The Granular Drag on Growth*

Juan Llavador Peralt†

IIES, Stockholm University

November 2, 2025

Job Market Paper. Click here for the latest version.

Abstract

This paper develops a simple theory linking firm granularity to productivity growth. I show that when production is concentrated in a few large firms, idiosyncratic shocks generate smaller reallocation gains, an effect that I term the *granular drag* on growth. I formalize this mechanism in a multi-sector model where granular firms experience stochastic productivity shocks. In efficient economies, higher sales concentration lowers expected sectoral growth by reducing the gains from reallocation. With distortions, the slowdown is amplified whenever sales are more concentrated than production cost shares. Using firm- and industry-level data from Sweden, the United States, and European economies, I find empirical evidence consistent with these predictions. The quantified version of the model predicts that a ten-percentage-point rise in the Herfindahl index of sales concentration reduces five-year productivity growth by about 0.6 percentage points in the efficient benchmark and that a similar rise in the gap between the Herfindahl indices of sales and costs reduces five-year growth by roughly 2 percentage points. The calibrated model also reproduces the observed size dependence of firm growth rates, matching the decline in volatility and skewness with firm size. These findings highlight the importance of micro-reallocation for understanding how market structure shapes productivity growth across sectors and, potentially, aggregate economies.

*I am indebted to my advisors, Per Krusell, Timo Boppart, Xavier Gabaix, and Joshua Weiss, for their guidance, mentorship, and support. 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 Thore Petersen and Jinglun Yao for checking the size-dependent skewness in firm growth in German and French firm data, respectively. I thank the "Jan Wallanders, Tom Hedelius, and Tore Browaldh Foundation" for financial support for my visit to Harvard University. All errors are my own.

†Web: https://juanlla.github.io/; Email: juan.llavadorperalt@iies.su.se

1 Introduction

Welfare improvements in an economy ultimately depend on its ability to generate sustained productivity growth, which arises from two sources: technological progress and the reallocation of resources toward more efficient producers. Empirically, the allocation of production across firms is highly uneven, with a few large firms accounting for a disproportionate share of output in many sectors. In 2024, for example, Nvidia captured roughly 90% of global GPU revenues, Amazon Web Services about one-third of the cloud-infrastructure market, and Tesco nearly 27% of UK grocery sales. This fact has been central to the study of the granular origins of aggregate fluctuations, which emphasizes how idiosyncratic shocks to large firms can generate aggregate volatility. Yet, whether such firm-level granularity also shapes *expected productivity growth* remains an open question.

This paper proposes a simple theory of micro-reallocation to answer this question. It builds on two premises. First, that some firms are granular—large enough to influence aggregate outcomes. Second, that firm productivity evolves stochastically in response 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. I characterize the stochastic dynamics of firms and aggregates and show that granular concentration drags down expected productivity growth by reducing the reallocation gains from idiosyncratic shocks. To evaluate these mechanisms empirically, I test the model's predictions using firm- and industry-level data from Sweden, complemented by industry data from the United States and a broad set of European countries. The results confirm the presence of granular effects on productivity growth consistent with the theory.

Many models of growth and firm dynamics acknowledge the fact that firm size distributions are highly skewed, and imply a firm size distribution with a Pareto tail. However, for tractability, they often rely on a continuum of infinitesimal firms, thus abstracting from the finite nature of granular firms.[1] I relax this assumption and develop a model with a finite number of sectors, each populated by a finite number of firms. To capture the key mechanisms, the model features a nested CES structure. Within sectors, goods are gross substitutes with an elasticity of substitution greater than one, whereas sectoral output exhibits unit elasticity, reflecting higher competition within than across sectors. Firms experience random, idiosyncratic productivity shocks, leading to

[1]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.

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 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 maximized: 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. These sectoral effects naturally propagate to the aggregate economy: when more sectors become concentrated, aggregate productivity growth slows as reallocation becomes less effective at the economy-wide level.

Granularity also shapes individual firm growth. Because the elasticity of substitution across sectors is lower than the elasticity within sectors, as a firm grows large within its sector, it saturates the market and has less room to grow. Consequently, its growth rate distribution becomes left-skewed and less volatile. In contrast, its smaller rivals face more business-stealing opportunities, 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.

I test the model's predictions using firm- and industry-level data. The theory implies two empirical relationships. First, in an efficiently allocated economy, higher concentration should be

followed by slower productivity growth, as reallocation becomes less effective when a few firms dominate production. Second, when resource allocation is distorted, for example when the largest firms charge higher markups than others, differences between sales- and cost-based concentration should further magnify this slowdown. I evaluate these predictions using administrative firm-level data from Sweden, together with harmonized industry-level data from the United States and the European Union. Across datasets, industries that become more concentrated, or display greater differences between sales- and cost-based concentration, experience significantly lower future productivity growth, consistent with the granular drag mechanism predicted by the model.

I quantify the model using the simulated method of moments (SMM). The parameters of the productivity process are disciplined by cross-sectional moments of firm growth that capture volatility, skewness, and kurtosis in the unconditional distribution. The model reproduces the observed size-dependence of volatility and skewness, even though these moments are not explicitly targeted. I also match the median level of concentration across 5-digit Swedish industries. Because sectors contain a finite number of firms, the model naturally generates a realistic dispersion in concentration levels across sectors. A temporary increase in concentration leads to an immediate rise in productivity, followed by a prolonged slowdown in growth. Quantitatively, in the efficient benchmark, a ten percentage point increase in concentration lowers five-year productivity growth by about 0.6 percentage points. When distortions are present, an equivalent increase in the difference between sales- and cost-based concentration reduces five-year growth by roughly 2 percentage points. These effects propagate to the aggregate economy and remain economically significant over extended time horizons, highlighting the importance of micro-reallocation effects for understanding productivity growth at medium and long-term horizons.

Related Literature This paper relates to three main strands of the literature. First, it is most closely related to the literature that studies how microeconomic shocks propagate to the macroeconomy. The seminal contribution by Gabaix (2011) shows that idiosyncratic shocks to large firms can generate aggregate fluctuations even in the absence of aggregate shocks. Subsequent work extended this idea to trade (di Giovanni et al., 2014, 2024), firm dynamics and aggregate volatility (Carvalho and Grassi, 2019), and markup fluctuations (Burstein et al., 2025). This research highlights the granular origins of aggregate fluctuations, showing that when economic activity is concentrated in a few large firms, idiosyncratic shocks do not average out and generate measurable macro-level variability. My paper builds on this insight but focuses on the implications of granularity for

expected productivity growth rather than short-term fluctuations. Gaubert and Itskhoki (2021) are closest to my approach. They show how firm granularity predicts reversals in trade flows, as large firms drive fluctuations in export dynamics across countries.

A closely related line of research studies how microeconomic shocks and distortions aggregate to affect macroeconomic outcomes, building on the aggregation theorems of Hulten (1978) and the general-equilibrium frameworks of Baqaee and Farhi (2019a) and Baqaee and Farhi (2019b), but my approach differs by focusing on the expected evolution of productivity growth given the probability distribution of firm-level shocks and the current market structure, rather than tracing the effect of a realized shock.

Second, 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.

Third, 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 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). My approach is complementary to this literature in that it emphasizes a different channel through which concentration affects productivity growth.

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 finite 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:

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

where ω_j for $j = 1, \ldots, N$ are non-negative preference weights satisfying $\sum_{j=1}^{N} \omega_j = 1$. This formulation defines a sector as a market with a fixed expenditure share ω_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.[2] 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:

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

[2] 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 $\int_0^1 \omega_j dj = 1$.

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 $P \equiv \prod_{j=1}^{N} P_j^{\omega_j}$ 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:

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

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.[3] 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.3 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

[3] 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.

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: $\sum_{j=1}^{N}\sum_{i=1}^{N_j} L_{ij} = L$.
- The government budget is balanced: $T = \sum_{j=1}^{N}\sum_{i=1}^{N_j} \tau_{ij} P_{ij} Y_{ij}$.

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.

2.4 Firm-Level Outcomes

Two main firm-level outcomes are of interest for the upcoming analysis: sectoral sales- and cost-shares, denoted as s_{ij} and $\widetilde{s}_{ij}$, 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:

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

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

$$S_{ij} := P_{ij} Y_{ij} = s_{ij}\, \omega_j P Y.$$

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:

$$\widetilde{s}_{ij} := \frac{WL_{ij}}{WL_j} = \frac{s_{ij}/\mu_{ij}}{\sum_{k=1}^{N_j} s_{kj}/\mu_{kj}}. \tag{7}$$

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 markup 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-share-weighted arithmetic average:

$$\mu_j = \sum_{i=1}^{N_j} \widetilde{s}_{ij}\mu_{ij} \tag{8}$$

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 toward 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 cost HHIs in sector j are defined as:

$$\mathcal{H}_j = \sum_{i=1}^{N_j} s_{ij}^2, \quad \widetilde{\mathcal{H}}_j = \sum_{i=1}^{N_j} \left(\widetilde{s}_{ij}\right)^2.$$

If markups are constant across firms, sales and cost shares coincide, and thus $\mathcal{H}_j = \widetilde{\mathcal{H}}_j$. 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: $\mathcal{H}_j > \widetilde{\mathcal{H}}_j$. 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 $\mu := \frac{P}{W/A}$, which equals the cost-share-weighted average of sectoral markups. Using this definition, we can express aggregate productivity as a function of sectoral productivities and markups:

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

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. I focus on these two channels in the next section.

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.[4] 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 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 is the common drift minus the volatility correction $\sigma^2/2$ due to Jensen's inequality, plus the expected jump contribution $\lambda \mathbb{E}[J]$.[5]

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

[4] 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, $\Pr(\Delta Q_{ijt} = 1) = \lambda \Delta t + o(\Delta t)$.

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

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.[6]

A critique of random growth models is that in the data, rather than being constant, growth volatility declines with size.[7] 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.

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:

Proposition 1 (Granular Drag in Efficient Sectors)**.** *In an efficient economy with firm productivity dynamics given by $dA_{ijt}/A_{ijt} = gdt + \sigma dW_{ijt}$, expected sectoral productivity growth $\gamma_{jt} := \mathbb{E}_t[\frac{1}{dt} d \ln A_{jt}]$ is given by:*

$$\gamma_{jt} = \underbrace{g - \frac{\sigma^2}{2}}_{\textit{Avg. Firm}} + \underbrace{(\varepsilon - 1)\frac{\sigma^2}{2}\left(1 - \mathcal{H}_{jt}\right)}_{\textit{Reallocation}}, \tag{13}$$

where $g - \frac{\sigma^2}{2}$ is the average firm productivity growth, and $(\varepsilon - 1)\frac{\sigma^2}{2}\left(1 - \mathcal{H}_{jt}\right)$ is a positive reallocation residual that scales with one minus the sales-HHI $\mathcal{H}_{jt} := \sum_{i=1}^{N_j} s_{ijt}^2$.

[6] 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.

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

Equation (13) shows that in an efficient economy with only diffusion shocks, the sales HHI $\mathcal{H}_{jt}$ is a sufficient statistic for how granularity affects expected productivity growth. I leave the proof to Appendix A.2 and focus here on building the intuition behind the result using two polar cases. First, consider the case of a monopolist that dominates the whole sector, such that $s_{1jt} = 1$ and $\mathcal{H}_{jt} = 1$. The expected growth rate reduces to the expected growth of the single firm:

$$\gamma^1 := \lim_{s_{1jt} \to 1} \gamma_{jt} = \underbrace{g - \frac{\sigma^2}{2}}_{\text{Avg. Firm}}. \tag{14}$$

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 law of large numbers holds and the growth rate is now deterministic. Since no single firm has a sizable market share, $\mathcal{H}_{jt} = 0$, and the growth rate can be written as the *average-firm* term plus a positive residual that captures *reallocation* gains:

$$\gamma^\infty := \lim_{N_j \to \infty} \gamma_{jt} = \underbrace{g - \frac{\sigma^2}{2}}_{\text{Avg. Firm}} + \underbrace{(\varepsilon - 1)\frac{\sigma^2}{2}}_{\text{Reallocation}}. \tag{15}$$

Where does the reallocation term in (15) 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:

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

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: $dA_{ijt}/A_{ijt} = (e^{J_{ijt}} - 1)dQ_{ijt}$.[8] The expected growth rate of sectoral productivity is now:

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

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

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

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

[8] The general case with both diffusion and jumps is just the sum of the two components, as they are independent. See Appendix A.2 for the general case.

The expected growth rate in the fully-diversified case can again be decomposed into an average-firm term plus a reallocation term. As Proposition 2 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 more complex, an approximation for small jumps $J \approx 0$ makes the role of granularity explicit:

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

Up to 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 $\mathcal{H}_{n,jt} := \sum_{i=1}^{N_j} s_{ijt}^n$. 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 $0 \leqslant 3\mathcal{H}_{jt} - 2\mathcal{H}_{3,jt} \leqslant 1$.

The following proposition formalizes the preceding results, showing that in efficient economies,

expected sectoral productivity growth is always bounded between the monopolist and fully diversified cases.

Proposition 2. *Consider an efficient economy where firm productivity follows the 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* γ^∞ *case and below by that of a monopolist* γ^1*:*

$$\gamma^1 \leqslant \gamma_{jt} < \gamma^\infty,$$

and the reallocation term with a continuum of firms $\gamma^\infty - \gamma^1$ *is increasing in the within sector elasticity of substitution* ε*.*

The proof for the case without jumps is immediate from (13). 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 2, it follows that in efficient economies, the reallocation residual is always positive.

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 ($\varepsilon > 2$). The following corollary to Proposition 2 formalizes this result.

Corollary 1 (Percentage Growth)**.** *Consider an efficient economy where firm productivity follows the jump-diffusion process in* (11)*, and suppose that the elasticity of substitution between sectors is greater than two (*$\varepsilon > 2$*). Then, the expected sectoral productivity percentage growth rate* $\Gamma_{jt} := \mathbb{E}_t[\frac{1}{dt} dA_{jt} / A_{jt}]$ *is bounded above by the growth rate in the fully diversified case and below by that of a monopolist:*

$$\Gamma^1 \leqslant \Gamma_{jt} < \Gamma^\infty.$$

If $\varepsilon < 2$, the inequalities are reversed.

The proof is in Appendix A. Intuitively, because firms are granular, sectoral productivity is volatile, and this volatility increases with concentration as in Gabaix (2011). Taking the logarithm of a volatile variable introduces a negative volatility correction due to Jensen's inequality. In the case with no jumps this correction term is $-\frac{\sigma^2}{2}\mathcal{H}_{jt}$. With percentage growth, the correction is not present, and so the reallocation gains need not always be positive. Reassuringly, as the number of firms goes to infinity, the two measures of growth converge: $\gamma^{\infty} = \Gamma^{\infty}$.

Aggregate Productivity Growth The previous analysis extends naturally to the aggregate economy with N sectors. Under efficient allocation across sectors, aggregate productivity is given by the Cobb-Douglas index $A_t = \prod_{j=1}^{N} A_{jt}^{\omega_{jt}}$, where ω_{jt} is sector j's expenditure share. The following corollary to Proposition 1 characterizes aggregate productivity growth.

Corollary 2. *In an efficient economy with firm productivity dynamics given by $dA_{ijt}/A_{ijt} = gdt + \sigma dW_{ijt}$, expected aggregate productivity growth $\gamma_t := \mathbb{E}_t[\frac{1}{dt} d\ln A_t]$ is given by:*

$$\gamma_t = g - \frac{\sigma^2}{2} + (\varepsilon - 1)\frac{\sigma^2}{2}\left(1 - \sum_{j=1}^{N} \omega_{jt}\mathcal{H}_{jt}\right). \tag{16}$$

The proof follows directly from Proposition 1 and the Cobb-Douglas aggregation across sectors. Since sectors are granular, aggregate productivity growth inherits a granular drag that scales with the sales-weighted average of sectoral sales HHIs. In analogy to the sectoral case, the aggregate economy is well diversified when there are infinitely many firms per sector, so that $\mathcal{H}_{jt} \to 0$ for all j, and aggregate productivity growth reaches its maximum. Conversely, if one sector dominates the economy and is itself dominated by a single firm, so that $s_{1t} = 1$ and $\mathcal{H}_{1t} = 1$, aggregate productivity growth reduces to the average-firm contribution $g - \sigma^2/2$.

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. The next proposition shows how misallocation also affects *growth* when firms are granular.

Proposition 3. *Consider an economy where firm productivity follows the process* $dA_{ijt}/A_{ijt} = gdt + \sigma dW_{ijt}$, *and firms have fixed markup heterogeneity* μ_{ij}. *Then, the expected sectoral productivity growth rate* $\gamma_{jt} = \mathbb{E}_t[d \ln A_{jt}/dt]$ *is given by:*

$$\gamma_{jt} = \underbrace{g - \frac{\sigma^2}{2}}_{\text{Avg. Firm}} + \underbrace{(\varepsilon - 1)\frac{\sigma^2}{2}\left[1 - \mathcal{H}_{jt} - (\varepsilon - 1)\left(\mathcal{H}_{jt} - \widetilde{\mathcal{H}}_{jt}\right)\right]}_{\text{Reallocation}}. \tag{17}$$

The proof is in Appendix A.5. In the efficient allocation, sales and cost shares coincide, and the sales HHI $\mathcal{H}_{jt}$ was, up to second order, a sufficient statistic for how granularity affected growth. Under misallocation, however, the difference between sales- and cost-based shares matters as well. If we compare equation (A.4) to the efficient case in equation (13), for a fixed sales concentration $\mathcal{H}_{jt}$, misallocation increases or decreases the expected growth rate by $(\varepsilon - 1)^2\sigma^2/2 \times (\mathcal{H}_{jt} - \widetilde{\mathcal{H}}_{jt})$. Intuitively, if more productive firms have high markups (relative to small firms), sales concentration $\mathcal{H}_{jt}$ will be high relative to cost concentration $\widetilde{\mathcal{H}}_{jt}$. 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. Empirically, the former case is more common, as large firms tend to have lower labor shares (Autor et al., 2020).

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 $(\mathcal{H}_{jt}, \widetilde{\mathcal{H}}_{jt} \to 0)$, 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.

Sectoral Markup Growth Even with firm markups fixed over time, sectoral markups μ_{jt} evolve endogenously as the sales shares of firms change. A simple Corollary to Proposition 3 characterizes expected sectoral markup growth.

Corollary 3. *Consider an economy where firm productivity follows the process* $dA_{ijt}/A_{ijt} = gdt + \sigma dW_{ijt}$, *and firms have fixed markup heterogeneity* μ_{ij}. *Then, expected sectoral markup growth is given by:*

$$\mathbb{E}_t\left[\frac{1}{dt}d\ln\mu_{jt}\right] = (\varepsilon-1)^2\frac{\sigma^2}{2}\left(\widetilde{\mathcal{H}}_{jt} - \mathcal{H}_{jt}\right). \tag{18}$$

The proof is in Appendix A.5. Comparing (18) to the misallocation growth expression in (A.4), we see that expected sectoral markup growth is proportional to the difference between cost and sales HHIs. If more productive firms have high markups, sales concentration will be high relative to cost concentration, and sectoral markups will tend to decline on average. Intuitively, as the most productive firms grow larger, they push up sectoral productivity, but their high markups limit their sales growth, leading to a decline in sectoral markups. Conversely, if more productive firms have low markups, sales concentration will be low relative to cost concentration, and sectoral markups will tend to rise on average.

Aggregate Growth For efficient economies, we saw that aggregate productivity growth inherited the granular drag from sectoral sales-weighted HHIs. With distortions, aggregate productivity growth also depends on the joint distribution of productivity and markups across firms and sectors. The next proposition shows that expected aggregate productivity $\gamma_t := \mathbb{E}_t[\frac{1}{dt}d\ln A_t]$ is given the sectoral sales-weighted difference between expected sectoral productivity growth and markup growth, plus expected aggregate markup growth:

$$\gamma_t = \sum_{j=1}^{N}\omega_{jt}\left(\gamma_{jt} - \mathbb{E}_t\left[\frac{1}{dt}d\ln\mu_{jt}\right]\right) + \mathbb{E}_t\left[\frac{1}{dt}d\ln\mu_t\right].$$

Proposition 4. *Consider an economy where firm productivity follows the process* $dA_{ijt}/A_{ijt} = gdt + \sigma dW_{ijt}$*, and firms have fixed markup heterogeneity* μ_{ij}*. Then, expected aggregate markup growth is given by:*

$$\mathbb{E}_t\left[\frac{1}{dt}d\ln\mu_t\right] = (\varepsilon-1)^2\frac{\sigma^2}{2}\sum_{j=1}^{N}\widetilde{\omega}_j\left(\widetilde{\mathcal{H}}_{jt}-\mathcal{H}_{jt}\right) - (\varepsilon-1)^2\frac{\sigma^2}{2}\sum_{j=1}^{N}\widetilde{\omega}_j(1-\widetilde{\omega}_j)\mathcal{V}_{jt}, \tag{19}$$

where $\widetilde{\omega}_j := \omega_{jt}\mu_{jt}^{-1}/\sum_{k=1}^{N}\omega_{kt}\mu_{kt}^{-1}$ *are cost shares at the sector level, and* $\mathcal{V}_{jt} := \sum_{i=1}^{N_j}\left(s_{ijt}-\tilde{s}_{ijt}\right)^2$ *is the variance of the difference between sales and cost shares within-sector* j*. Expected aggregate productivity growth is given by:*

$$\begin{aligned}\gamma_t = {} & g - \frac{\sigma^2}{2} + (\varepsilon-1)\frac{\sigma^2}{2}\sum_{j=1}^{N}\omega_{jt}\left(1-\mathcal{H}_{jt}\right) \\ & + (\varepsilon-1)^2\frac{\sigma^2}{2}\sum_{j=1}^{N}\widetilde{\omega}_{jt}\left(\widetilde{\mathcal{H}}_{jt}-\mathcal{H}_{jt}\right) - (\varepsilon-1)^2\frac{\sigma^2}{2}\sum_{j=1}^{N}\widetilde{\omega}_{jt}(1-\widetilde{\omega}_{jt})\mathcal{V}_{jt}.\end{aligned} \tag{20}$$

The proof is in Appendix A.5. The aggregate markup is the sectoral cost-share weighted average of sectoral markups. Thus, aggregate markup growth depends on the sectoral cost-share weighted average of the difference between cost and sales HHIs. The additional term captures the dependence of sectoral cost shares on sectoral markups and is large and negative when the gap between sales and cost shares is large.

For aggregate productivity growth, the first line of the expected growth expression mirrors the efficient economy case in Corollary 2, with a granular drag that scales with the sales-weighted average of sectoral sales HHIs. The second line captures the impact of misallocation on aggregate productivity growth. The first term is similar to the sectoral case in Proposition 3, capturing how the joint distribution of productivity and markups across firms and sectors affects growth. The second term is always negative, capturing an additional drag on growth from the variance of the difference between sales and cost shares within sectors.

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

fan out over time.[9] 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):

$$0 \;=\; -\mu_x(\eta)\,\phi'(x) \;+\; \frac{\sigma^2}{2}\,\phi''(x) \;+\; \lambda\,\mathbb{E}[\phi(x-J)-\phi(x)] \;-\; \delta\,\phi(x), \qquad x \in \mathbb{R}\backslash\{x_e\}, \tag{21}$$

with an inflow of mass at $x_e := \ln A_e$ at rate ν. A standard implication of (21) 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 (21) with jumps. Guessing $\phi(x) \propto e^{-\alpha x}$ away from x_e and substituting into (21) yields a mapping between the traveling-wave speed η and the tail index α

$$\eta \;=\; g \;+\; \frac{\alpha-1}{2}\,\sigma^2 \;+\; \lambda\,\frac{\mathbb{E}\left[e^{\alpha J}\right]-1}{\alpha} \;-\; \frac{\delta}{\alpha}. \tag{22}$$

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.[10] 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

[9] 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.

[10] 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 $\mathcal{H}_{jt}$ 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:

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

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.

Taking Stock In summary, the model predicts that micro-level reallocation gains are limited by granularity. As concentration increases, the ability to reallocate resources toward the most productive firms diminishes, dragging down expected sectoral productivity growth. This granular drag on growth is present both in efficient and misallocated economies, and it extends naturally to aggregate productivity growth. I test these predictions empirically in the next section.

4 Data and Estimation

In this section, I describe the data sources and present reduced-form evidence on the link between granularity and sectoral productivity growth. Building on the previous section's theory, firm-level idiosyncratic shocks aggregate into sectoral dynamics: under efficient allocation, greater concentration attenuates reallocation gains and lowers expected productivity growth. The relevant concentration metric is the sales-based Herfindahl-Hirschman Index (HHI); when markups are

dispersed, the relevant object is the gap between sales- and cost-based HHIs. I test these predictions using Swedish firm-level data, complemented by industry-level evidence from CompNet and the U.S. Census. This section concludes with the calibration strategy used to discipline the model for the quantitative analysis in section 5.

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.[11] The exact construction of the final sample is detailed in Appendix B.3.

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.1.

CompNet CompNet is 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 in a broader cross-country context. Results are reported in Appendix B.2.

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 in Appendix B, including specifications that control for future concentration. 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 I find a negative relationship between concentration and productivity growth consistent with the model.

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

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 carry over 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-by-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 estimate two regression specifications corresponding to the efficient and inefficient economy cases described in the theory, as outlined in equations (23) and (24). First, I regress one-year ahead labor productivity growth on the Herfindahl-Hirschman Index (HHI) of sales shares. In an efficient economy with constant markups, this is the relevant concentration measure, and the theory predicts a negative coefficient $\beta < 0$. Second, I regress productivity growth on both sales and cost share HHIs. When markups differ across firms, the relevant concentration measure is the gap between sales and cost HHIs, which I include as an additional regressor in equation (24). The theory predicts that both coefficients will be negative and that the coefficient on the gap will be larger in magnitude $\beta_2 < \beta_1 < 0$.

$$\ln A_{jt+\Delta t} - \ln A_{jt} = \alpha + \beta \mathcal{H}_{jt} + \beta_2 \ln A_{jt} + \epsilon_{jt+\Delta t} \tag{23}$$

$$\ln A_{jt+\Delta t} - \ln A_{jt} = \alpha + \beta_1 \mathcal{H}_{jt} + \beta_2 \left(\mathcal{H}_{jt} - \widetilde{\mathcal{H}}_{jt} \right) + \beta_3 \ln A_{jt} + \epsilon_{jt+\Delta t} \tag{24}$$

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 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×Year FE	x	x	x	x
Observations	7218	7218	7218	7218
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.

I run similar regressions for U.S. and CompNet data in Appendix B.1 and B.2, respectively, finding results consistent with those reported here. For the U.S., I find that whenever an industry is more concentrated in sales than its long-run average, it experiences lower total factor productivity growth in the following five years. This finding is robust to controlling for future concentration, suggesting that the results are not driven by industry fixed effects. For CompNet, I find that the gap between sales and cost HHIs negatively predicts five-year ahead productivity growth, while sales HHI alone shows no systematic effect. Quantitatively, a one-percentage-point increase in the difference between the two measures predicts a reduction in 5-year productivity growth of about 0.5 percentage points, consistent with a granular drag operating through misallocation.

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 nonparametric relationship between concentration and productivity growth, with the linear fit from columns (2) and (4) of Table 1 overlaid.[12] Panel (A) shows the relationship between sales HHI and five-year

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

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, a negative linear relationship is clearly visible, supporting the model predictions that granularity affects sectoral dynamics through misallocation when markups differ across firms.

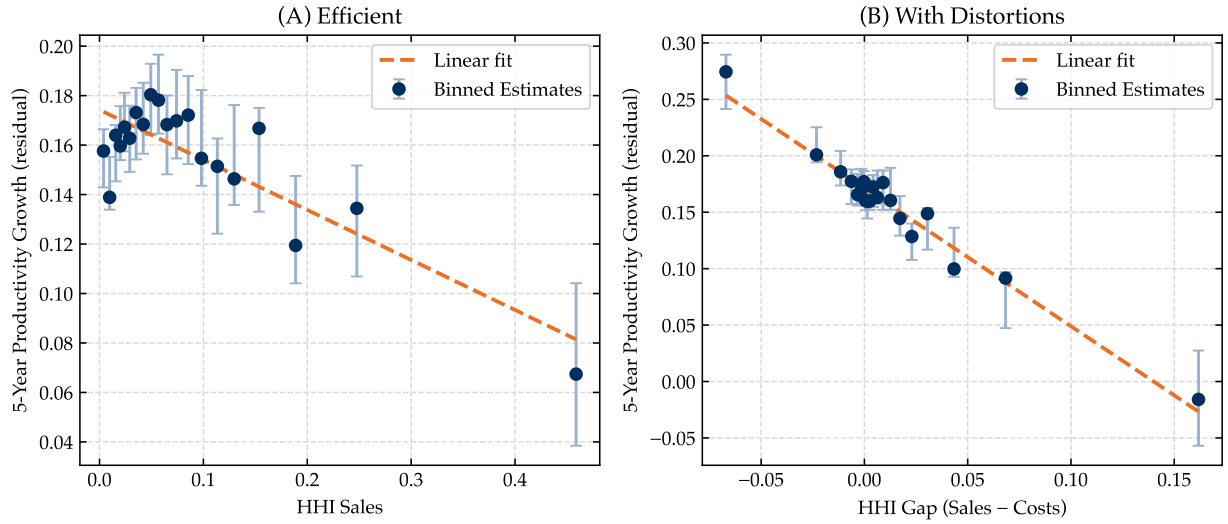

Figure 1: Binscatter of productivity growth on concentration measures, binned by ventiles for a total of 7,218 industry-year pairs, using the method in Cattaneo et al. (2024). Panel (A) shows sales HHI, further controlling for current productivity and fixed effects as in column (2) of Table 1. Panel (B) shows the relationship between productivity growth and the difference between sales and cost HHIs, controlling for current productivity, sales HHI, and fixed effects as in column (4) of Table 1.

Since my measure of productivity is labor productivity rather than TFP, I cannot rule out that part or all of the estimated correlation is driven by changes in markups rather than productivity. This, however, would not invalidate the mechanism proposed in the model. As shown in equation (18), the expected growth rate of sectoral markups is also decreasing in the gap between sales and cost HHIs. Thus, even if the estimated coefficients are driven by changes in markups rather than efficiency, the model fits the data well.

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.[13] If, for example,

[13]See Berry et al. (2019) for a comprehensive discussion of the role of market definition in empirical industrial organization.

the relevant market is narrower, then the estimated coefficients will be larger in magnitude than the true effect, as I show later when comparing data to model results. Furthermore, the regression specification is likely misspecified. For example, firm dynamics might themselves depend on concentration. If large firms 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.[14]

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 (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:[15]

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

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

[14] 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.

[15] An empirical regularity observed in firm growth rates is that the unconditional distribution of firm growth rates follows a double-exponential (Laplace) distribution; see Stanley et al. (1996).

for asymmetry in the jumps allows matching skewness, while kurtosis is controlled by the jump intensity λ.[16] Finally, the model includes an exogenous exit rate δ and a parameter η that governs the speed of the firm size distribution's traveling 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 $\frac{\text{P97.5}-\text{P2.5}}{\text{P75}-\text{P25}} - 2.91$.[17] 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). The elasticity of substitution ε is fixed at 5 in the baseline, standard in the literature, and consistent with micro-level estimates from Boppart et al. (2023) who estimate within-industry elasticities around 4.5 for manufacturing and 5.5 for service industries.

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.

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

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 to the jump component, indicating that large shocks play a significant role in firm

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

[17] 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.

productivity dynamics. The tail index α_{tail} is estimated at 3.96, implying Zipf's law for sales shares within industries.[18]

5 Quantitative Results

This section evaluates the model's quantitative performance at all levels of aggregation. Starting at the micro level, I first assess how well the calibrated model matches the size-dependent features of the firm growth distribution. Second, I examine whether the model can replicate the empirical relationship between changes in concentration and future productivity growth at the industry/sector level, and describe the dynamic impact of an idiosyncratic concentration shock on sectoral productivity growth. Finally, I show that the granular drag has implications for aggregate productivity growth in the medium to long run.

5.1 Firm Growth Distribution by Firm Size

How and why the firm growth distribution varies with size has been a long-standing puzzle in the literature. Two empirical regularities stand out. First, the mean growth rate is roughly constant for medium to large firms, while small firms tend to grow faster on average. Second, the volatility of growth rates declines with size. These patterns are puzzling because they are difficult to reconcile with the empirical regularity that the firm size distribution exhibits a Pareto tail. For a stochastic process of proportional growth to have a Pareto tail, the ratio of the mean to the variance of growth rates must be asymptotically constant with size (Gabaix, 2009). Firm granularity provides a simple answer to this puzzle, and can account for additional features of the growth distribution, such as its skewness.

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 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. In the data, small firms exit more frequently, which mechanically raises the average growth rate of small surviving firms. However, for medium to large firms, the exit hazard is low and approximately constant with size, such that model and data are directly comparable.[19] The right panel shows the volatility profile,

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

[19] See appendix B.4 for evidence on the exit hazard.

which declines with size in both the data and the model. In the latter, small firms are exposed to their own idiosyncratic shocks as well as shocks to large firms, which amplifies their growth volatility. As firms grow larger, they saturate their market of operation and are constrained by the lower elasticity of substitution across sectors, such that identical idiosyncratic shocks translate into smaller sales growth fluctuations. The model matches the level and slope of the volatility profile well, showing that granularity can account for this important empirical regularity.

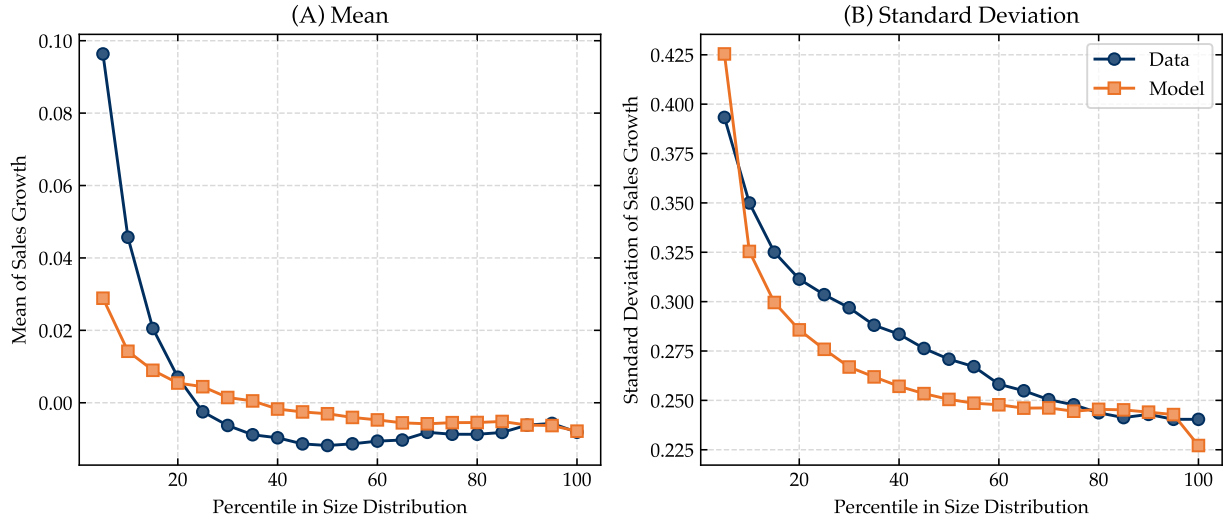

Figure 2: Mean and standard deviation of sales growth across size bins (data and model).

The mechanism behind the decline in volatility can also account for the size-dependent skewness of firm growth rates. 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. Since the standard skewness measure is sensitive to outliers, I also consider an outlier-robust measure, the Kelly-skewness, defined as $(P_{90} + P_{10} - 2P_{50})/(P_{90} - P_{10})$. 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.

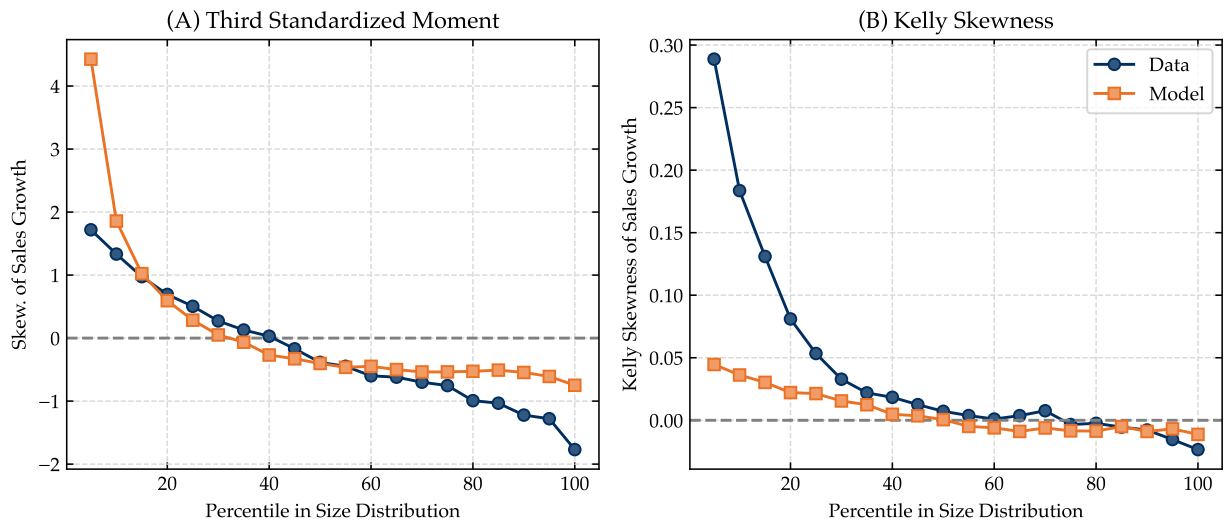

Figure 3: Skewness and Kelly skewness of sales growth across size bins (data and model).

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.[20] 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 equivalent Kelly-skewness measures, left tail, and right tail in the model are shown in panels (D), (E), and (F) of figure 4, respectively. The model matches the size-dependent patterns of skewness well. The right tail drive the decline in skewness with size, as small firms benefit from large positive growth opportunities when dominant firms contract, whereas large firms have less room to grow as they saturate their market. A notable difference is that the model generates a longer right tail for small firms than in the data. The explanation is that in the model, the exit hazard is constant with size, whereas in the data, small firms exit more frequently. In the model, some small firms benefit from waiting until the large firms contract to capture a larger market share, leading to a longer right tail.

Overall, granularity provides a natural explanation for these size-dependent patterns in the growth distribution. Large firms have less room to grow within their sector, leading to a lower volatility and skewness of growth rates, even when firm productivity follows a random walk. These findings further reinforce the validity of assuming idiosyncratic random growth processes for firm productivity, which serve as the foundation for the theoretical predictions for sectoral dynamics which I examine next.

[20] Formally, the Kelly-skewness with tail threshold τ is defined as $(P_{1-\tau} + P_{\tau} - 2P_{50})/(P_{1-\tau} - P_{\tau})$, where P_x is the x-th percentile of the distribution.

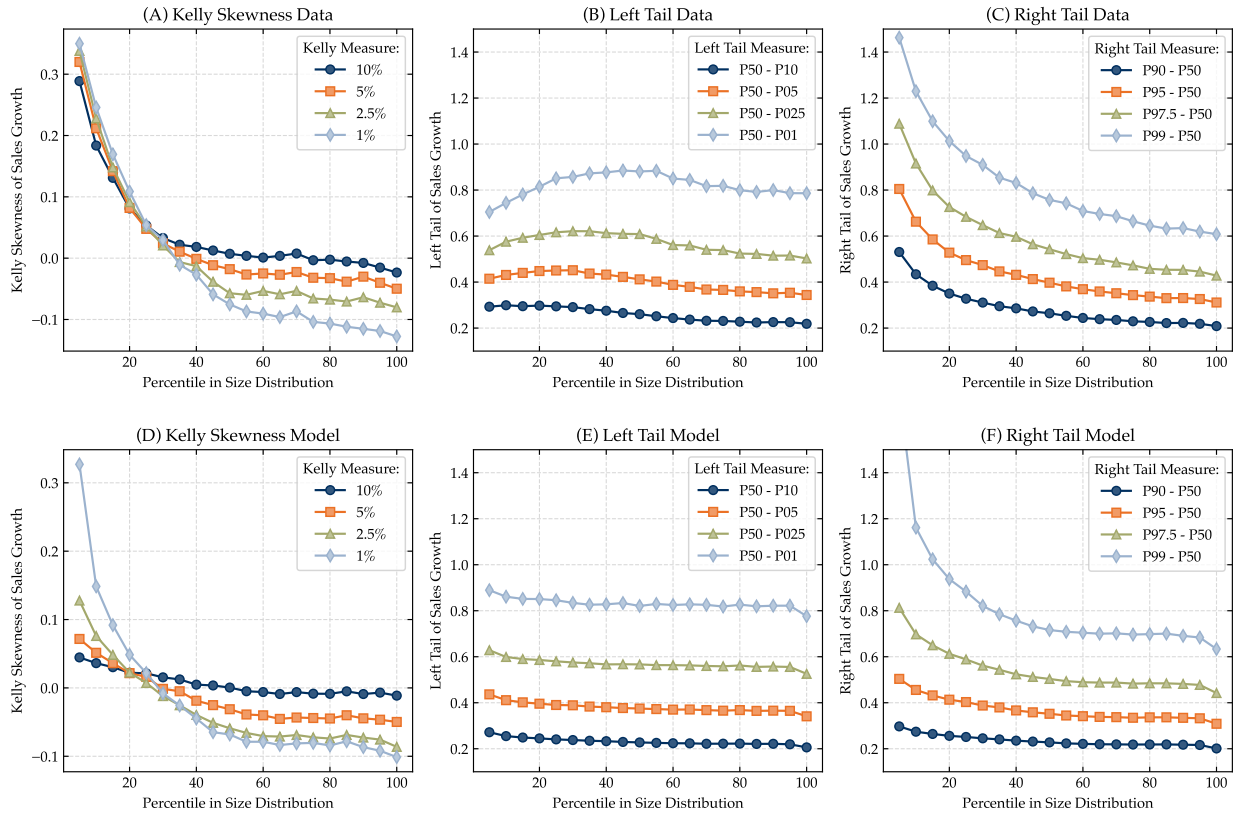

Figure 4: Kelly skewness across size bins (data). Notes: Each line corresponds to a tail definition (10%, 5%, 2.5%, 1%).

5.2 The Granular Drag at the Industry Level

As shown in section 4.2, there is strong empirical evidence that increases in concentration and markup dispersion lead to lower future productivity growth at the industry/sector level. I now assess whether the calibrated model can replicate this empirical relationship. To do so, I simulate a stationary economy with a large number of sectors and estimate industry-level regressions. To allow for distortions, I follow Restuccia and Rogerson (2008) and allow for i.i.d. taxes and subsidies on firm sales of $\pm 20\%$. Note that these distortions are not calibrated to the data, but rather chosen to generate variation in sales- and cost-based firm shares. Table 3 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 3: Sector-Industry Productivity Growth and Granular Drag

	Data		Model	
	$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$		$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$	
	(1)	(2)	(3)	(4)
HHI_t sales	-0.215**	-0.013	-0.059***	-0.060***
	(0.062)	(0.065)	(0.006)	(0.005)
$\ln(\text{Prod}_t)$	-0.106***	-0.070***	-0.066***	-0.068***
	(0.017)	(0.016)	(0.008)	(0.007)
HHI_t sales $-$ HHI_t costs		-1.164***		-0.215***
		(0.198)		(0.026)
Observations	7218	7218	200000	200000
R^2	0.318	0.352	0.064	0.068
R^2 Within	0.062	0.109	-	-

In data: SEs clustered by 5-digit industry and year, 2-digit times year FE.

It is worth noting that the empirical estimates may be upward biased due to the definition of industries in the data. Suppose that a "mega industry" in the data contains K underlying true "micro industries" (CES nests) each with the same true concentration H. The measured mega-industry concentration will then be $H^{\text{mega}} = H \cdot h$, where $h = \sum_{k=1}^{K} (s_k^{\text{micro}})^2$ captures a form of "sub-industrial HHI" based on the relative shares s_k^{micro} of each micro industry within the mega industry. In this case, the empirical regression coefficient satisfies $\beta^{\text{empirical}} = \beta^{\text{model}}/h$, implying that the empirical estimate is upward biased by a factor $1/h$. For example, if the mega industry consists of K equal-sized sub-industries, then $h = 1/K$ and the empirical coefficient is K times larger than the theoretical one. The evidence is consistent with $K \approx 5$.

The model generates a quantitatively significant granular drag in the sectoral cross-section. However, all theoretical derivations have been over an infinitesimal time horizon. It could be that concentration and sectoral productivity respond immediately to idiosyncratic shocks, such that the drag is only relevant at very short horizons. To assess the quantitative importance of the granular drag over longer horizons, I next examine the transitional dynamics of productivity growth following a concentration shock. Since the only shocks in the model are idiosyncratic firm-level productivity shocks, concentration shocks arise endogenously from the aggregation of these shocks. To trace the impulse response of sectoral productivity to a change in concentration,

I use local projections (Jordà, 2005). Specifically, I estimate the following equation for 5-year horizons $h = 0, 5, 10, \ldots 140$:

$$\Delta_h \ln A_{j,t} \;=\; \beta_h \, \Delta_h \mathcal{H}_{j,t} + \theta_h \Delta_h \left(\mathcal{H}_{j,t} - \widetilde{\mathcal{H}}_{j,t} \right) + \beta_A \, \ln A_{j,t-1} \;+\; \alpha_j \;+\; \tau_t \;+\; \epsilon_{j,t+h}, \tag{25}$$

where $\Delta_h X_{j,t} = X_{j,t+h} - X_{j,t}$ denotes the h-step-ahead change in variable X, and α_j and τ_t are sector and time fixed effects, respectively. The coefficient β_h captures the impulse response of the h-period change in log productivity to a one-unit change in concentration, while θ_h captures the effect of a change in the gap between sales- and cost-based HHIs.

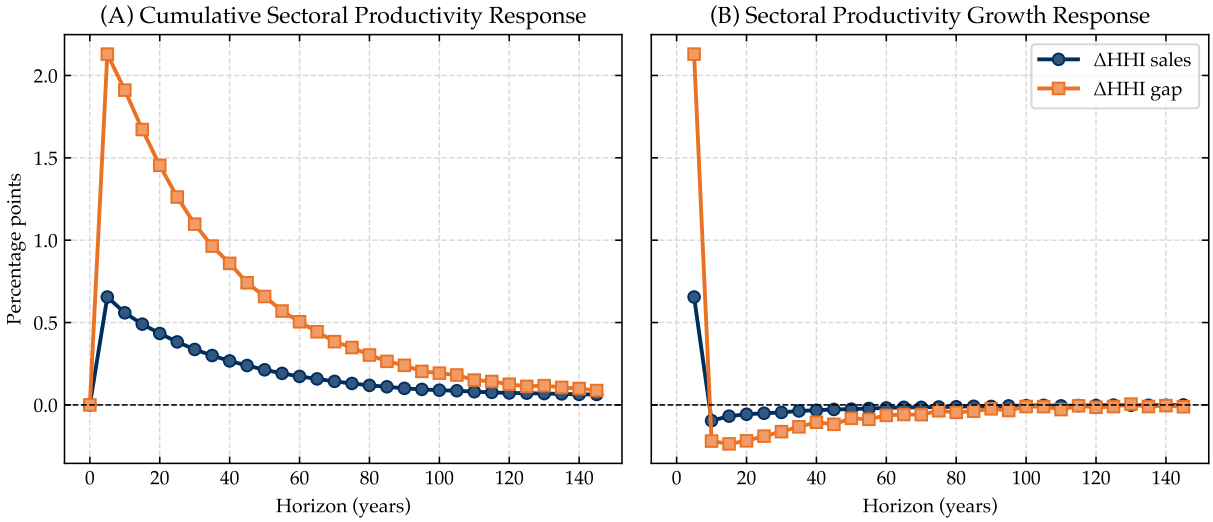

Figure 5: Local-projection IRFs to one-percentage-point increases in sales HHI (blue) and HHI gap (orange): cumulative (A), 5-year growth (B).

Panel (A) of figure 5 shows the sectoral productivity response to a one-percentage-point increase in concentration. The blue line with circles corresponds to a one-percentage-point change in sales-based HHI, holding the gap between sales- and cost-based HHIs constant, while the orange line with squares isolates the effect of the HHI gap between sales- and cost-shares. A one-percentage-point increase in sales concentration raises cumulative productivity by roughly 0.7 percentage points on impact, whereas a similar increase in the HHI gap generates an immediate gain of about 2.2 percentage points. Both effects gradually decay over time: after around 50 years, sectoral productivity is still about 0.2 and 0.5 percentage points higher, respectively. Panel (B) reports the corresponding 5-year growth responses. The initial burst in productivity growth, most pronounced for the HHI gap, reflects short-run reallocation gains as resources shift toward firms

operating below their socially optimal scale. Over time, however, higher concentration dampens reallocation from idiosyncratic shocks, leading to a persistent slowdown in growth.

5.3 Aggregates and Persistence

Finally, I assess the implications of the granular drag for aggregate productivity growth. I assume that there are 600 sectors in the economy, each with a Poisson number of firms with mean 140. According to equation (16), aggregate productivity growth can be decomposed into a weighted sum of sectoral productivity growth rates, where the weights are given by the sectoral sales shares:

$$\gamma_t = g - \frac{\sigma^2}{2} + (\varepsilon - 1)\frac{\sigma^2}{2}\left(1 - \sum_{j=1}^{N} \omega_j \mathcal{H}_{jt}\right).$$

Thus, aggregate productivity growth inherits the drag from the sectoral level, and the relevant measure of granularity is the sales-weighted average sectoral HHI, $\sum_{j=1}^{N} \omega_j \mathcal{H}_{jt}$. I choose a conservative calibration with $\omega_j = 1/N$ for all sectors.[21] Panel (A) of figure 6 plots binned scatter plots of annualized 10-year aggregate productivity growth against the current sales-weighted aggregate HHI. There is a clear linear relationship, with a 5 percentage point increase in the sales-weighted aggregate HHI leading to a decline in 10-year productivity growth of about 1.1 percentage points. Panel (B) illustrates the decay of this effect over different horizons. It plots the annualized absolute value of the regression coefficient of aggregate productivity growth on the sales-weighted aggregate HHI for horizons from 1 to 30 years. For example, the point at horizon 10 shows the same absolute value as in panel (A). The effect decays slowly over time, with a 5 percentage point increase in the sales-weighted aggregate HHI reducing 30-year productivity growth by about 1.65 percentage points.

Panel (B) also plots the corresponding effect when controlling for the current level of aggregate productivity. Since granularity is the only source of sector heterogeneity in the model, more concentrated sectors are also more productive on average. Thus, controlling for current productivity removes part of the variation in concentration that drives future growth. Nevertheless, the granular drag remains quantitatively significant even after controlling for current productivity, with a 5 percentage point increase in the sales-weighted aggregate HHI reducing 10-year productivity

[21] In practice, the distribution of sectoral sales shares is substantially more skewed, and has a statistically insignificant correlation with sectoral HHIs. The calibration here likely understates the dispersion of sales-weighted aggregate HHIs and the quantitative importance of the granular drag at the aggregate level.

growth by about 0.5 percentage points.

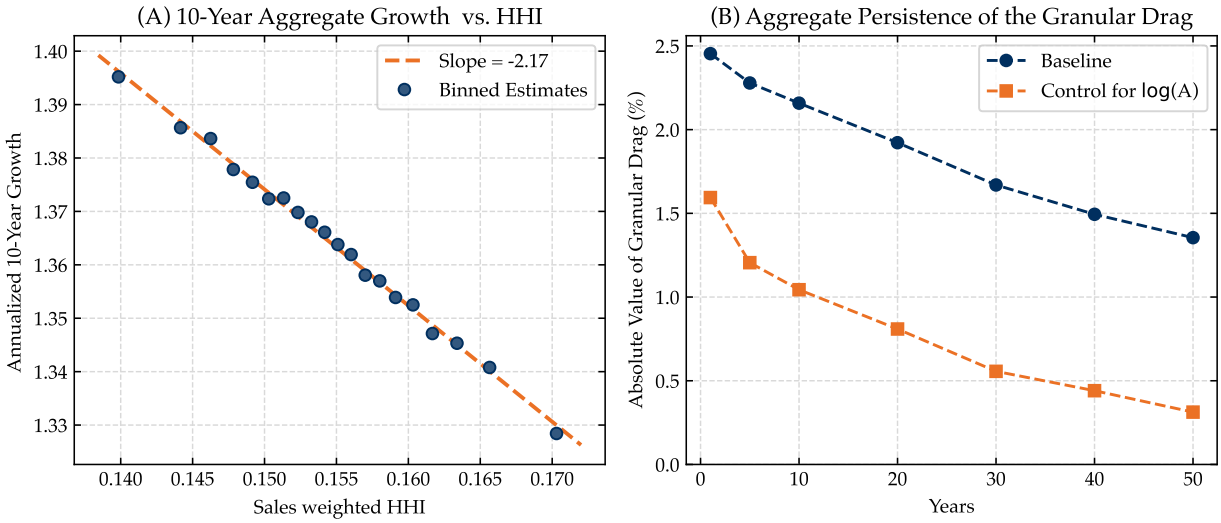

Figure 6: Model simulated long-run effects of aggregate concentration on productivity growth. (A) Binscatter of annualized 10-year aggregate productivity growth vs. sales-weighted aggregate HHI. (B) Absolute value of the annualized effect of sales-weighted aggregate HHI on aggregate productivity growth across different horizons (1-50 years).

The analysis in this subsection focuses on the aggregate implications of the granular drag under an efficient allocation of resources. As shown in the preceding theoretical results, the impact of granularity on productivity growth is amplified in the presence of misallocation. Incorporating these effects quantitatively is a natural next step, which I will address in a subsequent version of the paper.

6 Conclusion

This paper has developed a unified framework linking firm granularity to productivity growth. Embedding idiosyncratic productivity shocks in a multi-sector model with finitely many firms, I show that market concentration shapes expected productivity growth. When firms hold non-negligible market shares, their shocks do not average out, and the reallocation of resources across producers becomes imperfect. As a result, higher concentration mechanically reduces the gains from micro-level reallocation, generating a *granular drag* on productivity growth.

At the firm level, granularity generates size-dependent patterns of growth: large firms exhibit lower volatility and left-skewed growth, while smaller firms display higher volatility and right-

skewed growth as they benefit from reallocations when dominant firms contract. At the sector level, concentration hampers reallocation gains from idiosyncratic shocks. Distortions in resource allocation further amplify this mechanism when the largest firms command disproportionately high markups

Consistent with these predictions, I find strong support for these mechanisms using firm- and industry-level data from Sweden, the United States, and Europe. Industries that become more concentrated, or where sales and cost concentration diverge, subsequently experience slower productivity growth. In the quantified version of the model, a ten-percentage-point rise in concentration reduces five-year productivity growth by roughly 0.6 percentage points in the efficient benchmark and by about 2 percentage points in the presence of distortions.

More broadly, the results suggest that micro-reallocation plays a central role in shaping how economies grow. By linking firm granularity to expected productivity growth, this paper highlights a structural channel through which market concentration influences long-run performance. The framework provides a foundation for future work exploring how entry, policy distortions, or endogenous innovation decisions interact with granular dynamics to shape productivity and growth at both the sectoral and aggregate levels.

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.

Baqaee, David Rezza and Emmanuel Farhi, "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem," *Econometrica*, 2019, *87* (4), 1155–1203.

_ and _ , "Productivity and Misallocation in General Equilibrium*," *The Quarterly Journal of Economics*, 09 2019, *135* (1), 105–163.

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.

Boppart, Timo, Mikael Carlsson, Markus Kondziella, and Markus Peters, "Micro PPI-Based Real Output Forensics," May 2023. Manuscript.

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.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda, "Who Creates Jobs? Small versus Large versus Young," *The Review of Economics and Statistics*, 05 2013, *95* (2), 347–361.

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.

Hulten, Charles R., "Growth Accounting with Intermediate Inputs," *The Review of Economic Studies*, 10 1978, *45* (3), 511–518.

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.

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.

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

Òscar Jordà, "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, March 2005, *95* (1), 161–182.

A Model Appendix

A.1 Itô's Lemma

Throughout the paper, I frequently use Itô's lemma to derive the stochastic differential equations (SDEs) governing firm, sector, and economy-wide productivity dynamics. Intuitively, Itô's lemma is the stochastic analogue of the chain rule. When a variable evolves randomly according to a diffusion or jump–diffusion process, its changes depend not only on the instantaneous drift and volatility of the underlying process, but also on how randomness propagates through nonlinear transformations of that variable. For example, if firm productivity follows a geometric Brownian motion, then the growth rate of its logarithm must correct for curvature (the $-\frac{1}{2}\sigma^2$ term) because expectations and nonlinear transformations do not commute. Itô's lemma formalizes this correction for general transformations. In this paper, it allows us to map micro-level stochastic processes for firms into consistent laws of motion for aggregates such as sectoral or economy-wide productivity indices.

Itô's Lemma (with Jumps) Let X_t follow

$$\frac{dX_t}{X_{t^-}} = \mu_t\, dt + \sigma_t\, dW_t + \left(e^{J_t} - 1\right) dQ_t,$$

where W_t is a Wiener process, Q_t is a Poisson process with intensity λ, and J_t is the (possibly random) jump size. For any $f \in C^{2,1}$ (continous, twice differentiable in X and once in t), Itô's lemma states that

$$\begin{aligned} df(X_t, t) = &\left(\partial_t f + \mu_t X_t\, \partial_X f + \tfrac{1}{2}\sigma_t^2 X_t^2\, \partial_{XX} f\right) dt + \sigma_t X_t\, \partial_X f\, dW_t \\ &+ \left[f(X_{t^-} e^{J_t}, t) - f(X_{t^-}, t)\right] dQ_t. \end{aligned} \tag{A.1}$$

For example, if $f(X_t, t) = \ln X_t$ this yields:

$$d \ln X_t = \left(\mu_t - \tfrac{1}{2}\sigma_t^2\right) dt + \sigma_t\, dW_t + J_t\, dQ_t$$

The traditional Itô's lemma is without jumps and can be recovered by setting the jump intensity λ to zero.

A.2 Proofs and Derivations for The Granular Drag in Efficient Economies

In this section I illustrate the main derivations of SDEs in the efficient allocation case. For notational convenience, I drop the subscript j in the following. The SDE for the productivity of firm i in sector j is given by:

$$\frac{dA_{ijt}}{A_{ijt}} = gdt + \sigma dW_{ijt} + \left(e^{J_{ijt}} - 1\right) dQ_{ijt}$$

where g is the drift, σ is the diffusion, W_{ijt} is a Wiener process, J_{ijt} is the jump size, and Q_{ijt} is a Poisson process with intensity λ. It will be useful to derive two related SDEs:

$$\frac{dA_{ijt}^{\varepsilon-1}}{A_{ijt}^{\varepsilon-1}} = (\varepsilon-1)\left(g + \frac{(\varepsilon-2)\sigma^2}{2}\right)dt + (\varepsilon-1)\sigma dW_{ijt} + \left(e^{(\varepsilon-1)J_{ijt}} - 1\right) dQ_{ijt}$$

$$d\ln A_{ijt} = \left(g - \frac{\sigma^2}{2}\right)dt + \sigma dW_{ijt} + J_{ijt} dQ_{ijt}$$

To derive the SDE for $A_{jt} = \left(\sum_{i=1}^{N_j} A_{ijt}^{\varepsilon-1}\right)^{\frac{1}{\varepsilon-1}}$, we first derive $dA_{jt}^{\varepsilon-1} = \sum_{i=1}^{N_j} dA_{ijt}^{\varepsilon-1}$. We have:

$$\frac{dA_{jt}^{\varepsilon-1}}{A_{jt}^{\varepsilon-1}} = (\varepsilon-1)\left(g + \frac{(\varepsilon-2)\sigma^2}{2}\right)dt + (\varepsilon-1)\sigma\sum_{i=1}^{N} s_{it} dW_{ijt} + \sum_{i=1}^{N} s_{it}\left(e^{(\varepsilon-1)J_{ijt}} - 1\right) dQ_{ijt}$$

where $s_{ijt} = A_{ijt}^{\varepsilon-1} / \left(\sum_{k=1}^{N_j} A_{kt}^{\varepsilon-1}\right)$. It will be useful to note that $\sum_{i=1}^{N} s_{ijt} dW_{ijt} \overset{d}{=} \sqrt{\mathcal{H}_{jt}} dW_{jt}$, where $\mathcal{H}_{jt} = \sum_{i=1}^{N} s_{it}^2$ is the Herfindahl index. That is, sector level volatility is proportional to the square-root of the sector sales-HHI, as in Gabaix (2011). Note that this is an artifact of ignoring sector level shocks, which would induce additional sector level volatility, but are not relevant for the main results of the paper.

By applying Itô's lemma, we can derive the SDE for A_t:

$$\frac{dA_{jt}}{A_{jt}} = \left(g + \frac{(\varepsilon-2)}{2}\sigma^2\left(1 - \mathcal{H}_{jt}\right)\right)dt + \sigma\sqrt{\mathcal{H}_{jt}} dW_{jt} + \sum_{i=1}^{N_j}\left[\left(1 + s_{ijt}\left(e^{(\varepsilon-1)J_{ijt}} - 1\right)\right)^{\frac{1}{\varepsilon-1}} - 1\right] dQ_{ijt}$$

using Ito's lemma, we get

$$d\ln A_{jt} = \left(g + \frac{\varepsilon-2}{2}\sigma^2\left(1-\mathcal{H}_{jt}\right) - \frac{\sigma^2}{2}\mathcal{H}_{jt}\right)dt + \sigma\sqrt{\mathcal{H}_{jt}}dW_{jt} + \frac{1}{\varepsilon-1}\sum_{i=1}^{N_j}\ln\left(1+s_{ijt}\left(e^{(\varepsilon-1)J_{ijt}}-1\right)\right)dQ_{ijt}$$

We are now ready to prove Proposition 1, and the case with jumps.

Proof of Proposition 1 and/or jumps. The focus of the paper is on expected productivity growth $\gamma_{jt} := \mathbb{E}_t[\frac{1}{dt}d\ln A_{jt}]$. We have $\mathbb{E}_t\left[dW_{jt}\right] = 0$ and $\mathbb{E}_t\left[dQ_{ijt}\right] = \lambda dt$, so taking expectations yields:

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

To recover the case without jumps, we set $\lambda = 0$, which yields equation (13). ■

A.2.1 Concentration and Growth over Finite Horizons

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

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

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.[22] It implies higher values for standard

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

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.

Proposition 5. *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 $\vec{s}_{jt} > \vec{s}_{kt}$, then*

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

A.3 Markups à 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:

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

Here s_{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

textbook treatment.

sectoral price index P_j.

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

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.

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

A.4 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, $dW_{\mu_{ij}} \, dW_{ij} = \rho_{A_{ij}\mu_{kj}} \, dt$.

$$\gamma_{jt} = \underbrace{g - \frac{\sigma^2}{2}}_{\text{Mean Productivity Change}} + \underbrace{\frac{(\varepsilon-1)}{2}\sigma^2\Big(1 - \mathcal{H}_{jt} + (\varepsilon-1)\big(\mathcal{H}^{\kappa}_{jt} - \mathcal{H}_{jt}\big)\Big)}_{\text{Reallocation due to technology}}$$

$$+ \underbrace{\varepsilon \sum_{i=1}^{N_j} (\kappa_{ij} - s_{ij})\, \mathbb{E}_t\left[\frac{1}{dt} d \ln \mu_{ij}\right]}_{\text{Mean Markup Change}}$$

$$+ \underbrace{\frac{1}{2}\left\{\varepsilon(\varepsilon-1)\left[\sum_i s_{ij}\, \sigma^2_{\mu_{ij}} - \sum_{i,k} s_{ij} s_{kj}\, \sigma_{\mu_{ij}} \sigma_{\mu_{kj}} \rho_{\mu_{ij}\mu_{kj}}\right] - \varepsilon^2\left[\sum_i \kappa_{ij}\, \sigma^2_{\mu_{ij}} - \sum_{i,k} \kappa_{ij}\kappa_{kj}\, \sigma_{\mu_{ij}} \sigma_{\mu_{kj}} \rho_{\mu_{ij}\mu_{kj}}\right]\right\}}_{\text{Reallocation due to markup changes (Jensen/variance terms)}}$$

$$+ \underbrace{\varepsilon(\varepsilon-1)\left[\sum_i (\kappa_{ij} - s_{ij})\, \sigma\, \sigma_{\mu_{ij}}\, \rho_{A_{ij}\mu_{ij}} + \sum_{i,k} (s_{ij} s_{kj} - \kappa_{ij}\kappa_{kj})\, \sigma\, \sigma_{\mu_{kj}}\, \rho_{A_{ij}\mu_{kj}}\right]}_{\text{Interaction between technology and markup changes (covariances)}}.$$

■

A.5 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:

$$A_{jt} = \frac{\left(\sum_{i=1}^{N_j} A_{ijt}^{\varepsilon-1}\, \mu_{ij}^{1-\varepsilon}\right)^{\frac{\varepsilon}{\varepsilon-1}}}{\sum_{i=1}^{N_j} A_{ijt}^{\varepsilon-1}\, \mu_{ij}^{-\varepsilon}}. \tag{A.2}$$

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

$$s_{ijt} := \frac{A_{ijt}^{\varepsilon-1} \mu_{ij}^{1-\varepsilon}}{\sum_{k=1}^{N_j} A_{kjt}^{\varepsilon-1} \mu_{kj}^{1-\varepsilon}}, \qquad \widetilde{s}_{ijt} := \frac{A_{ijt}^{\varepsilon-1} \mu_{ij}^{-\varepsilon}}{\sum_{k=1}^{N_j} A_{kjt}^{\varepsilon-1} \mu_{kj}^{-\varepsilon}} = \frac{L_{ijt}}{\sum_{k=1}^{N_j} L_{kjt}},$$

and let $\mathcal{H}_{jt} := \sum_i s_{ijt}^2$ and $\widetilde{\mathcal{H}}_{jt} := \sum_i (\widetilde{s}_{ijt})^2$ 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

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

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

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

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 $s_{ijt} := X_{ijt}^{(N)}/N_{jt}$ and $\widetilde{s}_{ijt} := X_{ijt}^{(D)}/D_{jt}$. 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 s_{ijt}) yields

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

The same calculation for D_{jt} (with weights $\widetilde{s}_{ijt}$) gives

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

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

A.6 Sectoral Markup Dynamics

I derive here the SDE for sectoral markups under the diffusion specification in (11) with $\lambda = 0$ (no jumps), and under the assumption that firm-level markups μ_{ij} are heterogenous but constant over time. The sectoral markup index is defined as

$$\mu_{jt} = \left(\sum_{i=1}^{N_j} s_{ijt} \mu_{ij}^{-1} \right)^{-1}$$

It will be convenient to rewrite this as

$$\mu_{jt} = \frac{\sum_{i=1}^{N_j} A_{ijt}^{\varepsilon-1} \mu_{ij}^{1-\varepsilon}}{\sum_{i=1}^{N_j} A_{ijt}^{\varepsilon-1} \mu_{ij}^{-\varepsilon}}.$$

Applying the results from Appendix A.5, we can write the SDE for sectoral markups as

$$d \ln \mu_{jt} = (\varepsilon - 1)^2 \frac{\sigma^2}{2} \left(\widetilde{\mathcal{H}}_{jt} - \mathcal{H}_{jt} \right) dt \; + \; (\varepsilon - 1) \sigma \left[\sum_{i=1}^{N_j} \left(s_{ijt} - \widetilde{s}_{ijt} \right) dW_{ijt} \right].$$

As with sectoral productivity, note that $\sum_{i=1}^{N_j} \left(s_{ijt} - \widetilde{s}_{ijt} \right) dW_{ijt} \overset{d}{=} \sqrt{\mathcal{V}_{jt}} dW_{\mu_{jt}}$, where $\mathcal{V}_{jt} := \sum_{i=1}^{N_j} \left(s_{ijt} - \widetilde{s}_{ijt} \right)^2$ and $W_{\mu_{jt}}$ is a standard Wiener process. Thus, the volatility of sectoral markup changes is $\sigma_{\mu_{jt}} = (\varepsilon - 1) \sigma \sqrt{\mathcal{V}_{jt}}$.

A.7 Aggregate Markup Dynamics

The aggregate markup index is defined as

$$\mu_t = \left(\sum_{j=1}^{N} \beta_j \mu_{jt}^{-1} \right)^{-1}$$

It will be convenient to define $X_j = \beta_j \mu_{jt}^{-1}$, with $d \ln X_j = -d \ln \mu_{jt}$. Define $\widetilde{\beta}_j = X_j / \sum_{k=1}^{N} X_k$, which are the sectoral cost shares. Applying Itô's lemma, we can write

$$\begin{aligned} d \ln(\mu_t^{-1}) &= \sum_{j=1}^{N} \widetilde{\beta}_j \, d \ln X_j + \frac{1}{2} \left(\sum_{j=1}^{N} \widetilde{\beta}_j \sigma_{X_j}^2 - \sum_{j=1}^{N} (\widetilde{\beta}_j)^2 \sigma_{X_j}^2 \right) dt \\ &= (\varepsilon - 1)^2 \frac{\sigma^2}{2} \sum_{j=1}^{N} \widetilde{\beta}_j \left(\mathcal{H}_{jt} - \widetilde{\mathcal{H}}_{jt} \right) dt - (\varepsilon - 1) \sigma \sum_{j=1}^{N} \widetilde{\beta}_j \sqrt{\mathcal{V}_{jt}} dW_{\mu_{jt}} \\ &\quad + \frac{(\varepsilon - 1)^2 \sigma^2}{2} \left(\sum_{j=1}^{N} \widetilde{\beta}_j \mathcal{V}_{jt} - \sum_{j=1}^{N} (\widetilde{\beta}_j)^2 \mathcal{V}_{jt} \right) dt \end{aligned}$$

A.8 Aggregate Productivity Dynamics with Misallocation

Using the results from Appendix A.5 and **??**, we can write the SDE for aggregate productivity under misallocation as

$$d \ln A_t = \sum_{j=1} \beta_j d \ln A_{jt} - \sum_{j=1} \beta_j d \ln \mu_{jt} + d \ln(\mu_t)$$

Substituting the expressions for $d \ln A_{jt}$, $d \ln \mu_{jt}$, and $d \ln(\mu_t)$ gives

$$\begin{aligned} d \ln A_t = &\left(g - \frac{\sigma^2}{2} \right) dt \; + \; (\varepsilon - 1) \frac{\sigma^2}{2} \sum_{j=1}^{N} \beta_j \left(1 - \mathcal{H}_{jt} + (\varepsilon - 1)(\widetilde{\mathcal{H}}_{jt} - \mathcal{H}_{jt}) \right) dt \\ &- \; (\varepsilon - 1)^2 \frac{\sigma^2}{2} \sum_{j=1}^{N} \beta_j \left(\widetilde{\mathcal{H}}_{jt} - \mathcal{H}_{jt} \right) dt \\ &+ \; (\varepsilon - 1)^2 \frac{\sigma^2}{2} \sum_{j=1}^{N} \widetilde{\beta}_j \left(\widetilde{\mathcal{H}}_{jt} - \mathcal{H}_{jt} \right) dt - \frac{(\varepsilon - 1)^2 \sigma^2}{2} \left(\sum_{j=1}^{N} \widetilde{\beta}_j \mathcal{V}_{jt} - \sum_{j=1}^{N} (\widetilde{\beta}_j)^2 \mathcal{V}_{jt} \right) dt \\ &+ \text{Martingale Terms} \end{aligned}$$

Taking expectations and simplifying gives the expected aggregate productivity growth rate under misallocation:

$$\begin{aligned} \gamma_t = &\, g - \frac{\sigma^2}{2} + (\varepsilon - 1) \frac{\sigma^2}{2} \sum_{j=1}^{N} \beta_j \left(1 - \mathcal{H}_{jt} \right) + (\varepsilon - 1)^2 \frac{\sigma^2}{2} \sum_{j=1}^{N} \widetilde{\beta}_j \left(\widetilde{\mathcal{H}}_{jt} - \mathcal{H}_{jt} \right) \\ &- \frac{(\varepsilon - 1)^2 \sigma^2}{2} \left(\sum_{j=1}^{N} \widetilde{\beta}_j \mathcal{V}_{jt} - \sum_{j=1}^{N} (\widetilde{\beta}_j)^2 \mathcal{V}_{jt} \right) \end{aligned}$$

where $\mathcal{H}_{jt} = \sum_i s_{ijt}^2$, $\widetilde{\mathcal{H}}_{jt} = \sum_i (\widetilde{s})_{ijt}^2$, and $\mathcal{V}_{jt} = \sum_i (s_{ijt} - \widetilde{s}_{ijt})^2$.

B Data

B.1 U.S. Data from Ganapati (2021)

I use the U.S. industry-level data from Ganapati (2021) to complement the analysis in Section 4.2. The dataset combines multiple administrative sources to construct consistent industry-level measures of concentration and productivity from 1972 to 2012. The main inputs are the U.S. Census Bureau's Economic Censuses, the NBER–CES Manufacturing Industry Database, and the Bureau of Economic Analysis (BEA) industry accounts, which together cover over 75% of private-sector gross output. Market concentration is measured using the market sales shares of the four largest firms and, 5-factor total factor productivity. The unit of observation is the 6-digit NAICS industry-year.

I regress 5-year productivity growth on current with industry and 2-digit times year fixed effects. To control for misallocation, I use the labor share as an additional regressor. Table B.1 reports the results. When an industry is above its historical average concentration, it experiences lower productivity growth over the next five years. The effect is economically significant: a 1-percentage-point increase in the CR4 index reduces five-year productivity growth by about 0.27 percentage points. A high labor share is associated with higher future productivity growth, consistent with the idea that misallocation dampens growth. The industry fixed effect might mechanically lead to a negative correlation between concentration and growth if concentration growth and productivity growth are positively correlated in the cross-section. To address this concern, I include the lead of concentration as an additional regressor. The negative effect of current concentration on future productivity growth remains significant, suggesting that the results are not driven by mechanical mean reversion.

	$\Delta_5 \ln(\text{TFP})$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$CR4_t$ Sales	-0.268*	-0.181*	0.039*	0.033	-0.337**	-0.254*	-0.190	-0.192
	(0.084)	(0.070)	(0.015)	(0.021)	(0.073)	(0.070)	(0.082)	(0.086)
ln(Labor Share)		0.226*		-0.007		0.221*		-0.006
		(0.071)		(0.017)		(0.073)		(0.016)
$CR4_{t+5}$ Sales					0.250*	0.254**	0.237*	0.235*
					(0.070)	(0.065)	(0.076)	(0.071)
2-Digit Industry×Year FE	x	x	x	x	x	x	x	x
Industry FE	x	x	-	-	x	x	-	-
Observations	2753	2753	2753	2753	2739	2739	2739	2739
R^2	0.411	0.427	0.071	0.071	0.421	0.437	0.085	0.085
R^2 Within	0.011	0.038	0.003	0.004	0.022	0.048	0.016	0.016

Standard errors clustered by industry and year in parentheses

Table B.1: U.S. Industry-Level Regressions of 5-Year Productivity Growth on Concentration from Ganapati (2021)

The TFP data is normalized to 1 in 1972, making cross-sectional comparisons including the level of TFP difficult. These regressions might therefore be affected by division bias if there is both measurement error in TFP and in sales. Click here to go back to Section 4.2.

B.2 CompNet

The CompNet dataset provides harmonized cross-country firm-level information aggregated to the 2-digit NACE industry level for European economies. It covers measures of productivity, concentration, and cost structures for a broad set of EU countries. Following the same specification as for the Swedish and U.S. data, I regress 5-year industry-level productivity growth on current sales and cost HHIs, controlling for initial productivity and country-by-year fixed effects. As shown in Table B.2, the gap in HHI between sales and costs enters negatively and significantly, while the sales HHI alone shows no systematic effect. Quantitatively, a one-percentage-point increase in the difference between the two measures predicts a reduction in 5-year productivity growth of about 0.5 percentage points, consistent with a granular drag operating through misallocation.

	$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$			
	(1)	(2)	(3)	(4)
HHI_t sales	-0.097	0.007	0.134	0.229*
	(0.126)	(0.125)	(0.077)	(0.099)
HHI_t gap (sales - costs)	-0.762***	-0.493**	-0.556***	-0.434**
	(0.180)	(0.139)	(0.107)	(0.136)
$\ln(\text{Prod}_t)$		-0.198***		-0.042*
		(0.026)		(0.015)
Country×Year FE	x	x	x	x
industry	x	x	-	-
Observations	9921	9921	9921	9921
R^2	0.210	0.279	0.142	0.152
R^2 Within	0.015	0.101	0.009	0.022

Industry refers to 2-digit SNI and 2-digit NACE codes for Sweden and CompNet, respectively.

Table B.2: CompNet Industry-Level Regressions of 5-Year Productivity Growth on Concentration

Click here to go back to Section 4.2.

B.3 Swedish Firm Data

B.4 Exit Hazard

In the model, firm exit is driven by an exogenous Poisson process with constant hazard rate δ. To assess the empirical plausibility of this assumption, I estimate the exit hazard as a function of firm size using the Swedish firm-level data. The Serrano database includes information on firm exits and re-registration of previously exited firms, allowing for accurate measurement of exit events. Figure B.1 displays the estimated exit hazard rate by log-sales, controlling for year fixed effects. Controlling for industry fixed effects yields similar results. The exit hazard declines with firm size, consistent with the notion that larger firms are less likely to exit the market. However, for medium and large firms, the exit hazard is relatively flat, supporting the model's assumption of a constant hazard rate for established firms. (Haltiwanger et al., 2013) also document similar patterns in U.S. data, where exit rates decline sharply for small firms but stabilize for larger firms, with the exception that for firms with more than 500 employees, exit rates decline sharply to 1% or lower. Of course, in the Swedish data, very large firms are much rarer than in the U.S. data, so the

flat hazard for large firms might reflect sample size limitations.

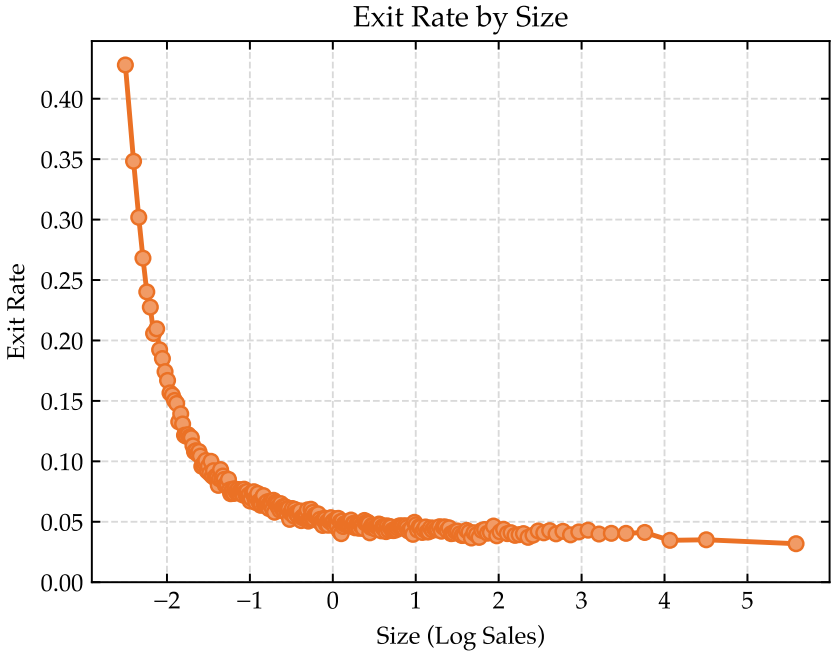

Figure B.1: Exit Hazard by Log-Sales, conditioning on year fixed effects