Granular Firms and the Concentration Drag on Growth*

Juan Llavador Peralt†

IIES, Stockholm University

September 30, 2025

Job Market Paper. Click here for the latest version.

Abstract

This paper studies the dynamic effect of market concentration on productivity growth. I develop a multisector model with granular firms and idiosyncratic productivity shocks and characterize the stochastic dynamics of firms and aggregates. At the sector level, higher concentration lowers future productivity growth by reducing the reallocation gains from idiosyncratic shocks under gross substitutability. I denote this negative effect the *concentration drag on growth*, which is amplified when firms set markups strategically. At the micro level, granularity generates size-dependent dynamics: small firms are more volatile but have higher growth potential, while the opposite is true for large firms. Using firm- and industry-level data, I provide empirical evidence consistent with these predictions and estimate the model. An increase in concentration due to idiosyncratic shocks is associated with a contemporaneous productivity growth burst, followed by a long-lasting slowdown in productivity growth. Quantitatively, the model predicts that in a typical industry, a 5 percentage-point increase in the Herfindahl index reduces five-year productivity growth by 0.7 percentage points.

*I am indebted to my advisors, Per Krusell, Timo Boppart, Xavier Gabaix, and Joshua Weiss. I also thank for useful comments, Mitchell Downey, Brian Higgins, Kieran Larkin, Arash Nekoei, Alessandra Peter, participants at the IIES Macro Group, and participants at the Harvard Macroeconomics Lunch Seminar. I thank the Hedelius Foundation for financial support.

†Web: ; Email: juan.llavadorperalt@iies.su.se

1 Introduction

This paper shows that firm granularity matters for understanding productivity growth. The term granularity, introduced by Gabaix (2011), refers to situations where production is concentrated in a few large units that carry sufficient weight to affect aggregate outcomes. In the context of firms, this notion is one of the best documented empirical regularities in economics across time and space. For example, Figure 1 plots the evolution of the top-four firm sales share across Swedish 5-digit industries. The average industry has a top-four share of around 44%, with the 90th percentile above 70%. Many models of growth and firm dynamics acknowledge this fact, and imply a heavy-tailed firm size distribution. However, for tractability, they often rely on a continuum of infinitesimal firms, thus abstracting from the finite nature of granular firms.[1] What are the growth consequences of relaxing this assumption and allowing for individual firms to be granular?

I answer this question by developing a multi-sector growth model with finitely many firms per sector. The model shows that concentration hampers reallocation and, in expectation, drags down sectoral productivity growth.

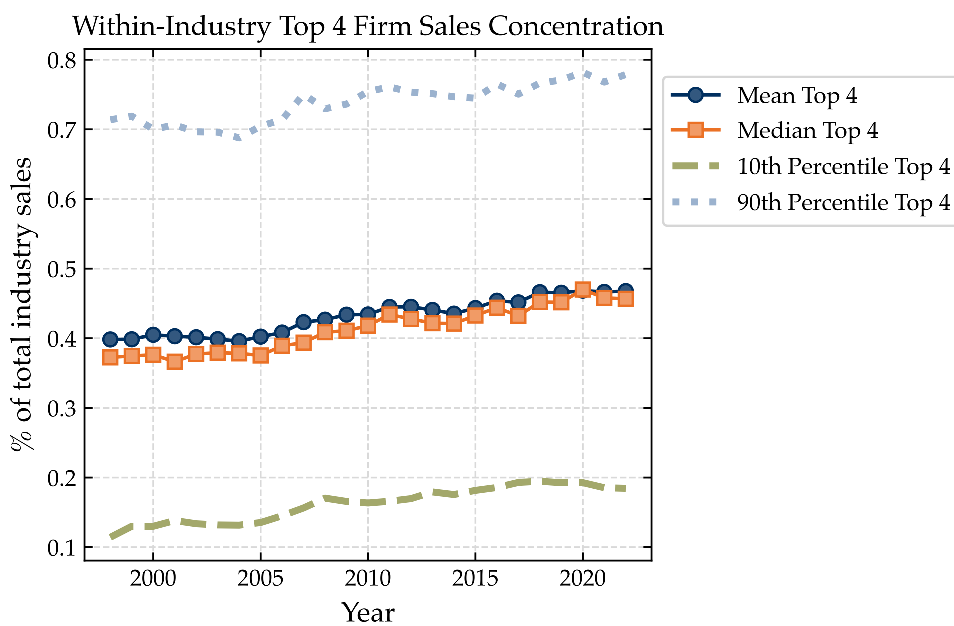

Figure 1: Mean, median, 10th percentile, and 90th percentile of the Top-4 firm sales share across Swedish 5-digit industries, by year.

[1]The term granular describes an irregular, discrete nature, in contrast to the "smoothness" of a continuum of infinitesimal agents. In the latter case, no single unit represents a sizeable share of aggregates.

To capture the key mechanisms, the model features a nested CES structure. Within sectors, good are gross substitutes, with an elasticity of substitution greater than one, while sectoral output exhibits unit elasticity, reflecting higher competition within than across sectors. Firms experience random, idiosyncratic productivity shocks, leading to a firm size distribution with a Pareto upper tail, consistent with empirical evidence (Axtell, 2001). The combination of a heavy-tailed firm size distribution and a finite number of firms generates granularity, with a few large firms accounting for a disproportionate share of sectoral production.

The main theoretical contribution is to characterize expected sectoral productivity growth as a function of the distribution of firm sales shares. Using continuous-time tools, I demonstrate that under gross substitutability, *sectoral* log growth in productivity exceeds the average log growth in *firm* productivity. The positive residual captures the gains from reallocating production toward firms with positive shocks and away from those with negative shocks. With a continuum of infinitesimal firms, this reallocation term is maximised: for every "unlucky" firm, there is a similarly sized "lucky" firm to reallocate to. With a finite number of firms, however, concentration hampers reallocation. For instance, positive shocks to small firms might not offset a negative shock to a large firm, or vice versa. In the extreme case of a monopolist, there is no reallocation at all. In expectation, concentration drags down future productivity growth.

Granularity also shapes individual firm growth. Because the elasticity of substitution across sectors is lower than the elasticity within sectors, a large firm faces diminishing business-stealing opportunities as it grows. Consequently, its growth rate distribution becomes left-skewed and less volatile with size. In contrast, its rivals face more business-stealing opportunities as the large firm grows, so their growth rate distribution becomes right-skewed, yet more volatile with the size of the large firm. Even with identical random growth shocks, granularity shapes how the firm growth distribution varies with size, generating size-dependent volatility and skewness profiles for firm growth.

The model delivers testable predictions which I test on administrative firm data from Sweden. Higher sectoral concentration predicts lower subsequent productivity growth, even when controlling for mean reversion in productivity. This finding aligns with evidence from Gaubert and Itskhoki (2021), who find that more concentrated sectors are more likely to experience mean reversion in exports.[2] At the firm level, it is well known that firm-growth volatility decreases with size (Stanley

[2] Gaubert and Itskhoki (2021) focus on export share growth rather than "domestic" productivity growth, but both are positively correlated in their model and mine.

et al., 1996). I also document that firm-growth skewness decreases with size, consistent with the model's predictions.

To quantify the model, I use the simulated method of moments (SMM). I discipline the productivity process using cross-sectional firm-growth moments that capture the volatility, skewness, and kurtosis of the unconditional firm growth distribution. The model matches the size-dependent volatility and skewness profiles well, even though these are not targeted. I further target the median level of concentration in 5-digit Swedish industries, but due to the finite number of firms per sector, the model generates a realistic range of concentration levels across sectors. A shock to concentration is associated with a contemporaneous increase in productivity, but this is followed by a prolonged slowdown in growth. The effect is quantitatively meaningful: for instance, a rise in the Herfindahl index from 0.05 to 0.10 is associated with about 0.5 to 1.2 percentage points lower productivity growth over a five-year horizon.

Since individual firms carry sufficient weight to influence aggregates, it is natural to consider that they internalize this effect when making pricing decisions. To capture such strategic interactions, I allow firms to compete in a market structure à la Atkeson and Burstein (2008), where markups are endogenous and increasing in market share. In this setting, resources are less responsive to idiosyncratic shocks, further reducing the scope for reallocation. Endogenous markups thus amplify the concentration drag on productivity growth.

Related Literature This paper relates to three main strands of the literature. First, it connects to the work on granularity and aggregate fluctuations pioneered by Gabaix (2011), who shows that idiosyncratic shocks to large firms can generate aggregate fluctuations. Subsequent work extended this idea to trade (di Giovanni et al., 2014; Gaubert and Itskhoki, 2021; di Giovanni et al., 2024), firm dynamics and aggregate fluctuations (Carvalho and Grassi, 2019), and markup fluctuations (Burstein et al., 2025). This paper differs by focusing on the growth implications of granularity, rather than short-term fluctuations.

Second, the paper contributes to the literature on market concentration and productivity growth. Several studies document an increase in market concentration in the US and other developed countries (Autor et al., 2020; Ganapati, 2021; Kwon et al., 2024; Ma et al., 2025). The endogenous growth literature has been linked this increase in concentration to the fall in productivity growth (Aghion et al., 2023). A complementary line of research emphasizes the Arrow replacement effect: in more concentrated industries, larger incumbents have weaker incentives to innovate because new innovations cannibalize their existing rents (Aghion et al., 2005; **?**; Cavenaile et al., 2025). In contrast, my mechanism shows that even holding firms' innovation incentives constant, higher concentration mechanically reduces the scope for reallocation, slowing down sectoral productivity growth. This paper provides a novel mechanism through which concentration affects growth *mechanically* by reducing the reallocation gains from idiosyncratic shocks.

Finally, the paper contributes to empirical and theoretical work on how firm growth varies with size. A natural benchmark is Gibrat's law, which states that firm growth is independent of size. This assumption has played a central role in the firm-dynamics literature because it helps explain both the stability of the firm size distribution and the emergence of a Pareto upper tail. Empirically, Gibrat's law is approximately valid for average growth rates, but it fails for higher moments. It is well documented that firm-growth volatility decreases with size (Stanley et al., 1996; Sutton, 1997; Yeh, 2025). I further document that firm-growth skewness decreases with size. On the theoretical side, Klette and Kortum (2004) emphasize that firms operate multiple products, so larger firms naturally diversify and become less volatile. Herskovic et al. (2020) are closest to my approach, showing how network linkages across firms shape the propagation of shocks and the distribution of firm-level volatility. Finally, Boehm et al. (2024) highlight how long-term contracting frictions in buyer–supplier networks can give rise to persistent deviations from Gibrat's law.

The remainder of the paper is organized as follows. Section 2 presents the static equilibrium, which holds at any point in time. Section 3 introduces the parsimonious dynamics of the model. It derives the stochastic dynamics of sectoral productivity under the efficient allocation, shows how more productive sectors exhibit higher concentration, and analyzes firm-level dynamics. Section 4 presents the data, tests the model's predictions, and estimates the model using the simulated method of moments. Section 5 presents the main results in a stationary economy. Section 6 studies the role of endogenous markups. Finally, Section 7 concludes.

2 Static Equilibrium

This section presents the static environment of the model, describing preferences, technology, and equilibrium outcomes under different market structures. The next section introduces the stochastic dynamics.

2.1 Preferences and Technology

Aggregate demand in the model is organized in two layers: an outer layer aggregates output across a continuum of infinitesimal sectors, while an inner layer aggregates a finite number of differentiated goods within each sector. The assumption of a finite number of goods per sector enables analysis of the consequences of granularity, while the continuum of sectors serves as a useful abstraction for aggregation. Preferences over consumption are formalized as follows. A representative household supplies L units of labor inelastically and consumes the final good Y, which is a Cobb-Douglas aggregator over a unit mass of sectors:

$$Y = \exp\left(\int_0^1 \beta_j \ln Y_j dj\right) \tag{1}$$

where Y_j is the output from sector $j \in [0, 1]$ and β_j represents sector-specific expenditure weights that integrate to 1. This formulation defines a sector as a market with a fixed expenditure share β_j in the aggregate consumption basket.

Within each sector, demand favors greater substitution than across sectors. Sectoral output Y_j is the result of combining $N_j \in \mathbb{N}_+$ differentiated varieties using a constant elasticity of substitution (CES) aggregator:

$$Y_j = \left(\sum_{i=1}^{N_j} Y_{ij}^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}} \tag{2}$$

where $\varepsilon > 1$ is the elasticity of substitution between goods in the same sector. Higher substitutability within sectors reflects the closer competition among firms producing similar goods. Given this preference structure, utility is maximized by choosing demand for each variety subject to the budget constraint:

$$\int_0^1 \sum_{i=1}^{N_j} P_{ij} Y_{ij} dj = WL + \Pi \tag{3}$$

where P_{ij} is the price charged by firm i in sector j, W is the wage rate, and Π represents aggregate profits, which are rebated lump-sum to households. Standard CES algebra delivers the familiar demand curve for variety i in sector j:

$$Y_{ij} = \left(\frac{P_{ij}}{P_j}\right)^{-\varepsilon} \frac{\beta_j}{P_j} PY \tag{4}$$

where $P_j \equiv \left(\sum_{i=1}^{N_j} P_{ij}^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}}$ is the price index for sector j, and $P \equiv \exp\left(\int_0^1 \beta_j \ln P_j dj\right)$ is the aggregate price index.

On the production side, each variety i in sector j is produced by a single firm using a constant returns to scale technology with labor as the only input:

$$Y_{ij} = A_{ij} L_{ij} \tag{5}$$

where A_{ij} is firm-specific productivity and L_{ij} is the labor input.

2.2 Market Structure

Firms set prices P_{ij} and quantities Y_{ij} to maximize profits $\Pi_{ij} = P_{ij}Y_{ij} - WL_{ij}$ subject to the system of demand curves (4). The optimal price is such that the markup $\mu_{ij} := \frac{P_{ij}}{W/A_{ij}}$ is given by the Lerner condition:

$$\mu_{ij} = \frac{\zeta_{ij}}{\zeta_{ij} - 1} \tag{6}$$

where ζ_{ij} is the *perceived price elasticity of demand* faced by firm i in sector j. Because the number of firms is finite, individual firms impact sector aggregates, such that their perceived price elasticities depend on sales shares within the sector. Given vectors of firm productivities A_{ij} and markups μ_{ij}, we can express the sales share of firm i in sector j as:

$$\omega_{ij} := \frac{P_{ij}Y_{ij}}{P_jY_j} = \frac{\left(A_{ij}/\mu_{ij}\right)^{\varepsilon-1}}{\sum_{k=1}^{N_j} \left(A_{kj}/\mu_{kj}\right)^{\varepsilon-1}}. \tag{7}$$

The nature of competition determines how firms internalize their impact on sector aggregates, and thus equilibrium markups and sales shares. I consider three scenarios that bracket the range of competitive forces: (i) monopolistic competition, where markups are constant; (ii) Bertrand competition, where firms strategically choose prices; and (iii) Cournot competition, where firms strategically choose quantities. For each of these market structures, the perceived price elasticity of demand ζ_{ij} takes the following form:

$$\zeta(\omega_{ij}) = \begin{cases} \varepsilon & \text{under monopolistic competition} \\ \varepsilon(1-\omega_{ij}) + \omega_{ij} & \text{under Bertrand competition} \\ \left(\frac{1}{\varepsilon}(1-\omega_{ij}) + \omega_{ij}\right)^{-1} & \text{under Cournot competition} \end{cases}$$

In Bertrand and Cournot competition, larger sales shares translate into higher markups, while under monopolistic competition markups remain constant and passthrough is complete. The dynamic analysis in the remainder of the paper focuses on two regimes that bracket the relevant forces. Monopolistic competition provides a baseline with constant markups, isolating the effects of granularity. In contrast, Cournot competition generates the greatest markup variability across firm sizes among the three market structures.

2.3 Sectoral and Aggregate Productivity

I now describe how firm level productivities and markups map to the two main objects of interest in the subsequent dynamic analysis: sectoral and aggregate productivity. Define the sectoral production function as $Y_j = A_j L_j$ where $L_j = \sum_{i=1}^{N_j} L_{ij}$ is the total labor input in sector j and A_j is the sector level productivity index. The sector level markup is then given by the sectoral price over the sectoral marginal cost: $\mu_j := \frac{P_j}{W/A_j}$. Using that μ_j is equal to the *cost-weighted markup* of the sector $\mu_j = \sum_{i=1}^{N_j} \frac{L_{ij}}{L_j}\mu_{ij}$, we can write the sector level productivity index as:

$$A_j = \left(\sum_{i=1}^{N_j} \left(\frac{\mu_{ij}}{\mu_j} \right)^{-\varepsilon} A_{ij}^{\varepsilon-1} \right)^{\frac{1}{\varepsilon-1}}. \tag{8}$$

Similarly, for the aggregate production function, define $Y = AL$, where $L = \sum_{j=1}^{N} L_j$ is the total labor input in the economy and A is the aggregate productivity index. The aggregate markup is defined as: $\mu := \frac{P}{W/A}$ and the aggregate productivity index is given by:

$$A = \exp\left(\int_0^1 \beta_j \left(\ln A_j - \ln\left(\mu_j/\mu\right)\right) \, dj \right). \tag{9}$$

It is dispersion in markups, rather than the average markup level, that lowers measured productivity by raising expenditure relative to output. When markups are constant across firms, either due to monopolistic competition or uniform productivity, the markup level is irrelevant; sector and aggregate productivity depend solely on the vectors of firm productivities. We have seen that under Bertrand and Cournot competition, firms with larger productivities (and hence larger market shares) charge higher markups, leading to markup dispersion within the sector, and lower sectoral productivity. In the next section, I will show how granularity will lead to the same situation at the aggregate level, with more productive sectors endogenously charging higher markups, a direct consequence of granularity and finite firm numbers.

2.4 Equilibrium Definition and Efficient Allocation

Given a choice of market structure and a sequence of firm productivity vectors $\{\vec{A}_j\}_{j\in[0,1]}$, a *static equilibrium* is (i) sequences of price and quantity vectors $\{\vec{P}_j, \vec{Y}_j, \vec{L}_j\}_{j\in[0,1]}$, (ii) sequences of sectoral prices and quantities $\{P_j, Y_j, L_j\}_{j\in[0,1]}$, and (iii) a sequence of aggregate prices and quantities $\{P, Y, L\}$ such that:

- Firms set prices and quantities to maximize profits given the demand curves (4) and the perceived price elasticity of demand $\zeta_{ij}(\omega_{ij})$.
- Y is the maximizer of (1) subject to (2) and the household budget constraint.
- The labor market clears: $\int_0^1 \sum_{i=1}^{N_j} L_{ij} dj = L$.

I normalize $W = 1$ and omit the goods market clearing due to Walras' law.

To finalize this section, I show that under constant markups, the allocation of resources is efficient. Consider a benevolent social planner that faces the technological constraints defined by (1), (2), and (5) and the resource constraint. The planner's objective is to maximize the aggregate output Y subject to these constraints. The maximized aggregate productivity A^* is given by:

$$A^* = \exp\left(\int_0^1 \ln A_j^* dj\right), \quad A_j^* = \left(\sum_{i=1}^{N_j} A_{ij}^{\varepsilon-1}\right)^{\frac{1}{\varepsilon-1}}$$

where A_j^* is the optimal sectoral productivity given the technological and resource constraints. The planner's allocation coincides with the decentralized equilibrium when markups are constant. This makes the monopolistic competition case a natural benchmark, isolating the role of granularity from any distortions due to misallocation.

In what follows, I first derive results under monopolistic competition, which isolates the effects of granularity. Section 6 then shows how allowing for endogenous markups further amplifies these effects.

3 Dynamics

Having characterized the static allocation of resources, I now turn to the dynamics of firm productivity and sectoral growth. The goal is to understand how idiosyncratic shocks to firms propagate to aggregate outcomes when markets are granular. I begin with the benchmark case of random growth, then analyze how concentration shapes reallocation, how large and small firms differ, and the role of stabilizing forces and mean reversion.

3.1 A Baseline Productivity Process: Random Growth

Random growth provides the canonical baseline for modeling firm dynamics. It has long been recognized, beginning with Gibrat (1931), that mean growth rates are approximately independent of firm size for medium and large firms. Throughout, "size" refers to sales (equivalently employment under CRS and a common wage), and growth is in logs. Combined with entry and exit, random growth also generates the highly skewed size distributions observed in the data, with a Pareto upper tail (Gabaix, 2009). I return to the formal mapping between random growth, entry/exit, and Pareto tails in Section 3.4. These two features make random growth the natural starting point for analyzing how idiosyncratic shocks propagate in granular economies.

Empirically, firm growth rates are not normally distributed: they exhibit fat tails and excess kurtosis (Stanley et al., 1996). To capture this, I model productivity as a continuous-time jump–diffusion process. The drift term captures systematic growth, the diffusion term captures Gaussian volatility, and the jump component generates the observed leptokurtic distributions:

$$\frac{dA_{ijt}}{A_{ijt}} = g dt + \sigma dW_{ijt} + \left(e^{J_{ijt}} - 1\right) dQ_{ijt}, \tag{10}$$

where g is the drift, σ the diffusion coefficient, W_{ijt} a Brownian motion, Q_{ijt} a Poisson process with intensity λ, and J a random variable with distribution F_J describing jump size.[3]

[3]Heuristically, over a short interval Δt, $\Delta W_{ijt} \sim \mathcal{N}(0, \Delta t)$, so that $\mathbb{E}[\Delta W_{ijt}] = 0$ and $\text{Var}(\Delta W_{ijt}) = \Delta t$; independently, the jump indicator $\Delta Q_{ijt} = 1$ with probability $\lambda \, \Delta t$ and 0 otherwise, $\Pr(\Delta Q_{ijt} = 1) = \lambda \, \Delta t + o(\Delta t)$.

Shocks are independent across firms and over time: W_{ijt} is independent of Q_{ijt} and J_{ijt}, and the jump sizes $\{J_{ijt}\}$ are i.i.d. with the finite moments used below. Diffusion governs the Gaussian core; jumps deliver excess kurtosis and, over finite horizons, skewness in sectoral growth.

This formulation nests several prominent models in the firm-dynamics literature. If $\lambda = 0$, (10) reduces to a geometric Brownian motion as in Luttmer (2007) and Arkolakis (2016). When $\lambda > 0$, the Poisson term naturally interprets as innovation in quality-ladder models such as Klette and Kortum (2004), Acemoglu and Cao (2015), and Garcia-Macia et al. (2019), with λ the innovation arrival rate and J the proportional step size.

Working in continuous time makes the analysis tractable. In discrete time, many firms can move at once, so tracking how simultaneous shocks reallocate demand across a finite set of producers becomes cumbersome. In continuous time, Brownian shocks scale smoothly with the interval and Poisson jumps arrive one at a time, so I can apply Itô's lemma to the CES aggregator and obtain closed-form drift and volatility terms for sectoral productivity. This will be essential in Section **??** when deriving the concentration–reallocation link.

In models with infinitesimal firms, random growth implies flat volatility profiles across the size distribution. In the data, however, growth volatility declines systematically with size.[4] Later in this section I show how introducing granularity reconciles these facts: even when all firms follow the same random growth process, finite sectors generate declining volatility with size. As a robustness check, I also show how modifying (10) to allow for volatility to decline with firm size remains consistent with the main results.

3.2 The Link Between Concentration and Reallocation

I now turn to the sector level and show how firm-level dynamics aggregate to sectoral productivity. To isolate the role of granularity, I begin with constant markups under monopolistic competition. Under Cobb-Douglas aggregation across sectors, the relevant object is expected sectoral log growth, $\gamma_{jt} := \mathbb{E}_t[d \ln A_{jt}/dt]$; aggregate productivity growth is the (expenditure-weighted) average of these sectoral rates.[5] I start by characterizing the stochastic dynamics of $\ln A_{jt}$ and then focus on decomposing the expected growth rate γ_{jt}.

Applying Itô's lemma to the sectoral productivity index $A_{jt} = \left(\sum_{i=1}^{N_j} A_{ijt}^{\varepsilon-1}\right)^{1/(\varepsilon-1)}$ gives the

[4]For example, Stanley et al. (1996) find that firm growth volatility declines like a power law with firm size, with an exponent around $-\frac{1}{6}$.

[5]I use $\mathbb{E}_t[d \ln A_{jt}/dt]$ as shorthand for $\lim_{\Delta t \to 0} \Delta t^{-1} \mathbb{E}_t[\ln A_{jt+\Delta t} - \ln A_{jt}]$.

stochastic differential equation (SDE) for sectoral productivity:

$$\begin{aligned} d\ln A_{jt} = &\left(g - \frac{\sigma^2}{2} + (\varepsilon-1)\frac{\sigma^2}{2}\left(1-\mathcal{H}_{jt}\right)\right)dt \\ &+ \sigma\sqrt{\mathcal{H}_{jt}}dW_{jt} \\ &+ \frac{1}{\varepsilon-1}\sum_{i=1}^{N_j}\ln\left(1+\omega_{ijt}\left(e^{(\varepsilon-1)J_{ijt}}-1\right)\right)dQ_{ijt}, \end{aligned} \tag{11}$$

and taking expectations gives the following expression for γ_{jt}:

$$\gamma_{jt} = g - \frac{\sigma^2}{2} + (\varepsilon-1)\frac{\sigma^2}{2}\left(1-\mathcal{H}_{jt}\right) + \frac{\lambda}{\varepsilon-1}\sum_{i=1}^{N_j}\mathbb{E}\left[\ln\left(1+\omega_{ijt}\left(e^{(\varepsilon-1)J_{ijt}}-1\right)\right)\right] \tag{12}$$

where $\mathcal{H}_{jt} := \sum_{i=1}^{N_j}\omega_{ijt}^2$ is the Herfindahl-Hirschman index (HHI) measure of concentration of sector j.[6] Equation (12) makes concentration explicit: the HHI enters the drift negatively, while jumps aggregate through revenue weights.

Two polar cases help build intuition in how concentration affects growth. First, consider the case of a monopolist that dominates the whole sector, such that $\omega_{1jt} = 1$ and $\mathcal{H}_{jt} = 1$. In this case, the expected growth rate reduces to the expected growth of the single firm:

$$\gamma^1 := \lim_{\omega_{1jt}\to 1}\gamma_{jt} = g - \frac{\sigma^2}{2} + \lambda\mathbb{E}[J] \tag{13}$$

I refer to this term as the *average-firm* contribution to growth, $\gamma^1 = \mathbb{E}_t[d\ln A_{ijt}/dt]$. Second, consider the polar opposite case of a sector with a continuum of infinitesimal firms. I refer to this setting as the *fully diversified* case since the growth rate is now deterministic, and can be written as the *average-firm* term plus a *reallocation* term that captures reallocation gains:[7]

$$\gamma^\infty := \lim_{N_j\to\infty}\gamma_{jt} = \underbrace{g - \frac{\sigma^2}{2} + \lambda\mathbb{E}[J]}_{\text{Average-firm}} + \underbrace{(\varepsilon-1)\frac{\sigma^2}{2} + \lambda\frac{\mathbb{E}[e^{(\varepsilon-1)J}-1] - (\varepsilon-1)\mathbb{E}[J]}{\varepsilon-1}}_{\text{Reallocation}}. \tag{14}$$

The first bracketed term in (14) captures the expected growth from average-firm dynamics, while the second bracketed term is the residual growth due to reallocation, which occurs through

[6] See Appendix A for the full derivation of the stochastic differential equation.

[7] Olley and Pakes (1996) provide a statistical decomposition of productivity growth into within-firm and between-firm components. Here, "average-firm" and "reallocation" arise endogenously from equilibrium aggregation and play an analogous role.

diffusion and jump shocks. To build intuition for the *reallocation* term in the case of infinitesimal firms, consider a simplified version of (10) without jumps (i.e., $\lambda = 0$). Heuristically, over a short interval Δt, half of the firms experience a positive productivity shock of magnitude $\sigma\sqrt{\Delta t}$, while the other half experience a negative shock of the same magnitude, $-\sigma\sqrt{\Delta t}$. Because goods are gross substitutes ($\varepsilon > 1$), consumers reallocate their expenditure $P_{ijt}Y_{ijt}$ toward the newly more productive firms and away from the less productive ones. When goods are sufficiently substitutable ($\varepsilon > 2$), more dispersion in idiosyncratic shocks (σ) leads to faster growth.[8]

Beyond the two benchmarks, a sector with finitely many firms inherits only part of the reallocation gains from the continuum case. Again, the reallocation terms consist of a diffusion and a jump component. The diffusion component falls linearly with the sector HHI:

$$\text{Diffusion Reallocation}_{jt} = (\varepsilon - 1)\frac{\sigma^2}{2}\left(1 - \mathcal{H}_{jt}\right),$$

reaching its maximum in the limit case of infinitesimal firms ($\mathcal{H}_{jt} \to 0$), and vanishing when the sector is dominated by a single firm ($\mathcal{H}_{jt} \to 1$). The jump component is more involved, but can be approximated with a Taylor expansion around $\mathbb{E}[J] = 0$:

$$\begin{aligned}
\text{Jump Reallocation}_{jt} &= \frac{\lambda}{\varepsilon - 1}\sum_{i=1}^{N_j}\mathbb{E}_t\left[\ln\left(1 + \omega_{ijt}\left(e^{(\varepsilon-1)J} - 1\right)\right)\right] - \lambda\mathbb{E}[J] \\
&\approx \lambda(\varepsilon - 1)\frac{\mathbb{E}[J^2]}{2}\left(1 - \mathcal{H}_{jt}\right) \\
&+ \lambda\left((\varepsilon - 1)^2\frac{\mathbb{E}[J^3]}{3!}\left(1 - 3\mathcal{H}_{jt} + 2\mathcal{H}_{3,jt}\right)\right) \\
&+ \lambda\left((\varepsilon - 1)^3\frac{\mathbb{E}[J^4]}{4!}\left(1 - 7\mathcal{H}_{jt} + 12\mathcal{H}_{3,jt} - 6\mathcal{H}_{4,jt}\right)\right) \\
&+ O(\mathbb{E}[J^5]),
\end{aligned}$$

where $\mathcal{H}_{k,jt} := \sum_{i=1}^{N_j}\omega_{ijt}^k$. Up to second order, the jump expansion mirrors the diffusion logic: the $\mathcal{H}_{jt}$ is a sufficient statistic for how concentration affects growth. Higher-order terms, involving $\mathbb{E}[J^3]$ and $\mathbb{E}[J^4]$, load on higher concentration moments ($\mathcal{H}_{3,jt}, \mathcal{H}_{4,jt}$) and capture skewness and kurtosis. Taken together, these contributions are non-negative and vanish at monopoly, reaching

[8] For small $a := \sigma\sqrt{\Delta t}$, the change in sectoral log productivity can be approximated as:

$$\Delta \ln A_j = \frac{1}{\varepsilon - 1}\ln\left[\frac{1}{2}(1 + a)^{\varepsilon-1} + \frac{1}{2}(1 - a)^{\varepsilon-1}\right] = \frac{\varepsilon - 2}{2}a^2 + o(a^2).$$

The average-firm log change is $-\frac{1}{2}a^2$, so the reallocation premium is $\frac{\varepsilon-1}{2}a^2$, which yields $(\varepsilon - 1)\sigma^2/2$ per unit time.

their maximum under full diversification. Proposition 1 formalizes these bounds.

Propositon 1. *$\gamma_{jt} = \mathbb{E}_t[d \ln A_{jt}/dt]$ is bounded above by the growth rate in the fully diversified case and below by that of a monopolist:*

$$\gamma^1 \leqslant \gamma_{jt} \leqslant \gamma^\infty.$$

Furthermore, the reallocation term $\gamma^\infty - \gamma^1$ is increasing in the within sector elasticity of substitution ε.

I leave the proof to the appendix and provide an intuition for the result here. We have seen that the difference between a single monopolist and a fully diversified sector is a positive reallocation term. The general case with finite firms lies between these two extremes. In the limit case of infinitesimal firms, reallocation occurs uniformly along the firm size distribution (e.g. half of the mass of firms experiences a positive shock while the other half experiences a negative shock). When goods are more substitutable, consumers reallocate expenditure more aggressively toward the most productive firms, increasing the reallocation premium due to idiosyncratic shocks. With finite firms, however, reallocation is limited: if the size distribution is very skewed, a negative shock to the largest firm may not be fully offset by positive shocks to smaller firms. In the limit case of a monopolist, reallocation vanishes.

Beyond Instantaneous Growth The results above characterize instantaneous log growth. Do these results change when considering growth over an arbitrary horizon Δt? To answer this, define

$$\Gamma_{jt}(\Delta t) := \frac{1}{\Delta t}\, \mathbb{E}_t\left[\ln A_{j,t+\Delta t} - \ln A_{jt}\right],$$

the expected sectoral *log* growth between t and $t + \Delta t$. Ranking $\Gamma_{jt}(\Delta t)$ across sectors requires a stronger notion of concentration than single-index measures such as the HHI or top-m shares. The relevant concept is *Lorenz concentration*:

Definition 1 (Lorenz Concentration)**.** *Let $\vec{\omega}_{jt}$ and $\vec{\omega}_{kt}$ be two sorted share vectors (padding with zeros if $N_j \neq N_k$). We say that $\vec{\omega}_{jt}$ is* more Lorenz-concentrated *than $\vec{\omega}_{kt}$, written $\vec{\omega}_{jt} \succ \vec{\omega}_{kt}$, if*

$$\sum_{i=1}^{m} \omega_{ijt} \geqslant \sum_{i=1}^{m} \omega_{ikt} \quad \textit{for all } m,$$

with strict inequality for some m.

Interpreting the ordered shares as an empirical distribution, Lorenz concentration is exactly first-order stochastic dominance (FOSD) of that distribution.[9] It implies higher values for all standard measures of concentration, including the HHI and top-m concentration ratios. With this notion of concentration in hand, we can establish a negative relationship between concentration and growth even over finite horizons.

Theorem 1. *Suppose $\varepsilon > 1$. For any $\Delta t > 0$, consider two sectors j and k. If sector j is more Lorenz concentrated than sector k, written $\vec{\omega}_{jt} > \vec{\omega}_{kt}$, then*

$$\Gamma_{jt}(\Delta t) < \Gamma_{kt}(\Delta t).$$

Proof sketch. Given that shocks are i.i.d. across firms, the expected growth rate $\Gamma_{jt}(\Delta t)$ is a symmetric function of the share vector $\vec{\omega}_{jt}$, which is also strictly (Schur-)concave for $\varepsilon > 1$, such that more dispersion in the share distribution leads to lower expected growth. The full proof is in Appendix A.

Percentage vs. Log Growth Because the aggregation across sectors is Cobb-Douglas, I have focused on expected sector *log* growth, for which we have seen that concentration is associated with lower growth. Do these results change when considering percentage growth $A_{jt+\Delta t}/A_{jt}$ instead? In this case a second condition is necessary: the elasticity of substitution between sectors must be greater than two ($\varepsilon > 2$). Intuitively, this is because taking the log makes the growth rate concave in the sector shares for all $\varepsilon > 1$, while percentage growth is only concave for $\varepsilon > 2$. See Appendix A for details.

3.3 Large and Small Firm Dynamics

Having established how sectoral concentration shapes aggregate growth and volatility, I now turn to the implications of granularity for the dynamics of individual firms. In particular, I show how the presence of a large firm fundamentally alters the growth volatility and skewness of both the leader and its smaller rivals, even when all firms face identical productivity shocks.

Consider a sector composed of a single "granular" leader with productivity A_{Lj} and a competitive fringe—a continuum of infinitesimal firms with productivities $\{A_{fj}\}_{f\in[0,1]}$. All firms

[9]Mathematically, Lorenz concentration is referred to as the majorization order. See Marshall et al. (2011) for a textbook treatment.

are subject to the same random growth process, and all charge the same constant CES markup $\mu = \varepsilon/(\varepsilon-1)$. The only difference is quantitative: the leader has a non-negligible sales share, while each fringe firm is infinitesimal. For clarity, I abstract from jumps in this section (the logic is unchanged if they are included):

$$\frac{dA_{Lj}}{A_{Lj}} = g\,dt + \sigma\,dW_{Lj}, \qquad \frac{dA_{fj}}{A_{fj}} = g\,dt + \sigma\,dW_{fj}.$$

Shocks W_{Lj} and $\{W_{fj}\}$ are independent across firms and over time. The CES aggregator for the fringe is $A_{Fj}^{\varepsilon-1} := \int_0^1 A_{fj}^{\varepsilon-1} df$, so the leader's sales share is

$$\omega_{Lj} = \frac{A_{Lj}^{\varepsilon-1}}{A_{Lj}^{\varepsilon-1} + A_{Fj}^{\varepsilon-1}},$$

so the leader's sales are $S_{Lj} = \omega_{Lj} P_j Y_j$, while a representative fringe unit has sales proportional to

$$S_{fj} \propto \left(\frac{A_{fj}}{A_{Fj}}\right)^{\varepsilon-1} (1-\omega_{Lj}) P_j Y_j.$$

To see how this generates size-dependent outcomes, consider a symmetric innovation to the leader's relative log-technology $x := \ln(A_{Lj}/A_{Fj})$, with $\Delta x = \pm a$, $a > 0$. Under constant markups, the induced log changes in sales are:

$$\Delta \ln S_f(\pm a) = -\ln\Big((1-\omega_{Lj}) + \omega_{Lj} e^{(\varepsilon-1)(\pm a)}\Big),$$

$$\Delta \ln S_L(\pm a) = (\varepsilon-1)(\pm a) - \ln\Big((1-\omega_{Lj}) + \omega_{Lj} e^{(\varepsilon-1)(\pm a)}\Big).$$

Let $h(u) := \ln\left((1-\omega_{Lj}) + \omega_{Lj} e^{(\varepsilon-1)u}\right)$. Since h is strictly convex for $\varepsilon > 1$, a negative leader shock reallocates demand to the fringe more than an equal positive shock reallocates it away. This convexity implies:

- $|\Delta \ln S_f(+a)| < \Delta \ln S_f(-a)$: *fringe sales are right-skewed*: they benefit more from negative leader shocks than they lose from positive ones.
- $|\Delta \ln S_L(+a)| < |\Delta \ln S_L(-a)|$: *leader sales are left-skewed*: downside shocks are amplified, while upside shocks are dampened.

Propositon 2 (Size-dependent firm dynamics under constant markups)**.** *Suppose $\varepsilon > 1$ and shocks* *are independent across firms and over time. In a sector with a granular leader and a competitive fringe:*

(i) As the leader's share ω_{Lj} increases, the leader's sales-growth distribution exhibits lower volatility and becomes more left-skewed over finite horizons.

(ii) As ω_{Lj} increases, a fringe firm's sales-growth distribution becomes more volatile and more right-skewed.

Intuition. Because the outer Cobb–Douglas aggregator implies a lower elasticity of substitution across sectors than within sectors ($1 < \varepsilon$), a large firm faces diminishing business-stealing opportunities as it grows: the slope $d\omega_{Lj}/dx = (\varepsilon - 1)\omega_{Lj}(1 - \omega_{Lj})$ shrinks with ω_{Lj}. Upside shocks therefore move its share less than equal-magnitude downside shocks reduce them, so leader growth becomes less volatile and left-skewed. When the leader takes a negative shock, spending shifts toward the fringe. Fringe firms then see bigger swings in sales with occasional large increases—right-skewed growth.

In sum, granularity endogenously ties moments of firm growth to size: large firms are less volatile and left-skewed, while small firms are more volatile and right-skewed. This mechanism is not present in models with a continuum of infinitesimal firms, and highlights how taking into account the discrete nature of firms can generate size-dependent dynamics even when all firms face identical shocks.

3.4 Stationarity and Mean Reversion

Having established the instantaneous link between concentration and growth, I now ask what the long-run cross section looks like. Random growth without frictions does not admit a stationary distribution: firm productivities fan out over time. I introduce a minimal "stabilizing force" in the spirit of Gabaix (2009) and characterize the resulting stationary distribution.

Birth–death stabilization and the forward equation. Each incumbent exits permanently at Poisson rate $\delta > 0$, and new firms enter at Poisson rate $\lambda_e > 0$ with initial productivity $A_e e^{\eta t}$ (or more generally from an entry distribution F_e). In the large–N_j limit, this birth–death mechanism pins down a stationary cross-sectional density for log productivity $x_{ijt} = \ln A_{ijt}$. Let $\phi(x)$ denote that density and write $\mu_x(\eta) := g - \frac{\sigma^2}{2} - \eta$. The stationary density solves the Kolmogorov forward equation

$$0 \;=\; -\mu_x(\eta)\,\phi'(x) \;+\; \frac{\sigma^2}{2}\,\phi''(x) \;+\; \lambda\,\mathbb{E}[\phi(x-J)-\phi(x)] \;-\; \delta\,\phi(x), \qquad x \in \mathbb{R}\backslash\{x_e\}, \tag{15}$$

with an inflow of mass at $x_e := \ln A_e$ at rate λ_e. This equation summarizes how drift, diffusion, jumps, and demographic flows jointly shape the stationary cross section.

A standard implication of (15) is that the stationary right tail is exponential in logs, or Pareto in levels. Guessing $\phi(x) \propto e^{-\alpha x}$ away from x_e and substituting into (15) yields a mapping between the traveling-wave speed η and the tail index α

$$\eta \;=\; g \;+\; \frac{\alpha-1}{2}\,\sigma^2 \;+\; \lambda\,\frac{\mathbb{E}\left[e^{\alpha J}\right]-1}{\alpha} \;-\; \frac{\delta}{\alpha}. \tag{16}$$

Here, η denotes the traveling-wave speed that sustains a growth. Intuitively, greater volatility σ^2 or more right-skewed jumps (larger $\mathbb{E}[e^{\alpha J}]$) thicken the tail (reduce α) unless offset by faster wave speed η or higher exit δ.

With finitely many firms, the empirical cross section fluctuates around the stationary density. Figure 2 illustrates two sectors: panel (a) shows a sector with a small number of firms, while panel (b) shows a sector with many firms. Both sectors have the same stationary density $\phi(x)$, but the realized cross section fluctuates more when there are fewer firms. The expected contributions of exit and entry to sectoral log productivity growth are

$$\begin{aligned}
\text{Exit Contribution}_{jt} &= \frac{\delta}{\varepsilon-1}\sum_{i=1}^{N_j}\ln\left(1-\omega_{ijt}\right) \;\approx\; -\frac{\delta}{\varepsilon-1}\left(1+\frac{1}{2}\,\mathcal{H}_{jt}\right), \\
\text{Entry Contribution}_{jt} &= \frac{\lambda_e}{\varepsilon-1}\ln\left(1+\left(\tfrac{A_e e^{\eta t}}{A_{jt}}\right)^{\varepsilon-1}\right),
\end{aligned}$$

using $\ln(1-\omega) \approx -(\omega+\frac{1}{2}\omega^2)$ with $\sum_i \omega_{ijt} = 1$ and $\sum_i \omega_{ijt}^2 = \mathcal{H}_{jt}$.

Propositon 3. *In the stationary economy, the cross-sectional distribution of firm productivities has a Pareto right tail. Within this environment, sectoral productivity A_{jt} and concentration $\mathcal{H}_{jt}$ are positively associated: sectors with higher productivity realizations tend to be more concentrated.*

Intuition. With a common Pareto right tail (same scale and tail index), the largest draw in a sector is typically far from the rest. When a sector happens to realize an unusually large top draw, two things occur mechanically: (i) the CES aggregate A_j, which is a log–sum–exp in $\{A_{ij}^{\varepsilon-1}\}$, moves up

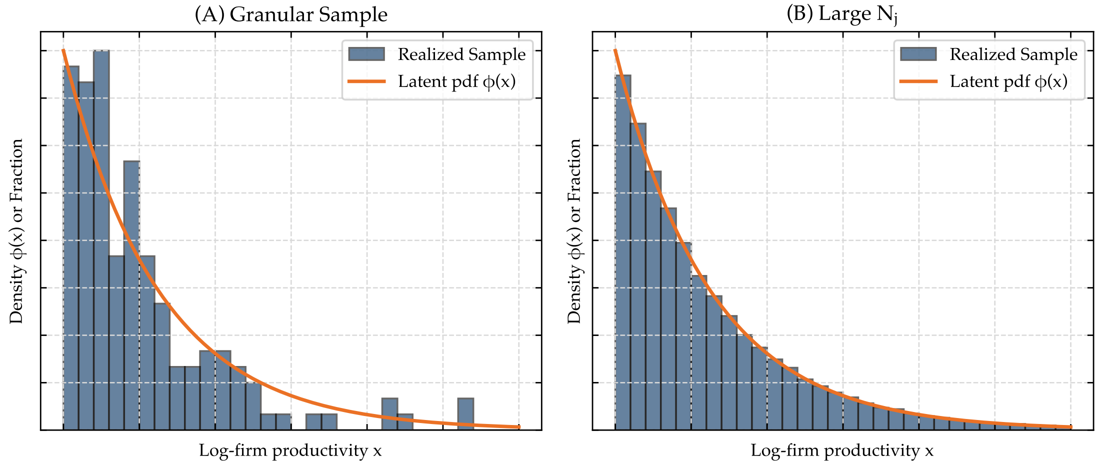

Figure 2: Stationary cross section of firm productivities has a Pareto right tail.

because the top order statistic receives disproportionate weight; and (ii) the realized HHI rises because the leader's share $\omega_{1j} \approx A_{1j}^{\varepsilon-1} / \sum_i A_{ij}^{\varepsilon-1}$ jumps. Both A_j and $\mathcal{H}_j$ are increasing in the gap between the first and the rest, so they co-move positively across sectors in the stationary cross section.

Stationarity further implies mean reversion in $\ln A_{jt}$: sectors that are currently in a high-productivity state tend to grow more slowly going forward. Together with the positive A_j–$\mathcal{H}_j$ association above, this yields a testable prediction: more concentrated sectors—those more likely to be in a temporarily high-productivity state—should, on average, exhibit lower subsequent productivity growth due to mean reversion.

4 Data and Estimation

4.1 Data

Swedish firm data I use administrative microdata on the universe of Swedish incorporated firms from the Serrano database. Compiled from the Swedish Companies Registration Office and Statistics Sweden, with group links from Dun & Bradstreet, Serrano provides firm–level financials from 1998 to 2022, covering 1,222,146 unique firms and 11,311,055 firm–years.[10] I restrict to active firms with positive sales, employment, and value added; I exclude finance, utilities, real estate, and public administration. Standard cleaning removes imputed/erroneous accounts and winsorizes

[10]See Weidenman (2016); data retrieved 15/10/2023.

extreme ratios; details and exact SNI codes are in Appendix A.

CompNet I complement the Swedish microdata with CompNet, a harmonized European dataset reporting industry–level indicators. I extract two–digit NACE measures of productivity growth and concentration to test the model's predictions outside Sweden. Results are reported in Appendix B. A fuller set of tables and robustness checks using CompNet are also provided there.

4.2 Calibration Strategy

Calibration strategy. For each industry-year pair, I compute cross-sectional moments of the distribution of one-year firm sales share growth, as well as industry-level moments of concentration, productivity growth, and related aggregates. For each moment, I compute the statistic within each industry and then take a sales-weighted average across industries (weights = total industry sales). Because all moments are based on sales shares, the calibration is not affected by common industry shocks. Results are robust to using simple medians instead of weights; the moments are quantitatively similar. I collect parameters in the vector θ, which includes all primitives governing the productivity process and the demographic block.

The productivity process (10) includes a common deterministic drift (g), a diffusion coefficient (σ) that reflects the standard deviation of thin-tailed shocks, and a jump component that captures the frequency and size of large shocks. For the jump distribution, I use an asymmetric Laplace distribution:

$$f_J(x; \mu_+, \mu_-) = \begin{cases} \frac{\mu_+\mu_-}{\mu_+ + \mu_-} e^{-\mu_-|x|}, & x < 0, \\ \frac{\mu_+\mu_-}{\mu_+ + \mu_-} e^{-\mu_+|x|}, & x \geqslant 0, \end{cases}$$

with mean $\frac{1}{\mu_+} - \frac{1}{\mu_-}$ and variance $\frac{1}{\mu_-^2} + \frac{1}{\mu_+^2}$. As we saw in Section 3, higher order moments like skewness and kurtosis might interact with granular concentration in a non-trivial way. Allowing for asymmetry in the jumps allows to match skewness, while kurtosis is controlled by the jump intensity λ.[11] Finally, the model includes an exogenous exit rate δ and a parameter η that governs the speed of the firm size distribution's travelling wave. Formally, I write $\theta \equiv (g, \sigma, \lambda, \mu_+, \mu_-, \eta, \delta; \varepsilon)$. In the baseline calibration I fix the within–sector elasticity of substitution at $\varepsilon = 5$ based on standard estimates for differentiated-product industries; the remaining parameters are chosen to match the moments below.

[11] A low rate λ makes jumps rare, leading to excess kurtosis.

While all parameters affect the distribution of sales-share growth, we can think of certain moments as being more sensitive to specific parameters, which aids in identification. Table 1 summarizes all the internally calibrated parameters. I discipline g using the median growth rate of industry labor productivity and identify σ from the difference between the 90th and 10th percentiles of sales share log changes, denoted by P90-P10. The left and right tail parameters (μ_+, μ_-) are identified from tail-sensitive moments: the upper and lower extreme spreads P99-P50 and P50-P01 respectively. The jump intensity λ is identified from the Crow-Siddiqui kurtosis measure $\frac{\text{P97.5}-\text{P2.5}}{\text{P75}-\text{P25}} - 2.91$.[12] The exit rate δ is set to match the average firm exit rate, while η is calibrated to match the median four-firm concentration ratio (CR4).

Objective function. Let $\theta_{\text{est}} = (g, \sigma, \lambda, \mu_+, \mu_-, \eta, \delta)$ denote the set of parameters estimated by simulated method of moments (SMM); ε is fixed at 5. I estimate θ_{est} by minimizing the distance between model and data moments according to the objective function:

$$L(\theta) = \sum_{m=1}^{M} \left(\frac{\text{model}_m(\theta) - \text{data}_m}{\frac{1}{2}\big(\text{model}_m(\theta) + \text{data}_m\big)} \right)^2 .$$

Table 1 summarizes the calibration targets and parameters. The calibrated model matches all targeted moments closely.

Table 1: Calibration targets and parameters

Parameter	Description	Value	Main Identifying moment	Data	Model
g	Common drift	0.015	P50 industry labor prod. growth	0.013	0.013
σ	Diffusion coeff.	0.027	P90–P10 of sales-share growth	0.45	0.46
λ	Jump rate	0.47	$\frac{\text{P97.5}-\text{P2.5}}{\text{P75}-\text{P25}} - 2.91$ of sales-share growth	3.26	3.03
μ_+	Right jump tail	18.7	P99-P50 of sales-share growth	0.020	0.021
μ_-	Left jump tail	13.9	P50-P01 of sales-share growth	0.018	0.019
α_{tail}	Tail thickness	3.6	CR4	0.46	0.46
δ	Exogenous exit	0.038	Firm exit	0.038	0.038

Note: The within–sector elasticity ε is fixed at 5 in the baseline and is therefore not listed among the estimated parameters.

[12] I use quantile based measures of the second, third, and fourth moments, rather than the standardized moments (standard deviation, skewness, and excess kurtosis coefficients) as the latter are less sensitive to outliers.

4.3 Model Fit: Size-Dependent Volatility and Skewness Profiles

I evaluate how well the calibrated model matches the growth distribution of firm size. This distribution is relevant for two reasons. First, it is well known that larger firms are less volatile. In Llavador (2025) I further show that larger firms are also more left-skewed, while smaller firms are more right-skewed. In section 3, I showed that granularity can qualitatively generate these patterns. Calibrating the model allows me to quantitatively assess how well it can match the size-dependent volatility and skewness profiles. Second, we also saw that concentration interacts with the firm productivity growth distribution to affect aggregate growth. Thus, matching the firm-level growth distribution is important to ensure that the model can reliably quantify the concentration-growth relationship.

Untargeted Moments: Volatility and Skewness Size Profiles For volatility, I consider the standard deviation of sales growth rates across firms within each size bin. This provides a measure of how much sales growth varies among firms of similar size. For an outlier robust measure I use the P90–P10 interquantile range. Figure 3 plots binned scatter plots of firm-level sales growth volatility against size (sales) in the data. The left panel shows the standard deviation, while the right panel shows the P90–P10 interquantile range, which is robust to outliers. Both measures show a clear negative relationship between size and volatility: smaller firms are more volatile, while larger firms are more stable. The model closely matches this pattern.

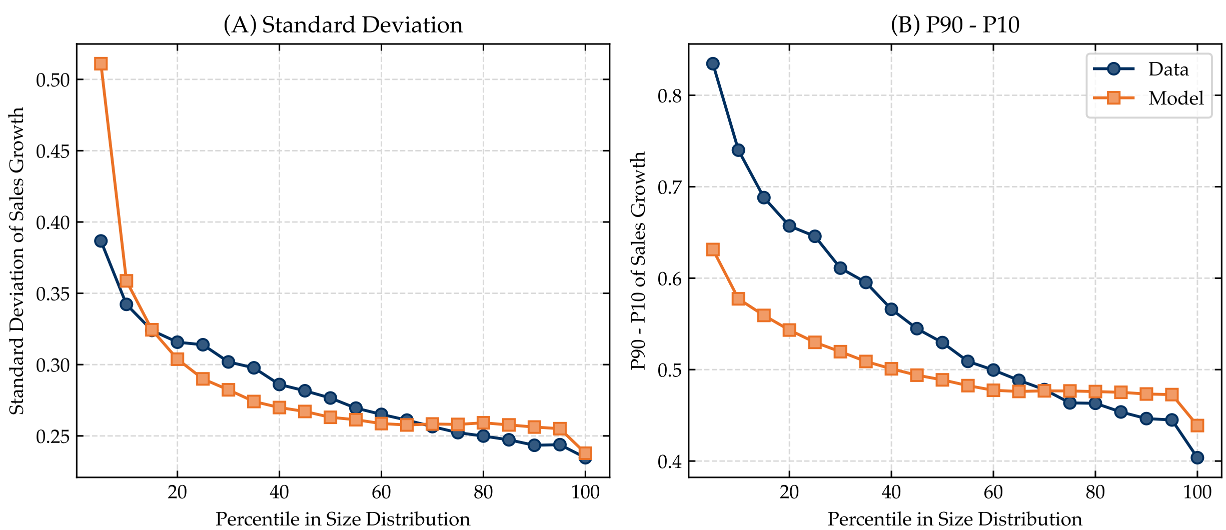

Figure 3: Volatility across size bins (data). Notes: Each line corresponds to a size bin.

For skewness, I consider the third standardized moment of sales growth rates across firms within each size bin. This captures the asymmetry of the distribution of sales growth. As an outlier robust measure, I consider the Kelly skewness, defined as $(P90 + P10 - 2P50)/(P90 - P10)$. Figure 4 plots binned scatter plots of firm-level sales growth skewness against size (sales) in the data. The left panel shows the standard skewness, while the right panel shows the Kelly skewness, which is robust to outliers. While the standard skewness measure declines fast with size, the kelly skewness measure shows a more gradual decline. This means that the change in skewness is driven by the tails of the distribution.

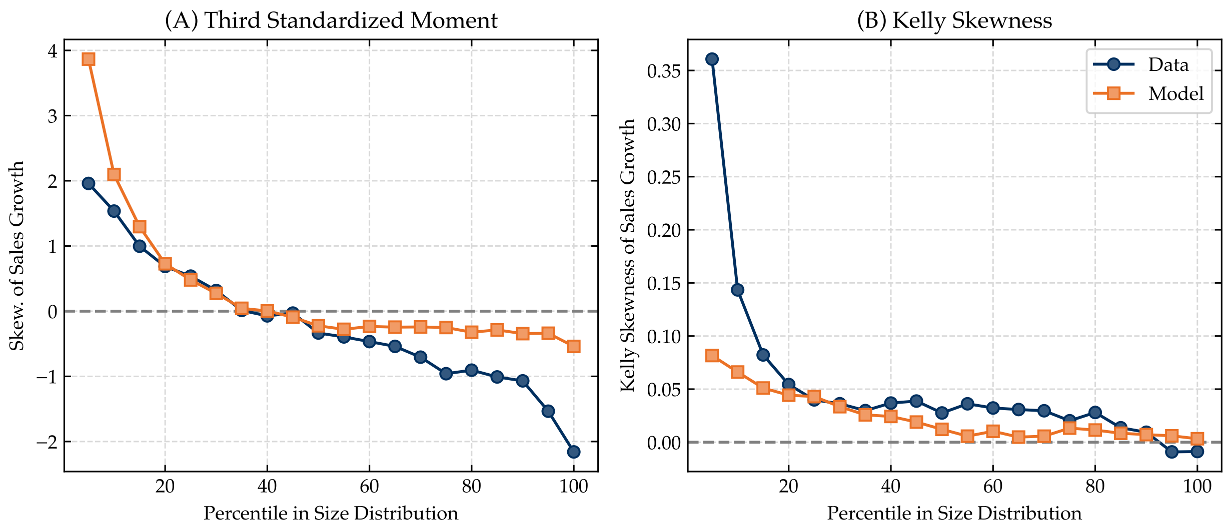

Figure 4: Kelly skewness across size bins (data). Notes: Each line corresponds to a tail definition (10%, 5%, 2.5%, 1%).

To further understand the decline in skewness, I plot different kelly skewness measures using different tail definitions (10%, 5%, 2.5%, 1%) in panel (A) of figure 5. Left or right skewness can come from the left or right tail. Panel (B) plots the left tail and panel (C) the right tail. We see that the decline in skewness in the data is mostly driven by the right tail. The model, shown in figure 6, matches this pattern well.

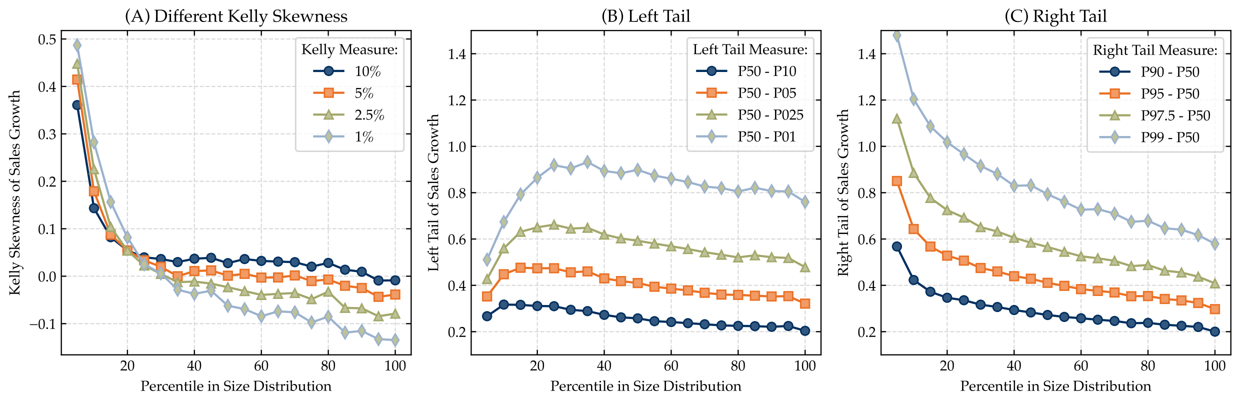

Figure 5: Kelly skewness across size bins (data). Notes: Each line corresponds to a tail definition (10%, 5%, 2.5%, 1%).

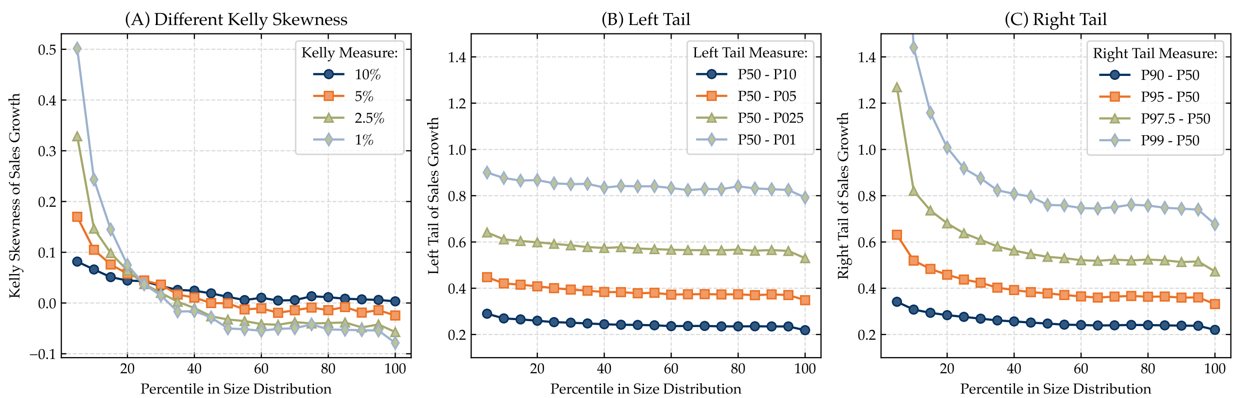

Figure 6: Kelly skewness across size bins (model). Notes: Same tail definitions as in the data figure.

5 Results

5.1 Reduced-Form Evidence and Model Validation

I begin with reduced-form evidence linking concentration to subsequent productivity growth, measured as $\ln(A_{j,t+h}) - \ln(A_{jt})$. As shown in Section 3, the Herfindahl-Hirschman Index (HHI_{jt}) is, up to a second-order approximation, a sufficient statistic for summarizing the firm size distribution when studying how concentration affects growth. To account for mean reversion in productivity, I estimate regressions of the form

$$\ln(A_{j,t+h}) - \ln(A_{jt}) = \beta_1 HHI_{jt} + \beta_2 \ln(A_{jt}) + \varepsilon_{j,t+h},$$

where $\ln(A_{j,t+h}) - \ln(A_{jt})$ is productivity growth in industry j from t to $t+h$, and $\ln(A_{jt})$ its log productivity at t. Table 2 shows a consistent negative relationship: industries with higher HHI_{jt} subsequently grow more slowly. The coefficient on concentration, β_1, is negative at both horizons. Including $\ln(A_{jt})$ via β_2 attenuates the estimates—consistent with mean reversion—but the effects remain robust. Quantitatively, a 5 percentage-point increase in HHI_{jt} predicts 0.5–0.7 percentage points lower growth over the following year, and 1–1.6 percentage points lower cumulative growth over five years. Lag-IV estimates that instrument HHI_{jt} with $HHI_{j,t-1}$ are reported in Appendix B and are similar, especially at the five-year horizon.[13] Taken together, these regressions establish a robust reduced-form link: more concentrated industries systematically grow more slowly, consistent with the mechanism of Section 3.

Table 2: Baseline regression: concentration and productivity growth

	1 Year		5 Year	
	$\ln(\text{Prod}_{t+1}) - \ln(\text{Prod}_t)$		$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$	
	(1)	(2)	(3)	(4)
HHI_t	-0.133***	-0.090**	-0.323***	-0.215**
	(0.028)	(0.025)	(0.062)	(0.062)
$\ln(\text{Prod}_t)$		-0.045***		-0.106***
		(0.006)		(0.017)
2-digit×Year FE	x	x	x	x
Observations	8964	8964	7218	7218
S.E. type	by: industry+year	by: industry+year	by: industry+year	by: industry+year
R^2	0.261	0.274	0.292	0.318
R^2 Within	0.013	0.029	0.026	0.062

SEs CRV1 clustered by 5-digit industry and year; FE: 2-digit×Year.

These reduced-form estimates provide a clear motivation for the next step: testing whether the calibrated model can reproduce the same empirical patterns.

[13] A classic "division bias" arises when the outcome and regressor share a noisy denominator. In this setting, such bias requires measurement error in sales to be serially correlated across the five-year window; thus it pertains primarily to column (4). The lag-IV specification in Appendix B mitigates this concern.

Model Regression Results Table 3 reports the corresponding regressions in simulated model panels. The model replicates both key features of the data: the negative effect of HHI_{jt} on future growth and the mean-reversion effect of initial productivity. This close alignment provides a disciplined validation of the mechanism.

Table 3: Model Regression Results

	1 Year		5 Year	
	$\ln(A_{jt+1}) - \ln(A_{jt})$		$\ln(A_{jt+5}) - \ln(A_{jt})$	
	(1)	(2)	(3)	(4)
HHI_{jt}	-0.031***	-0.020***	-0.144***	-0.086***
	(0.002)	(0.002)	(0.008)	(0.014)
$\ln(A_{jt})$		-0.011***		-0.060***
		(0.002)		(0.013)
Year FE	x	x	x	x
Observations	240000	240000	200000	200000
S.E. type	by: sector+year	by: sector+year	by: sector+year	by: sector+year
R^2	0.014	0.015	0.065	0.068
R^2 Within	0.014	0.015	0.065	0.068

Standard errors in parentheses, clustered by sector and year.

5.2 Stationary Sectoral Transition Dynamics

I next examine transitional dynamics: how concentration shocks translate into sectoral productivity paths, even in a stationary equilibrium. To do so, I estimate local projections that trace out the impulse response of productivity growth to changes in concentration.

I simulate a stationary economy with 1,100 sectors and 140 firms per sector, and estimate local projections of the form

$$\underbrace{\left(\ln A_{j,t+h} - \ln A_{j,t+h-1}\right)}_{\text{one-year growth at horizon } h} = \beta_h \Delta \ln HHI_{jt} + \gamma \ln A_{jt} + \alpha_j + \tau_t + \varepsilon_{j,t+h}, \tag{17}$$

where $\Delta \ln HHI_{jt}$ is the (optionally standardized) one-year change in concentration at time t,

$\ln A_{jt}$ controls for mean reversion, and α_j and τ_t denote sector and year fixed effects, respectively. Standard errors are clustered by sector and year. Equation (17) traces the impulse response of *one-year* productivity growth to an innovation in concentration at t.

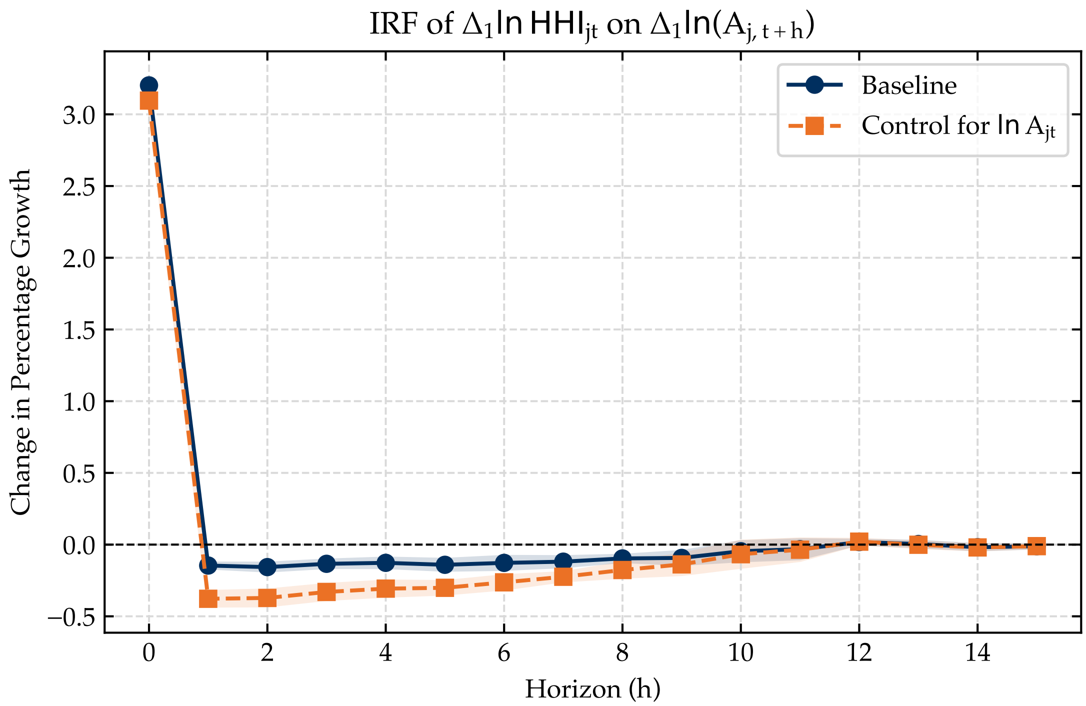

Figure 7: Local projection impulse responses to a change in concentration $\Delta \ln HHI_t$. Top panel: cumulative log productivity $\ln A_{j,t+h} - \ln A_{jt}$. Bottom panel: one-year growth $\ln A_{j,t+h} - \ln A_{j,t+h-1}$ (specification in Eq. (17)). Shaded bands are 95% confidence intervals; dashed lines add $\ln A_{jt}$ to absorb mean reversion.

The impulse responses show a clear pattern: a positive concentration shock initially boosts productivity growth, but this effect quickly reverses into a prolonged slowdown. These transitional dynamics match the model's theoretical predictions: granularity generates an immediate reallocation gain that fades as concentration drags down longer-run growth.

6 Endogenous Market Power and Markups

I now extend the constant-markup benchmark by allowing firms to set prices strategically à la Atkeson and Burstein (2008). This extension introduces an additional channel: as markups increase with market share, the firm's passthrough of productivity shocks to prices are decreasing in firm market share. This mechanism further dampens the reallocation effects from idiosyncratic shocks, thereby increasing the concentration drag on growth. I start by

In this section, I first illustrate the mechanism in the specialized model with one large firm and a competitive fringe under Cournot competition. Bertrand competition in prices can be thought of as an intermediate case between the constant-markup benchmark and Cournot competition.[14] Then, I recalibrate the full model under Cournot and Bertrand competition to compare the results with the constant-markup benchmark.

6.1 Illustration in a Simplified Cournot Model

Consider again the specialized model with a single sector j populated by one large firm and a continuum of infinitesimal fringe firms, as in Section 3.4. To introduce endogenous markups, I replace monopolistic competition with Cournot competition in quantities. Fringe firms are infinitesimal and thus charge the constant markup $\mu_F = \frac{\varepsilon}{\varepsilon-1}$. The large firm internalizes its market power; under Cournot competition, its markup μ_L depends on its sales share $\omega_L := \frac{P_{Lj}Y_{Lj}}{P_jY_j}$ through the reduced form

$$\mu_L(\omega_L) = \frac{\mu_F}{1-\omega_L}, \quad \text{where} \quad \omega_L = \frac{\left(A_{Lj}/\mu_L\right)^{\varepsilon-1}}{\left(A_{Lj}/\mu_L\right)^{\varepsilon-1} + \left(A_{Fj}/\mu_F\right)^{\varepsilon-1}}.$$

In this two-block sector, the Herfindahl–Hirschman Index (HHI) equals ω_L^2. The dynamics of sectoral productivity A_j can be derived as in Section 3.2, yielding

$$d\ln A_j^{\text{Cournot}} = \left[g - \frac{\sigma^2}{2} + \frac{\sigma^2}{2}(\varepsilon-1)(1-\omega_L^2) + \Delta^{\text{Cournot}}(\omega_L)\right]dt + \sigma\nu^{\text{Cournot}}(\omega_L)dW_{Lj}, \tag{18}$$

where $\nu^{\text{Cournot}}(\omega_L)$ is the volatility loading on the leader's shock and $\Delta^{\text{Cournot}}(\omega_L)$ is the drift adjustment induced by endogenous markups. The expected rate of productivity growth can be decomposed as

$$\gamma_{jt} = \underbrace{g - \frac{\sigma^2}{2}}_{\text{average drift}} + \underbrace{\frac{\sigma^2}{2}(\varepsilon-1)(1-\omega_L^2)}_{\text{reallocation under constant markups}} + \underbrace{\Delta^{\text{Cournot}}(\omega_L)}_{\text{markup drag}}.$$

The first term is the average drift, the second term is the reallocation gain from idiosyncratic shocks (as in the constant-markup benchmark), and the third term is the adjustment due to endogenous

[14]The price elasticity of demand faced by a firm under Cournot is more sensitive to its own market share than under Bertrand competition. Thus, Cournot competition generates stronger endogenous markup effects than Bertrand competition. Mathematically, the harmonic mean is more senstive to large values than the arithmetic mean.

markups.

Propositon 4 (Effects of Endogenous Markups)**.** *For any elasticity of substitution $\varepsilon > 1$ and any interior leader share $\omega_L \in (0, 1)$, the following inequality holds:*

$$\Delta^{Cournot}(\omega_L) < 0.$$

That is, compared to the constant-markup benchmark, endogenous Cournot markups reduce the expected productivity growth rate, thus amplifying the concentration drag on growth.

6.2 Quantitative Results with Endogenous Markups

I now reestimate the full model under Cournot and Bertrand competition to compare the results with the constant-markup benchmark. I follow the same estimation procedure as in Section 4, but now allowing for endogenous markups. Table 5 reports the regression results under the three modes of competition. The model replicates the negative effect of concentration on future productivity growth under all modes of competition. However, the magnitude of the effect varies: it is smallest under constant markups, larger under Bertrand competition, and largest under Cournot competition. This ranking is consistent with the strength of the endogenous markup channel across competition modes.

Table 4: Estimated Parameters under Different Modes of Competition

Parameter	Constant Markups	Cournot	Bertrand
α_{tail}	3.630	2.983	3.396
g	0.0146	0.0156	0.0158
σ	0.0264	0.0277	0.0273
λ	0.470	0.482	0.418
μ_+	18.700	17.167	17.800
μ_-	13.860	13.602	13.633
δ	0.0379	0.0383	0.0406

In the recalibration exercise, the parameters governing the unconditional firm growth distribution remain remarkably similar across modes of competition. Drift, diffusion, and jump intensity are largely unchanged, indicating that the basic shape of firm growth dynamics is stable. The main difference lies in the tail thickness of the firm productivity distribution. Intuitively, when large firms charge higher markups, their sales shares respond less strongly to productivity shocks. To rationalize the observed concentration levels, the productivity gap between leaders and followers must be larger. This requirement thickens the right tail of the distribution. As expected, the tail is thickest under Cournot competition, somewhat thinner under Bertrand, and thinnest under the constant-markup benchmark.

Table 5: Model Regressions under Different Modes of Competition

	Constant Markups		Cournot		Bertrand	
	$\ln(A_{jt+5}) - \ln(A_{jt})$		$\ln(A_{jt+5}) - \ln(A_{jt})$		$\ln(A_{jt+5}) - \ln(A_{jt})$	
	(1)	(2)	(3)	(4)	(5)	(6)
HHI_{jt}	-0.144***	-0.086***	-0.361***	-0.251***	-0.249***	-0.132***
	(0.008)	(0.014)	(0.017)	(0.029)	(0.012)	(0.019)
$\ln(A_{jt})$		-0.060***		-0.043***		-0.075***
		(0.013)		(0.011)		(0.014)
Year FE	x	x	x	x	x	x
Observations	200000	200000	200000	200000	200000	200000
S.E. type	by: sector+year	by: sector+year	by: sector+year	by: sector+year	by: sector+year	by: sector+year
R^2	0.065	0.068	0.068	0.069	0.075	0.078
R^2 Within	0.065	0.068	0.068	0.069	0.075	0.078

Standard errors in parentheses, clustered by sector and year.

Turning to the regression results in Table 5, all three models reproduce the negative effect of concentration on future productivity growth. The strength of the drag, however, depends on the mode of competition. With constant markups, the effect is present but moderate. Allowing for endogenous markups amplifies it: the coefficient on HHI_{jt} is larger in magnitude under Bertrand, and largest under Cournot. Quantitatively, the estimated coefficient on HHI_{jt} is approximately -0.12 under constant markups, increases to about -0.18 under Bertrand, and reaches nearly -0.25 under Cournot competition. This ordering is consistent with theory. Under Cournot, strategic pricing makes markups most sensitive to market share, which mutes reallocation gains and strengthens the concentration drag. Bertrand sits between the two extremes, closer to Cournot but with somewhat weaker effects.

7 Conclusion

This paper has developed a multisector model to analyze how market concentration shapes productivity growth. The framework combines granular firms with idiosyncratic productivity shocks, making it possible to study both sectoral and aggregate dynamics when firms hold non-negligible market shares. Calibrated to micro data, the model is paired with new empirical evidence on the relationship between concentration and growth. Together, theory and data provide a unified account of how firm granularity translates into aggregate outcomes.

The main finding is a *concentration drag*: higher concentration reduces the reallocation gains from idiosyncratic productivity shocks, leading to lower expected productivity growth. At the firm level, granularity generates systematic size-dependent patterns in the growth distribution. Large firms face diminishing business-stealing opportunities as they expand, so their growth becomes less volatile and left-skewed. Small firms, by contrast, experience higher volatility and right-skewed growth because they benefit from reallocations when dominant firms stumble. At the sector level, concentration and productivity are positively associated in the stationary cross section, which implies mean reversion in sectoral growth rates: more concentrated sectors tend to grow more slowly in the future. These transitional dynamics are a direct consequence of granularity and arise even in stationary equilibrium. Finally, when firms set prices strategically à la Atkeson–Burstein, concentration drags are amplified: endogenous markups further dampen the reallocation channel.

Beyond aggregate productivity, the framework can also be applied to other settings where a few actors dominate outcomes. For instance, it can help explain why certain cities or regions grow faster than others, or how multiproduct firms shape aggregate dynamics by internalizing reallocation across their product lines. More broadly, the results highlight that market structure and firm granularity are central to understanding growth.

References

D. Acemoglu and D. Cao. Innovation by entrants and incumbents. *Journal of Economic Theory*, 157:255–294, 2015. ISSN 0022-0531. doi: https://doi.org/10.1016/j.jet.2015.01.001. URL https://www.sciencedirect.com/science/article/pii/S0022053115000034.

P. Aghion, N. Bloom, R. Blundell, R. Griffith, and P. Howitt. Competition and innovation: An inverted-u relationship. *Quarterly Journal of Economics*, 120(2):701–728, 2005. doi: 10.1162/0033553053970214.

P. Aghion, A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li. A theory of falling growth and rising rents. *The Review of Economic Studies*, 90(6):2675–2702, 2023. doi: 10.1093/restud/rdad016. URL https://academic.oup.com/restud/article/90/6/2675/7048489.

C. Arkolakis. A unified theory of firm selection and growth *. *The Quarterly Journal of Economics*, 131(1):89–155, 02 2016. ISSN 0033-5533. doi: 10.1093/qje/qjv039. URL https://doi.org/10.1093/qje/qjv039.

A. Atkeson and A. Burstein. Pricing-to-market, trade costs, and international relative prices. *American Economic Review*, 98(5):1998–2031, December 2008. doi: 10.1257/aer.98.5.1998. URL https://www.aeaweb.org/articles?id=10.1257/aer.98.5.1998.

D. Autor, D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen. The fall of the labor share and the rise of superstar firms*. *The Quarterly Journal of Economics*, 135(2):645–709, 02 2020. ISSN 0033-5533. doi: 10.1093/qje/qjaa004. URL https://doi.org/10.1093/qje/qjaa004.

R. L. Axtell. Zipf distribution of u.s. firm sizes. *Science*, 293(5536):1818–1820, 2001. doi: 10.1126/science.1062081. URL https://www.science.org/doi/abs/10.1126/science.1062081.

J. Boehm, R. South, E. Oberfield, and M. Waseem. The network origins of firm dynamics: Contracting frictions and dynamism with long-term relationships. Technical report, 2024. URL https://ezraoberfield.github.io/NetworkDynamics.pdf. Preliminary draft, July 8 2024.

A. T. Burstein, V. M. Carvalho, and B. Grassi. Bottom-up markup fluctuations*. *The Quarterly Journal of Economics*, page qjaf029, 06 2025. ISSN 0033-5533. doi: 10.1093/qje/qjaf029. URL https://doi.org/10.1093/qje/qjaf029.

V. Carvalho and B. Grassi. Large firm dynamics and the business cycle. *American Economic Review*, 109(4):1375–1425, April 2019. doi: 10.1257/aer.20151317. URL https://www.aeaweb.org/articles?id=10.1257/aer.20151317.

L. Cavenaile, M. A. Celik, and X. Tian. Are markups too high? competition, strategic innovation, and industry dynamics. University of Toronto and University of Georgia, Working Paper, 2025.

J. di Giovanni, A. A. Levchenko, and I. Méjean. Firms, destinations, and aggregate fluctuations. *Econometrica*, 82(4):1303–1340, 2014. doi: 10.3982/ECTA11041.

J. di Giovanni, A. A. Levchenko, and I. Méjean. Foreign shocks as granular fluctuations. *Journal of Political Economy*, 132(2):463–514, 2024. doi: 10.1086/726235. URL https://www.journals.uchicago.edu/doi/10.1086/726235.

X. Gabaix. Power laws in economics and finance. *Annual Review of Economics*, 1(1):255–294, 2009. doi: 10.1146/annurev.economics.050708.142940. URL https://doi.org/10.1146/annurev.economics.050708.142940.

X. Gabaix. The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772, 2011. doi: https://doi.org/10.3982/ECTA8769. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA8769.

S. Ganapati. Growing oligopolies, prices, output, and productivity. *American Economic Journal: Microeconomics*, 13(3):309–27, August 2021. doi: 10.1257/mic.20190029. URL https://www.aeaweb.org/articles?id=10.1257/mic.20190029.

D. Garcia-Macia, C.-T. Hsieh, and P. J. Klenow. How destructive is innovation? *Econometrica*, 87 (5):1507–1541, 2019. doi: https://doi.org/10.3982/ECTA14930. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA14930.

C. Gaubert and O. Itskhoki. Granular comparative advantage. *Journal of Political Economy*, 129(3): 871–939, 2021. doi: 10.1086/712444. URL https://doi.org/10.1086/712444.

R. Gibrat. *Les Inégalités Économiques*. Librairie du Recueil Sirey, Paris, France, 1931. Original formulation of Gibrat's law.

B. Herskovic, B. Kelly, H. Lustig, and S. Van Nieuwerburgh. Firm volatility in granular networks. *Journal of Political Economy*, 128(11):4097–4162, 2020. doi: 10.1086/710345. URL https://www.journals.uchicago.edu/doi/full/10.1086/710345.

T. Klette and S. Kortum. Innovating firms and aggregate innovation. *Journal of Political Economy*, 112(5):986–1018, 2004. doi: 10.1086/422563. URL https://doi.org/10.1086/422563.

S. Y. Kwon, Y. Ma, and K. Zimmermann. 100 years of rising corporate concentration. *American Economic Review*, 114(7):2111–40, July 2024. doi: 10.1257/aer.20220621. URL https://www.aeaweb.org/articles?id=10.1257/aer.20220621.

J. Llavador. Skewed firm dynamics, 2025.

E. G. J. Luttmer. Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics*, 122(3):1103–1144, 2007. ISSN 00335533, 15314650. URL http://www.jstor.org/stable/25098869.

Y. Ma, M. Zhang, and K. Zimmermann. Business concentration around the world: 1900–2020. Draft, February 2025, 2025. URL https://www.dropbox.com/scl/fi/bzp4dhxft9vxvklhy391a/draft_0225.pdf.

A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer Series in Statistics. Springer, New York, 2 edition, 2011. doi: 10.1007/978-0-387-68276-1.

G. S. Olley and A. Pakes. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297, 1996. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/2171831.

M. H. R. Stanley, L. A. N. Amaral, S. V. Buldyrev, et al. Scaling behaviour in the growth of companies. *Nature*, 379:804–806, 1996. doi: 10.1038/379804a0.

J. Sutton. Gibrat's legacy. *Journal of Economic Literfature*, 35(1):40–59, 1997.

P. Weidenman. The serrano database for analysis and register-based statistics. Swedish House of Finance Research Data Center, 2016. URL https://www.hhs.se/en/houseoffinance/data-center/. Accessed: 2023-10-15.

C. Yeh. Revisiting the origins of business cycles with the size-variance relationship. *The Review of Economics and Statistics*, 107(3):864–871, 2025. doi: 10.1162/rest_a_01374. URL https://direct.mit.edu/rest/article/107/3/864/117914/Revisiting-the-Origins-of-Business-Cycles-With-the.