Why Risky Sectors Grow Fast*

Basile Grassi† Jean Imbs‡

February 1, 2016

[Preliminary]

Abstract

Because they are populated with large firms. We construct a model of idea flows in which growth and volatility both depend on the prevalence of large firms in a sector. There is a finite number of firms that choose whether to imitate or to experiment. Experimenting means producing using a random technology, given by a discrete Markov deviation from its earlier value. In the limit, experimenting firms define an expanding technology frontier. Imitating means drawing technology from the pool of existing producers. In equilibrium, only large enough firms experiment, and growth increases in their share. Since experimenting has stochastic consequences, so does volatility. The model's key predictions are born out in US firm-level data: growth and volatility both increase in the share of large firms. The dispersion in tails can explain about 40% of the positive link between growth and volatility at the 4-digit level. As the data are aggregated, growth and volatility cease to correlate significantly: We argue this is consistent with our model, as structural change reallocates factors across sectors from high to low technology growth. In the aggregate, the link between large, experimenting firms, growth, and volatility is broken.

Keywords: Growth, Volatility, Idea Flows, Granularity, Random Growth, Aggregation, Structural Transformation.

*The authors would like to acknowledge helpful comments from Axelle Arquié, Florin Bilbiie, Vasco M. Carvalho, Xavier Gabaix, Boyan Jovanovic, Robert Lucas, Claire Lelarge and Isabelle Mejean.

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

‡Paris School of Economics, Centre National de la Recherche Scientifique. jean.imbs@psemail.eu

1 Introduction

Risk is compensated by returns. While this tradeoff is a central tenet of finance, surprisingly little is known about its counterpart in the real economy: Are volatile activities growing fast because they are risky? Of course, there is an extensive and venerable literature discussing the relation between growth and volatility going back at least to Schumpeter. But it is largely silent about the risk-taking behavior of firms, and how it translates into produced quantities; it has focused instead on models of causal links, going from growth to volatility, or from volatility to growth. Here we argue theoretically and empirically that growth and volatility arise from the risk taking behavior of firms that are active in a given sector. In our model, risk in a sector can be summarized by the characteristics of the distribution of firms productivities. It is because they contain a few very productive firms that risky sectors grow fast.

We construct a growth model based on ideas flow à la Lucas (2009), where firms can either imitate or experiment. Imitators delay production by one period, and draw their technology from the set of incumbents. This strategy carries high risk, but potentially high return as the technological jump can be substantial. Experimenters produce every period, but after experimentation draw technology from a Markov chain process. The new technology can be identical, lower, or higher with exogenous probabilities. This strategy carries low risk, and low return. The initial distribution of ideas (and of firms) is finite. In equilibrium, the technology frontier expands thanks to experimenting firms coming up with new ideas, and potentially better technologies. Since only highly productive (large) firms find it worthwhile to experiment, large firms are the engine of growth: growth is high in a sector populated by a few large firms that come up with new technologies, that are in turn copied by a large mass of risk-taking, small firms.

With a finite distribution of firms, the endogenous churning of productivities engenders volatility. This comes from a slightly modified granularity argument: as in Carvalho and Grassi (2015), volatility increases with the share of large firms in the economy, because they are the ones that tend to experiment. But imitating firms also contribute to volatility: we show their contribution also increases in the share of large firms, since these are the ones expanding the support of existing technologies. The risk taken by small, imitating firms increases with the prevalence of large firms, since it expands the support of technologies available to imitate. Thus, both growth and volatility increase in the share of large firms in the economy.

In US firm-level data, we show sector-level growth and volatility correlate positively and significantly with the prevalence of large firms, estimated in a variety of ways. The relation exists both in the short and the long term, and is robust to a battery of alternative specifications. In the data, risky sectors grow fast because they tend to be populated by a few, very large firms, and these happen to have good ideas. This is consistent with the well known trade off between risk and return: sectors with a large mass of small, risk-taking firms display both high growth and high volatility. The result is also consistent with the cross-sector evidence documented in Imbs (2007).

We also show the positive link between growth, volatility, and the share of large firms tends to dissipate as the data are aggregated. Our model can explain why. In our theory, the dispersion across sectors in growth rates -and volatilities- reflects differences in the growth of technology, determined endogenously by the prevalence of large firms. With heterogeneous growth rates in technology, we have known since Baumol (1967), or Ngai and Pissarides (2007) that structural change will happen. In particular, in equilibrium, factors of production move from high technology to low technology growth sectors, in order to serve the continuing demand for goods produced with low productivity: as shown by Ngai and Pissarides (2007), the result requires some complementarities in demand across sectors. If instead goods are substitutes, then of course factors of production move to high technology sectors.

In a sector defined as an aggregate of substitutable goods, the dynamics of production reflect therefore the characteristics of the high technology goods. But in a sector defined as an aggregate of complementary goods, the dynamics of production are instead dominated by the low technology goods. It is well known - and intuitive - that goods are substitutes at a granular level, but complements in the aggregate. As data are aggregated, the dispersion in measured growth and volatility at sector level is decreasingly reflective of the growth of technology, because factors are increasingly allocated in low technology activities. We show this in an extended version of our idea flows model to a multi-sector model with nested CES demand.[1]

Our paper builds on the recent literature on "idea flows" pioneered by Lucas (2009) and Lucas and Moll (2014). It is most closely related with Perla and Tonetti (2014), who showed imitation could translate into long run productivity growth. We augment their

[1]The same logic applies to the time pattern of volatility at country level. Countries at the early stages of structural transformation are concentrated in sectors with high technology growth, and display high volatility. With structural transformation, factors are reallocated towards activities with relatively low technology growth, and volatility falls.

model with experimenting firms whose optimal decisions create an expanding technology frontier. Thus, we can work with a finite distribution of firms and of technology, as Romer (2015) recommends.

Because we have a finite number of firms, our paper is also related with the recent literature on "granularity". As in Gabaix (2011), firm specific disturbances have consequences on aggregate volatility, because of the fat-tailed distribution of firm productivity. And as in Carvalho and Grassi (2015), an aggregate business cycle arises endogenously as individual firm productivities vary. Here they vary endogenously, in response to the firm's optimal decision to imitate or to experiment.

The endogenous emergence of growth and volatility, and the fact they both relate to an underlying cause (with the same sign) is related to the venerable Schumpeterian view that growth is inherently a destructive process. So that volatility is but a manifestation of the growth process, as in for instance in models of creative destruction introduced by Aghion and Howitt (1992), or growth is itself caused by volatility, as in the view that recessions have cleansing effects (see for instance Caballero and Hammour, 1994 or Aghion and Saint-Paul, 1991). Here, there is no causal link from growth to volatility, or in the other direction: both are caused by the distribution of firm productivities, and its evolution over time.

The existence of an empirical relation between growth and volatility is equally well known, with early contributions by Ramey and Ramey (1995), Saint-Paul (1993), or Imbs (2007). Interestingly, aggregation matters for the estimated relation: in granular sector data, growth and volatility correlate positively, but the relation vanishes with aggregation, to change signs across countries. It is as if the axiomatic tradeoff between risk and return dissipates in the aggregate. Here we provide an explanation for this fact, based on recent models of structural transformation. Since growth and volatility are both created by the process of technological change at sector level, we model aggregation following the intuition in Ngai and Pissarides (2007). The argument is inspired by Baumol's (1967) "cost disease", arguing that during structural transformation, factors of production are reallocated from high productivity growth to low productivity growth activities (i.e., from high volatility to low volatility sectors in our model). Finally, the paper is related to the empirical literature concerned with volatility and development. For instance, Koren and Tenreyro (2007) document that countries productive structure moves from more volatile to less volatile sectors as they develop. This is a direct prediction of our model.

The rest of the paper is structured as follows. Section 2 presents our model. Section 3 discusses our empirics. Section 4 concludes.

2 Model

We present a model populated by a finite number of firms, where the growth rate of output and its volatility both depend on the cross-sectional moments of the distribution of firm size. Growth happens as firms can either experiment on their own technology or imitate their peers *à la* Perla and Tonetti (2014). In equilibrium the distribution of firms is follows a Pareto distribution: since there is a finite number of firms, aggregate uncertainty arises, just like in Carvalho and Grassi (2015). Aggregate fluctuations arise purely from uncorrelated idiosyncratic shocks at the firm level, following the "Granular Hypothesis" introduced in the seminal paper by Gabaix (2011).

We first describe the model abstracting from any aggregate uncertainty arising from idiosyncratic shocks. We derive a Balanced Growth Path (BGP) that is very similar to the one described in Perla and Tonetti (2014). We next take into account the aggregate uncertainty arising from firm level uncertainty, and we characterize the stochastic process of output growth. We show that both the conditional expectation and the volatility of output growth is determined by cross-sectional moments of the firm distribution.

2.1 Overview

Time is discrete and infinite and there in an integer number of firms, N, that are heterogeneous in term of their productivity levels. Firms produce perfectly substitutable good in a perfect competition environment. The productivity level evolves on a discrete (an infinite) evenly distributed grid: $\Phi = \{\varphi, \varphi^2, \ldots, \varphi^s, \ldots\} = \{\varphi^s | \forall s \in \mathbb{N}^*\}$ where $\varphi^{s+1}/\varphi^s = \varphi > 1$. Each period firms face a binary choice. They can either (i) produce and experiment or (ii) postpone production and imitate their peers.

Experimenting Firms

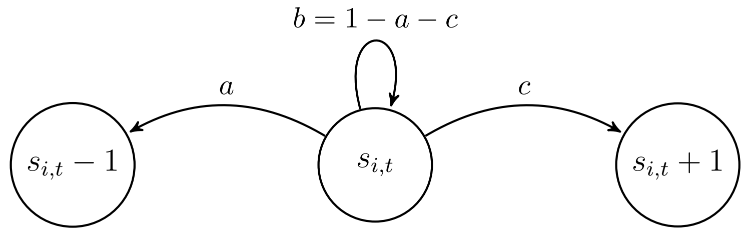

Figure 1: Description of experimenting firms productivity process.

If a firm decides to experiment, it produces φ^s units of goods today. However, its next period productivity follows a Markov Chain described in figure 1 by $P_{s,s'} = \mathbb{P}\{\varphi^{s'}|\varphi^s\}$ over Φ where:

$$P_{s,s'} = \begin{cases} a & \text{if} \quad s' = s-1 \quad \text{Move down the grid} \\ b & \text{if} \quad s' = s \quad \text{Do not move} \\ c & \text{if} \quad s' = s+1 \quad \text{Move up the grid} \\ 0 & \text{if} \quad \text{otherwise} \end{cases} \tag{1}$$

This Markovian process taken from Cordóba (2008) is a discretization of Gibrat's law which states that the growth rate is independent of the current level. For a productivity level $s_{i,t}$ and conditional on experimentation we have $\mathbb{E}\left[\frac{\varphi^{s_{i,t+1}} - \varphi^{s_{i,t}}}{\varphi^{s_{i,t}}}|\varphi^{s_{i,t}}\right] = a(\varphi^{-1} - 1) + c(\varphi - 1)$ and $\mathbb{V}ar\left[\frac{\varphi^{s_{i,t+1}} - \varphi^{s_{i,t}}}{\varphi^{s_{i,t}}}|\varphi^{s_{i,t}}\right] = \sigma_e^2$ where σ_e^2 is a constant independent of $\varphi^{s_{i,t}}$. Define $V_t(s)$ the value of a firm t productivity level φ^s at date t, and $V_t^E(s)$ the value of experimenting for that firm at date t. We have:

$$V_t^E(s) = \varphi^s + \beta \sum_{s'} V_{t+1}(s') P_{s,s'}$$

where β is the discount factor and $P_{s,s'}$ is defined in equation 1.

Imitating Firms

If a firm decides to imitate at date t, production is delayed by one period. Productivity next period is drawn from the distribution of producers at date t. Imitating firms inherit the level of productivity of one of the producers active at date t. Define $\mu_{s,t}$ the number of firms at productivity level φ^s for a given $s \in \mathbb{N}^*$. The sequence $\mu_t = \{\mu_{s,t}\}_{s \in \mathbb{N}^*}$ describes the distribution of productivities across the productivity grid Φ.

Define S_t the number of imitating firms at date t. The value of imitating $V_t^I(s)$ for a firm with productivity level φ^s at date t is given by:

$$V_t^I(s) = \beta \sum_{\text{firms at } s' \text{ produce}} V_{t+1}(s') \frac{\mu_{s',t}}{N - S_t}$$

The choice between experimentation and imitation is similar to the one described in Perla and Tonetti (2014), but for two important differences. First, producing firms are subjected to idiosyncratic random productivity shocks, defined by equation 1 and illustrated in figure 1. This is an important ingredient to prove the existence of a balanced growth path. Second, there is a finite number of firms rather than a continuum. Together with the existence of a balanced growth path, this assumption will generate aggregate uncertainty, arising from granularity as in Gabaix (2011) and Carvalho and Grassi (2015).

Each period, firms face a choice between being a imitator or a experimentator. The firm problem can be written as follow:

$$V_t(s) = \text{Max}\left\{ \varphi^s + \beta \sum_{s' \in \mathbb{N}^*} V_{t+1}(s') P_{s,s'} \quad ; \quad \beta \sum_{s' \text{ produce}} V_{t+1}(s') \frac{\mu_{s',t}}{N - S_t} \right\}$$

where the first term is the value of experimentation and the second the value of imitation. Such a problem yields a threshold rule

$$\begin{cases} s < s_t, & V_t^I(s) > V_t^E(s) \quad \text{the firm decides to imitate} \\ s \geq s_t, & V_t^I(s) \leq V_t^E(s) \quad \text{the firm decides to experiment} \end{cases} \tag{2}$$

where s_t is such that the value of producing is above the value of imitating, i.e. s_t is such that $V_t^E(s_t) - V_t^I(s_t) \geq 0$ and $V_t^E(s_t - 1) - V_t^I(s_t - 1) < 0$.

Focusing on rational expectations equilibria implies that $s_t - 1 = m_{t+1} = \min \mathit{support}\{\mu_{t+1}\}$. The intuition is straightforward: All firms strictly below s_t imitate, and draw from the producing firms that have productivity weakly above s_t. Their next period productivity will thus be weakly above s_t. All firms weakly above s_t engage in production, their productivity next period will be higher than $s_t - 1$. The firm with the lowest productivity level will be the one whose productivity was s_t and that sees its productivity decreases by one level to $s_t - 1$ because of a bad draw. It follows that $s_t - 1 = m_{t+1} = \min \mathit{support}\{\mu_{t+1}\}$.

The Evolution of the Productivity Distribution

Recall that $\mu_{s,t}$ is the number of firms with productivity level φ^s at date t, and that the sequence $\mu_t = \{\mu_{s,t}\}_{s\in\mathbb{N}^*}$ is the distribution of firms across productivity levels Φ. We now describe how the distribution of firms across productivity levels evolves over time.

For $s > s_t$, the number of firms with productivity s at date $t+1$ is the sum of producing firms and imitating firms. The former are the sum of three terms: (i) the producing firms at productivity $s+1$ at date t with a bad draw, of which there are $a\mu_{s+1,t}$ (ii) the producing firms at productivity s at date t with unchanged productivity, of which there are $b\mu_{s,t}$, and (iii) the producing firms with productivity $s-1$ at date t that experience a good draw, of which there are $c\mu_{s-1,t}$. There are, at date t, S_t imitating firms that have a probability $\frac{\mu_{s,t}}{N-S_t}$ to imitate a date t producing firms with productivity level φ^s. The total number of firms at date $t+1$ with productivity level φ^s is therefore:

$$\mu_{s,t+1} = a\mu_{s+1,t} + b\mu_{s,t} + c\mu_{s-1,t} + S_t\frac{\mu_{s,t}}{N-S_t} \tag{3}$$

For $s \in \{s_t - 1, s_t\}$, the argument is similar except that there are no producing firms at $s_t - 1$ at date t and thus:

$$\mu_{s,t+1} = \begin{cases} a\mu_{s_t+1,t} + b\mu_{s_t,t} + S_t\frac{\mu_{s_t,t}}{N-S_t} & \text{if} \quad s = s_t \\ a\mu_{s_t,t} & \text{if} \quad s = s_t - 1 \end{cases} \tag{4}$$

Finally, for $s < s_t - 1$, there are no firm at date t at the productivity level φ^s.

Given current productivity distribution μ_t and the solution of the firm problem described by the threshold rule 2, equations 3 and 4 describe the evolution of the productivity distribution over time.

Figure 2 displays the evolution of the distribution of firms productivities when all firms start with the same level. At date $t = 0$, all firms have identical productivity, and there is no gain to imitate. All firms decide to experiment. At the following period $t = 1$, some firms will therefore be below the threshold s_t and thus will imitate. This process goes on until, the number of firms above the threshold is high enough to increase the value of the threshold s_t. In this example it happens at $t = 7$. The same example in Perla and Tonetti (2014) will lead to a very different outcome: since there is no experimentation, when all the firms are identical initially then there is no growth because firms will never have any incentive to imitate their peers. The system will stay at the top left panel

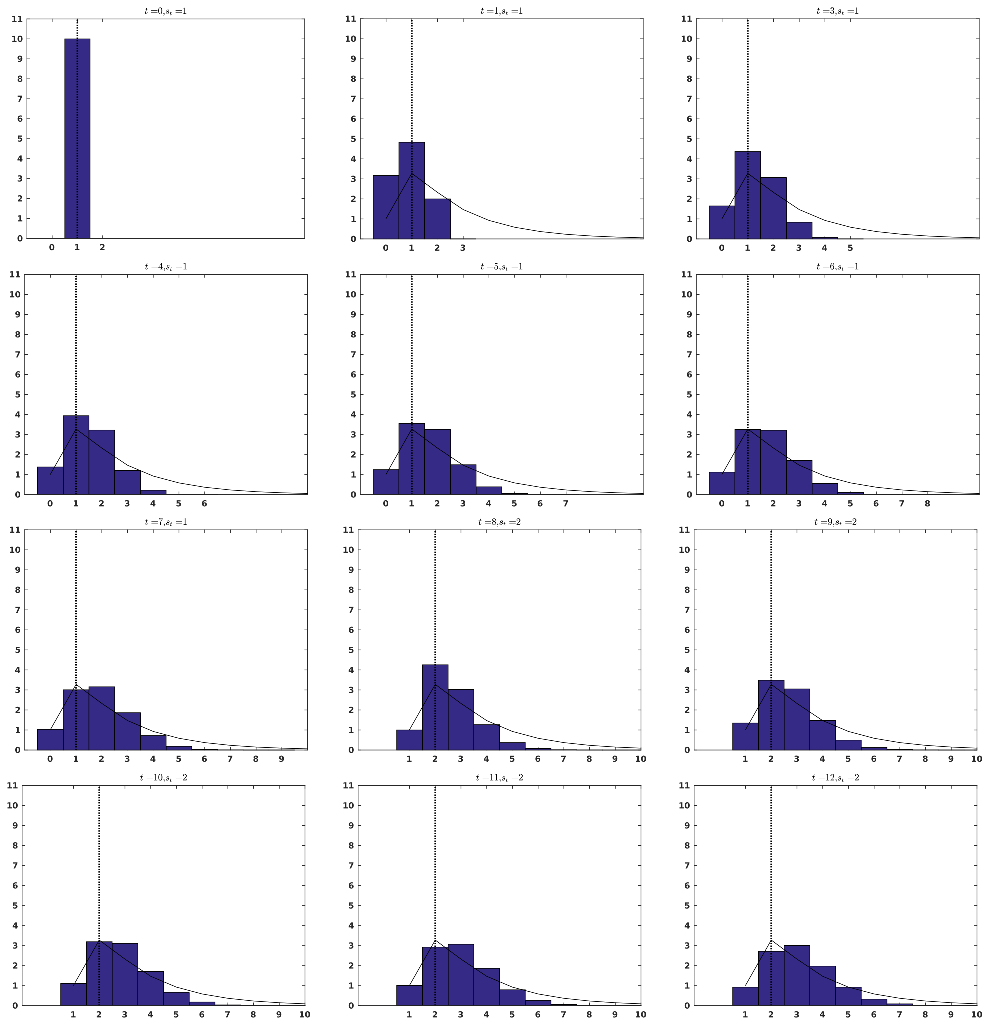

Figure 2: The evolution of the distribution of firms when all the firms starts at the same level.

for ever. Introducing some randomness at the firm level allows the imitation process of Perla and Tonetti (2014) to create endogenous growth.

2.2 Balanced Growth Path

This section introduces the Balanced Growth Path of this economy. Each firm growth will not necessarily be balanced, but the distribution of firms - which ultimately matters for the aggregate - is scale invariant. The section open with definitions, we then move to the characterization of the productivity distribution of firms along a balanced growth path (BGP). Finally, we prove the existence of a BGP.

Aggregate output Y_t is defined as the sum of firm level output, $Y_t = \sum_{i=1}^{N} \varphi^{s_{i,t}} = \sum_{s=1}^{\infty} \varphi^s \mu_{s,t}$. Along a BGP the productivity distribution of is fixed as long as its support grows at the appropriate rate. This notion is defined formally below.

Definition 1 *(Scale invariance) Given a scale parameter $g > 1$ such that $g = \varphi^\eta$. We have $g^t \varphi^s = \varphi^{\eta t} \varphi^s = \varphi^{s+\eta t}$. A sequence of distribution $\mu_t = \{\mu_{s,t}\}_{s \in \mathbb{N}}$ is scale invariant for $g = \varphi^\eta$ iff*

$$\mu_{s+\eta t,t} \text{ is identical for all } t$$

In other words the distribution remains unchanged if the grid is scaled by a factor g. The concept is common to models of growth with heterogeneous agents, e.g., Alvarez, Buera, Lucas (2008) and Luttmer (2007).

We are now ready to define the BGP.

Definition 2 *(Balanced Growth Path) A BGP with a constant $g = \varphi^\eta$ is a competitive equilibrium such that*

i The sequence of productivity distributions and the growth rate, $\{\mu_t, g\}$, are scale invariant.

ii Output grows geometrically at rate g: $Y_{t+1} = gY_t$

iii The minimum of the support of μ_t grows arithmetically at rate η: $s_{t+1} = s_t + \eta$

With these definitions, we can solve for the productivity distribution of firms.

Proposition 1 *Given a growth rate $g = \varphi^\eta$ with $\eta > 0$, the following scale invariant distribution is a solution to the system formed by equations (3) and (4) describing the evolution of the productivity distribution:*

$$\mu_{s,t} = \begin{cases} N(1-\varphi^{-\delta}) \left(\frac{\varphi^s}{\varphi^{s_{t-1}}}\right)^{-\delta} & \text{for} \quad s > s_{t-1} \\ N(1-\varphi^{-\delta}) \left(1 - c\varphi^{\delta}\varphi^{-\delta\eta}\right) & \text{for} \quad s = s_{t-1} \\ N(1-\varphi^{-\delta}) a\varphi^{-\delta\eta} & \text{for} \quad s = s_{t-1} - 1 \\ 0 & \text{for} \quad s < s_{t-1} \end{cases} \tag{5}$$

where $\delta = \frac{\log \frac{a}{c}}{\log \varphi}$. For $s > s_{t-1}$ this distribution is Pareto with tail index δ. The tail index is a function of the Markovian process generating firm level productivity.

Proof:

It is easy to see that this distribution sum to N, the number of firms in the economy:

$$\begin{aligned}
\sum_{s=m_t=s_{t-1}-1}^{\infty} \mu_{s,t} &= N(1-\varphi^{-\delta}) \left(a\varphi^{-\delta\eta} + \left(1 - c\varphi^{\delta}\varphi^{-\delta\eta}\right) + (\varphi^{\delta})^{s_{t-1}} \sum_{s=s_{t-1}+1}^{\infty} (\varphi^{-\delta})^{s} \right) \\
&= N(1-\varphi^{-\delta}) \left(a\varphi^{-\delta\eta} + \left(1 - c\varphi^{\delta}\varphi^{-\delta\eta}\right) + (\varphi^{\delta})^{s_{t-1}} \frac{(\varphi^{-\delta})^{s_{t-1}}(\varphi^{-\delta})}{1-\varphi^{-\delta}} \right) \\
&= N\left(\left(a\varphi^{-\delta\eta} + 1 - c\varphi^{\delta}\varphi^{-\delta\eta}\right)(1-\varphi^{-\delta}) + \varphi^{-\delta} \right) \\
&= N\left(a\varphi^{-\delta\eta} + 1 - c\varphi^{\delta}\varphi^{-\delta\eta} - a\varphi^{-\delta}\varphi^{-\delta\eta} - \varphi^{-\delta} + c\varphi^{-\delta}\varphi^{\delta}\varphi^{-\delta\eta} + \varphi^{-\delta} \right) \\
&= N\left(1 + a\varphi^{-\delta\eta} - c\varphi^{\delta}\varphi^{-\delta\eta} - a\varphi^{-\delta}\varphi^{-\delta\eta} + c\varphi^{-\delta\eta} \right)
\end{aligned}$$

By definition of δ, $\varphi^{-\delta} = \frac{c}{a}$ and $\varphi^{\delta} = \frac{a}{c}$. Substituting:

$$\begin{aligned}
\sum_{s=m_t}^{\infty} \mu_{s,t} &= N\left(1 + a\varphi^{-\delta\eta} - c\frac{a}{c}\varphi^{-\delta\eta} - a\frac{c}{a}\varphi^{-\delta\eta} + c\varphi^{-\delta\eta} \right) \\
&= N\left(1 + a\varphi^{-\delta\eta} - a\varphi^{-\delta\eta} - c\varphi^{-\delta\eta} + c\varphi^{-\delta\eta} \right) \\
&= N
\end{aligned}$$

This distribution sum to N the number of firms in the economy. We now show this distribution satisfies the system formed by equations (3) and (4).

For $s > s_t$ we have

$$\mu_{s,t+1} = a\mu_{s+1,t} + b\mu_{s,t} + c\mu_{s-1,t} + S_t \frac{\mu_{s,t}}{N - S_t} \tag{6}$$

By definition, $s_t = s_{t-1} + \eta > s_{t-1}$ since $\eta > 1$. Therefore, for $s > s_t > s_{t-1}$, we have:

$$\begin{aligned}
a\mu_{s+1,t} + b\mu_{s,t} + c\mu_{s-1,t} &= N(1-\varphi^{-\delta})\left(\varphi^{s_{t-1}}\right)^{\delta} \left(a\left(\frac{c}{a}\right)^{s+1} + b\left(\frac{c}{a}\right)^{s} + c\left(\frac{c}{a}\right)^{s-1} \right) \\
&= N(1-\varphi^{-\delta})\left(\varphi^{s_{t-1}}\right)^{\delta} \left(a^{-s}c^{s+1} + (1-a-c)c^{s}a^{-s} + cc^{s}a^{1-s} \right) \\
&= N(1-\varphi^{-\delta})\left(\varphi^{s_{t-1}}\right)^{\delta} \left(a^{-s}c^{s} \right) \\
&= N(1-\varphi^{-\delta})\left(\varphi^{s_{t-1}}\right)^{\delta} \left(\varphi^{s}\right)^{-\delta} \qquad = \mu_{s,t}
\end{aligned}$$

Substituting in equation 6 yields

$$a\mu_{s+1,t} + b\mu_{s,t} + c\mu_{s-1,t} + S_t \frac{\mu_{s,t}}{N - S_t} = \frac{N}{N - S_t}\mu_{s,t} \tag{7}$$

The mass of producers at time t is given by:

$$N - S_t = \sum_{s=s_t}^{\infty} \mu_{s,t} = \sum_{s=s_t}^{\infty} N(1-\varphi^{-\delta}) \left(\frac{\varphi^s}{\varphi^{s_{t-1}}} \right)^{-\delta} = N(1-\varphi^{-\delta}) \left(\varphi^{s_{t-1}}\right)^{\delta} \frac{(\varphi^{s_t})^{-\delta}}{1-\varphi^{-\delta}}$$

Substituting back in equation 7, we get

$$\frac{N}{N-S_t}\mu_{s,t} = N(1-\varphi^{-\delta}) \left(\frac{\varphi^s}{\varphi^{s_t}} \right)^{-\delta} = \mu_{s,t+1}$$

This completes the proof of Proposition 1 for $s > s_t$. For $s \leq s_t$, since $s_t = s_{t-1} + \eta > s_{t-1}$ we have

$$\begin{aligned}
a\mu_{s_t+1,t} + b\mu_{s_t,t} + S_t \frac{\mu_{s_t,t}}{N-S_t} &= -cN(1-\varphi^{-\delta})\varphi^{\delta} \left(\frac{\varphi^{s_t}}{\varphi^{s_{t-1}}} \right)^{-\delta} + \frac{N}{N-S_t}\mu_{s_t,t} \\
&= -cN(1-\varphi^{-\delta})\varphi^{\delta} \left(\frac{\varphi^{s_t}}{\varphi^{s_{t-1}}} \right)^{-\delta} + N(1-\varphi^{-\delta}) \\
&= -cN(1-\varphi^{-\delta})\varphi^{\delta} \left(\varphi^{\eta}\right)^{-\delta} + N(1-\varphi^{-\delta}) \\
&= N(1-\varphi^{-\delta}) \left(1 - c\varphi^{\delta}\varphi^{-\eta\delta}\right) = \mu_{s_t,t+1}
\end{aligned}$$

Finally since $s_t = s_{t-1} + \eta > s_{t-1}$,

$$a\mu_{s_t,t} = aN(1-\varphi^{-\delta}) \left(\frac{\varphi^{s_t}}{\varphi^{s_{t-1}}} \right)^{-\delta} = aN(1-\varphi^{-\delta}) \left(\varphi^{\eta}\right)^{-\delta} = \mu_{s_t-1,t+1}$$

This completes the proof of Proposition 1. As soon as the system formed by equations (3) and (4) admits one solution, then it satisfies a BGP. □

Figure 3 illustrates the productivity distribution of firms along a BGP. The distribution of firms is Pareto on the right tail with tail δ. This results is similar to the one of Perla and Tonetti (2014), except that these authors have to assume enough initial heterogeneity, i.e., a fat enough right tail, so as to ensure that the BGP is well defined. Here, firms are experimenting and the assumption is not necessary. Heuristically, the Perla and Tonetti (2014) imitation process truncates the initial distribution at the BGP growth rate. Here the existence of the Markovian process generate the distribution.

Armed with the distribution of firms along a BGP, we can solve for the number of imitating firms on the BGP.

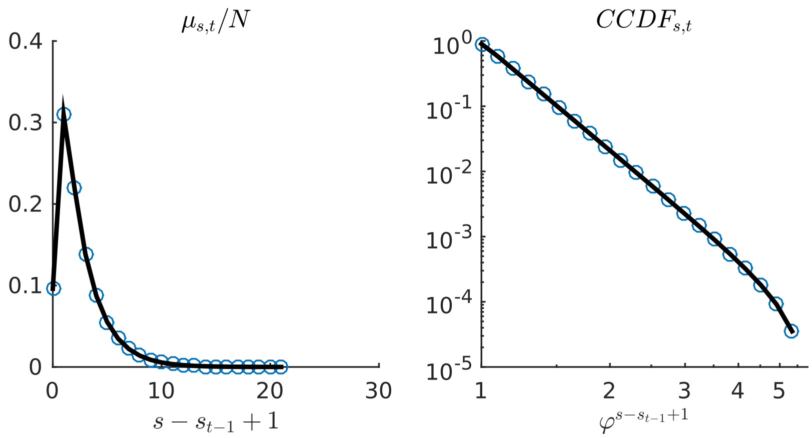

Figure 3: The firm size distribution along a BGP.

Proposition 2 *The number of firms S that imitate is constant along a BGP and equal to $S = N(1 - g^{-\delta})$*

Proof: Since $s_t = s_{t-1} + \eta$ along a BGP, we have

$$
\begin{aligned}
S_t = \sum_{s=s_{t-1}-1}^{s_t-1} \mu_{s,t} &= N(1-\varphi^{-\delta})\left(a\varphi^{-\delta\eta} + \left(1 - c\varphi^{\delta}\varphi^{-\delta\eta}\right) + (\varphi^{s_{t-1}})^{\delta} \sum_{s=s_{t-1}+1}^{s_{t-1}+\eta-1} (\varphi^{s})^{-\delta}\right) \\
&= N(1-\varphi^{-\delta})\left(a\varphi^{-\delta\eta} + \left(1 - c\varphi^{\delta}\varphi^{-\delta\eta}\right) + (\varphi^{s_{t-1}})^{\delta} \frac{(\varphi^{s_{t-1}})^{-\delta}\varphi^{-\delta} - (\varphi^{s_{t-1}})^{-\delta}\varphi^{-\delta\eta}}{1-\varphi^{-\delta}}\right) \\
&= N(1-\varphi^{-\delta})\left(a\varphi^{-\delta\eta} + \left(1 - c\varphi^{\delta}\varphi^{-\delta\eta}\right) + \frac{\varphi^{-\delta} - \varphi^{-\delta\eta}}{1-\varphi^{-\delta}}\right) \\
&= N\left(a\varphi^{-\delta\eta}(1-\varphi^{-\delta}) + \left(1 - c\varphi^{\delta}\varphi^{-\delta\eta}\right)(1-\varphi^{-\delta}) + \varphi^{-\delta} - \varphi^{-\delta\eta}\right) \\
&= N\left(a\varphi^{-\delta\eta} - a\varphi^{-\delta}\varphi^{-\delta\eta} + 1 - c\varphi^{\delta}\varphi^{-\delta\eta} - \varphi^{-\delta} + c\varphi^{-\delta}\varphi^{\delta}\varphi^{-\delta\eta} + \varphi^{-\delta} - \varphi^{-\delta\eta}\right) \\
&= N\left(a\varphi^{-\delta\eta} - c\varphi^{-\delta\eta} + 1 - a\varphi^{-\delta\eta} + c\varphi^{-\delta\eta} - \varphi^{-\delta\eta}\right) \\
&= N\left(1 - \varphi^{-\delta\eta}\right) = N\left(1 - g^{-\delta}\right)
\end{aligned}
$$

where the last equality used the fact that $\varphi^{-\delta} = \frac{c}{a}$. S does not depend on time.

□

We are now ready to show the existence of the BGP. The proof is very similar to Perla and Tonetti (2014) except that producing firms are subjected to the idiosyncratic Markovian process.

Proposition 3 *A BGP exist with growth rate*

$$g = \left(\beta \frac{1-\varphi^{-\delta}}{1-\varphi^{1-\delta}}\right)^{\frac{1}{\delta-1}}$$

Proof:

We form the following guess: $\forall s \leq s_t, V_t(s) = \varphi^{s_{t-1}}W$. This means that for firms below the threshold s_t (i.e imitating firms) the value function is proportional to $\varphi^{s_{t-1}}$. We then need to solve for the constant W and the growth rate $g = \varphi^{\eta}$.

Given the guess, the value functions for firm s_t indifferent between imitating and experimenting must verify:

$$V_t(s_t) = \varphi^{s_{t-1}}W = \varphi^{s_t} + \beta\left(aV_{t+1}(s_t-1) + bV_{t+1}(s_t) + cV_{t+1}(s_t+1)\right)$$

where the first equality comes from the guess and the second comes from the fact that the value function has to be equal to value of experimenting. Since $s_t < s_t + \eta = s_{t+1}$ (we have positive growth, i.e $\eta > 0$) we have $s_t + 1 \leq s_{t+1}$ and thus $s_t - 1 < s_t < s_t + 1 \leq s_{t+1}$. It follows from the guess at date $t+1$ that $aV_{t+1}(s_t-1) + bV_{t+1}(s_t) + cV_{t+1}(s_t+1) = \varphi^{s_t}W(a+b+c) = \varphi^{s_t}W = g\varphi^{s_{t-1}}W$. Substituting in the previous equation:

$$\varphi^{s_{t-1}}W = g\varphi^{s_{t-1}} + \beta g \varphi^{s_{t-1}}W$$

which implies:

$$W = \frac{g}{1-g\beta} \tag{8}$$

The number of producing firms is $N - S = N - N(1-g^{-\delta}) = Ng^{-\delta}$.[2] At the indifference point s_t, the value of experimenting should be equal to the value of imitating:

$$g\varphi^{s_{t-1}} + \beta g \varphi^{s_{t-1}}W = \beta \sum_{s' \geq s_t} V_{t+1}(s') \frac{\mu_{s',t}}{g^{-\delta}N} = \beta N^{-1} g^{\delta} \left(\sum_{s'=s_t}^{s_{t+1}-1} V_{t+1}(s')\mu_{s',t} + \sum_{s'=s_{t+1}}^{\infty} V_{t+1}(s')\mu_{s',t} \right)$$

where we used the fact that $N - S = Ng^{-\delta}$. For $s' \leq s_{t+1} - 1 < s_{t+1}$, firms at $t+1$ imitate and their value will be $V_{t+1}(s') = \varphi^{s_t}W$ according to our guess. For $s' \geq s_t > s_{t-1}$, we

[2] This is similar to equation (10) in Perla and Tonetti (2014).

have $\mu_{s',t} = N(1-\varphi^{-\delta})\left(\frac{\varphi^{s'}}{\varphi^{s_{t-1}}}\right)^{-\delta}$.[3] It follows that:

$$\begin{aligned}
\sum_{s'=s_t}^{s_{t+1}-1} V_{t+1}(s')\mu_{s',t} &= \varphi^{s_t} W \sum_{s'=s_t}^{s_{t+1}-1} \mu_{s',t} = \varphi^{s_t} W N(1-\varphi^{-\delta})(\varphi^{s_{t-1}})^{\delta} \sum_{s'=s_t}^{s_{t+1}-1} \varphi^{-s'\delta} \\
&= \varphi^{s_t} W N(1-\varphi^{-\delta})(\varphi^{s_{t-1}})^{\delta} \frac{\varphi^{-s_t\delta} - \varphi^{-s_{t+1}\delta}}{1-\varphi^{-\delta}} \\
&= \varphi^{s_t} W N(\varphi^{s_{t-1}})^{\delta}(\varphi^{-s_t\delta} - \varphi^{-s_{t+1}\delta}) \\
&= \varphi^{s_t} W N g^{-\delta}(1-g^{-\delta})
\end{aligned}$$

At $t+1$, for $s' \geq s_{t+1}$ firms produce and thus

$$\sum_{s'=s_{t+1}}^{\infty} V_{t+1}(s')\mu_{s',t} = \sum_{s'=s_{t+1}}^{\infty} \left[\varphi^{s'} + \beta\left(aV_{t+2}(s'-1) + bV_{t+2}(s') + cV_{t+2}(s'+1)\right)\right]\mu_{s',t} \tag{9}$$

For $s' \geq s_{t+1} > s_{t-1}$, we have $\mu_{s',t} = N(1-\varphi^{-\delta})\left(\frac{\varphi^{s'}}{\varphi^{s_{t-1}}}\right)^{-\delta}$. We now analyze each of the two terms at the right hand side. Rearranging the first term:

$$\begin{aligned}
\sum_{s'=s_{t+1}}^{\infty} \varphi^{s'}\mu_{s',t} &= N(1-\varphi^{-\delta})(\varphi^{s_{t-1}})^{\delta} \sum_{s'=s_{t+1}}^{\infty} \varphi^{s'(1-\delta)} = N(1-\varphi^{-\delta})(\varphi^{s_{t-1}})^{\delta} \frac{\varphi^{s_{t+1}(1-\delta)}}{1-\varphi^{1-\delta}} \\
&= N\frac{1-\varphi^{-\delta}}{1-\varphi^{1-\delta}}(\varphi^{s_{t-1}})^{\delta}\varphi^{-s_{t+1}\delta}\varphi^{s_{t+1}} \\
&= \varphi^{s_{t+1}} N g^{-2\delta} \frac{1-\varphi^{-\delta}}{1-\varphi^{1-\delta}}
\end{aligned} \tag{10}$$

[3]This is similar to equation (13) in Perla and Tonetti (2014).

Rearranging the second term:

$$
\begin{aligned}
&\beta \sum_{s'=s_{t+1}}^{\infty} \Big(aV_{t+2}(s'-1) + bV_{t+2}(s') + cV_{t+2}(s'+1)\Big)\mu_{s',t} \\
&= \beta N(1-\varphi^{-\delta})(\varphi^{s_{t-1}})^{\delta} \sum_{s'=s_{t+1}}^{\infty} \Big(aV_{t+2}(s'-1) + bV_{t+2}(s') + cV_{t+2}(s'+1)\Big)\varphi^{-s'\delta} \\
&= \beta N(1-\varphi^{-\delta})\varphi^{s_{t-1}\delta} \left(a \sum_{s'=s_{t+1}}^{\infty} V_{t+2}(s'-1)\varphi^{-s'\delta} + b \sum_{s'=s_{t+1}}^{\infty} V_{t+2}(s')\varphi^{-s'\delta} + c \sum_{s'=s_{t+1}}^{\infty} V_{t+2}(s'+1)\varphi^{-s'\delta}\right) \\
&= \beta N(1-\varphi^{-\delta})\varphi^{s_{t-1}\delta} \left(a \sum_{s'=s_{t+1}-1}^{\infty} V_{t+2}(s')\varphi^{-s'\delta}\varphi^{-\delta} + b \sum_{s'=s_{t+1}}^{\infty} V_{t+2}(s')\varphi^{-s'\delta} + c \sum_{s'=s_{t+1}+1}^{\infty} V_{t+2}(s')\varphi^{-s'\delta}\varphi^{\delta}\right) \\
&= \beta N(1-\varphi^{-\delta})\varphi^{s_{t-1}\delta} \left(\sum_{s'=s_{t+1}}^{\infty} \Big(a\varphi^{-\delta} + b + c\varphi^{\delta}\Big) V_{t+2}(s')\varphi^{-s'\delta} + aV_{t+2}(s_{t+1}-1)\varphi^{-\delta}\varphi^{-(s_{t+1}-1)\delta} - cV_{t+2}(s_{t+1})\varphi^{\delta}\varphi^{-(s_{t+1})\delta}\right)
\end{aligned}
$$

By definition we have $\varphi^{-\delta} = \frac{c}{a}$. Therefore, on the one hand we have $a\varphi^{-\delta} + b + c\varphi^{\delta} = a\frac{c}{a} + b + c\frac{a}{c} = c + b + a = 1$, and on the other hand we have $a = c\varphi^{\delta}$. Given our guess, and since $s_{t+1} - 1 < s_{t+1} < s_{t+1} + \eta = s_{t+2}$, we have $V_{t+2}(s_{t+1} - 1) = V_{t+2}(s_{t+1}) = \varphi^{s_{t+1}}W = g\varphi^{s_t}W$. Using all these results leads to:

$$
\begin{aligned}
&\beta \sum_{s'=s_{t+1}}^{\infty} \Big(aV_{t+2}(s'-1) + bV_{t+2}(s') + cV_{t+2}(s'+1)\Big)\mu_{s',t} \\
&= \beta N(1-\varphi^{-\delta})\varphi^{s_{t-1}\delta} \left(\sum_{s'=s_{t+1}}^{\infty} V_{t+2}(s')\varphi^{-s'\delta} + a\varphi^{-\delta}\varphi^{\delta}g\varphi^{s_t}W\varphi^{-s_{t+1}\delta} - c\varphi^{\delta}g\varphi^{s_t}W\varphi^{-s_{t+1}\delta}\right) \\
&= \beta N(1-\varphi^{-\delta})\varphi^{s_{t-1}\delta} \left(\sum_{s'=s_{t+1}}^{\infty} V_{t+2}(s')\varphi^{-s'\delta} + ag\varphi^{s_t}W\varphi^{-s_{t+1}\delta} - ag\varphi^{s_t}W\varphi^{-s_{t+1}\delta}\right) \\
&= \beta N(1-\varphi^{-\delta})\varphi^{s_{t-1}\delta} \sum_{s'=s_{t+1}}^{\infty} V_{t+2}(s')\varphi^{-s'\delta}
\end{aligned}
$$

The indifference condition between imitating and experimenting can be rewritten at $t+1$:

$$
g\varphi^{s_t} + \beta g\varphi^{s_t}W = \beta \sum_{s' \geq s_{t+1}} V_{t+2}(s')\frac{\mu_{s',t+1}}{Ng^{-\delta}} = g^{\delta}\beta(1-\varphi^{-\delta})\varphi^{s_t\delta} \sum_{s' \geq s_{t+1}} V_{t+2}(s')\varphi^{-s'\delta}
$$

Where we have used the fact that for $s' \geq s_{t+1} > s_t$ we have $\mu_{s',t+1} = N(1 - \varphi^{-\delta})\varphi^{s_t\delta}\varphi^{-s'\delta}$. Rearranging:

$$\sum_{s' \geq s_{t+1}} V_{t+2}(s')\varphi^{-s'\delta} = \frac{g\varphi^{s_t} + \beta g\varphi^{s_t}W}{g^{\delta}\beta(1-\varphi^{-\delta})\varphi^{s_t\delta}}$$

Substituting back:

$$\begin{aligned}\beta \sum_{s'=s_{t+1}}^{\infty} \Big(aV_{t+2}(s'-1) + bV_{t+2}(s') + cV_{t+2}(s'+1)\Big)\mu_{s',t} &= \beta N(1-\varphi^{-\delta})\varphi^{s_{t-1}\delta}\frac{g\varphi^{s_t} + \beta g\varphi^{s_t}W}{g^{\delta}\beta(1-\varphi^{-\delta})\varphi^{s_t\delta}} \\ &= \varphi^{s_t}Ng^{-2\delta}g(1+\beta W) \qquad (11)\end{aligned}$$

Substituting equations 10 and 11 back in equation 9 yields:

$$\begin{aligned}\sum_{s'=s_{t+1}}^{\infty} V_{t+1}(s')\mu_{s',t} &= \varphi^{s_{t+1}}Ng^{-2\delta}\frac{1-\varphi^{-\delta}}{1-\varphi^{1-\delta}} + \varphi^{s_t}Ng^{-2\delta}g(1+\beta W) \\ &= g\varphi^{s_t}Ng^{-2\delta}\frac{1-\varphi^{-\delta}}{1-\varphi^{1-\delta}} + \varphi^{s_t}Ng^{-2\delta}g(1+\beta W) \\ &= N\varphi^{s_t}g^{1-2\delta}\left(\frac{1-\varphi^{-\delta}}{1-\varphi^{1-\delta}} + 1 + \beta W\right)\end{aligned}$$

At the indifference point at time t we thus have:

$$\begin{aligned}g\varphi^{s_{t-1}} + \beta g\varphi^{s_{t-1}}W &= \beta N^{-1}g^{\delta}\left(\sum_{s'=s_t}^{s_{t+1}-1} V_{t+1}(s')\mu_{s',t} + \sum_{s'=s_{t+1}}^{\infty} V_{t+1}(s')\mu_{s',t}\right) \\ &= \beta N^{-1}g^{\delta}\left(\varphi^{s_t}WNg^{-\delta}(1-g^{-\delta}) + N\varphi^{s_t}g^{1-2\delta}\left(\frac{1-\varphi^{-\delta}}{1-\varphi^{1-\delta}} + 1 + \beta W\right)\right) \\ &= \beta\varphi^{s_t}\left(W(1-g^{-\delta}) + g^{1-\delta}\left(\frac{1-\varphi^{-\delta}}{1-\varphi^{1-\delta}} + 1 + \beta W\right)\right) \\ &= \beta g\varphi^{s_{t-1}}\left(W(1-g^{-\delta}) + g^{1-\delta}\left(\frac{1-\varphi^{-\delta}}{1-\varphi^{1-\delta}} + 1 + \beta W\right)\right)\end{aligned}$$

Rearranging

$$
\begin{aligned}
1+\beta W &= \beta\left(W(1-g^{-\delta})+g^{1-\delta}\left(\frac{1-\varphi^{-\delta}}{1-\varphi^{1-\delta}}+1+\beta W\right)\right)\\
&= \beta W - g^{-\delta}\beta W + \beta g^{1-\delta}\frac{1-\varphi^{-\delta}}{1-\varphi^{1-\delta}} + \beta g^{1-\delta} + \beta^2 g^{1-\delta} W\\
g^{\delta} &= -\beta W + \beta g\frac{1-\varphi^{-\delta}}{1-\varphi^{1-\delta}} + \beta g + \beta^2 g W
\end{aligned}
$$

If the following system of two equations and two unknowns W and g has a solution, then the existence of the BGP is shown.

$$
W = \frac{g}{1-g\beta} \tag{12}
$$

$$
\frac{1}{\beta}g^{\delta} = -W + g\frac{1-\varphi^{-\delta}}{1-\varphi^{1-\delta}} + g(1+\beta W) \tag{13}
$$

Eliminating W and solving for the growth rate g yields

$$
g = \left(\beta\frac{1-\varphi^{-\delta}}{1-\varphi^{1-\delta}}\right)^{\frac{1}{\delta-1}}
$$

As φ goes to zero, then g goes to $\left(\beta\frac{\delta}{\delta-1}\right)^{\frac{1}{\delta-1}}$ which is exactly the growth rate in Perla and Tonetti (2014).

□

The economy's growth rate g is decreasing in δ, the tail of the productivity distribution of firms. The fatter the tail, the smaller δ, the higher the growth rate g. The intuition is straightforward. Growth is led by the imitation of high productive firms by low productive firms. Thus if the distribution is fatter then the probability for a low productive firm to imitate a better firm is higher and the growth rate is higher. Figure 4 illustrates the relation between the growth rate and the tail of firms distribution.

2.3 Out of BGP Dynamics under Aggregate Uncertainty

The previous section assumed away any sampling variations arising from a finite number of firms, which can lead to aggregate uncertainty. With a finite number of firms that are distributed according to a Pareto distribution along a BGP, Gabaix (2011) shows

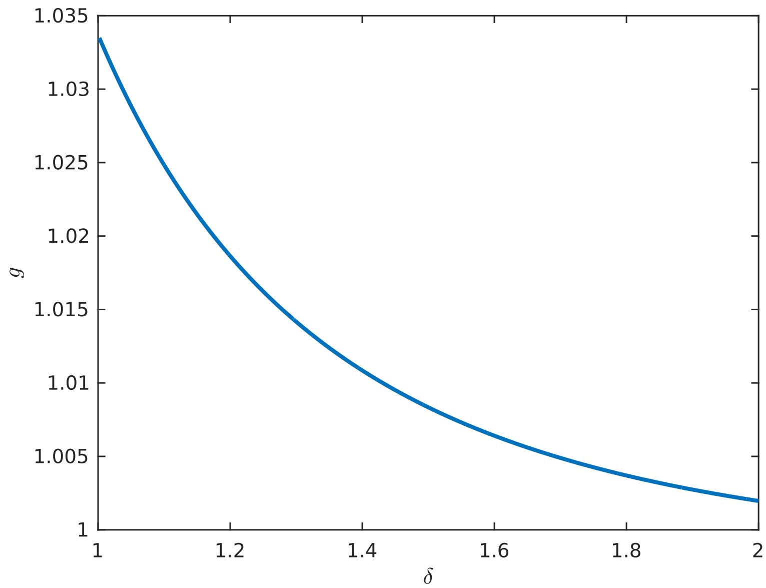

Figure 4: An illustration of the relation between the tail of the productivity distribution along a BGP and the growth rate.

that fat tails imply sizable aggregate fluctuations in response to firm-level disturbances. We now describe the nature and origins of aggregate volatility in our framework.

Figure 5 provides an intuitive illustration. There are only three productivity levels, and all firms start at the intermediate one. Each firm has probability 1/4 of going up or down and 1/2 of staying put. The top panel displays the case of a continuum of firms: at the next period exactly 1/2 of the firms are at intermediate level, 1/4 at top and 1/4 at low level. With a continuum the law of large number holds and there is no sampling variation. The future distribution of firms is a deterministic object. But if there is a finite number of firms - say four - then any arrangement of firms among the three productivity levels is possible with a positive probability. The triplet of the number of firms at each productivity level follows a multinomial distribution, where the number of trials is four and the vector of probabilities is$(1/4, 1/2, 1/4)'$. The future distribution of firms is a stochastic object. If the distribution of firms is a stochastic object, so are any aggregate quantities since there will be a moment of that distribution.

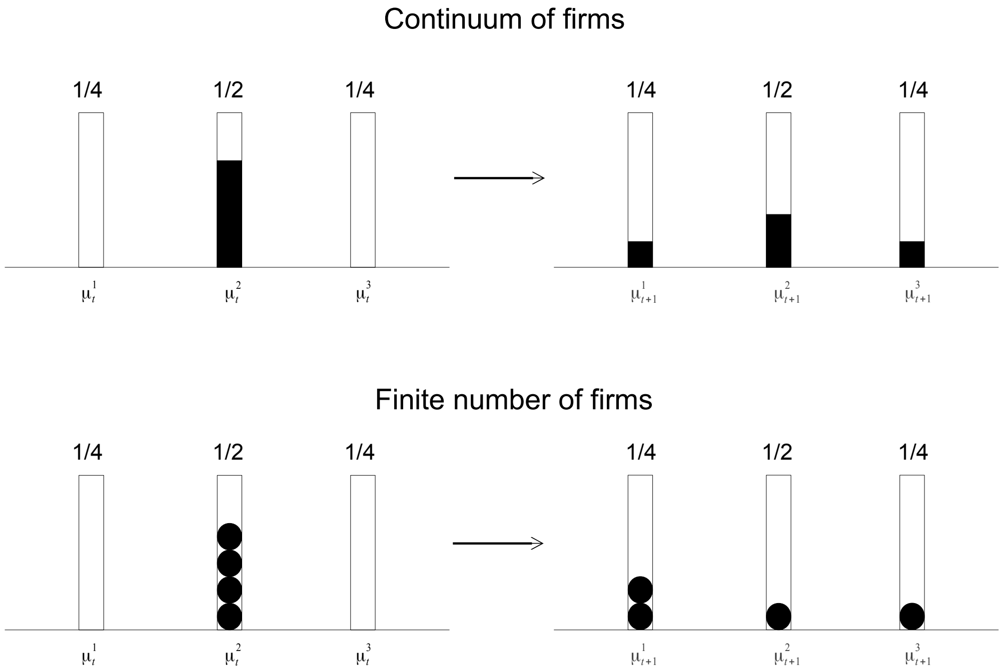

Figure 5: Example of sampling variation in the case of three productivity levels. The top panel has a continuum of firms, the bottom panel has a finite number of firms.

Firm Problem

The firms' dynamic problem must take into account the failure of the law of large numbers. The firm problem becomes

$$V(s,\mu_t) = \text{Max}\left\{\varphi^s + \beta\mathbb{E}_t\left[\sum_{s'} V(s',\mu_{t+1})P_{s,s'}\right] \quad ; \quad \beta\mathbb{E}_t\left[\sum_{s'\geq s_t} V(s',\mu_{t+1})\frac{\mu_{s',t}}{N-S_t}\right]\right\}$$

The key difference is the expectation operator $\mathbb{E}_t$ for the future value of firms. The problem also yields a thresholds rule as above.

$$\begin{cases} s < s_t, \quad V_t^I(s) > V_t^E(s) & \text{the firm decides to imitate} \\ s \geq s_t, \quad V_t^I(s) \leq V_t^E(s) & \text{the firm decides to experiment} \end{cases}$$

Stochastic Properties of the Productivity Distribution

This Section derives moments of μ_{t+1} conditional on μ_t. In the case of a finite number of firms, the evolution of the productivity distribution of firms is described by

$$\mu_{s,t+1} = \begin{cases} a\mu_{s+1,t} + b\mu_{s,t} + c\mu_{s-1,t} + S_t\frac{\mu_{s,t}}{N-S_t} + \varepsilon_{s,t} & \text{if} \quad s > s_t \\ a\mu_{s_t+1,t} + b\mu_{s_t,t} + S_t\frac{\mu_{s_t,t}}{N-S_t} + \varepsilon_{s_t,t} & \text{if} \quad s = s_t \\ a\mu_{s_t,t} + \varepsilon_{s_t-1,t} & \text{if} \quad s = s_t - 1 \\ 0 & \text{if} \quad s < s_t \end{cases}$$

where the variance-covariance structure of $\{\varepsilon_{s,t}\}_s$ is a function of $\{\mu_{s,t}\}_s$. The shocks $\{\varepsilon_{s,t}\}_s$ are the only difference with the continuum case and make the distribution of firms $\{\mu_{s,t}\}_s$ a stochastic object. The rest of this section is dedicated to the description of this stochastic structure.

Define two vectors for every s and k in $\{\varphi, \ldots, \varphi^s\}$,

$f_{t+1}^{s,k}$ = number of **experimenting** firms in state s at $t+1$ that were in state k at t

g_{t+1}^{s} = number of **imitating** firms in state s at $t+1$

It is straightforward to show that: $\mu_{s,t+1} = \sum_{k=s_t}^{\infty} f_{t+1}^{s,k} + g_{t+1}^{s}$. Given the structure of P we know that $f_{t+1}^{s,k} = 0$ for $(s,k) \notin \{(s,s-1),(s,s),(s,s+1)\}$ and thus

$$\mu_{s,t+1} = f_{t+1}^{s,s-1} + f_{t+1}^{s,s} + f_{t+1}^{s,s+1} + g_{t+1}^{s}$$

Define $\mu_{s,t+1}^{P} = \mu_{s,t+1} - g_{t+1}^{s} = f_{t+1}^{s,s-1} + f_{t+1}^{s,s} + f_{t+1}^{s,s+1}$ the dynamics due only to time t producers.

The vector $f_{t+1}^{\cdot,k} = \{f_{t+1}^{k-1,k}, f_{t+1}^{k,k}, f_{t+1}^{k+1,k}\}'$ (size 3×1) follows a multinomial distribution with number of trials $\mu_{k,t}$ and a variance-covariance matrix $W = diag(V) - VV'$ where $V = \{a,b,c\}'$. The $f_{t+1}^{\cdot,k}$ are independent for different k. It follows that

$$\mathbb{E}_t[\mu_{s,t+1}^{P}] = c\mu_{s-1,t} + b\mu_{s,t} + a\mu_{s+1,t}$$
$$\mathbb{V}ar_t[\mu_{s,t+1}^{P}] = c(1-c)\mu_{s-1,t} + b(1-b)\mu_{s,t} + a(1-a)\mu_{s+1,t}$$

Let us compute $\mathbb{C}ov_t[\mu^P_{s,t+1}, \mu^P_{k,t+1}]$. It is clear since the $f^{\cdot,s}_{t+1}$ are independent for different s and have support $\{s-1, s, s+1\}$ that $\mathbb{C}ov_t[\mu^P_{s,t+1}, \mu^P_{k,t+1}] = 0$ for $|s-k| > 2$. For a given s, we have

$$\begin{aligned}
\mathbb{C}ov_t[\mu^P_{s,t+1}, \mu^P_{s+1,t+1}] &= \mathbb{C}ov_t\left[f^{s,s-1}_{t+1} + f^{s,s}_{t+1} + f^{s,s+1}_{t+1}, f^{s+1,s}_{t+1} + f^{s+1,s+1}_{t+1} + f^{s+1,s+2}_{t+1}\right] \\
&= \mathbb{C}ov_t\left[f^{s,s}_{t+1}, f^{s+1,s}_{t+1}\right] + \mathbb{C}ov_t\left[f^{s,s+1}_{t+1}, f^{s+1,s+1}_{t+1}\right] \\
&= -\mu_{s,t} bc - \mu_{s+1,t} ab
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{C}ov_t[\mu^P_{s,t+1}, \mu^P_{s+2,t+1}] &= \mathbb{C}ov_t\left[f^{s,s-1}_{t+1} + f^{s,s}_{t+1} + f^{s,s+1}_{t+1}, f^{s+2,s+1}_{t+1} + f^{s+2,s+2}_{t+1} + f^{s+2,s+3}_{t+1}\right] \\
&= \mathbb{C}ov_t\left[f^{s,s+1}_{t+1}, f^{s+2,s+1}_{t+1}\right] \\
&= -ac\mu_{s+1,t}
\end{aligned}$$

This completes the description of the distribution of the random vector μ^P_{t+1}.

Focus now on the random vector $g_{t+1} = \{g^s_{t+1}\}$: it follows a multinomial distribution with number of trials S_t (the number of imitating firms) over the support $\{s_t, \ldots, S^{max}_t\}$ where S^{max}_t is the productivity level of the most productive producing firms at date t, and with the event probabilities vector $\frac{\mu_t}{N-S_t}$.[4]. It follows that for a given $s, k \in [s_t, \ldots, S^{max}_t]$ with $k \neq s$:

$$\begin{aligned}
\mathbb{E}_t\left[g^s_{t+1}\right] &= \frac{S_t}{N-S_t}\mu_{s,t} \\
\mathbb{V}ar_t\left[g^s_{t+1}\right] &= S_t\frac{\mu_{s,t}}{N-S_t}\left(1 - \frac{\mu_{s,t}}{N-S_t}\right) \\
\mathbb{C}ov_t\left[g^s_{t+1}, g^k_{t+1}\right] &= -S_t\frac{\mu_{s,t}\mu_{k,t}}{(N-S_t)^2}
\end{aligned}$$

The vector g_{t+1} is independent of the vectors $f^{\cdot,s}_{t+1}$.

Putting all results together, for $s \geq s_{t+1}$ and k such that $|s-k| > 2$, the moments of

[4] We abuse the notation between the sequence (or the size infinite vector) μ_t and the vector of size $(S^{max}_t - s_t + 1 \times 1)$ that stops where $\mu_{s,t} = 0 \forall s > S^{max}_t$.

the vector μ_{t+1} are given by:

$$
\begin{aligned}
\mathbb{E}_t\left[\mu_{s,t+1}\right] &= c\mu_{s-1,t} + b\mu_{s,t} + a\mu_{s+1,t} + \frac{S_t}{N-S_t}\mu_{s,t} \\
\mathbb{V}ar_t\left[\mu_{s,t+1}\right] &= c(1-c)\mu_{s-1,t} + b(1-b)\mu_{s,t} + a(1-a)\mu_{s+1,t} + S_t\frac{\mu_{s,t}}{N-S_t}\left(1-\frac{\mu_{s,t}}{N-S_t}\right) \\
\mathbb{C}ov_t\left[\mu_{s,t+1},\mu_{s+1,t+1}\right] &= -\mu_{s,t}bc - \mu_{s+1,t}ab - S_t\frac{\mu_{s,t}\mu_{s+1,t}}{(N-S_t)^2} \\
\mathbb{C}ov_t\left[\mu_{s,t+1},\mu_{s+2,t+1}\right] &= -\mu_{s+1,t}ac - S_t\frac{\mu_{s,t}\mu_{s+2,t}}{(N-S_t)^2} \\
\mathbb{C}ov_t\left[\mu_{s,t+1},\mu_{k,t+1}\right] &= -S_t\frac{\mu_{s,t}\mu_{k,t}}{(N-S_t)^2}
\end{aligned}
$$

Stochastic Properties of Aggregate Output

This section derives the stochastic properties of aggregate output. We first derive the dynamic process followed by aggregate output, which we use to derive its conditional variance.

Proposition 4 *Out of a Balanced Growth Path, output evolves according to (if $s_{t+1} > s_t$)*

$$
Y_{t+1} = \rho Y_t + \frac{S_t}{N-S_t}Y_t - \left(\rho + \frac{S_t}{N-S_t}\right)\sum_{s=s_t}^{s_{t+1}-1}\varphi^s\mu_{s,t} + O_t^Y + \sigma_t\epsilon_{t+1} \tag{14}
$$

where $\rho = a\varphi^{-1} + b + c\varphi$ *and* $O_t^Y = -a\varphi^{s_{t+1}-1}\mu_{s_{t+1},t} + c\varphi^{s_{t+1}}\mu_{s_{t+1}-1,t}$

Proof: For $s > s_t$, the productivity distribution at date $t+1$ is $a\mu_{s+1,t} + b\mu_{s,t} + c\mu_{s-1,t} + S_t \frac{\mu_{s,t}}{N-S_t}$. Let us assume that $s_{t+1} > s_t$. Aggregate output at date $t+1$ is

$$\begin{aligned}
Y_{t+1} &= \sum_{s=s_{t+1}}^{\infty} \varphi^s \mu_{s,t+1} \\
&= \sum_{s=s_{t+1}}^{\infty} \varphi^s \left(a\mu_{s+1,t} + b\mu_{s,t} + c\mu_{s-1,t} + \frac{S_t}{N-S_t}\mu_{s,t} \right) \\
&= a \sum_{s=s_{t+1}}^{\infty} \varphi^s \mu_{s+1,t} + b \sum_{s=s_{t+1}}^{\infty} \varphi^s \mu_{s,t} + c \sum_{s=s_{t+1}}^{\infty} \varphi^s \mu_{s-1,t} + \frac{S_t}{N-S_t} \sum_{s=s_{t+1}}^{\infty} \varphi^s \mu_{s,t} \\
&= a \sum_{s=s_{t+1}+1}^{\infty} \varphi^{s-1} \mu_{s,t} + b \sum_{s=s_{t+1}}^{\infty} \varphi^s \mu_{s,t} + c \sum_{s=s_{t+1}-1}^{\infty} \varphi^{s+1} \mu_{s,t} + \frac{S_t}{N-S_t} \sum_{s=s_{t+1}}^{\infty} \varphi^s \mu_{s,t} \\
&= \left(a\varphi^{-1} + b + c\varphi\right) \sum_{s=s_{t+1}}^{\infty} \varphi^s \mu_{s,t} - a\varphi^{s_{t+1}-1}\mu_{s_{t+1},t} + c\varphi^{s_{t+1}}\mu_{s_{t+1}-1,t} + \frac{S_t}{N-S_t} \sum_{s=s_{t+1}}^{\infty} \varphi^s \mu_{s,t}
\end{aligned}$$

We know that aggregate output at date t is $Y_t = \sum_{s=s_t}^{\infty} \varphi^s \mu_{s,t} = \sum_{s=s_t}^{s_{t+1}-1} \varphi^s \mu_{s,t} + \sum_{s=s_{t+1}}^{\infty} \varphi^s \mu_{s,t}$. Substitute in the previous equation:

$$Y_{t+1} = \left(a\varphi^{-1} + b + c\varphi + \frac{S_t}{N-S_t} \right) \left(Y_t - \sum_{s=s_t}^{s_{t+1}-1} \varphi^s \mu_{s,t} \right) - a\varphi^{s_{t+1}-1}\mu_{s_{t+1},t} + c\varphi^{s_{t+1}}\mu_{s_{t+1}-1,t}$$

$\square$

Equation 14 describes the evolution of aggregate output. Y_{t+1} can be decomposed in five terms. The first one is the contribution of time t experimenting firms, whose productivity is subject to the Markovian process. The second term is the contribution of time t imitating firms. They are producing at time $t+1$ and thus contribute to time $t+1$ aggregate output. The third term is the cost of time $t+1$ imitation. The firms that decided to imitate at time $t+1$ are not producing, which reduces aggregate output at time $t+1$. The fourth term is a correction term due to the discrete number of firms. Finally the last term is the stochastic perturbation on aggregate output due to idiosyncratic perturbations as described in figure 5.

Equation 14 can be solved for the evolution of output growth:

$$\frac{Y_{t+1}}{Y_t} = \rho \left(\frac{\sum_{s=s_{t+1}}^{\infty} \varphi^s \mu_{s,t}}{Y_t} \right) + \left(\frac{S_t}{N-S_t} \right) \left(\frac{\sum_{s=s_{t+1}}^{\infty} \varphi^s \mu_{s,t}}{Y_t} \right) + \frac{O_t^Y}{Y_t} + \frac{\sigma_t}{Y_t}\epsilon_{t+1}$$

Output growth responds to two terms: the share of the output of the largest firm $\left(\frac{\sum_{s=s_{t+1}}^{\infty} \varphi^s \mu_{s,t}}{Y_t}\right)$, and the ratio of small to large firms. The first term captures persistence: persistence increases in the output share of large firms. The second term captures convergence: growth increases in the number of small relative to large firms, conditional on the output share of large firms. We next characterize the conditional variance of output growth rate.

Proposition 5 *Out of a Balanced Growth Path, the conditional variance of the output growth is*

$$\frac{\sigma_t^2}{Y_t^2} = \mathbb{V}ar_t\left[\frac{Y_{t+1}}{Y_t}\right] = \varrho \frac{\sum_{s=s_{t+1}} \varphi^{2s}\mu_{s,t}}{Y_t^2} + S_t \frac{\mathbb{V}ar_{i\in\mathcal{P}_t}[\varphi_{s_{i,t}}]}{Y_t^2} + \frac{O_t^\sigma}{Y_t^2} \tag{15}$$

where $\mathcal{P}_t$ is the set of large firms ($s_{i,t} > s_{t+1}$) at t and $\varrho = a\varphi^{-2} + b + c\varphi^2 - \rho^2$.

Proof: See Appendix A.1. □

The first term on the right hand side of equation 15 is akin to Carvalho and Grassi (2015). Aggregate volatility increases in the dispersion among large firms - measured by the second moment of the cross-sectional distribution of large firms - because (experimenting) large firms create persistence. The second term is due to imitating firms: volatility also increases in the dispersion of large firms because this dispersion captures the potential churning created by (imitating) small firms. The intuition becomes clearer if the second term is rewritten as

$$\frac{S_t}{Y_t^2}\mathbb{V}ar_{i\in\mathcal{P}_t}[\varphi_{s_{i,t}}] = \frac{S_t}{N-S_t}\frac{\sum_{s=s_{t+1}} \varphi^{2s}\mu_{s,t}}{Y_t^2} - \frac{S_t}{(N-S_t)^2}\left(\frac{\sum_{s=s_{t+1}} \varphi^{s}\mu_{s,t}}{Y_t}\right)^2$$

Where the dispersion of large firms $\frac{\sum_{s=s_{t+1}} \varphi^{2s}\mu_{s,t}}{Y_t^2}$ now enters specifically.

The following proposition describe how each term of equation 15 evolves as the number of firms N increases along a BGP.

Proposition 6 *Along a BGP with a finite number of firms, when $N \to \infty$:*

$$\left(\varrho + \frac{S_t}{N-S_t}\right)\frac{\sum_{s=s_{t+1}} \varphi^{2s}\mu_{s,t}}{Y_t^2} \underset{N\to\infty}{\sim} \frac{K_1}{N^{2(1-\frac{1}{\delta})}}$$

and

$$\frac{S_t}{(N-S_t)^2}\left(\frac{\sum_{s=s_{t+1}}\varphi^s\mu_{s,t}}{Y_t}\right)^2 \underset{N\to\infty}{\sim} \frac{K_2}{N}$$

When $1<\delta<2$ *(fat tail distribution) then* $2(1-\frac{1}{\delta})<1$*, which implies*

$$\frac{\sigma_t}{Y_t}=\sqrt{\mathbb{V}ar_t\left[\frac{Y_{t+1}}{Y_t}\right]}\underset{N\to\infty}{\sim}\frac{\sqrt{K_1}}{N^{1-\frac{1}{\delta}}} \tag{16}$$

Proof: See Appendix A.1. □

This result is related to Gabaix (2011) and Carvalho and Grassi (2015). In the thin tail case aggregate volatility decays at rate $1/\sqrt{N}$, i.e., faster than $1/N^{1-\frac{1}{\delta}}$ for $1<\delta<2$. Therefore, aggregate volatility arising from pure idiosyncratic shocks can be quantitatively important.

Equation 16 is an approximation when the number of firms is large; it defines a negative relation between output growth volatility and the tail index δ of the distribution of firms along a BGP. A tail index of the distribution of firm determines both aggregate volatility and the long term growth rate of output. A fat productivity distribution of firm, i.e. a low tail index, implies both high aggregate volatility and high long term growth rate.

3 Growth, Volatility and Firm Size Distribution

In this section, we are testing the relationship between the fatness of the firm size distribution, a proxy for the distribution of firms across productivity levels and the long term output growth and the volatility of the output growth rate. We are first describing the data sources before moving to the main results. We then describe how sectoral tails are estimated on firm-level data. We investigate the link between sectoral growth and sectoral tail, and perform some robustness check. Finally, we study the link between sectoral volatility and sectoral tail, also with some robustness checks.

3.1 Data Description

In the empirical part of this paper, we use two sources of information: *i*) firm-level data collected by Compustat, and *ii*) sector-level data, collected by the NBER-CES Manufacturing Industry Database. The first database is used to estimate tails at sector level; the second database is used to compute sectoral growth and volatility. The rest of this section is dedicated to the description of these data.

Firm Level Data

Compustat data are collected from the mandatory forms that each listed firm in the US fills each year. This is a firm-level yearly panel database with balance sheet information. For each 4-digit SIC category, we collect three variables between 1958 and 2009: employment, sales and total assets. The latter two are nominal. They are deflated using the price deflator given by the NBER-CES Manufacturing Industry Database for shipment (PISHIP). These deflators are computed by the Bureau of Economic Analysis, for their GDP-by-Industry data.

Sector Level Data

The NBER-CES Manufacturing Industry Database collects sector-level data. The majority of these data are extracts from the Annual Survey of Manufacturing that samples approximatively 50,000 establishments selected from the approximatively 330,000 establishments included in the Census of Manufacturing. The variables are only available for the manufacturing sector. The data are annual and cover the period 1958-2009[5]. For each 4-digit SIC category, we use information on sales (VSHIP), total factor productivity (TFP), and employment. Sales are deflated following the method recommended by the NBER using the provided associated deflator (PISHIP). For the TFP we use the 5-factor TFP index (TFP5) computed by the NBER.

Armed with these variables, we compute the growth rate of sales and TFP. For robustness, we also filter the data using the Hodrick-Prescott filter (with smoothing parameter 100). We compute mean growth rate and variance, on a centered rolling window of 11 years. For example, the mean growth rate in 1985 is the mean of the growth rates of years 1980 to 1991. These mean are interpreted as long-term growth rate along a SBGP.

[5]Both 6-digits NAICS and 4-digit SIC level data are available. For consistency, we use the 4-digit SIC information.

3.2 Estimation of Tails

The tails of the sector-level size distributions are estimated following Clauset, Shalizi and Newman (2009). This uses a maximum likelihood estimator related to Hill (1975), in order to estimate the tail parameter of a distribution. Importantly, the method also allows to estimate the threshold above which the distribution can be well described by a Pareto distribution.

Traditionally, tail indexes are estimated using an OLS regression of firms' log-rank on their log-size, for firms larger than a arbitrarily chosen threshold.[6] While convenient, the arbitrariness of the threshold is a drawback. A higher threshold increases the fit of the truncated distribution to a Pareto, but it also reduces the sample of firms that are used in the regression. Clauset, Shalizi and Newman (2009) estimate the threshold from the data, where is is chosen optimally so that the fit of the truncated distribution to a Pareto is maximized.

Firm size is measured using sales, employment or total assets. Compustat data imply too few firms in a given sector and in a given to perform the Clauset, Shalizi and Newman (2009) estimation. We pool firms in a given sector over a centered 11 years rolling window. For example, the tail of the firm size distribution in sector 2011 (Meat Packing Plants) in 1985 is estimated on the sample formed firms in that sectors in years 1980 to 1991. The window length is the same as what was used to compute mean growth rate and volatility.

We obtain a panel formed by the estimated tails of firm size distribution, at the sector and year levels. The sample is winsorized , and we keep sector year observations with tails that are estimated on at least 20 firms, and where the estimated tails are statistically different from zero at a 5 percent confidence level.

Figure 6 displays the density of the pooled panel of estimated tails, using sales, employment and total assets as a measure of firm size. There is considerable sector-level heterogeneity in tail estimates, but the mass of estimates is below one irrespective how firm size is measured. A tail parameter below one means that, at least at the right tail, sector-level size distributions follow a Zipf law. Therefore, a majority of sectors have a size distribution fatter than Zipf law. They are highly concentrated sectors. Of course Compustat data are not representative as they consider listed firms only. They tend to be larger. But Compustat is presumably representative of large firms, i.e. firms at the right end of the firm size distribution, which is the basis for these estimations.

[6]See Gabaix and Ibragimov (2011) for example.

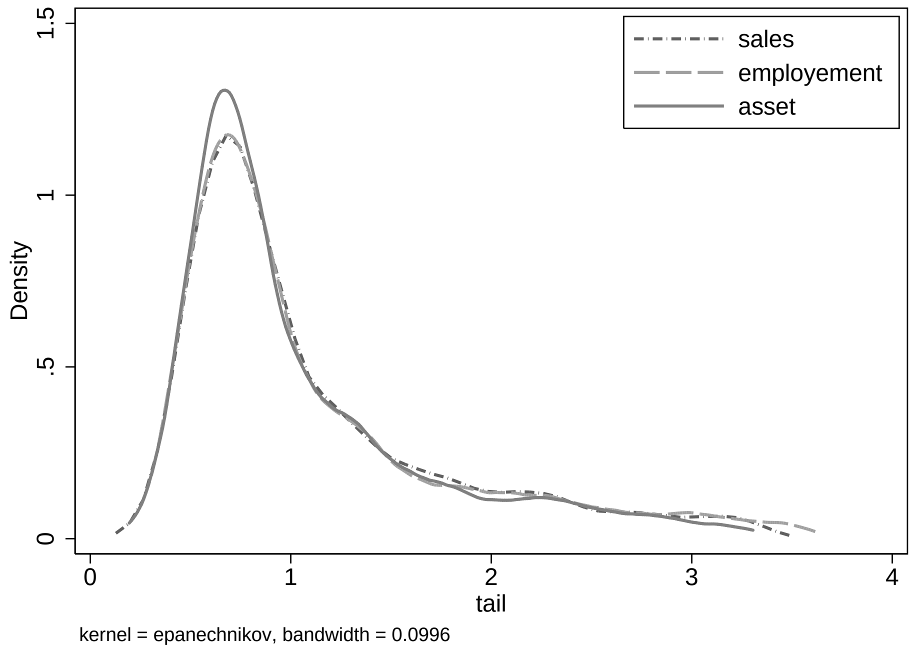

Figure 6: Density of the panel of distribution tails estimates of sector*year. DATA: Compustat. METHOD: following Clauset, Shalizi and Newman (2009).

	Sales	Employment	Assets
Mean	1.11	1.13	1.05
Std.	0.67	0.71	0.62
Min	0.23	0.25	0.19
Max	3.40	3.62	3.31
Mean # firms	105.42	123.02	112.42
Std. # firms	114.11	148.20	135.59
Min # firms	21	21	21
Max # firms	1144	1321	1398
Observations	2391	2152	2311

Table 1: Descriptive statistics on the tails estimates of a panel of sector.

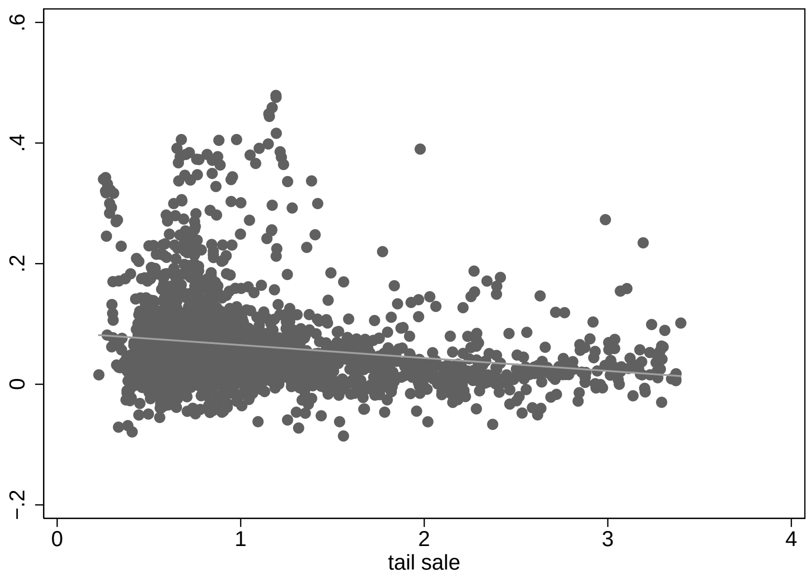

Figure 7: Scatter plot of mean growth rate of sales (y-axis) and tail estimates for sales (x-axis) for sector year observations.

Table 1 displays summary statistics for the panel of estimated tails. On average, tail estimates are close to, but above one. Given figure 6, this is consistent with right skewness. The bottom panel of the table reports descriptive statistics for the number of firms in each sector for which the tail index is estimated. This number is endogenously determined by the estimation. Since observations are dropped whenever tails are estimated on fewer than 20 firms, the minimum number of firms is 21 for all measures of size. The mean number of firms goes from 105.42 to 123.02 depending on the measure of firm size considered. The number of sector year observations is above 2,150.

3.3 Tail and Growth

This sub-section documents the empirical link between long-term growth and the tail of firm size distribution, δ, both at sector level. Theory suggests the link should be negative. Figure 7 displays a scatter of the mean growth rate of sales against the corresponding estimated tail, across sectors and years. The figure suggests a negative relationship.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ln(gr.y)	ln(gr.y)	ln(gr.y)	ln(gr.y)	ln(gr.y)	ln(gr.y)	ln(gr.y)	ln(gr.y)
$\ln(1+\delta\_sale)$	-0.0467***	0.0109	-0.0511***	0.00395	-0.0485***	0.00952	-0.0542***	-0.000873
	(0.0154)	(0.0108)	(0.0157)	(0.0101)	(0.0156)	(0.0116)	(0.0162)	(0.0101)
Employment					0.0000422	0.0000629	0.0000389	0.000102
					(0.000172)	(0.000113)	(0.000177)	(0.000104)
Observations	2115	2115	2115	2115	1917	1917	1917	1917
Adjusted R^2	0.035	0.002	0.055	0.129	0.041	0.003	0.069	0.142
FE_year	No	No	Yes	Yes	No	No	Yes	Yes
FE_sector	No	Yes	No	Yes	No	Yes	No	Yes

Robust standard errors clustered at the sector level in parenthesis
***p<0.01, **p<0.05, *p<0.1

Table 2: Baseline specification for growth and tail relationship.

To test the prediction formally, we estimate the following model:

$$\log(1+growth_{t,s}) = \kappa + sector_s + year_t + \beta \log(1+tail_{t,s}) + \gamma X_{t,s} + \epsilon_{t,s} \tag{17}$$

κ is a constant, $year_t$ is a time effect, $sector_s$ is a sector fixed effect and $X_{t,s}$ is a set of controls. The estimations are performed with and without time and/or sector fixed effect. The coefficient β is expected to be negative. In the baseline specification, it is the mean growth rate of sales at the 4-digit SIC level that measures $growth_{t,s}$. We also control for the size of the sector, as measured by the total number of employees there.

Table 2 reports the result. Columns (1) to (4) display the result without control at the sectoral level, columns (5) to (8) include sector-level employment. The negative relation between long-term growth and the tail index is statistically significant and negative in the absence of sector fixed effects, whether year effects are controled for or not. Estimates of β become insignificant once sector fixed effects are added. In the cross section of sectors, fat tails are associated with high growth. But over time, fluctuations within-sector in the tail of size distribution are not associated with any difference in growth. This latter results is consistent with the above theory. Indeed, the results derive in Proposition 3 is valide along a BGP and thus our theory only predict that this relation holds between sectors.

3.4 Tail and Volatility

This section analyzes the empirical relationship between volatility and tail indexes, both at sector level. Theory suggests the link should be negative.

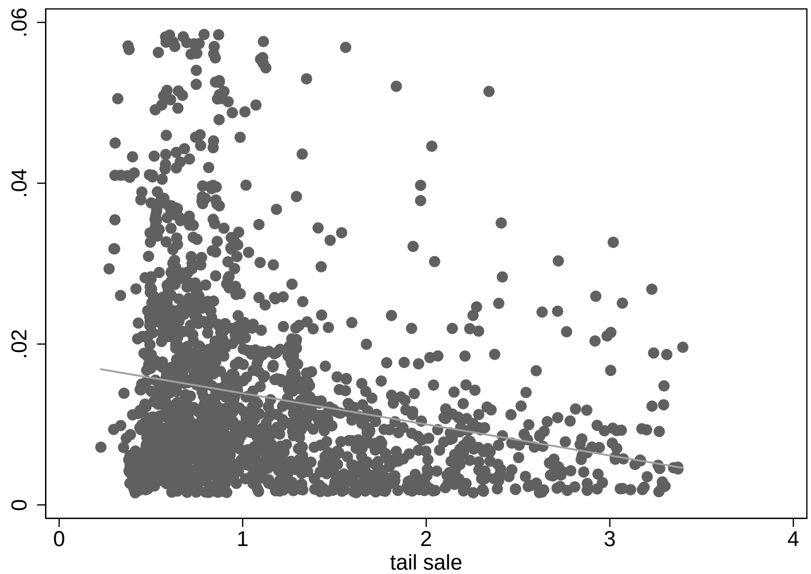

Figure 8: Scatter plot of variance of growth rate of sales (y-axis) and tail estimates (x-axis) for sales for sector year observations.

Figure 8 displays a scatter of the variance in the growth rate of sales against the estimates of the tail index, across both sectors and years. The figure suggests a negative relation.

To test this prediction formally, we estimate the following empirical model:

$$vol_{t,s} = \kappa + sector_s + year_t + \beta(1 + tail_{t,s}) + \gamma N_{t,s} + \epsilon_{t,s} \tag{18}$$

Once again, κ is a constant, $year_t$ is a time-fixed effect, $sector_s$ is a sector fixed effect, $vol_{t,s}$ is a measure of sectoral volatility, $tail_{t,s}$ a measure of the tail index, and $N_{t,s}$ a measure of the number of firms in a given sector and a given year. We expect the estimate of β to be negative as it is in the equation 16. In this equation the sign of γ remains indeterminate.

In the baseline specification of the model 18, we use the 11-year rolling window variance of sector-level growth rate as a measure of sectoral volatility, $vol_{t,s}$. The total number of employees is a proxy for the number of firms.

Table 3 displays the results for the baseline specification. Columns (5) to (8) include a control for the number of firms in a sector, whereas columns (1) to (4) do not. The table

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Var(gr.y)	Var(gr.y)	Var(gr.y)	Var(gr.y)	Var(gr.y)	Var(gr.y)	Var(gr.y)	Var(gr.y)
$1 + \delta\_sale$	-0.00387***	-0.00334***	-0.00408***	-0.00379***	-0.00386***	-0.00306***	-0.00413***	-0.00368***
	(0.000785)	(0.00101)	(0.000795)	(0.000865)	(0.000687)	(0.00116)	(0.000700)	(0.00106)
Employment					-0.0000350*	-0.00000848	-0.0000386*	-0.0000106
					(0.0000197)	(0.0000218)	(0.0000199)	(0.0000251)
Observations	1817	1817	1817	1817	1649	1649	1649	1649
Adjusted R^2	0.046	0.040	0.077	0.148	0.058	0.035	0.085	0.117
FE_year	No	No	Yes	Yes	No	No	Yes	Yes
FE_sector	No	Yes	No	Yes	No	Yes	No	Yes

Robust standard errors clustered at the sector level in parenthesis

***p<0.01, **p<0.05, *p<0.1

Table 3: Baseline specification for volatility and tail relationship (model 18)

shows that the negative relationship between volatility and the tail index is statistically different from zero and robust to the introduction of sectoral and time fixed effects. Furthermore, including the total number of employees barely affects estimates of β. The estimated value for γ is barely significant at 10 percent confidence level across specifications.

The fact that estimates of β are negative and statistically significant irrespective of fixed effects means the relation between sector volatility and tails holds both between and within sectors. The former is consistent with the model derived in the revious section. In the model, volatility decreases with the tail of the stationary firm size distribution along a BGP: the correlation is negative between sectors. In other words the model predict a *between* sectors negative relationship between volatility and tail index.

But the fact that the estimate of β is negative and statistically significant also with sector fixed effects also implies a negative correlation *within* sector. The intuition is that when large firms, i.e firms with good ideas, experiment higher idiosyncratic shocks, these firms become larger. Hence the firm size distribution becomes fatter and the volatility is higher: the negative (resp. positive) relation between tail index (resp. tail fatness) and sector volatility holds also *within* sectors.

To ensure robustness, we considere other specifications. We use tail indexes estimated on firm employment and total assets, and use other measures of sector volatility based on TFP, or HP-filtered data instead of growth rates. The results are presented in appendix B Table 5 shows the negative significance of β prevails in all cases. Finally, table 6 displays the results for clustered standard errors in alternative dimensions. The results are once again robust.

3.5 Tails and the Growth-Volatility Relation

The framework presented in section 2 rationalizes the existence of a positive relationship between the tail indexes and both sectoral growth and volatility. This implies the dispersion in tail indexes can account for at least a fraction of the positive relation between growth and volatility. In this section, we investigate how large a fraction. We proceed in two steps. First, the cross-section of sectoral volatility is regressed on the estimated tail indexes. We then explore how much of the dispersion in sector growth rates can be explained by the fitted values of sector volatilities, explained by tail indexes only. Since the tails explain only long run differences in sector-level growth, the regressions include time effects only.

	(IV)	(non IV)
	gr.y	gr.y
Var(gr.y)	3.488***	1.806***
	(0.757)	(0.248)
Observations	1817	1817
R^2	0.120	0.297
FE_year	Yes	Yes
FE_sector	No	No

Robust standard errors clustered at the sector level in parenthesis
***p<0.01, **p<0.05, *p<0.1

Table 4: IV vs non IV estimation of Growth and Volatility relationship

Table 4 compares the result of this regression (column (IV)) with one that simply regresses sector growth on its volatility (column (non-IV)). Sectoral volatility is measured by the variance of the sales growth rates, while sectoral growth is its mean, both computed on 11-year rolling windows. The dispersion in sectoral volatility can account for about 29 percent of sectoral growth. When volatility is instrumented by tail indexes, it can still account for about 12 percent of sectoral growth. Therefore, tail indexes explain about 40% of the growth-volatility relationship between sectors.

4 Conclusion

This paper develops a sector-level analysis of the relation between the distribution of firms, growth, and volatility. We develop an "idea flows" model where firms can either imitate existing technologies (a risky strategy), or experiment with new, random technologies (a conservative strategy). In equilibrium only relatively large firms experiment, which engenders an expanding technology frontier. Small firms tend to imitate, which, provided a few large firms are in existence, generates sector risk, growth, and volatility. The implied relation between growth, volatility, and the share of large firms holds in US sector data.

We embed our theory in an aggregate model of structural transformation. The model can account for the consequences of aggregation on the growth-volatility link, provided one assumes goods are substitutes between granular sectors, but complements at a more aggregate level. A six-digit ”sector” is an aggregate of substitutable goods; both growth and volatility in that sector reflect the characteristics of high technology firms, since they command most factors of production. But a one-digit ”sector” is an aggregate of complements; both growth and volatility now reflect the features of low technology firms, which constitute the bulk of the factors of production. The direct implication is that growth, volatility, and the distribution of firms should stop correlating as the data are aggregated. The prediction is borne out in US data.

References

[1] Philippe Aghion and Peter Howitt. A Model of Growth through Creative Destruction. *Econometrica*, 60(2):323–51, March 1992.

[2] Philippe Aghion and Gilles Saint-Paul. On the Virtue of Bad Times: An Analysis of the Interaction between Economic Fluctuations and Productivity Growth. CEPR Discussion Papers 578, C.E.P.R. Discussion Papers, September 1991.

[3] Fernando E. Alvarez, Francisco J. Buera, and Jr. Robert E. Lucas. Models of idea flows. NBER Working Papers 14135, National Bureau of Economic Research, Inc, June 2008.

[4] William J. Baumol. Macroeconomics of unbalanced growth: The anatomy of urban crisis. *The American Economic Review*, 57(3):415–426, 1967.

[5] Ricardo J Caballero and Mohamad L Hammour. The Cleansing Effect of Recessions. *American Economic Review*, 84(5):1350–68, December 1994.

[6] Vasco M. Carvalho and Basile Grassi. Large firm dynamics and the business cycle. *CEPR*, (DP10587), May 2015.

[7] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, Vol. 51(No. 4):pp. 661–703, 2009.

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

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

[10] 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.

[11] Bruce M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, Vol. 3(No. 5):pp. 1163–1174, Sept. 1975.

[12] Jean Imbs. Growth and volatility. *Journal of Monetary Economics*, 54(7):1848–1862, October 2007.

[13] Miklós Koren and Silvana Tenreyro. Volatility and development. *The Quarterly Journal of Economics*, 122(1):243–287, 02 2007.

[14] Robert E. Lucas. Ideas and growth. *Economica*, 76(301):1–19, 02 2009.

[15] Robert E. Lucas and Benjamin Moll. Knowledge growth and the allocation of time. *Journal of Political Economy*, 122(1):1 – 51, 2014.

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

[17] L. Rachel Ngai and Christopher A. Pissarides. Structural Change in a Multisector Model of Growth. *American Economic Review*, 97(1):429–443, March 2007.

[18] Jesse Perla and Christopher Tonetti. Equilibrium Imitation and Growth. *Journal of Political Economy*, 122(1):52 – 76, 2014.

[19] Garey Ramey and Valerie A Ramey. Cross-country evidence on the link between volatility and growth. *American Economic Review*, 85(5):1138–51, December 1995.

[20] Paul M. Romer. Mathiness in the Theory of Economic Growth. *American Economic Review*, 105(5):89–93, May 2015.

[21] Gilles Saint-Paul. Productivity growth and the structure of the business cycle. *European Economic Review*, 37(4):861–883, May 1993.

A Proofs

A.1 Proof of Proposition 5

By definition $Y_{t+1} = \sum_{s=s_{t+1}}^{\infty} \varphi^s \mu_{s,t+1}$ and $\mu_{s,t+1} = f_{t+1}^{s,s-1} + f_{t+1}^{s,s} + f_{t+1}^{s,s+1} + g_{t+1}^{s}$ for $s \geq s_{t+1}$[7]. It follows:

[7] Note that there are no infinite sum here, since there are a finite number of firms and thus $\mu_{s,t+1} = 0$ for s large enough. The summations in the text converge.

$$
\begin{aligned}
\mathbb{V}ar_t[Y_{t+1}] &= \mathbb{V}ar_t\left[\sum_{s=s_{t+1}}^{\infty}\varphi^s\mu_{s,t+1}\right] \\
&= \sum_{s=s_{t+1}}^{\infty}\mathbb{V}ar_t\left[\varphi^s\mu_{s,t+1}\right] + 2\sum\sum_{s_{t+1}\leq s<k\leq\infty}\mathbb{C}ov_t\left[\varphi^s\mu_{s,t+1},\varphi^k\mu_{k,t+1}\right] \\
&= \sum_{s=s_{t+1}}^{\infty}\varphi^{2s}\left(c(1-c)\mu_{s-1,t} + b(1-b)\mu_{s,t} + a(1-a)\mu_{s+1,t} + S_t\frac{\mu_{s,t}}{N-S_t}\left(1-\frac{\mu_{s,t}}{N-S_t}\right)\right) + \dots \\
&\quad \dots + 2\sum\sum_{s_{t+1}\leq s<k\leq\infty}\varphi^s\varphi^k\mathbb{C}ov_t\left[\mu^P_{s,t+1}+g^s_{t+1},\mu^P_{k,t+1}+g^k_{t+1}\right] \\
&= \sum_{s=s_{t+1}}^{\infty}\varphi^{2s}\left(c(1-c)\mu_{s-1,t} + b(1-b)\mu_{s,t} + a(1-a)\mu_{s+1,t} + S_t\frac{\mu_{s,t}}{N-S_t}\left(1-\frac{\mu_{s,t}}{N-S_t}\right)\right) + \dots \\
&\quad \dots + 2\sum\sum_{s_{t+1}\leq s<k\leq\infty}\varphi^s\varphi^k\mathbb{C}ov_t\left[\mu^P_{s,t+1},\mu^P_{k,t+1}\right] + \dots \\
&\quad \dots + 2\sum\sum_{s_{t+1}\leq s<k\leq\infty}\varphi^s\varphi^k\mathbb{C}ov_t\left[g^s_{t+1},g^k_{t+1}\right] \\
&= \sum_{s=s_{t+1}}^{\infty}\varphi^{2s}\left(c(1-c)\mu_{s-1,t} + b(1-b)\mu_{s,t} + a(1-a)\mu_{s+1,t} + S_t\frac{\mu_{s,t}}{N-S_t}\left(1-\frac{\mu_{s,t}}{N-S_t}\right)\right) + \dots \\
&\quad \dots + 2\sum_{s=s_{t+1}}^{\infty}\varphi^{2s+1}\mathbb{C}ov_t\left[\mu^P_{s,t+1},\mu^P_{s+1,t+1}\right] + \varphi^{2s+2}\mathbb{C}ov_t\left[\mu^P_{s,t+1},\mu^P_{s+2,t+1}\right]\dots \\
&\quad \dots - \frac{2S_t}{(N-S_t)^2}\sum\sum_{s_{t+1}\leq s<k\leq\infty}\varphi^s\varphi^k\mu_{s,t}\mu_{k,t} \\
&= \sum_{s=s_{t+1}}^{\infty}\varphi^{2s}\left(c(1-c)\mu_{s-1,t} + b(1-b)\mu_{s,t} + a(1-a)\mu_{s+1,t} + S_t\frac{\mu_{s,t}}{N-S_t}\left(1-\frac{\mu_{s,t}}{N-S_t}\right)\right) + \dots \\
&\quad \dots + 2\sum_{s=s_{t+1}}^{\infty}\varphi^{2s}\left(-\mu_{s,t}bc - \mu_{s+1,t}ab\right) + \varphi^{2s+1}\left(-ac\mu_{s+1,t}\right)\dots \\
&\quad \dots - \frac{2S_t}{(N-S_t)^2}\sum\sum_{s_{t+1}\leq s<k\leq\infty}\varphi^s\varphi^k\mu_{s,t}\mu_{k,t} \\
&= \sum_{s=s_{t+1}}^{\infty}\varphi^{2s}\left(c(1-c)\mu_{s-1,t} + b(1-b)\mu_{s,t} + a(1-a)\mu_{s+1,t}\right) + \dots \\
&\quad \dots + 2\sum_{s=s_{t+1}}^{\infty}\varphi^{2s+1}\left(-\mu_{s,t}bc - \mu_{s+1,t}ab\right) + \varphi^{2s+2}\left(-ac\mu_{s+1,t}\right) + \dots \\
&\quad \dots + \sum_{s=s_{t+1}}^{\infty}\varphi^{2s}S_t\frac{\mu_{s,t}}{N-S_t}\left(1-\frac{\mu_{s,t}}{N-S_t}\right) - \frac{2S_t}{(N-S_t)^2}\sum\sum_{s_{t+1}\leq s<k\leq\infty}\varphi^s\varphi^k\mu_{s,t}\mu_{k,t}
\end{aligned}
$$

Focus on the first term

$$
\begin{aligned}
&\sum_{s=s_{t+1}}^{S_{t+1}^{max}} \varphi^{2s}\left(c(1-c)\mu_{s-1,t} + b(1-b)\mu_{s,t} + a(1-a)\mu_{s+1,t}\right) \\
=&c(1-c)\sum_{s=s_{t+1}}^{S_{t+1}^{max}} \varphi^{2s}\mu_{s-1,t} + b(1-b)\sum_{s=s_{t+1}}^{S_{t+1}^{max}} \varphi^{2s}\mu_{s,t} + a(1-a)\sum_{s=s_{t+1}}^{S_{t+1}^{max}} \varphi^{2s}\mu_{s+1,t} \\
=&c(1-c)\sum_{s=s_{t+1}-1}^{S_{t+1}^{max}-1} \varphi^{2s+2}\mu_{s,t} + b(1-b)\sum_{s=s_{t+1}}^{S_{t+1}^{max}} \varphi^{2s}\mu_{s,t} + a(1-a)\sum_{s=s_{t+1}+1}^{S_{t+1}^{max}+1} \varphi^{2s-2}\mu_{s,t} \\
=&\left(c(1-c)\varphi^2 + b(1-b) + a(1-a)\varphi^{-2}\right)\sum_{s=s_{t+1}}^{S_{t+1}^{max}} \varphi^{2s}\mu_{s,t} + c(1-c)\varphi^2\left(\varphi^{2(s_{t+1}-1)}\mu_{s_{t+1}-1,t} - \varphi^{2(S_{t+1}^{max})}\mu_{S_{t+1}^{max},t}\right) + \dots \\
&\dots + a(1-a)\varphi^{-2}\left(\varphi^{2(S_{t+1}^{max}+1)}\mu_{S_{t+1}^{max}+1,t} - \varphi^{2s_{t+1}}\mu_{s_{t+1},t}\right)
\end{aligned}
$$

Focus now on the second term

$$
\begin{aligned}
&2\sum_{s=s_{t+1}}^{S_{t+1}^{max}} \varphi^{2s+1}\left(-\mu_{s,t}bc - \mu_{s+1,t}ab\right) + \varphi^{2s+2}\left(-ac\mu_{s+1,t}\right) \\
=&2\left(-bc\sum_{s=s_{t+1}}^{S_{t+1}^{max}} \varphi^{2s+1}\mu_{s,t} - ab\sum_{s=s_{t+1}}^{S_{t+1}^{max}} \varphi^{2s+1}\mu_{s+1,t} - ac\sum_{s=s_{t+1}}^{S_{t+1}^{max}} \varphi^{2s+2}\mu_{s+1,t}\right) \\
=&-2\left(bc\varphi\sum_{s=s_{t+1}}^{S_{t+1}^{max}} \varphi^{2s}\mu_{s,t} + ab\varphi^{-1}\sum_{s=s_{t+1}+1}^{S_{t+1}^{max}+1} \varphi^{2s}\mu_{s,t} + ac\sum_{s=s_{t+1}+1}^{S_{t+1}^{max}+1} \varphi^{2s}\mu_{s,t}\right) \\
=&-2\left(bc\varphi + ac + ab\varphi^{-1}\right)\sum_{s=s_{t+1}}^{S_{t+1}^{max}} \varphi^{2s}\mu_{s,t} - 2ab\varphi^{-1}\left(\varphi^{2(S_{t+1}^{max}+1)}\mu_{S_{t+1}^{max}+1,t} - \varphi^{2(s_{t+1}+1)}\mu_{s_{t+1}+1,t}\right) + \dots \\
&\dots - 2ac\left(\varphi^{2(S_{t+1}^{max}+1)}\mu_{S_{t+1}^{max}+1,t} - \varphi^{2(s_{t+1}+1)}\mu_{s_{t+1}+1,t}\right)
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\mathbb{V}ar_t[Y_{t+1}] = &\left(c(1-c)\varphi^2 + b(1-b) + a(1-a)\varphi^{-2} - 2\left(bc\varphi + ac + ab\varphi^{-1}\right)\right)\sum_{s=s_{t+1}}^{S_{t+1}^{max}} \varphi^{2s}\mu_{s,t} + \dots \\
&\dots + \sum_{s=s_{t+1}}^{\infty} \varphi^{2s} S_t \frac{\mu_{s,t}}{N-S_t}\left(1 - \frac{\mu_{s,t}}{N-S_t}\right) - \frac{2S_t}{(N-S_t)^2}\sum\sum_{s_{t+1}\le s<k\le\infty} \varphi^s\varphi^k\mu_{s,t}\mu_{k,t} + \dots \\
&\dots + c(1-c)\varphi^2\left(\varphi^{2(s_{t+1}-1)}\mu_{s_{t+1}-1,t} - \varphi^{2(S_{t+1}^{max})}\mu_{S_{t+1}^{max},t}\right) + a(1-a)\varphi^{-2}\left(\varphi^{2(S_{t+1}^{max}+1)}\mu_{S_{t+1}^{max}+1,t} - \varphi^{2s_{t+1}}\mu_{s_{t+1},t}\right). \\
&\dots - 2ab\varphi^{-1}\left(\varphi^{2(S_{t+1}^{max}+1)}\mu_{S_{t+1}^{max}+1,t} - \varphi^{2(s_{t+1}+1)}\mu_{s_{t+1}+1,t}\right) - 2ac\left(\varphi^{2(S_{t+1}^{max}+1)}\mu_{S_{t+1}^{max}+1,t} - \varphi^{2(s_{t+1}+1)}\mu_{s_{t+1}+1,t}\right)
\end{aligned}
$$

Note that $\varrho = c(1-c)\varphi^2 + b(1-b) + a(1-a)\varphi^{-2} - 2\left(bc\varphi + ac + ab\varphi^{-1}\right) = a\varphi^{-2} + b + c\varphi^2 - \rho^2$. Define $D_t = \sum_{s=s_t}^{\infty} \varphi^{2s}\mu_{s,t}$ the second moment of the firm size distribution, and $O_t^{\sigma} = c(1-c)\varphi^2\left(\varphi^{2(s_{t+1}-1)}\mu_{s_{t+1}-1,t} - \varphi^{2(S_{t+1}^{max})}\mu_{S_{t+1}^{max},t}\right) + a(1-a)\varphi^{-2}\left(\varphi^{2(S_{t+1}^{max}+1)}\mu_{S_{t+1}^{max}+1,t} - \varphi^{2s_{t+1}}\mu_{s_{t+1},t}\right) - 2ab\varphi^{-1}\left(\varphi^{2(S_{t+1}^{max}+1)}\mu_{S_{t+1}^{max}+1,t} - \varphi^{2(s_{t+1}+1)}\mu_{s_{t+1}+1,t}\right) - 2ac\left(\varphi^{2(S_{t+1}^{max}+1)}\mu_{S_{t+1}^{max}+1,t} - \varphi^{2(s_{t+1}+1)}\mu_{s_{t+1}+1,t}\right)$ a term due to the discrete grid. The volatility of the growth rate of output becomes:

$$\mathbb{V}ar_t\left[\frac{Y_{t+1}}{Y_t}\right] = \varrho\left(\frac{\sum_{s=s_{t+1}}^{\infty}\varphi^{2s}\mu_{s,t}}{Y_t^2}\right) + \frac{O_t^{\sigma}}{Y_t^2} + \frac{S_t}{(N-S_t)^2}\left(\sum_{s=s_{t+1}}^{\infty}\frac{\varphi^{2s}\mu_{s,t}}{Y_t^2}\left(N - S_t - \mu_{s,t}\right) - 2\sum\sum_{s_{t+1}\leq s<k\leq\infty}\frac{\varphi^s\varphi^k\mu_{s,t}\mu_{k,t}}{Y_t^2}\right) \tag{19}$$

□

B Robustness Checks

	(1) Var(gr.y)	(2) Var(gr.y)	(3) Var(gr.y)	(4) Var(gr.y)	(5) Var(gr.tfp)	(6) Var(gr.tfp)	(7) Var(gr.tfp)	(8) Var(gr.tfp)	(9) Var(HPgr.y)	(10) Var(HPgr.y)	(11) Var(HPgr.y)	(12) Var(HPgr.y)
$1+\delta\_emp$	-0.00316***	-0.00208*	-0.00335***	-0.00225**					-0.00535***	-0.00106	-0.00553***	-0.00142*
	(0.000741)	(0.00106)	(0.000729)	(0.000987)					(0.00152)	(0.000863)	(0.00157)	(0.000847)
$1+\delta\_asset$					-0.000759***	-0.000433***	-0.000682***	-0.000358**				
					(0.000145)	(0.000141)	(0.000149)	(0.000159)				
Employment	-0.0000325	-0.0000153	-0.0000347*	-0.0000126	-0.00000745	-0.00000413	-0.00000791*	-0.00000613	-0.0000529	-0.000234**	-0.0000513	-0.000207***
	(0.0000204)	(0.0000234)	(0.0000209)	(0.0000276)	(0.00000469)	(0.00000815)	(0.00000472)	(0.00000802)	(0.0000346)	(0.0000950)	(0.0000360)	(0.0000767)
Observations	1491	1491	1491	1491	1607	1607	1607	1607	1784	1784	1784	1784
Adjusted R^2	0.046	0.018	0.053	0.069	0.050	0.014	0.082	0.074	0.034	0.051	0.045	0.063
FE_year	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
FE_sector	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Robust standard errors clustered at the sector level in parenthesis
***p<0.01, **p<0.05, *p<0.1

Table 5: Other specification for sectoral volatility and tail relationship

	(1) Var(gr.y)	(2) Var(gr.y)	(3) Var(gr.y)	(4) Var(gr.y)	(5) Var(gr.y)	(6) Var(gr.y)	(7) Var(gr.y)	(8) Var(gr.y)	(9) Var(gr.y)	(10) Var(gr.y)	(11) Var(gr.y)	(12) Var(gr.y)
$1+\delta\_sale$	-0.00386***	-0.00306***	-0.00413***	-0.00368***	-0.00386***	-0.00306***	-0.00413***	-0.00368***	-0.00386***	-0.00306***	-0.00413***	-0.00368***
	(0.000687)	(0.00116)	(0.000700)	(0.00106)	(0.000327)	(0.000311)	(0.000358)	(0.000322)	(0.000687)	(0.00116)	(0.000777)	(0.00108)
Employment	-0.0000350*	-0.00000848	-0.0000386*	-0.0000106	-0.0000350***	-0.00000848	-0.0000386***	-0.0000106	-0.0000350*	-0.00000848	-0.0000386*	-0.0000106
	(0.0000197)	(0.0000218)	(0.0000199)	(0.0000251)	(0.00000666)	(0.0000121)	(0.00000713)	(0.0000128)	(0.0000195)	(0.0000238)	(0.0000231)	(0.0000275)
Observations	1649	1649	1649	1649	1649	1649	1649	1649	1649	1649	1649	1649
Adjusted R^2	0.058	0.035	0.085	0.117	0.059	0.606	0.109	0.648	0.059	0.606	0.109	0.648
FE_year	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
FE_sector	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Robust standard errors clustered at the sector level in parenthesis
***p<0.01, **p<0.05, *p<0.1

Table 6: Different clustering for sectoral volatility and tail relationship