The Granular Drag on Growth*

Juan Llavador Peralt†

IIES, Stockholm University

October 26, 2025

Job Market Paper. Click here for the latest version.

Abstract

This paper develops a unified theory that accounts for cross-sectional differences in productivity growth across firms and aggregates. I develop a multi-sector model in which granular firms experience idiosyncratic productivity shocks, and I characterize the resulting stochastic dynamics of firms and aggregates. In efficient economies, higher concentration drags future sectoral productivity growth by limiting the reallocation gains from idiosyncratic shocks under gross substitutability. In inefficient economies, distortions can amplify or dampen the granular drag depending on the joint distribution of sales and cost shares. 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 structurally estimate the model. An increase in concentration due to idiosyncratic shocks generates a contemporaneous burst in productivity growth, followed by a persistent slowdown. Quantitatively, in the efficient benchmark, a 10-percentage-point increase in the Herfindahl index reduces five-year productivity growth by 0.6 percentage points. In the presence of distortions, a similar increase in the gap between sales- and cost-based Herfindahl indices reduces five-year growth by about 2.1 percentage points.

*I am indebted to my advisors, Per Krusell, Timo Boppart, Xavier Gabaix, and Joshua Weiss, for their guidance, mentorship, and support. I thank Thore Petersen and Jinglun Yao for checking the size-dependent skewness in firm growth in German and French firm data, respectively. I also thank Tessa Bold, Monir Bounadi, Konrad Buchardi, Mitchell Downey, Brian Higgins, Martin Koenen, Kieran Larkin, Kurt Mitman, Arash Nekoei, Alessandra Peter, Verónica Salazar-Restrepo, Jinglun Yao, participants at the IIES Macro Group, and participants at the Harvard Macroeconomics Lunch Seminar for useful comments. I thank the "Jan Wallanders, Tom Hedelius, and Tore Browaldh Foundation" for financial support for my visit to Harvard University.

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

1 Introduction

A central question in economics is why productivity growth differs so widely across firms, sectors, and countries. While some sustain rapid growth, others stagnate for decades. Understanding the sources of these persistent differences is crucial for identifying the mechanisms and potential policies that shape aggregate productivity growth, and thus welfare, in modern economies.

This paper develops a simple *joint* theory of cross-sectional growth differences across firms and aggregates. It builds on two premises: first, that some firms are *granular*–large enough to influence aggregate outcomes. Second, that firms' productivity evolves stochastically due to idiosyncratic shocks. The interaction between these shocks and market structure determines how factors of production are reallocated across firms, shaping productivity growth at different levels of aggregation. This interaction gives rise to what I term the *granular drag*: a slowdown in expected productivity growth as firms become more dominant.

The term granularity, introduced by Gabaix (2011), refers to situations where production is concentrated in a few large firms that carry sufficient weight to affect aggregate outcomes. Such concentration is not an abstract possibility but a pervasive feature of modern economies. For instance, in 2024, Nvidia roughly accounted for 90% of global revenues in discrete GPUs, Amazon Web Services controlled about one-third of the global cloud infrastructure market, and Tesco captured roughly 27% of UK grocery sales.[1]

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.[2] 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.

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

[1] Sources: YahooFinance (2024), Statista (2025b), Statista (2025a)

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

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 and cost 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. In the extreme case of a monopolist, there is no reallocation at all.

With a finite number of firms, however, granularity shapes how effectively resources can be reallocated in response to idiosyncratic shocks. In an efficient allocation, concentration hampers reallocation. For instance, positive shocks to small firms might not offset a negative shock to a large firm, or vice versa. In expectation, concentration drags down future productivity growth. With distortions, the effect depends on the joint distribution of sales and cost shares. If more productive firms also have lower cost shares, then distortions amplify the concentration drag because resources are misallocated away from the most productive firms. Conversely, if more productive firms have higher cost shares, distortions can mitigate the concentration drag by reallocating resources toward more productive firms.

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.[3] At the firm level, it is well known that firm-growth volatility decreases with size (Stanley

[3] Gaubert and Itskhoki (2021) focus on export share growth rather than "domestic" productivity growth, but both are

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. Quantitatively, the model predicts that in the efficient benchmark, a 10-percentage-point increase in the Herfindahl index reduces five-year productivity growth by about 0.6 percentage points. In the presence of distortions, an equivalent increase in the gap between sales- and cost-based concentration measures reduces five-year growth by roughly 2 percentage points.

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; Olmstead-Rumsey, 2019; Cavenaile et al., 2025). In contrast, my mechanism shows that even holding firms' innovation incentives constant, higher concentration reduces the scope for reallocation, slowing down sectoral

positively correlated in their model and mine.

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. Finally, Section 6 concludes.

2 Static Equilibrium

This section presents how production in the economy is organized at any point in time. A representative household derives utility from consuming a discrete set of differentiated goods. A discrete number of firms produce these goods with a good-specific productivity. Since the equilibrium holds at a point in time, I refer to this setting as the *static* equilibrium. In the next section, I introduce dynamics by allowing firm productivities to evolve stochastically over time.

2.1 Preferences and Technology

There are a *finite* number of sectors $N \in \mathbb{N}_+$, each populated by a *finite* number of differentiated goods $N_j \in \mathbb{N}_+$. A representative household supplies L units of labor inelastically and derives utility from consumption over the discrete set of goods $\{\{Y_{ij}\}_{i=1}^{N_j}\}_{j=1}^{N}$, where Y_{ij} is the consumption of variety i in sector j. In particular, the representative household has Cobb-Douglas preferences over sectoral output consumption Y_j:

$$Y = \prod_{j=1}^{N} Y_j^{\beta_j} \tag{1}$$

where β_j for $j = 1, \ldots, N$ are non-negative preference weights satisfying $\sum_{j=1}^{N} \beta_j = 1$. This formulation defines a sector as a market with a fixed expenditure share β_j in the aggregate consumption basket.

Within each sector, preferences favor greater substitution than across sectors. Sectoral output Y_j is the result of combining the N_j differentiated goods in sector j with 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 than across sectors reflects a greater degree of similarity, and thus competition, among goods within a sector.

A single firm produces variety i in sector j with a constant-returns-to-scale technology specific to that good:

$$Y_{ij} = A_{ij} L_{ij}. \tag{3}$$

Here A_{ij} is firm-specific productivity and L_{ij} is the single input in labor. In reality, firms may operate in several sectors or produce multiple varieties within a sector. My setting is analogous to assuming that multi-product firms within the same sector have identical productivities for their products. Multi-sector firms can be seen as a sum of independent single-sector subsidiaries.

The preference formulation over a discrete set $\{\{Y_{ij}\}_{i=1}^{N_j}\}_{j=1}^{N}$ contrasts with the common assumption of a continuum of infinitesimal sectors, each populated by a continuum of infinitesimal firms.[4] With finitely many sectors and firms, shocks to individual firms generate sectoral and aggregate fluctuations. The quantitative relevance of these fluctuations depends on the joint size distribution of firms and sectors, the number of firms and sectors, the elasticity of substitution, and the probability distribution of firm shocks. For empirically plausible distributions, these fluctuations can be quantitatively relevant (Gabaix, 2011).

Given this preference structure, the representative household maximizes utility by choosing demand for each variety subject to the sum of expenditure in each variety ($P_{ij}Y_{ij}$) not exceeding the total sum of labor income (WL), profits (Π), and government transfers (T). Solving the household's problem gives the demand system for each variety i in sector j:

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

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 $P \equiv \prod_{j=1}^{N} P_j^{\beta_j}$ is the aggregate price index.

2.2 Sector Market Structure

Given the demand curves (4) and the production technology (3), firms choose prices P_{ij} and quantities Y_{ij} to maximize profits:

$$\max_{P_{ij}Y_{ij}} \left\{(1-\tau_{ij})P_{ij}Y_{ij} - WL_{ij}\right\}.$$

Here, $\tau_{ij} \in (-1,1)$ is a firm-specific tax/subsidy rate on sales that distorts firm incentives. The government runs a balanced budget and rebates total tax revenue from firms lump-sum to households: $T = \sum_{j=1}^{N}\sum_{i=1}^{N_j} \tau_{ij}P_{ij}Y_{ij}$. The optimal firm price is such that the markup $\mu_{ij} := \frac{P_{ij}}{W/A_{ij}}$ is given by the Lerner condition:

$$\mu_{ij} = \frac{\zeta_{ij}}{\zeta_{ij}-1}\frac{1}{1-\tau_{ij}} \tag{5}$$

[4] In which case we could write $Y = \exp\left(\int_0^1 \ln Y_j dj\right)$ and $Y_j = \left(\int_0^1 Y_{ij}^{\frac{\varepsilon-1}{\varepsilon}} di\right)^{\frac{\varepsilon}{\varepsilon-1}}$, for some probability measures over i and j with $\int_0^1 \beta_j dj = 1$.

where $\zeta_{ij} := -d \ln Y_{ij} / d \ln P_{ij}$ is the *perceived price elasticity of demand* faced by firm i in sector j. In the main body of the paper, I focus on monopolistic competition, where each firm takes the sectoral price index P_j as given. In this case, the perceived price elasticity of demand is constant and equal to the elasticity of substitution: $\zeta_{ij} = \varepsilon$. All heterogeneity in markups is driven by government distortions τ_{ij}.

Given that firms are granular, it is natural to consider that firms internalize their impact on sector aggregates.[5] I consider oligopolistic market structures with endogenous markups à la Atkeson and Burstein (2008) both under Bertrand and Cournot competition. Under oligopolistic competition, the perceived price elasticity of demand depends on the market share of the firm. See Appendix A.1 for details.

2.3 Equilibrium Definition and Efficient Allocation

I normalize the labor wage to $W = 1$ and define a static equilibrium as follows. Given a choice of market structure for the perceived elasticity of demand ζ_{ij}, and a sequence of firm productivity vectors $\{\{A_{ij}\}_{i=1}^{N_j}\}_{j=1}^{N}$, a *static equilibrium* is (i) vectors of prices and quantities $\{\{P_{ij}, Y_{ij}, L_{ij}\}_{i=1}^{N_j}\}_{j=1}^{N}$, (ii) vectors of sectoral prices and quantities $\{P_j, Y_j, L_j\}_{j=1}^{N}$, and (iii) 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 ζ_{ij}.
- Y is the maximizer of (1) subject to (2) and the household budget constraint.
- The labor market clears: $\sum_{j=1}^{N} \sum_{i=1}^{N_j} L_{ij} = L$.

When markups are constant across firms and sectors ($\mu_{ij} = \mu$), the decentralized equilibrium allocation is *efficient*; it coincides with the choice of a benevolent social planner who maximizes aggregate output subject to the technological and resource constraints. This makes the monopolistic competition case a natural benchmark, as in the absence of government distortions, the allocation of labor across firms and sectors is efficient.

[5] I abstract from the possibility of firms internalizing their impact on the aggregate price index P. See Appendix A.8 in Burstein et al. (2025) for a case in which this assumption is relaxed.

2.4 Firm Level Outcomes

Two main firm-level outcomes are of interest for the upcoming analysis: sectoral sales- and cost-shares, denoted as ω_{ij} and κ_{ij}, respectively. Given vectors of firm productivities A_{ij} and markups μ_{ij}, we can express the sales share of firm i in sector j as a composite of markup adjusted productivities in the sector:

$$\omega_{ij} := \frac{P_{ij}Y_{ij}}{P_jY_j} = \frac{\left(A_{ij}/\mu_{ij}\right)^{\varepsilon-1}}{\sum_{k=1}^{N_j}\left(A_{kj}/\mu_{kj}\right)^{\varepsilon-1}}. \tag{6}$$

Size of a firm is often measured by its total sales S_{ij}. Using (6), we can write firm as the product of its sales share, sector share expenditure, and aggregate expenditure:

$$S_{ij} := P_{ij}Y_{ij} = \omega_{ij}\,\beta_j PY.$$

The cost-share of firm i in sector j is defined as the share of total labor costs in sector j incurred by firm i:

$$\kappa_{ij} := \frac{WL_{ij}}{WL_j} = \frac{\omega_{ij}/\mu_{ij}}{\sum_{k=1}^{N_j}\omega_{kj}/\mu_{kj}}. \tag{7}$$

When markups are constant across firms, sales shares and cost shares coincide. A gap between sales and cost shares reflects markup dispersion within the sector, and thus an inefficient allocation of resources.

2.5 Sector Level Outcomes

The objects of interest at the sector level are productivity, and concentration in sales and costs. Sector level productivity is defined as labor productivity at the sector level: $A_j := Y_j/L_j$, where $L_j = \sum_{i=1}^{N_j} L_{ij}$ is the total labor input in sector j. The sector level markups is defined as the sectoral price over the sectoral marginal cost: $\mu_j := \frac{P_j}{W/A_j}$, and can be written as the cost-weighted arithmetic average:

$$\mu_j = \sum_{i=1}^{N_j}\kappa_{ij}\mu_{ij} \tag{8}$$

which we can use to express sector level productivity as a function of firm level productivities and markups:

$$A_j = \left(\sum_{i=1}^{N_j} \left(\frac{\mu_{ij}}{\mu_j} \right)^{-\varepsilon} A_{ij}^{\varepsilon-1} \right)^{\frac{1}{\varepsilon-1}}. \tag{9}$$

When $\mu_{ij} = \mu_j$ for all firms in sector j, sectoral productivity coincides with the efficient sectoral allocation. If firm i charges a higher markup than the sector markup, i.e., $\mu_{ij} > \mu_j$, its sales share is larger than its cost share, and thus its productivity contribution to sectoral productivity is down weighted relative to its standalone productivity A_{ij}. Firm i is thus smaller than socially optimal, it would be beneficial to reallocate labor towards it. By the same logic, if instead $\mu_{ij} < \mu_j$, firm i is larger than socially optimal, such that reallocating labor away from it would be beneficial. Thus, markup dispersion within the sector leads to an inefficient allocation of resources, and lower measured sectoral productivity.

The other main sectoral outcomes of interest are concentration in sales and costs, measured by the Herfindahl-Hirschman index (HHI). The sales and costs HHI in sector j is defined as:

$$\mathcal{H}_j^{\omega} = \sum_{i=1}^{N_j} \omega_{ij}^2, \quad \mathcal{H}_j^{\kappa} = \sum_{i=1}^{N_j} \kappa_{ij}^2.$$

If markups are constant across firms, sales and cost shares coincide, and thus $\mathcal{H}_j^{\omega} = \mathcal{H}_j^{\kappa}$. A gap between sales and cost concentration reflects markup dispersion within the sector, and thus an inefficient allocation of resources. If there is a perfect ordering of firms by productivity and markups, such that more productive firms charge higher markups, then sales concentration will be higher than cost concentration: $\mathcal{H}_j^{\omega} > \mathcal{H}_j^{\kappa}$. The opposite holds if more productive firms charge lower markups. Intuitively, when more productive firms charge higher markups, the largest firm(s) in sales will have a markup above the sector average, and thus have a sales share larger than its cost share, leading to higher sales concentration relative to cost concentration.

2.6 Aggregate Level Outcomes

Similarly, for the aggregate production function, define $A = Y/L$, 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 cost-share weighted average of sectoral markups. Using this definition, we can express aggregate productivity as a function of sectoral productivities and markups:

$$\ln A = \sum_{j=1}^{N} \beta_j \left(\ln A_j + \ln(\mu_j/\mu)\right) \tag{10}$$

A direct implication is that shocks to individual firms can affect aggregate productivity through two channels: (i) by changing sectoral productivity A_j and (ii) by changing sectoral markups μ_j relative to the aggregate markup μ.

3 Dynamics

Having characterized the static allocation, I now introduce parsimonious firm-level productivity shocks. Firms are granular both at the sector and aggregate levels, so idiosyncratic shocks have an impact on sector and economy-wide aggregates. This section illustrates how idiosyncratic shocks to firms propagate to sectoral and aggregate productivity, and how the concentration of sales and costs within sectors affects these dynamics.

3.1 A Baseline Productivity Process: Random Growth

I assume that firm-level productivity follows a proportional random growth process. Specifically, firm productivity features: (i) a trend component g which is common across all firms, (ii) an i.i.d. Brownian motion component W_{ijt} with volatility σ, capturing thin-tailed, frequent shocks, and (iii) an i.i.d. jump component driven by a Poisson process Q_{ijt} with intensity λ and i.i.d. jump size $J_{ijt} \sim F_J$, capturing rare and potentially asymmetric large shocks.[6] Formally, firm productivity evolves according to the following stochastic differential equation:

$$\frac{dA_{ijt}}{A_{ijt}} = g dt + \sigma dW_{ijt} + \left(e^{J_{ijt}} - 1\right) dQ_{ijt}. \tag{11}$$

It will later be useful to distinguish between sectoral productivity growth and average firm productivity growth. Since the shocks are i.i.d. across firms, average firm productivity growth is

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

the expected growth rate of an individual firm:

$$\mathbb{E}_t\left[\frac{1}{dt}d\ln A_{ijt}\right] = g - \frac{\sigma^2}{2} + \lambda\mathbb{E}[J], \tag{12}$$

that is, expected average firm productivity the common drift minus the volatility correction $\sigma^2/2$ due to Jensen's inequality, plus the expected jump contribution $\lambda\mathbb{E}[J]$.[7]

Proportional random growth is the canonical baseline for modeling firm dynamics for two reasons. First, it has long been recognized, beginning with Gibrat (1931), that mean growth rates are approximately independent of firm size for medium to large firms. Second, combined with a "stabilizing force" (Gabaix, 2009) such as entry and exit, random growth generates a steady-state size distribution with a Pareto upper tail, consistent with the data. I return to the formal mapping between random growth, entry/exit, and Pareto tails in subsection 3.4.

Empirically, firm growth rates deviate from normality. As first shown by Stanley et al. (1996) and confirmed by many subsequent studies, the unconditional distribution of log firm-size growth rates is well approximated by a Laplace (double-exponential) distribution, characterized by sharp peaks and heavy tails. The jump component in (11) captures this feature, generating heavy tails in the distribution of firm productivity growth rates.[8]

A critique of random growth models is that in the data, rather than being constant, growth volatility declines with size.[9] However, while firm productivity follows Gibrat's law, firms are granular, implying that firm size and productivity are not perfectly correlated. As I will show later, granularity generates declining volatility with size even with a random growth process for productivity.

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 intractable. In continuous time, Brownian motions generate continuous productivity paths, and over an infinitesimal interval dt at most one Poisson jump can occur. These properties make it possible to study how individual firm shocks propagate to sectoral productivity, which is the focus of the next subsections.

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

[8] In analogy to the Central Limit Theorem, where the sum of independent finite-variance shocks converges to a Gaussian, the sum of independent shocks arriving according to a Poisson process converges to a distribution of the Laplace family. See Kotz et al. (2001) for a comprehensive review of the Laplace distribution and its emergence.

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

3.2 The Granular Drag in Efficient Economies

I now show how firm-level productivity dynamics aggregate to sectoral productivity in efficient economies. For expositional simplicity, I begin with the case without jumps ($\lambda = 0$), such that (11) reduces to $dA_{ijt}/A_{ijt} = gdt + \sigma dW_{ijt}$. Applying Itô's lemma to the sectoral productivity index (9) gives the following stochastic differential equation (SDE) for sectoral productivity:

$$d \ln A_{jt} = \left(g - \frac{\sigma^2}{2} + (\varepsilon - 1)\frac{\sigma^2}{2}\left(1 - \mathcal{H}^{\omega}_{jt}\right)\right) dt + \sigma\sqrt{\mathcal{H}^{\omega}_{jt}} dW_{jt} \tag{13}$$

where $\mathcal{H}^{\omega}_{jt} := \sum_{i=1}^{N_j} \omega^2_{ijt}$ is the Herfindahl-Hirschman index (HHI) measure of concentration of sector j, and W_{jt} is a standard Brownian motion.[10] In an efficient economy, sector level volatility is proportional to the square-root of the sector sales-HHI, as in Gabaix (2011). The focus of this paper is on expected productivity growth $\gamma_{jt} := \mathbb{E}_t[\frac{1}{dt} d \ln A_{jt}]$, which is given by average firm growth plus a residual term that captures reallocation gains and depends on concentration:

$$\gamma_{jt} = \underbrace{g - \frac{\sigma^2}{2}}_{\text{Avg. Firm}} + \underbrace{(\varepsilon - 1)\frac{\sigma^2}{2}\left(1 - \mathcal{H}^{\omega}_{jt}\right)}_{\text{Reallocation}}. \tag{14}$$

Equation (14) shows that in an efficient economy with only diffusion shocks, the sales HHI $\mathcal{H}^{\omega}_{jt}$ is a sufficient statistic for how granularity affects expected productivity growth. To build intuition on why higher concentration reduces expected growth, consider two polar cases. First, in the case of a monopolist that dominates the whole sector, such that $\omega_{1jt} = 1$ and $\mathcal{H}^{\omega}_{jt} = 1$, the expected growth rate reduces to the expected growth of the single firm:

$$\gamma^1 := \lim_{\omega_{1jt} \to 1} \gamma_{jt} = \underbrace{g - \frac{\sigma^2}{2}}_{\text{Avg. Firm}}. \tag{15}$$

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 ($\mathcal{H}^{\omega}_{jt} = 0$), and can be

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

written as the *average-firm* term plus a positive residual that captures *reallocation* gains:

$$\gamma^{\infty} := \lim_{N_j \to \infty} \gamma_{jt} = \underbrace{g - \frac{\sigma^2}{2}}_{\text{Avg. Firm}} + \underbrace{(\varepsilon - 1)\frac{\sigma^2}{2}}_{\text{Reallocation}}. \tag{16}$$

Where does the reallocation term in (16) come from? Heuristically, with only diffusion shocks, 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 ($\varepsilon > 1$), workers are reallocated toward the newly more productive firms and away from the less productive ones:

$$\mathbb{E}_t[\Delta \ln A_{jt}] = \frac{1}{\varepsilon - 1} \ln\left[\tfrac{1}{2}(1 + \sigma\sqrt{\Delta t})^{\varepsilon - 1} + \tfrac{1}{2}(1 - \sigma\sqrt{\Delta t})^{\varepsilon - 1}\right] = \underbrace{-\tfrac{\sigma^2}{2}\Delta t}_{\text{Avg. Firm}} + \underbrace{(\varepsilon - 1)\tfrac{\sigma^2}{2}\Delta t}_{\text{Reallocation}} + o(\Delta t^2)$$

Note that the underlying productivity distribution for the continuum of firms does not play a role in the reallocation gains: for every "unlucky" firm that experiences a negative shock, there is a similarly sized "lucky" firm that experiences a positive shock to which resources are reallocated. Expected reallocation increases with the elasticity of substitution ε, as the response of labor is stronger, and the volatility of idiosyncratic shocks σ, as bigger responses will be profitable. Due to Jensen's inequality, higher dispersion in firm-level shocks also lowers average firm productivity growth by $\sigma^2/2$. However, when workers are reallocated in a more than one-to-one fashion ($\varepsilon > 2$), the reallocation gains dominate the volatility drag, leading to higher expected sectoral productivity growth. This logic extends to more general idiosyncratic shocks, as I show in the case with jumps.

Beyond the two benchmarks, a sector with finitely many firms inherits only part of the reallocation gains from the continuum case: reallocation gains scale with one minus the sales HHI:

$$\text{Reallocation}_{jt} = (\varepsilon - 1)\frac{\sigma^2}{2} \underbrace{\left(1 - \mathcal{H}_{jt}^{\omega}\right)}_{\text{Granular Drag}}.$$

I refer to the term in parentheses as the *granular drag*. Intuitively, in a sector with infinitely many firms, for every firm that experiences a negative shock, there is always a similarly sized firm that experiences a positive shock. However, with finitely many firms, a negative shock to a large firm might not be offset by positive shocks to other firms, and vice versa. Because firm output is gross substitutable, in expectation granularity reduces reallocation gains, leading to lower expected productivity growth. In the extreme case of a monopolist, there are no reallocation gains at all.

Jumps I now extend the previous analysis to include jumps ($\lambda > 0$) in firm productivity. To keep the algebra light, I assume now that there are no common trend or diffusion components ($g = 0$, $\sigma = 0$), so that firm productivity evolves purely through jumps: $dA_{ijt}/A_{ijt} = (e^{J_{ijt}} - 1)dQ_{ijt}$.[11] The expected growth rate of sectoral productivity is now:

$$\gamma_{jt} = \frac{\lambda}{\varepsilon - 1} \sum_{i=1}^{N_j} \mathbb{E}\left[\ln\left(1 + \omega_{ijt}\left(e^{(\varepsilon-1)J_{ijt}} - 1\right)\right)\right].$$

While more complex than in the diffusion case, the role of granularity is explicit: jumps aggregate through sales shares ω_{ijt}. Consider again the two polar cases. With jumps only, the monopolist case ($\omega_{1jt} = 1$) and the fully diversified case ($N_j \to \infty$) give respectively the following expected growth rates:

$$\gamma^1 = \underbrace{\lambda\mathbb{E}[J]}_{\text{Avg. Firm}},$$

$$\gamma^\infty = \underbrace{\lambda\mathbb{E}[J]}_{\text{Avg. Firm}} + \underbrace{\lambda\frac{\mathbb{E}[e^{(\varepsilon-1)J} - 1] - (\varepsilon - 1)\mathbb{E}[J]}{\varepsilon - 1}}_{\text{Reallocation}}.$$

The expected growth rate in the fully-diversified case can again be decomposed into an average-firm term plus a reallocation term. As Proposition 1 will show later, the reallocation term is always positive for any well-behaved jump distribution. The intuition for the reallocation term is similar to the diffusion case. Over a short interval Δt, a fraction $\lambda\Delta t$ of firms experience a jump. For any jump distribution, there will be winners and losers. For example, if the jump distribution is a positive constant, like in quality ladder models (Grossman and Helpman, 1991; Aghion and Howitt, 1992), winners are firms that jump, while losers are firms that do not. Because goods are gross substitutes ($\varepsilon > 1$), workers are reallocated toward the more productive firms that jumped, and away from the ones that did not. Since there are infinitely many firms, for every fraction $\lambda\Delta t$ of firms that jump, there is a fraction $1 - \lambda\Delta t$ of similarly sized firms that do not jump, so the distribution of firm productivity does not matter for reallocation gains.

With finitely many firms, granularity again reduces reallocation gains. While the expression is

[11] The general case with both diffusion and jumps is just the sum of the two components, as they are independent. See Appendix **??** for the general case.

more complex, an approximation for small jumps $J \approx 0$ makes the role of granularity explicit:

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

Up to a second order, the reallocation term mirrors the diffusion case, with reallocation gains scaling with one minus the sales HHI. Higher-order terms depend on higher-order generalized HHIs $\mathcal{H}^{\omega}_{n,jt} := \sum_{i=1}^{N_j} \omega^n_{ijt}$. For example, the third order term depends on the skewness of the jump distribution $\mathbb{E}[J^3]$ and captures asymmetries in firm productivity growth. If the jump distribution is left-skewed ($\mathbb{E}[J^3] < 0$), concentration reduces reallocation gains further; as concentration rises, the sector productivity inherits the negative skewness of the large firms, which cannot be offset by the smaller firms. Conversely, if the jump distribution is right-skewed ($\mathbb{E}[J^3] > 0$), concentration mitigates reallocation gains less; as concentration rises, the sector productivity inherits the positive skewness of the large firms, which dominate sector performance. However, the granular drag for the third order term is always bounded between 0 and 1, since $0 \leqslant 3\mathcal{H}^{\omega}_{jt} - 2\mathcal{H}^{\omega}_{3,jt} \leqslant 1$.

The previous discussion illustrates how, in efficient economies, granularity reduces reallocation gains from idiosyncratic shocks and gross substitution, dragging down expected sectoral productivity growth. This result holds more generally, as stated in the following proposition.

Propositon 1. *Consider an efficient economy where firm productivity follows the jump-diffusion process in* (11)*. Then, the expected sectoral productivity growth rate $\gamma_{jt} = \mathbb{E}_t[d \ln A_{jt}/dt]$ 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.

The proof for the case without jumps is immediate from (14). For the general case with jumps, see Appendix A. 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. 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. From Proposition 1, it follows that in efficient economies, the reallocation residual is always positive.

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. 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.[12] It implies higher values for 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*

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

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 The Granular Drag under Misallocation

The preceding analysis assumes that the economy is efficient. In reality, however, there is ample evidence of misallocation of resources across firms, e.g. Hsieh and Klenow (2009). To understand the role of misallocation for sectoral productivity growth, I first allow for firm-specific markup heterogeneity that is fixed over time. For example, such heterogeneity could arise from government distortions τ_{ij}. For expositional clarity, I focus on the case without jumps $dA_{ijt}/A_{ijt} = gdt + \sigma dW_{ijt}$, and leave the general case with jumps to Appendix A.

With markup heterogeneity, sales and cost shares differ. If a firm has a higher than average markup, it will have a higher sales share relative to its cost share. This firm is employing fewer workers than in the efficient allocation. Reallocating labor toward this firm would increase sectoral productivity. The converse is true for firms with lower than average markups. Thus, misallocation reduces the *level* of sectoral productivity relative to the efficient allocation. I show next how misallocation also affects *growth* when firms are granular. Specifically, the expected growth rate under fixed markup heterogeneity is given by:[13]

$$\gamma_{jt} \;=\; \underbrace{g - \frac{\sigma^2}{2}}_{\text{Avg. Firm}} \;+\; \underbrace{(\varepsilon-1)\frac{\sigma^2}{2}\Big[1 - \mathcal{H}^{\omega}_{jt} - (\varepsilon-1)\big(\mathcal{H}^{\omega}_{jt} - \mathcal{H}^{\kappa}_{jt}\big)\Big]}_{\text{Reallocation}}. \tag{17}$$

In the efficient allocation, sales and cost shares coincided, and the sales HHI $\mathcal{H}^{\omega}_{jt}$ was, up to second order, a sufficient statistic for how granularity affected growth. Under misallocation, however, the difference between sales and cost shares matters as well. If we compare equation (A.4) to the efficient case in equation (14), for a fixed sales concentration $\mathcal{H}^{\omega}_{jt}$, misallocation increases or decreases the expected growth rate by $(\varepsilon-1)^2\sigma^2/2 \times (\mathcal{H}^{\omega}_{jt} - \mathcal{H}^{\kappa}_{jt})$. Intuitively, if more productive firms have high markups (relative to small firms), sales concentration $\mathcal{H}^{\omega}_{jt}$ will be high relative to cost concentration $\mathcal{H}^{\kappa}_{jt}$. In this case, the granular drag on growth is amplified; as resources are misallocated away from the most productive firms, it would be beneficial to reallocate workers toward these firms. However, these firms are the largest ones, and granularity limits the potential reallocation gains, further dragging down growth. If the opposite is true, and more productive firms have low markups, sales concentration will be low relative to cost concentration, making reallocation gains easier to achieve despite granularity, mitigating the drag on growth.

Note that the impact of misallocation on growth vanishes in the two polar cases of monopoly and full diversification. In the monopolist case, there is a single firm, so there is no misallocation. In the fully diversified case ($\mathcal{H}^{\omega}_{jt}, \mathcal{H}^{\kappa}_{jt} \to 0$), granularity vanishes, and so does the impact of misallocation on growth. This result follows the same logic as in the efficient economy: with infinitely many firms, for every firm that experiences a negative shock, there is always a similarly sized- *and with a similar markup* firm that experiences a positive shock, so the distribution of firm productivity and markups does not matter for reallocation gains. Hence, misallocation matters for growth only in granular economies: when firms are discrete rather than infinitesimal, the joint distribution of productivity and markups shapes the granular drag on growth.

3.4 Stationarity and Mean Reversion

The analysis so far has focused on applying a random growth process to firm productivity and studying its implications for instantaneous sectoral productivity growth. Without a "stabilizing force" (Gabaix, 2009), random growth does not admit a stationary distribution: firm productivities

[13] See AppendixA.3 for the full derivation.

fan out over time.[14] To address this, I introduce firm entry and exit. I first characterize the stationary distribution with a continuum of infinitesimal firms, and then discuss how granularity affects the realized cross-section with finitely many firms.

Suppose we are in the large–N_j limit with a continuum of infinitesimal firms. Firm productivity follows the jump-diffusion process in (11). Each incumbent exits permanently at Poisson rate $\delta > 0$, and new firms enter at Poisson rate $\nu > 0$ with initial productivity $A_e e^{\eta t}$ (or more generally from an entry distribution F_e). Under some mild conditions on η, there exists a unique traveling-wave distribution that is shape invariant over time. Denote $x_{ijt} := \ln A_{ijt} - \eta t$ the productivity of firm i in sector j relative to the traveling wave, so that x_{ijt} is stationary over time. Let $\phi(x)$ denote the stationary density and write $\mu_x(\eta) := g - \frac{\sigma^2}{2} - \eta$. The stationary density solves the Kolmogorov forward equation (KFE):

$$0 \;=\; -\mu_x(\eta)\,\phi'(x) \;+\; \frac{\sigma^2}{2}\,\phi''(x) \;+\; \lambda\,\mathbb{E}[\phi(x-J)-\phi(x)] \;-\; \delta\,\phi(x), \qquad x \in \mathbb{R}\backslash\{x_e\}, \tag{18}$$

with an inflow of mass at $x_e := \ln A_e$ at rate ν. A standard implication of (18) is that the stationary right tail is exponential in logs, or Pareto in levels. See Appendix D in Gabaix et al. (2016) for the details for the same KFE equation (18) with jumps. Guessing $\phi(x) \propto e^{-\alpha x}$ away from x_e and substituting into (18) yields a mapping between the traveling-wave speed η and the tail index α

$$\eta \;=\; g \;+\; \frac{\alpha-1}{2}\,\sigma^2 \;+\; \lambda\,\frac{\mathbb{E}\left[e^{\alpha J}\right]-1}{\alpha} \;-\; \frac{\delta}{\alpha}. \tag{19}$$

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 sectoral distribution fluctuates around the stationary density.[15] As the number of firms N_j increases, the empirical distribution converges to the stationary density $\phi(x)$. If instead we let the number of sectors N go to infinity, the density of sectoral productivities converges to a stationary distribution as well, which depends on the stationary firm productivity distribution $\phi(x)$. In this cross-section of sectors, sectoral productivity

[14]For the stationary distribution to have a Pareto tail consistent with the data, the mean reversion induced by the stabilizing force must be "small", such that the previous analysis without mean reversion remains a good approximation for the upper tail. See Gabaix (1999, 2009) for details.

[15]One necessary modification with finitely many firms is that, to have on average $\bar{N}_j$ incumbents, the rate of entry $\nu = \delta \bar{N}_j$.

A_{jt} and concentration $\mathcal{H}^{\omega}_{jt}$ are positively associated.

Granular Entry and Exit With finitely many firms, entry and exit also interact with granularity. In particular, we can model entry and exit as a jump process, in which firms' productivity goes to zero upon exit, and new entrants arrive with initial productivity $A_e e^{\eta t}$. In an efficient economy, the expected growth contribution to sectoral productivity from entry and exit is:

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

Intuitively, exogenous exit reduces expected sectoral growth more when concentration is high, as the largest firms account for a disproportionate share of sales. Entry increases expected growth depending on the initial productivity of entrants relative to incumbents. When the economy is misallocated, the exit contribution is modified similarly to the diffusion case above. If a high-productivity-high-markup firm exits, the drag on growth is amplified by granularity, as that firm should have received more resources in an efficient allocation. Conversely, if a low-productivity-low-markup firm exits, the drag on growth is mitigated by granularity, as that firm was too large from a social perspective.

4 Data and Estimation

In this section I first describe the data sources, then I present reduced form evidence on the relationship between granularity and sectoral productivity growth, and finally I outline the calibration to estimate the model.

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.[16] I restrict to active

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

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

U.S. Census As further robustness, I use U.S. industry-level data from the replication package from Ganapati (2021), who constructs TFP and concentration measures at the 6-digit NAICS level. Results are reported in Appendix B.

4.2 Reduced Form Evidence

I define a sector as a 5-digit industry (SNI 2007, which maps to NACE Rev. 2) and compute firm market shares from nominal sales within each sector-year. Unfortunately, there are no measures of TFP at the 5-digit level. I proxy sectoral productivity using labor productivity (nominal output over labor). Labor productivity is an imperfect measure of TFP, as it confounds changes in markups with changes in efficiency. I present robustness checks, also controlling for future concentration in Appendix B. I further test the model's predictions using CompNet and U.S. data from Ganapati (2021). For the latter, industry-level TFP and concentration measures at the 6-digit NAICS level are available, and find a negative relationship between concentration and productivity growth consistent with the model; details are in Appendix B.

In practice, industries might differ in the deep parameters of the model, like the elasticity of substitution, as well as in the primitives of the productivity process. To control for such heterogeneity, ideally I would use industry fixed effects. I report such regressions in Appendix B, which show a clear negative relationship between concentration and productivity growth within industries. However, given my imperfect proxy for productivity, a high level of concentration today might be mechanically correlated with a high level of labor productivity today. Including an industry fixed effect makes that mechanical correlation also to future labor productivity, leading to a spurious negative correlation between concentration and productivity growth. To avoid this issue, I instead include current labor productivity as a control, and use 2-digit industry times year fixed effects to control for broad industry trends. The current labor productivity control captures the mechanical correlation between concentration and productivity levels, isolating the effect of concentration on future productivity growth.

I run two specifications. First, I regress one-year ahead labor productivity growth on the Herfindahl-Hirschman Index (HHI) of sales shares. This would be the relevant concentration measure if all firms charged the same markup. Second, I include both sales and cost share HHIs, which would be the relevant measures if markups differed across firms. The model predicts a negative coefficient on sales concentration, and a positive coefficient on cost concentration.

$$\ln A_{jt+\Delta t} - \ln A_{jt} = \alpha + \beta_1 \mathcal{H}^{\omega}_{jt} + \beta_2 \ln A_{jt} + \epsilon_{jt+\Delta t}$$
$$\ln A_{jt+\Delta t} - \ln A_{jt} = \alpha + \beta_1 \mathcal{H}^{\omega}_{jt} + \beta_2 \left(\mathcal{H}^{\omega}_{jt} - \mathcal{H}^{\kappa}_{jt}\right) + \beta_3 \ln A_{jt} + \epsilon_{jt+\Delta t}$$

Table 1 reports results. Columns (1) and (2) show that higher sales concentration is associated with lower 5-year ahead labor productivity growth. In particular, an increase in the HHI of 1 percentage point is associated with with a 0.2% lower five-year ahead labor productivity growth. Columns (3) and (4) include both sales and cost concentration. The coefficient on sales concentration remains negative, but shrinks substantially and is no longer statistically significant. The gap between sales and cost concentration drives the results: an increase in the difference between sales and cost concentration of 1 percentage point is associated with a 1% lower five-year ahead labor productivity growth. These results are consistent with the model's predictions.

Table 1: Reduced form evidence: concentration and productivity growth

	Efficient		With Distortions	
	$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$		$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$	
	(1)	(2)	(3)	(4)
HHI_t sales	-0.323***	-0.215**	-0.046	-0.013
	(0.068)	(0.070)	(0.075)	(0.074)
$\ln(\text{Prod}_t)$		-0.106***		-0.070***
		(0.019)		(0.018)
HHI_t sales $-$ HHI_t costs			-1.349***	-1.164***
			(0.215)	(0.217)
2-digit$\times$Year FE	x	x	x	x
Observations	7218	7218	7218	7218
S.E. type	by: industry	by: industry	by: industry	by: industry
R^2	0.292	0.318	0.342	0.352
R^2 Within	0.026	0.062	0.094	0.109

SEs clustered by 5-digit industry and year.

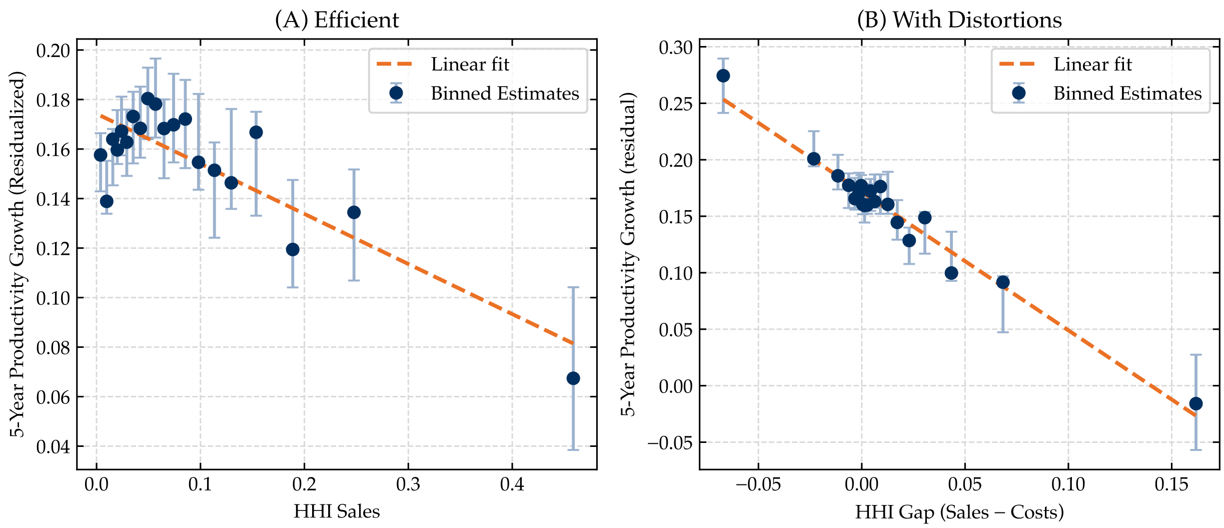
The theory predicts further that the relationship between productivity growth and the HHIs should be, up to second order, linear. Figure 1 shows binned scatter plots of the non-parametric relationship between concentration and productivity growth, with the linear fit from columns (2) and (4) of Table 1 overlaid.[17] Panel (A) shows the relationship between sales HHI and five-year ahead productivity growth, controlling for current productivity and fixed effects. While the relationship is negative, linearity is not apparent. Panel (B) shows the relationship productivity growth and the difference between sales and cost HHIs. As the theory predicts, the relationship is and apparently linear and negative.

I conclude this subsection with a note of caution on the magnitude of the reduced form estimates. For example, I use 5-digit industries to map firm-level data to sectors, but 5-digit industries may not correspond to the relevant market boundaries for competition (see Berry et al. (2019) for a comprehensive discussion). If for example, the relevant market is narrower, then the estimated coefficients will be larger in magnitude than the true effect, as I show in Appendix B. Furthermore, firm dynamics might be a function of concentration itself. If large firms

[17] I use the methodology developed by Cattaneo et al. (2024) to residualize and estimate the confidence intervals for the binned scatter plots.

in more concentrated sectors invest less in productivity-enhancing activities, there will be more mean reversion in productivity among large firms, which will mechanically generate a negative correlation between concentration and productivity growth.[18]

Figure 1: Non-parametric relationship between concentration and productivity growth



4.3 Calibration Strategy

To discipline the model, I calibrate the parameters governing the productivity process and firm demographics using simulated method of moments (SMM). I match cross-sectional moments of the distribution of firm sales share growth within industries, as well as industry-level moments of concentration and productivity growth.

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 (11) includes a common deterministic drift (g), a diffusion coefficient

[18]In practice however, this effect is unlikely to be very strong, as if it were the case, we would not observe the Pareto tail in firm size. See Gabaix (2009) for the details.

(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 matching skewness, while kurtosis is controlled by the jump intensity λ.[19] 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.

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 2 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 $\frac{\text{P97.5}-\text{P2.5}}{\text{P75}-\text{P25}} - 2.91$.[20] 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).

Let $\theta_{\text{est}} = (g, \sigma, \lambda, \mu_+, \mu_-, \alpha_{\text{tail}}, \delta)$ 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:

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

Table 2 summarizes the calibration results. The tail index and the exit rate are estimated separately, while the productivity process parameters are estimated jointly given α_{tail} and δ. The estimated productivity process features a jump roughly every three years, with left skewed jumps that are larger on average than right-skewed jumps. The diffusion component is relatively small compared

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

[20] 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.

to the jump component, indicating that large shocks play a significant role in firm productivity dynamics. The tail index α_{tail} is estimated at 3.96, implying Zipf's law for sales shares within industries.[21]

Table 2: Calibration targets and estimated parameters

Parameter	Description	Value	Main Identifying moment	Data	Model
g	Common drift	0.019	TFP growth 1998-2019	0.013	0.014
σ	Diffusion coeff.	0.025	P90–P10 of sales growth	0.45	0.46
λ	Jump rate	0.36	$\frac{\text{P97.5}-\text{P2.5}}{\text{P75}-\text{P25}} - 2.91$ of sales growth	3.26	3.18
μ_+	Right jump tail	19.6	P99-P50 of sales growth	0.73	0.75
μ_-	Left jump tail	15.0	P50-P01 of sales growth	0.84	0.83
α_{tail}	Tail thickness	3.96	CR4	0.46	0.46
δ	Exogenous exit	0.034	Firm exit	0.033	0.033

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

5 Results

5.1 Model Fit: Size-Dependent Volatility and Skewness Profiles

First, I evaluate how well the calibrated model matches how the firm growth distribution varies with firm size measured in sales. This growth distribution is relevant for two reasons. First, I have assumed that firm productivity follows a random growth process, which was critical to derive the model's theoretical predictions. A common critique of random growth models is that, in models with a continuum of firms, they fail to match key empirical features of firm growth dynamics, like a slowly declining volatility of sales growth with size. Second, matching how the firm growth distribution varies by size, while maintaining the Pareto size distribution, is an unresolved puzzle in the literature. Granularity provides an answer.

Untargeted Moments: Mean Growth and Volatility Profiles I start by plotting how the first two moments of firm-level sales growth vary with size in the data and the model. Figure 2 plots binned scatter plots by quantile of the sales distribution of mean and standard deviation of sales

[21] Since $\varepsilon = 5$, Zipf's law for sales shares implies a tail index of $\alpha_{\text{tail}} = \varepsilon - 1 = 4$.

growth rates. The left panel shows the mean growth profile, which is roughly flat for medium to large firms in both the data and the model. The right panel shows the volatility profile, which declines with size in both the data and the model. The model matches the level and slope of the volatility profile well, showing that granularity can account for this important empirical regularity.

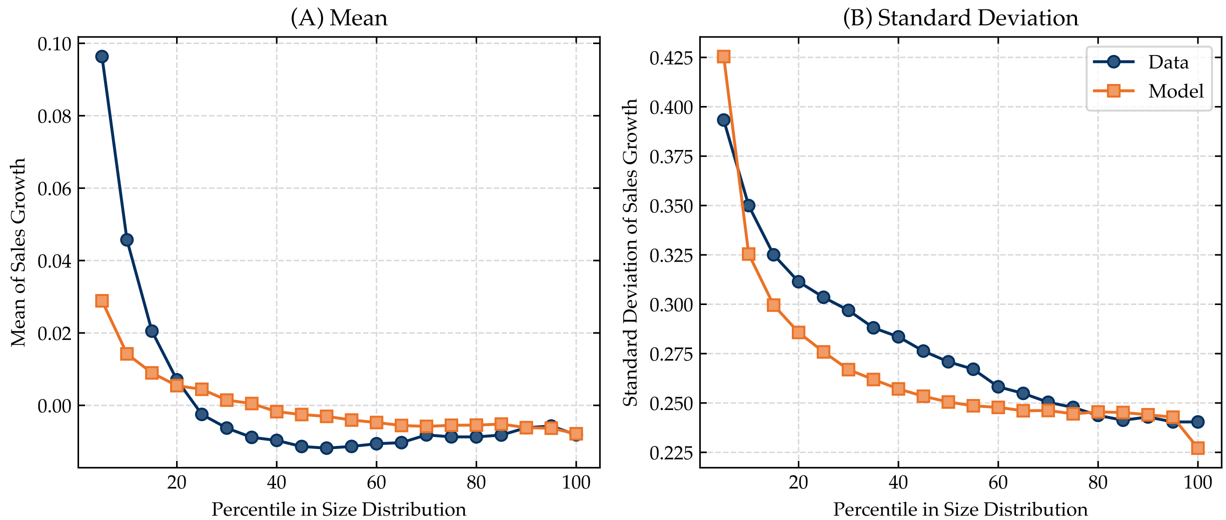


Figure 2: Volatility across size bins (data). Notes: Each line corresponds to a size bin.

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 3 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.

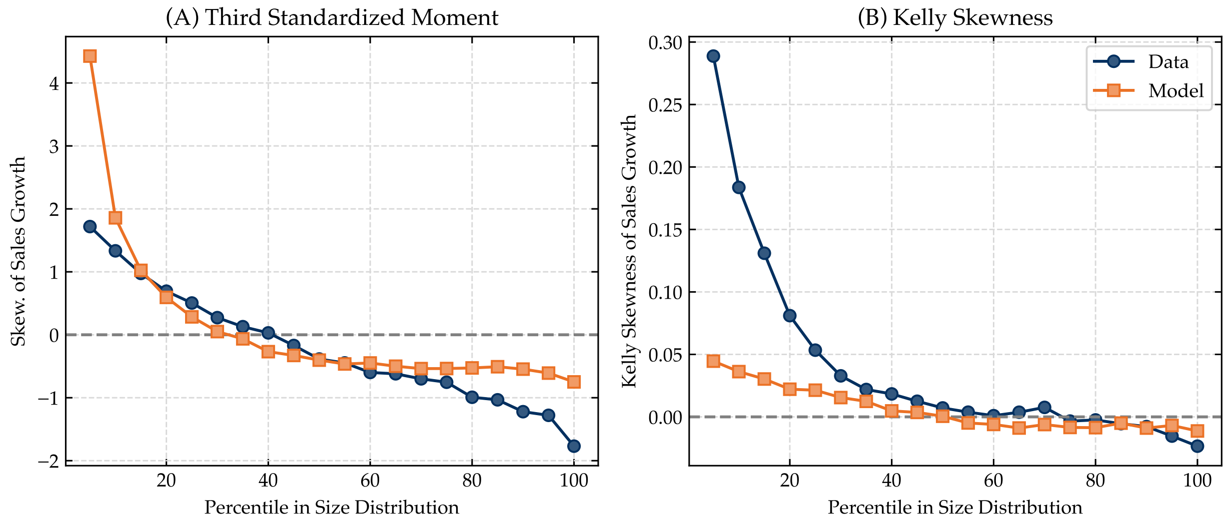


Figure 3: Kelly skewness across size bins (data). Notes: Each line corresponds to a tail definition (10%, 5%, 2.5%, 1%).

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 4. 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 5, matches this pattern well.

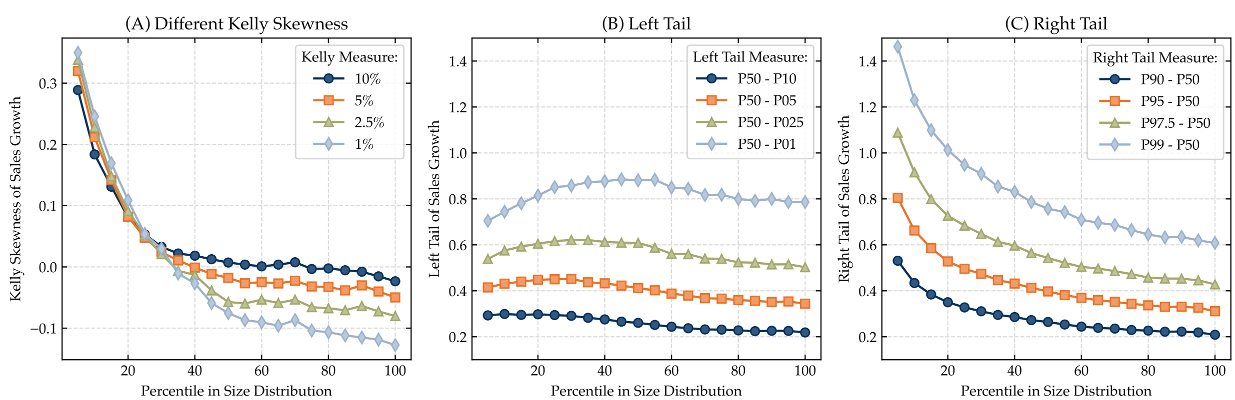


Figure 4: Kelly skewness across size bins (data). Notes: Each line corresponds to a tail definition (10%, 5%, 2.5%, 1%).

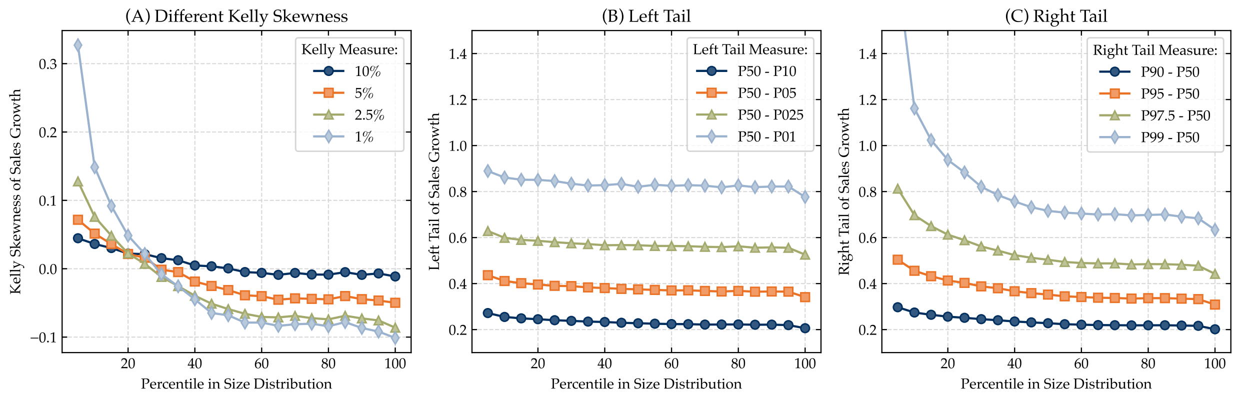


Figure 5: Kelly skewness across size bins (model). Notes: Same tail definitions as in the data figure.

Granularity thus helps the model match key empirical features of firm growth dynamics, including the size-dependent volatility and skewness profiles. These results support the assumption of idiosyncratic random growth processes for firm productivity, which underpins the model's theoretical predictions.

5.2 The Granular Drag in the Model

I next assess whether the model can replicate the empirical relationship between changes in concentration and future productivity growth at the industry level. To do so, I simulate a stationary economy with a large number of sectors (200000) and estimate industry-level regressions. To allow for distortions, I follow the Restuccia and Rogerson (2008) and allow for i.i.d. taxes and subsidies on firm sales of $\pm 10\%$. I then estimate regressions both from the model predictions with and without distortions. Table 3 shows the results when using the HHI based on sales. Both in the efficien and distorted model economies, an increase in concentration leads to a decline in future productivity growth, roughly of explaining 20-25% of the empirical coefficient.

Table 3: Model Meets Data: Industry-Level Regressions of Productivity Growth on Concentration Changes

	Data	Model Efficient	Model IID Wedges
	$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$	$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$	$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$
	(1)	(2)	(3)
HHI_t sales	-0.215**	-0.059***	-0.042***
	(0.062)	(0.006)	(0.004)
$\ln(\text{Prod}_t)$	-0.106***	-0.066***	-0.090***
	(0.017)	(0.008)	(0.006)
2-digit×Year FE	x	-	-
Observations	7218	200000	200000
S.E. type	by: industry+year	by: sector	by: sector
R^2	0.318	0.064	0.066
R^2 Within	0.062	-	-

SEs clustered by 5-digit industry and year.

Table 4 shows the results when using the HHI gap based on sales minus costs. In the model with distortions, an increase in the HHI gap leads to a decline in future productivity growth, roughly explaining 20% of the empirical coefficient.

Table 4: Model Meets Data: Industry-Level Regressions of Productivity Growth on Concentration Changes (HHI Gap)

	Data	Model Efficient	Model IID Wedges
	$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$	$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$	$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$
	(1)	(2)	(3)
HHI_t sales	-0.013	-0.059***	-0.060***
	(0.065)	(0.006)	(0.005)
HHI_t sales $-$ HHI_t costs	-1.164***		-0.215***
	(0.198)		(0.026)
$\ln(\text{Prod}_t)$	-0.070***	-0.066***	-0.068***
	(0.016)	(0.008)	(0.007)
2-digit$\times$Year FE	x	-	-
Observations	7218	200000	200000
S.E. type	by: industry+year	by: sector	by: sector
R^2	0.352	0.064	0.068
R^2 Within	0.109	-	-

SEs clustered by 5-digit industry and year.

Again, it is worth to take these results with caution, as the empirical estimates may be biased due to mismeasurement of the right sector boundaries.

5.3 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

$$\underbrace{\left(\ln A_{j,t+h} - \ln A_{j,t+h-1}\right)}_{\text{one-year growth at horizon } h} = \beta_h \, \Delta \ln HHI_{jt} + \gamma \ln A_{jt} + \alpha_j + \tau_t + \varepsilon_{j,t+h}, \tag{20}$$

where $\Delta \ln HHI_{jt}$ is the (standarized) 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 (20) traces the impulse response of *one-year* productivity growth to an innovation in concentration at t.

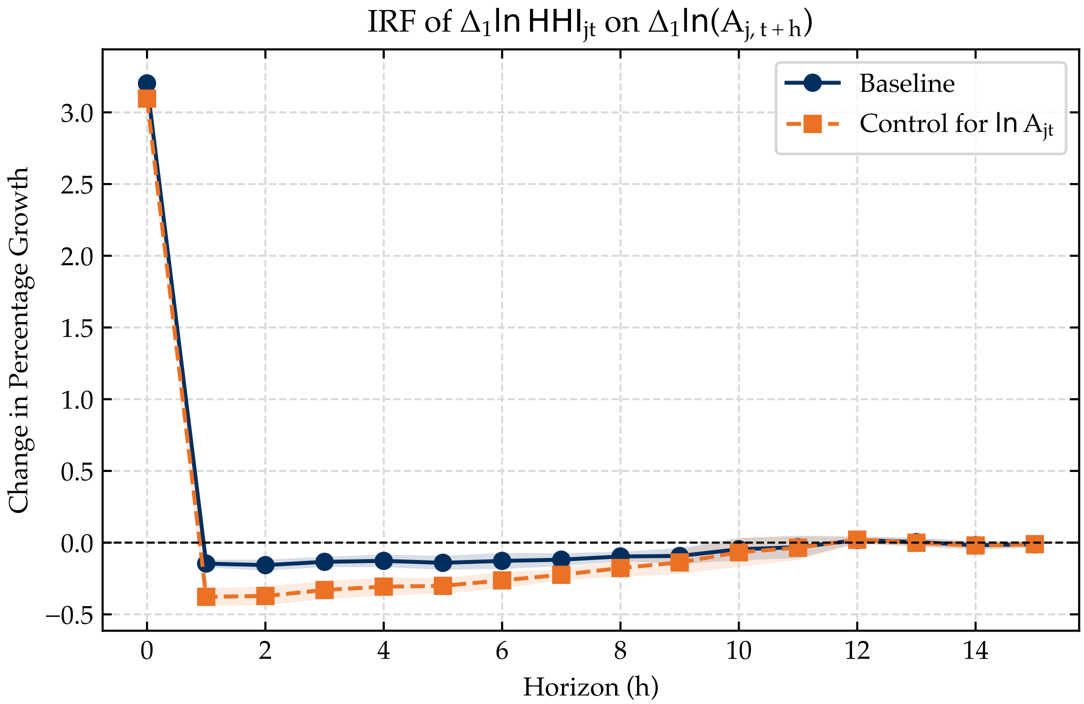


Figure 6: Local projection impulse responses to a change in concentration $\Delta \ln HHI_t$. Top panel: cumulative log productivity $\ln A_{j,t+h} - \ln A_{jt}$. Bottom panel: one-year growth $\ln A_{j,t+h} - \ln A_{j,t+h-1}$ (specification in Eq. (20)). Shaded bands are 95% confidence intervals; dashed 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 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

Aghion, Philippe and Peter Howitt, "A Model of Growth Through Creative Destruction," *Econometrica*, 1992, *60* (2), 323–351.

__ **, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li**, "A Theory of Falling Growth and Rising Rents," *The Review of Economic Studies*, 2023, *90* (6), 2675–2702.

__ **, Nicholas Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt**, "Competition and Innovation: An Inverted-U Relationship," *Quarterly Journal of Economics*, 2005, *120* (2), 701–728.

Atkeson, Andrew and Ariel Burstein, "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, December 2008, *98* (5), 1998–2031.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms*," *The Quarterly Journal of Economics*, 02 2020, *135* (2), 645–709.

Axtell, Robert L., "Zipf Distribution of U.S. Firm Sizes," *Science*, 2001, *293* (5536), 1818–1820.

Berry, Steven, Martin Gaynor, and Fiona Scott Morton, "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization," *Journal of Economic Perspectives*, August 2019, *33* (3), 44–68.

Boehm, Johannes, Ruairidh South, Ezra Oberfield, and Mazhar Waseem, "The Network Origins of Firm Dynamics: Contracting Frictions and Dynamism with Long-Term Relationships," Technical Report 2024. Preliminary draft, July 8 2024.

Burstein, Ariel T, Vasco M Carvalho, and Basile Grassi, "Bottom-Up Markup Fluctuations*," *The Quarterly Journal of Economics*, 06 2025, p. qjaf029.

Carvalho, Vasco and Basile Grassi, "Large Firm Dynamics and the Business Cycle," *American Economic Review*, April 2019, *109* (4), 1375–1425.

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng, "On Binscatter," *American Economic Review*, May 2024, *114* (5), 1488–1514.

Cavenaile, Laurent, Murat Alp Celik, and Xu Tian, "Are Markups Too High? Competition, Strategic Innovation, and Industry Dynamics," 2025. University of Toronto and University of Georgia, Working Paper.

di Giovanni, Julian, Andrei A. Levchenko, and Isabelle Méjean, "Firms, Destinations, and Aggregate Fluctuations," *Econometrica*, 2014, *82* (4), 1303–1340.

__ **,** __ **, and** __ , "Foreign Shocks as Granular Fluctuations," *Journal of Political Economy*, 2024, *132* (2), 463–514.

Gabaix, Xavier, "Zipf's Law and the Growth of Cities," *American Economic Review*, May 1999, *89* (2), 129–132.

__ , "Power Laws in Economics and Finance," *Annual Review of Economics*, 2009, *1* (1), 255–294.

__ , "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 2011, *79* (3), 733–772.

— , **Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll**, "The Dynamics of Inequality," *Econometrica*, 2016, *84* (6), 2071–2111.

Ganapati, Sharat, "Growing Oligopolies, Prices, Output, and Productivity," *American Economic Journal: Microeconomics*, August 2021, *13* (3), 309–27.

Gaubert, Cecile and Oleg Itskhoki, "Granular Comparative Advantage," *Journal of Political Economy*, 2021, *129* (3), 871–939.

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

Grossman, Gene M. and Elhanan Helpman, "Quality Ladders in the Theory of Growth," *The Review of Economic Studies*, 1991, *58* (1), 43–61.

Herskovic, Bernard, Bryan Kelly, Hanno Lustig, and Stijn Van Nieuwerburgh, "Firm Volatility in Granular Networks," *Journal of Political Economy*, 2020, *128* (11), 4097–4162.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India*," *The Quarterly Journal of Economics*, 11 2009, *124* (4), 1403–1448.

Klette, Tor Jakob and Samuel Kortum, "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 2004, *112* (5), 986–1018.

Kotz, Samuel, Tomaz J. Kozubowski, and Krzysztof Podgórski, *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance* Springer Book Archive, 1 ed., MA: Birkhäuser Boston, 2001.

Kwon, Spencer Y., Yueran Ma, and Kaspar Zimmermann, "100 Years of Rising Corporate Concentration," *American Economic Review*, July 2024, *114* (7), 2111–40.

Ma, Yueran, Mengdi Zhang, and Kaspar Zimmermann, "Business Concentration around the World: 1900–2020," 2025. Draft, February 2025.

Marshall, Albert W., Ingram Olkin, and Barry C. Arnold, *Inequalities: Theory of Majorization and Its Applications* Springer Series in Statistics, 2 ed., New York: Springer, 2011.

Olmstead-Rumsey, Jane, "Market Concentration and the Productivity Slowdown," Technical Report 107000, Munich Personal RePEc Archive (MPRA) 2019.

Restuccia, Diego and Richard Rogerson, "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic Dynamics*, 2008, *11* (4), 707–720.

Stanley, Michael H. R., Luís A. N. Amaral, Sergey V. Buldyrev et al., "Scaling behaviour in the growth of companies," *Nature*, 1996, *379*, 804–806.

Statista, "Grocery market share of leading retailers in the United Kingdom (UK)," https://www.statista.com/statistics/280208/grocery-market-share-in-the-united-kingdom-uk/ 2025. Accessed: 2025-10-02.

— , "Worldwide Market Share of Leading Cloud Infrastructure Service Providers," https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/ 2025. Accessed: 2025-10-02.

Sutton, John, "Gibrat's Legacy," *Journal of Economic Literfature*, 1997, *35* (1), 40–59.

Weidenman, Per, "The Serrano Database for Analysis and Register-Based Statistics," Swedish House of Finance Research Data Center 2016. Accessed: 2023-10-15.

YahooFinance, "Nvidia crushes rivals, secures unprecedented market share in GPUs," https://finance.yahoo.com/news/nvidia-crushes-rivals-secures-unprecedented-102235255.html 2024. Accessed: 2025-10-02.

Yeh, Chen, "Revisiting the Origins of Business Cycles With the Size-Variance Relationship," *The Review of Economics and Statistics*, 2025, *107* (3), 864–871.

A Model Appendix

A.1 Markups a la Atkeson and Burstein (2008)

I extend the market structure presented in subsection 2.2 to allow for endogenous markups following Atkeson and Burstein (2008). 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:

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

Here ω_{ij} is the sales share of firm i in sector j. In Bertrand and Cournot competition, larger sales shares translate into higher markups, while under monopolistic competition markups remain constant and passthrough is complete. 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.

Bertrand Competition The firm takes competitors' prices $\{P_{kj}\}_{k\neq i}$ as given. The elasticity is derived from the log-differentiated demand curve, recognizing that a firm's price P_{ij} affects the sectoral price index P_j.

$$\zeta_{ij} \equiv -\frac{\partial \ln Y_{ij}}{\partial \ln P_{ij}}$$

$$\text{Given } \ln Y_{ij} = -\varepsilon \ln P_{ij} + (\varepsilon - 1)\ln P_j + C$$

$$\zeta_{ij} = \varepsilon - (\varepsilon - 1)\frac{\partial \ln P_j}{\partial \ln P_{ij}}$$

$$\text{Since } \frac{\partial \ln P_j}{\partial \ln P_{ij}} = \frac{P_{ij}}{P_j}\frac{\partial P_j}{\partial P_{ij}} = \left(\frac{P_{ij}}{P_j}\right)^{1-\varepsilon} = \omega_{ij}$$

$$\implies \zeta_{ij} = \varepsilon - (\varepsilon - 1)\omega_{ij} = \varepsilon(1-\omega_{ij}) + \omega_{ij}.$$

Cournot Competition The firm takes competitors' quantities $\{Y_{kj}\}_{k\neq i}$ as given. We derive the inverse elasticity from the log-differentiated inverse demand curve, recognizing that a firm's quantity Y_{ij} affects sectoral output Y_j.

$$\begin{aligned}
\frac{1}{\zeta_{ij}} &\equiv -\frac{\partial \ln P_{ij}}{\partial \ln Y_{ij}} \\
\text{Given } \ln P_{ij} &= -\frac{1}{\varepsilon}\ln Y_{ij} + \left(\frac{1}{\varepsilon}-1\right)\ln Y_j \\
\frac{1}{\zeta_{ij}} &= \frac{1}{\varepsilon} - \left(\frac{1}{\varepsilon}-1\right)\frac{\partial \ln Y_j}{\partial \ln Y_{ij}} \\
\text{Since } \frac{\partial \ln Y_j}{\partial \ln Y_{ij}} &= \frac{Y_{ij}}{Y_j}\frac{\partial Y_j}{\partial Y_{ij}} = \left(\frac{Y_{ij}}{Y_j}\right)^{(\varepsilon-1)/\varepsilon} = \omega_{ij} \\
\implies \frac{1}{\zeta_{ij}} &= \frac{1}{\varepsilon} - \left(\frac{1}{\varepsilon}-1\right)\omega_{ij} = \frac{1}{\varepsilon}(1-\omega_{ij}) + \omega_{ij}.
\end{aligned}$$

A.2 The Concentration Drag with Endogenous Markups

Proof. We postulate that

$$d\ln\mu_{ij} = \mathbb{E}_t\left[\frac{1}{dt}d\ln\mu_{ij}\right]dt + \sigma_{\mu_{ij}}\,dW_{\mu_{ij}},$$

where $W_{\mu_{ij}}$ is a standard Wiener process with $dW_{\mu_{ij}}\,dW_{\mu_{kj}} = \rho_{\mu_{ij}\mu_{kj}}\,dt$, and $\sigma_{\mu_{ij}}$ is the volatility of the markup process. Furthermore, $dW_{\mu_{ij}}\,dW_{ij} = \rho_{A_{ij}\mu_{kj}}\,dt$.

$$\begin{aligned}
\gamma_{jt} = {} & \underbrace{g - \frac{\sigma^2}{2}}_{\text{Mean Productivity Change}} + \underbrace{(\varepsilon-1)\frac{\sigma^2}{2}\Big(1-\mathcal{H}_{jt} + (\varepsilon-1)\big(\mathcal{H}^{\kappa}_{jt}-\mathcal{H}_{jt}\big)\Big)}_{\text{Reallocation due to technology}} \\
& + \underbrace{\varepsilon\sum_{i=1}^{N_j}(\kappa_{ij}-\omega_{ij})\,\mathbb{E}_t\left[\frac{1}{dt}d\ln\mu_{ij}\right]}_{\text{Mean Markup Change}} \\
& + \underbrace{\frac{1}{2}\left\{\varepsilon(\varepsilon-1)\left[\sum_i \omega_{ij}\,\sigma^2_{\mu_{ij}} - \sum_{i,k}\omega_{ij}\omega_{kj}\,\sigma_{\mu_{ij}}\sigma_{\mu_{kj}}\rho_{\mu_{ij}\mu_{kj}}\right] - \varepsilon^2\left[\sum_i \kappa_{ij}\,\sigma^2_{\mu_{ij}} - \sum_{i,k}\kappa_{ij}\kappa_{kj}\,\sigma_{\mu_{ij}}\sigma_{\mu_{kj}}\rho_{\mu_{ij}\mu_{kj}}\right]\right\}}_{\text{Reallocation due to markup changes (Jensen/variance terms)}} \\
& + \underbrace{\varepsilon(\varepsilon-1)\left[\sum_i(\kappa_{ij}-\omega_{ij})\,\sigma\,\sigma_{\mu_{ij}}\,\rho_{A_{ij}\mu_{ij}} + \sum_{i,k}(\omega_{ij}\omega_{kj}-\kappa_{ij}\kappa_{kj})\,\sigma\,\sigma_{\mu_{kj}}\,\rho_{A_{ij}\mu_{kj}}\right]}_{\text{Interaction between technology and markup changes (covariances)}}.
\end{aligned}$$

■

A.3 Sectoral Productivity Growth with Misallocation

Throughout this subsection I focus on the case without jumps $dA_{ijt}/A_{ijt} = gdt + \sigma dW_{ijt}$. It is not hard to add the jumps, perhaps show in appendix.

The sectoral productivity index can be written as:

$$A_{jt} \;=\; \frac{\left(\sum_{i=1}^{N_j} A_{ijt}^{\varepsilon-1}\,\mu_{ij}^{1-\varepsilon}\right)^{\frac{\varepsilon}{\varepsilon-1}}}{\sum_{i=1}^{N_j} A_{ijt}^{\varepsilon-1}\,\mu_{ij}^{-\varepsilon}}. \tag{A.1}$$

Recall the *revenue weights* and the *cost (labor) shares* are defined as

$$\omega_{ijt} := \frac{A_{ijt}^{\varepsilon-1}\mu_{ij}^{1-\varepsilon}}{\sum_{k=1}^{N_j} A_{kjt}^{\varepsilon-1}\mu_{kj}^{1-\varepsilon}}, \qquad \kappa_{ijt} := \frac{A_{ijt}^{\varepsilon-1}\mu_{ij}^{-\varepsilon}}{\sum_{k=1}^{N_j} A_{kjt}^{\varepsilon-1}\mu_{kj}^{-\varepsilon}} = \frac{L_{ijt}}{\sum_{k=1}^{N_j} L_{kjt}}, \tag{A.2}$$

and let $\mathcal{H}_{jt}^{\omega} := \sum_i \omega_{ijt}^2$ and $\mathcal{H}_{jt}^{\kappa} := \sum_i \kappa_{ijt}^2$ be their concentration indices.

Result. Under the diffusion specification in (11) with $\lambda = 0$ (no jumps), the stochastic differential equation for sectoral log productivity is

$$\begin{aligned} d\ln A_{jt} = \left(g - \frac{\sigma^2}{2}\right) dt \;+\; \frac{\sigma^2}{2}\Big[\varepsilon(\varepsilon-1)\big(1-\mathcal{H}_{jt}^{\omega}\big) - (\varepsilon-1)^2\big(1-\mathcal{H}_{jt}^{\kappa}\big)\Big]\, dt \\ +\; \sigma\left[\varepsilon\sum_{i=1}^{N_j}\omega_{ijt}^{\omega}\, dW_{ijt} \;-\; (\varepsilon-1)\sum_{i=1}^{N_j}\kappa_{ijt}\, dW_{ijt}\right]. \end{aligned} \tag{A.3}$$

Taking expectations and simplyfying gives the expected growth rate under misallocation:

$$\gamma_{jt} := \mathbb{E}_t\left[\frac{d\ln A_{jt}}{dt}\right] \;=\; \left(g - \frac{\sigma^2}{2}\right) \;+\; (\varepsilon-1)\frac{\sigma^2}{2}\Big[1 - \mathcal{H}_{jt}^{\omega} - (\varepsilon-1)\big(\mathcal{H}_{jt}^{\omega} - \mathcal{H}_{jt}^{\kappa}\big)\Big]. \tag{A.4}$$

Proof. Write

$$N_{jt} := \sum_{i=1}^{N_j} A_{ijt}^{\varepsilon-1}\mu_{ij}^{1-\varepsilon}, \qquad D_{jt} := \sum_{i=1}^{N_j} A_{ijt}^{\varepsilon-1}\mu_{ij}^{-\varepsilon}, \qquad \ln A_{jt} = \frac{\varepsilon}{\varepsilon-1}\ln N_{jt} - \ln D_{jt}.$$

Since μ_{ij} are constants, for each summand $X_{ijt}^{(N)} := A_{ijt}^{\varepsilon-1}\mu_{ij}^{1-\varepsilon}$ and $X_{ijt}^{(D)} := A_{ijt}^{\varepsilon-1}\mu_{ij}^{-\varepsilon}$,

$$d\ln X_{ijt}^{(N)} = (\varepsilon-1)\,d\ln A_{ijt}, \qquad d\ln X_{ijt}^{(D)} = (\varepsilon-1)\,d\ln A_{ijt}.$$

Then we have $\omega_{ijt} := X_{ijt}^{(N)}/N_{jt}$ and $\kappa_{ijt} := X_{ijt}^{(D)}/D_{jt}$. For any positive sum $U = \sum_i X_i$ with weights $\varpi_i := X_i/U$ and independent Brownians, Itô's formula for $\ln U$ gives

$$d\ln U \;=\; \sum_i \varpi_i\, d\ln X_i \;+\; \frac{1}{2}\left(\sum_i \varpi_i b_i^2 \;-\; \sum_i \varpi_i^2 b_i^2\right) dt,$$

where b_i is the diffusion loading in $d\ln X_i$. Applying this identity to N_{jt} (with $b_i = (\varepsilon-1)\sigma$ and weights ω_{ijt}) yields

$$d\ln N_{jt} = (\varepsilon-1)\left(g - \frac{\sigma^2}{2}\right) dt + (\varepsilon-1)\sigma \sum_i \omega_{ijt} dW_{ijt} + \frac{(\varepsilon-1)^2\sigma^2}{2}\left(1 - \mathcal{H}_{jt}^{\omega}\right) dt,$$

The same calculation for D_{jt} (with weights κ_{ijt}) gives

$$d\ln D_{jt} = (\varepsilon-1)\left(g - \frac{\sigma^2}{2}\right) dt + (\varepsilon-1)\sigma \sum_i \kappa_{ijt} dW_{ijt} + \frac{(\varepsilon-1)^2\sigma^2}{2}\left(1 - \mathcal{H}_{jt}^{\kappa}\right) dt,$$

with $\mathcal{H}_{jt}^{\kappa} := \sum_i (\kappa_{ijt})^2$. Combining via $\ln A_{jt} = \frac{\varepsilon}{\varepsilon-1}\ln N_{jt} - \ln D_{jt}$ yields (A.3); taking expectations gives (A.4). ■

B Data

B.1 Swedish Firm Data

B.2 CompNet

B.3 U.S. Data from Ganapati (2021)