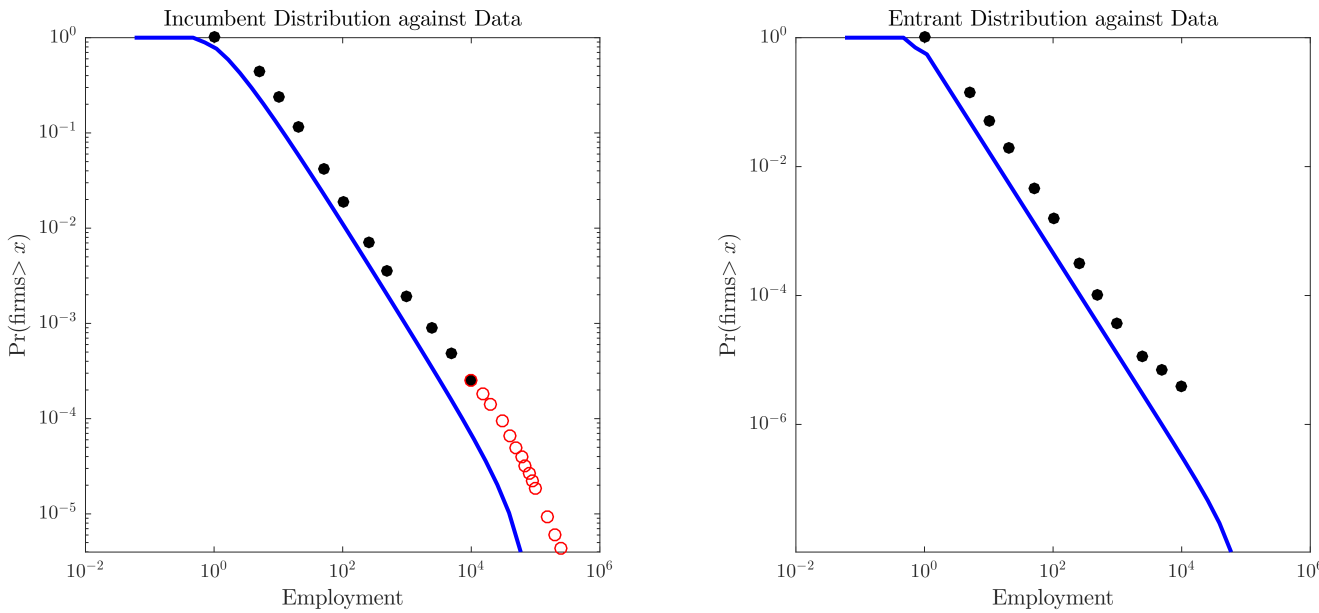

Figure 2: Counter Cumulative Distribution Functions (CCDF) of the firms size distribution of incumbents (left) and entrants (right) in the model (blue solid line) against data (circles).

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

5.2.1 Numerical Strategy

The characterization of the law of motion of the aggregate productivity in Theorem 2 and its generalization to the case with entry and exit (summarized in Theorem 3 in Online Appendix B.5), are 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 variable A_t. As Theorem 3 renders clear, the dynamics of A_t in turn depend on $E_t(\varphi)$ and σ_t and thus, on D_t and $E_t(\varphi^2)$. Note that D_t is proportional to the second moment of the firm size distribution and that, 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 maintain Assumption 3 that $\frac{E_t(\varphi)}{A_t}$, $\frac{O_t^A}{A_t}$ and $\frac{\sigma_t}{A_t}$ are perceived by firms to be fixed at its steady-state values. Given this assumption, the firms' problem can be solved by standard value function iteration methods. We discuss this numerical algorithm in detail in the Online Appendix D.[29]

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 are able to characterize the law of motion of the first moment of this distribution and

[29] It's also worth noting that under our baseline calibration, equilibrium wage volatility and grid coarseness jointly imply that the entry/exit threshold does not vary with aggregate fluctuations. That is, while our numerical algorithm allows this threshold to be time-varying, given our calibration, the stochastic domain of A_t does not visit regions of the state space that would induce the entry and exit threshold to change over the course of simulation (See Numerical Appendix D).

	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
Aggregate TFP	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 the Online Appendix C.

hence, in our numerical solution, agents know the law of motion of A_t, up to higher order moments. In Krusell and Smith (1998) this law of motion has to be solved for, which imposes a computationally costly simulation step. 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.

5.2.2 Business cycle statistics

Using the calibration in Table 2 and the numerical algorithm describes in the Online Appendix D, we compute the business cycle statistics. We simulate time series for output, hours and aggregate TFP using the law of motion (Equation 5) of the productivity distribution. These statistics are presented in Table 3.[30]

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 TFP 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.

[30]Note that given our choice of notation aggregate TFP refers to $A_t^{1-\alpha}$ while aggregate productivity refers to A_t.

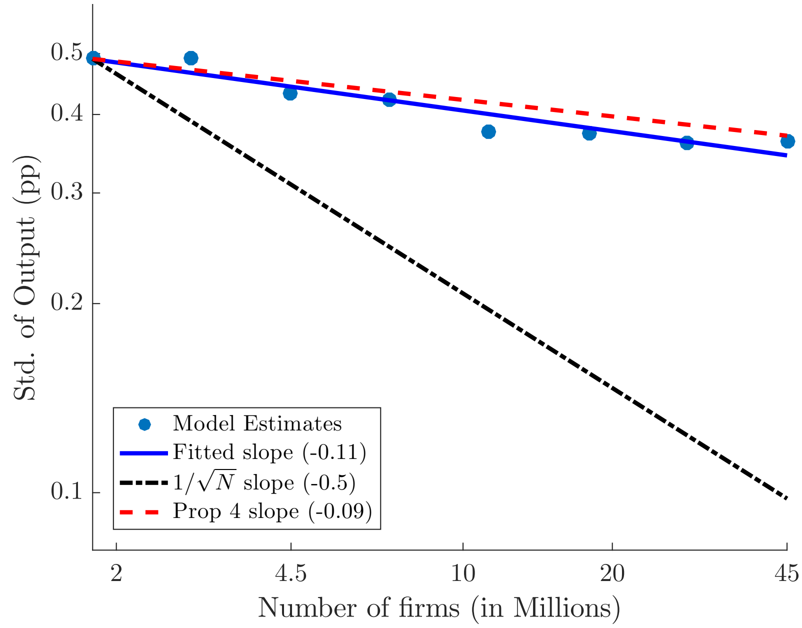

Figure 3: Decay of Volatility of Aggregate Output

NOTE: The blue (filled) circles give the standard deviation of aggregate output for different values of the equilibrium number of firms; the blue (solid) line shows the fit to the model estimates (blue circles); the red (dashed) line shows the slope implied by Proposition 4 for the case with no entry and exit; the black (dash-dotted) line shows the slope implied by a standard central limit theorem argument.

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 5.[31]

To see this, we implement numerically the same thought experiment that underlies Propositions 4 and 5. Namely, we investigate quantitatively the rate of decay of aggregate volatility when we vary the number of firms, relative to our baseline calibration. The number of firms across these different laboratory economies varies from 1.8 million, roughly three times smaller than the number of firms in the U.S., to 45 million, roughly 10 times that number. To discipline our exercise, we require that the targets in our baseline calibration are maintained across these different sized economies (in particular, we maintain our baseline targets for the entry rate, the idiosyncratic volatility and the tails of the entrants and incumbents firm size distributions). The exceptions are (i) the number of firms and, (ii) the share of the economic activity commanded by the largest firm, which, as instructed by Assumption 2 should decay as we move to economies with larger and larger number of firms.

[31]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 (2016), Khan and Thomas (2008) and Veracierto (2002).

Figure 3 summarizes our results. The first thing to note is that the rate of decay of aggregate volatility (given by the blue solid line) is much slower than what a standard central limit theorem argument would predict (as represented by the slope of the black dash-dotted line). In particular, the decline of aggregate volatility we observe across these economies is very close to the one predicted in Proposition 4 for the case without entry and exit (represented by the red dashed line). We can also see that there is a slight discrepancy between these two slopes. This is as it should be. Recall, that in our quantitative section we are studying an economy with entry and exit and therefore, according to Proposition 5, the decay of aggregate volatility should be slightly faster than what is predicted under no entry and exit. This is because the size distribution of entrants is thinner tailed (as disciplined by data). Therefore when the number of entrants increases (such that the entry rate is equalized across these different sized economies) diversification within the group of entrants should be stronger. Nevertheless, as the Figure renders clear, this is a second order effect precisely because, as predicted by Proposition 5, the decay is predominantly ruled by the size distribution of incumbents (which is fatter tailed).

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 80% of aggregate sales growth volatility 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).

Finally, we assess whether these benchmark aggregate volatility results are driven by an unrealistic high volatility of very large firms. Recall that our baseline choice of σ_e was on the lower end of the typical numbers reported in the literature for the *average* firm in the economy. However, as is well known, large firms are less volatile than the average firm in the economy, thus raising the possibility that our quantitative results for aggregate volatility are being driven by excessively volatile large firms. To address this concern, we now choose σ_e to match a standard deviation of annual employment volatility of 15%, corresponding to that of the largest 10% of firms (measured by number of employees) present in Compustat. While this number is in agreement with previously reported estimates for large firms (see e.g. Comin and Phillipon, 2006 and Davis et al, 2007) note that, relative to our baseline calibration, it now grossly understates the employment volatility of the typical firm in the economy, reported by Davis et al (2007) to be between between 40 and 50%. Based on this very conservative calibration, we recompute the aggregate volatility statistics discussed above. We find that our quantitative results remain largely unchanged: the standard deviation of output in the model is now 0.41% or 22% of annual GDP volatility observed in the data.

5.2.3 Inspecting the mechanism

As Propositions 4 and 5 render 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 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 productivity growth rate of the largest firm. Figure 4 shows the response of aggregate output, hours, TFP along with the average marginal productivity of labor across firms.[32]

The top panel of Figure 4 shows the responses of aggregate output and aggregate hours to this large-firm shock. The dynamics of both variables closely mirror that of aggregate TFP, as displayed in the bottom left panel. Thus, after a one standard deviation negative shock to the largest firm, aggregate TFP decreases by 0.016%. In turn, this 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 4 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

[32]From the structure of the model, computing the impulse response is straightforward. From Equation 5, 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 5 the result is exactly the same.

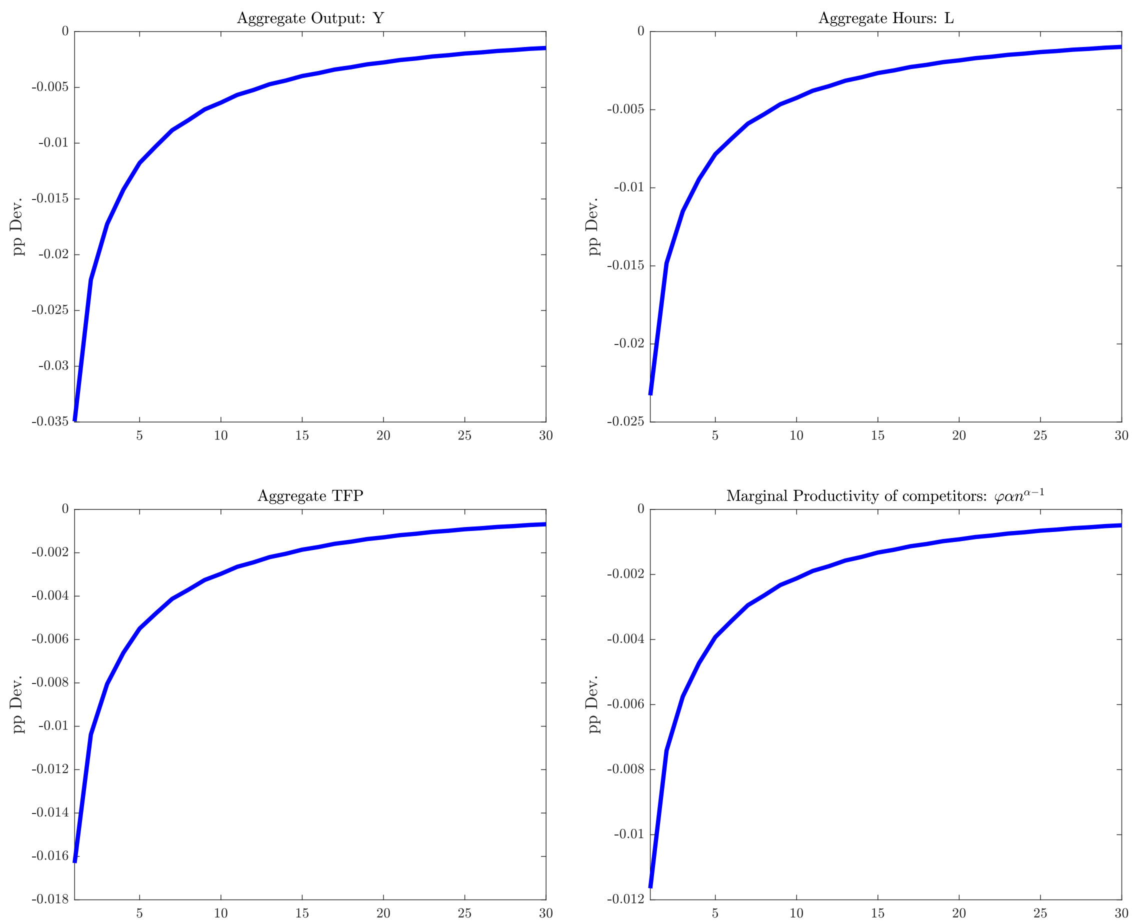

Figure 4: Impulse response to a one standard deviation negative productivity shock on the largest firm.

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.

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.

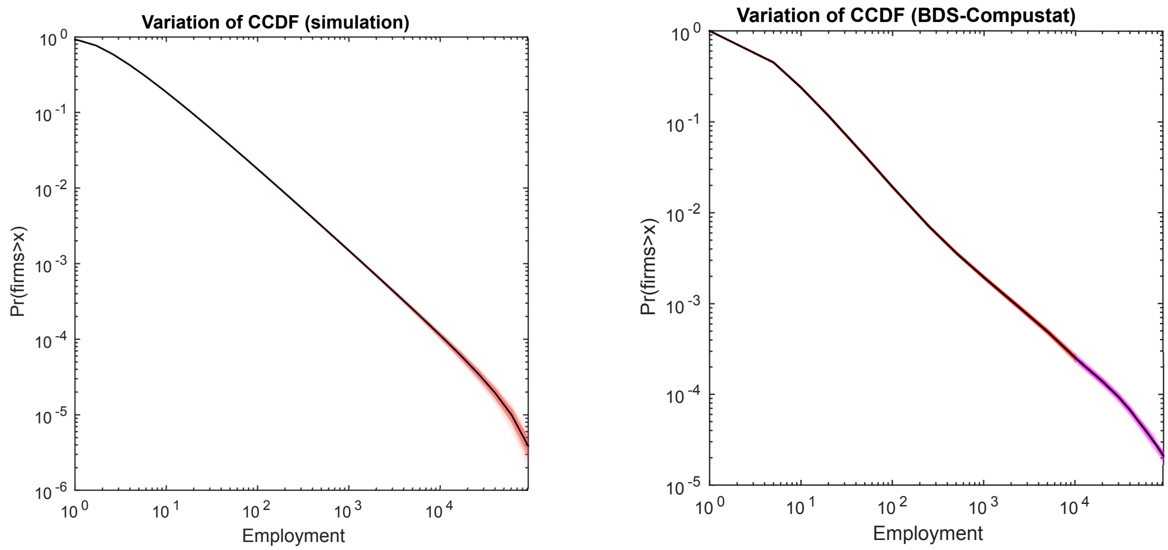

Figure 5: Variation of the Counter Cumulative Distribution Function (CCDF) in simulated data (left) and in the BDS data (right).

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 black line is the mean of each sample. For each period, we plot a transparent red line for that year CCDF. Darker regions of the plot imply that the economy spends more periods in this region of the state space. The transparency is chosen such that these two graphs are comparable.

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 and in the data) large firms comove with the business cycle.[33] 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 5 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. 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.

[33]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 (0.000)	−0.57 (0.000)	−0.48 (0.000)
	Correlation in growth rate	−0.41 (0.000)	−0.42 (0.000)	−0.44 (0.000)
Data	Correlation in (HP filtered) level	−0.34 (0.008)	−0.51 (0.000)	−0.46 (0.000)
	Correlation in growth rate	−0.33 (0.011)	−0.43 (0.001)	−0.38 (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.

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.

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.[34] 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:

[34] 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). Note that we cannot use the same estimator in the data and the model: given that we have a discrete state space, the tail index of the model-generated firm size distribution will necessarily have to be estimated from binned 'data'. An alternative would be to bin Compustat data and apply the Virkar and Clauset (2014) estimator to the data. By doing so, our results are qualitatively unaffected.

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 Theorem 2. 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 more we focus on the behavior of very large firms the stronger is the correlation in the data.[35] 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 Propositions 4 and 5 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 its analytical characterization in Theorem 3 in the Online Appendix B.5. We then correlate this series with 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 C 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

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

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 TFP (Equation 1), aggregate output (Equation 3) and aggregate volatility (Equation 10) in the model.[36] As a result of this exercise, we obtain from our model three aggregate time series - for aggregate output, aggregate TFP and the volatility of aggregate output - whose time variation reflects movements in the size distribution over time *alone*. Figure 6 plots these three series (solid lines) against their HP filtered data counterparts (dashed lines).

As Figure 6 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

[36] 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 the 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 $(ii) w_t^{BDS} = \left(\alpha^{\frac{1}{1-\alpha}} \frac{\sum_{s=1}^{S^{BDS}} \left(\varphi_{s,t}^{BDS}\right)^{\frac{1}{1-\alpha}} \mu_{s,t}^{BDS}}{M}\right)^{\frac{1-\alpha}{\gamma(1-\alpha)+1}}$.

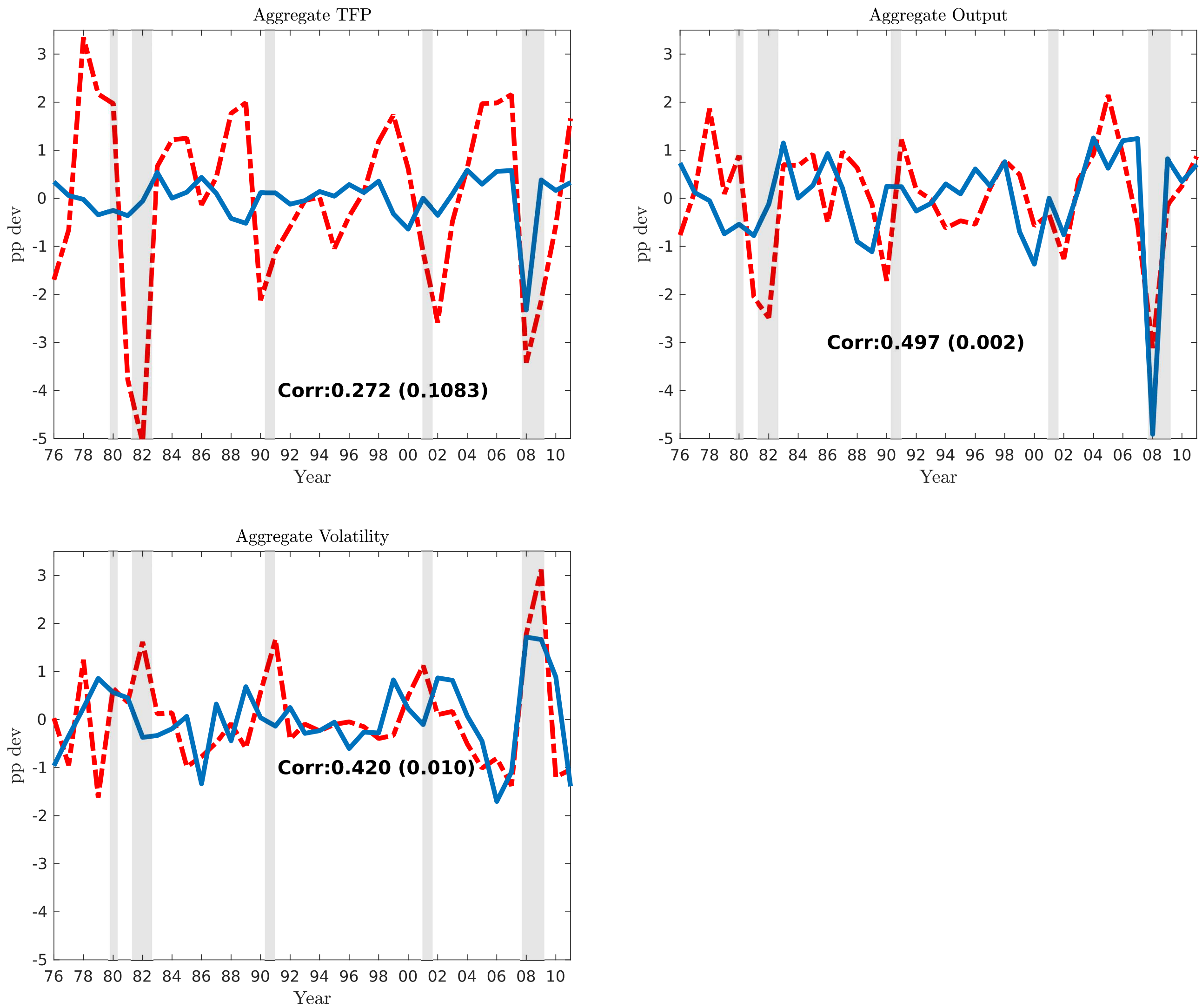

Figure 6: Dynamics of aggregate variable from firm size distribution data (BDS).

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 10 to compute aggregate TFP, 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).

the evolution of aggregate output (in deviations from trend) particularly well with a highly significant correlation of 0.533. For aggregate TFP 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 as in Grassi (2017).

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 Kaae Jensen. Robust Comparative Statics in Large Dynamic Economies. *Journal of Political Economy*, 123(3):587 – 640, 2015.

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

[4] Ruediger Bachmann, Ricardo J. Caballero, and Eduardo M. R. A. Engel. Aggregate Implications of Lumpy Investment: New Evidence and a DSGE Model. *American Economic Journal: Macroeconomics*, 5(4):29–67, October 2013.

[5] David Rezza Baqaee. Cascading failures in production networks. 2016.

[6] David Rezza Baqaee and Emmanuel Farhi. The macroeconomic impact of microeconomic shocks: Beyond hulten's theorem. 2017.

[7] 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.

[8] 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.

[9] 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.

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

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

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

[13] Rui Castro, Gian Luca Clementi, and Yoonsoo Lee. Cross sectoral variation in the volatility of plant level idiosyncratic shocks. *The Journal of Industrial Economics*, 63(1):1–29, 2015.

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

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

[16] Gian Luca Clementi and Berardino Palazzo. Entry, exit, firm dynamics, and aggregate fluctuations. *American Economic Journal: Macroeconomics*, 8(3):1–41, July 2016.

[17] 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.

[18] Diego Comin and Sunil Mulani. Diverging Trends in Aggregate and Firm Volatility. *The Review of Economics and Statistics*, 88(2):374–383, May 2006.

[19] 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.

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

[21] 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.

[22] 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.

[23] 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.

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

[25] Julian di Giovanni, Andrei A. Levchenko, and Isabelle Mejean. The Micro Origins of International Business Cycle Comovement. NBER Working Papers 21885, National Bureau of Economic Research, Inc, January 2016.

[26] 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.

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

[28] 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.

[29] 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.

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

[31] 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.

[32] Basile Grassi. Io in i-o: Competition and volatility in input-output networks. 2017.

[33] 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.

[34] John Haltiwanger, Ron S. Jarmin, and Javier Miranda. Who Creates Jobs? Small versus Large versus Young. *The Review of Economics and Statistics*, 95(2):347–361, May 2013.

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

[36] 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.

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

[38] 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.

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

[40] 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.

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

[42] Yoonsoo Lee and Toshihiko Mukoyama. Entry and exit of manufacturing plants over the business cycle. *European Economic Review*, 77(C):20–27, 2015.

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

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

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

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

[47] 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.

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

[49] Sophie Osotimehin. Aggregate productivity and the allocation of resources over the business cycle. 2016.

[50] 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.

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

[52] Thomas A. Severini. *Elements of Distribution Theory*. Cambridge University Press, 2005.

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

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

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

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

Appendix to "Large Firm Dynamics and the Business Cycle"

Vasco M. Carvalho and Basile Grassi

A Proof Appendix

A.1 Proof of Theorem 1

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 $f^{\cdot,s}_{t+1}$. This $(S \times 1)$ vector is such that for each state k in $\{1, \dots, S\}$, $f^{k,s}_{t+1}$, the k^{th} element of $f^{\cdot,s}_{t+1}$ gives the number of incumbents in state k at $t+1$ which were in state φ^s at t.

Similarly, let us define $g^{\cdot,s}_{t+1}$ 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 $\{1, \dots, S\}$, $g^{k,s}_{t+1}$, the k^{th} element of $g^{\cdot,s}_{t+1}$ 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:

$$\mu_{t+1} = \sum_{s=s^*(\mu_t)}^{S} f^{\cdot,s}_{t+1} + \sum_{s=s^*(\mu_t)}^{S} g^{\cdot,s}_{t+1} \tag{18}$$

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

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

Recall that the mean and variance-covariance matrix of a multinomial distribution with a number of trials m and event probabilities given by the $(S \times 1)$ vector h is respectively the $(S \times 1)$ vector mh and the $(S \times S)$ matrix $mH = m(diag(h) - hh')$. So let us define $W_s = diag(P_{s,.}) - P'_{s,.}P_{s,.}$. From the right hand side of Equation 18, using the fact that the $f^{\cdot,s}_{t+1}$ and $g^{\cdot,s}_{t+1}$ follow multinomials, μ_{t+1} has a mean $m(\mu_t)$ and a variance-covariance matrix $\Sigma(\mu_t)$ where

$$m(\mu_t) := \sum_{s=s^*(\mu_t)}^{S} \left[\mu^s_t P'_{s,.} + MG_s P'_{s,.}\right] = (P^*_t)'(\mu_t + MG)$$

$$\Sigma(\mu_t) := \sum_{s=s^*(\mu_t)}^{S} (MG_s + \mu^s_t) W_s$$

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

Equation 18 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.
□

A.2 Proof Sketch of Theorem 2

To understand the essence of the argument, consider the following proof sketch, which ignores the boundary effects arising from assuming a bounded state space. A full proof is given in the Online Appendix B.5.

The first thing to note is that aggregate productivity, A_t, is the sum of firm-level productivity (up to a Cobb-Douglas power). Aggregate productivity at $t+1$ is thus:

$$A_{t+1} = \sum_{i=1}^{N} \varphi^{\frac{s_{i,t+1}}{1-\alpha}} = \sum_{i=1}^{N} \left(\varphi^{\frac{s_{i,t+1}-s_{i,t}}{1-\alpha}} \right) \left(\varphi^{\frac{s_{i,t}}{1-\alpha}} \right) \tag{19}$$

so by taking expectations, we get

$$\mathbb{E}_t \left[A_{t+1} \right] = \sum_{i=1}^{N} \mathbb{E}_t \left[\varphi^{\frac{s_{i,t+1}-s_{i,t}}{1-\alpha}} \right] \left(\varphi^{\frac{s_{i,t}}{1-\alpha}} \right)$$

Under Assumption 1, $s_{i,t+1} - s_{i,t}$ takes only the values $(-1, 0, 1)$ with probability (a, b, c) and thus $\mathbb{E}_t \left[\varphi^{\frac{s_{i,t+1}-s_{i,t}}{1-\alpha}} \right] = a\varphi^{\frac{-1}{1-\alpha}} + b + c\varphi^{\frac{1}{1-\alpha}} = \rho$ which is independent of $s_{i,t}$. Using this last result in the above equation leads to $\mathbb{E}_t \left[A_{t+1} \right] = \rho A_t$. Similarely by taking the conditional variance of Equation 19, we have

$$\mathbb{V}ar_t \left[A_{t+1} \right] = \sum_{i=1}^{N} \mathbb{V}ar_t \left[\varphi^{\frac{s_{i,t+1}-s_{i,t}}{1-\alpha}} \right] \left(\varphi^{\frac{2s_{i,t}}{1-\alpha}} \right) = \varrho \sum_{i=1}^{N} \varphi^{\frac{2s_{i,t}}{1-\alpha}}$$

where $\varrho = \mathbb{V}ar_t \left[\varphi^{\frac{s_{i,t+1}-s_{i,t}}{1-\alpha}} \right] = a\varphi^{\frac{-2}{1-\alpha}} + b + c\varphi^{\frac{2}{1-\alpha}} - \rho^2$. □