

The Granular Drag on Growth*

Juan Llavador[†]

Job Market Paper

This version: January 9, 2026

Updated regularly; please click [here](#) for the latest version.

Abstract

This paper uncovers a novel mechanism through which market structure shapes future productivity growth. Using a micro-founded exogenous growth model in which granular firms experience random productivity shocks, I characterize sectoral and economy-wide productivity growth conditional on the current market structure. I test the model's predictions using firm-level data from Sweden, complemented by industry data from the United States and other European economies. In efficient industries, the model predicts that higher sales concentration lowers expected productivity growth due to limited reallocation. In the data, a 10-percentage-point increase in the Herfindahl index of sales concentration is associated with a 3-percentage-point lower productivity growth rate over a five-year period. Moreover, in line with the model's predictions for distorted economies, a similar increase in the gap between the Herfindahl indices of sales and cost shares is linked to a stronger decline of about 13 percentage points. The model generates persistent cross-sectional growth heterogeneity consistent with the empirical evidence, even though all firms follow identically distributed productivity processes. I conclude that the interaction between micro-reallocation and market concentration, a mechanism I term the granular drag, is important for understanding productivity growth across industries and possibly entire economies.

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

[†]Institute for International Economic Studies (IIES), Stockholm University. Email: juan.llavadorperalt@iies.su.se.

1 Introduction

Individual firm productivity growth is highly dispersed. Even within narrowly defined industries, some firms experience rapid improvements while others fall behind. Market economies respond to this dispersion by reallocating production away from less productive firms and toward newly more productive ones. These improvements in allocative efficiency lead to *aggregate* productivity growth, which can therefore exceed average *firm* productivity growth. Empirically, this mechanism is quantitatively important: [Baqae and Farhi \(2020\)](#) estimate that *within-industry* improvements in allocative efficiency accounted for about half of U.S. productivity growth over the period 1997-2015.

Why can aggregate productivity grow faster than average firm productivity growth? A simple way to see this is to consider an industry with only one firm: in that case, industry productivity growth cannot exceed firm productivity growth, because no reallocation is possible. By contrast, in industries with many firms, aggregate productivity rises not only through average firm productivity growth but also through reallocation: as productivity differences emerge, production moves toward improving firms and away from those that fall behind. This contrast motivates the central question of the paper: Does market concentration shape expected productivity growth *independently* of market power and innovation incentives?

This paper answers this question by developing a micro-founded exogenous growth model. It builds on two premises: (i) there is a *finite* number of firms within a *finite* number of sectors, such that firms are *granular*—large enough to influence aggregate outcomes—and (ii) individual firm productivity follows a random growth process, identical across all firms. Using continuous-time tools, I derive expected productivity growth at the sectoral and economy-wide levels, conditional on current market structure. In the baseline model, the Herfindahl indices of concentration for sales- and cost-based firm shares are sufficient statistics for the role of granularity in shaping expected productivity growth. The key insight is that granular concentration lowers the expected gains from reallocation. I refer to this mechanism as the *granular drag* on growth. In an efficient economy, expected sectoral growth is the sum of average firm growth and a positive reallocation term that scales with one minus the Herfindahl index of sales shares:

$$\text{Expected Growth} = \text{Average Firm Growth} + \underbrace{K \times (1 - \text{Herfindahl of Sales Shares})}_{\text{Expected Reallocation}}$$

where K is a positive constant that is *increasing* in the variance of firm-level shocks and the elasticity

of substitution across firms. Recall that the Herfindahl index is defined as the sum of squared market shares, and ranges from $1/N$ in an industry with N equally sized firms to 1 in the extreme case of a monopolist. Thus, as concentration increases, the Herfindahl index rises, reducing the reallocation term and slowing expected productivity growth. When the allocation of resources is distorted, the relevant statistic is the gap between the Herfindahl indices of sales and cost shares.

I empirically test the model's predictions using administrative firm-level data from Sweden, complemented by industry data from the United States and a broad set of European countries. The theory implies two main testable predictions. First, in an efficiently allocated economy, higher concentration should be followed by slower productivity growth. Across Swedish 5-digit industries, a 10-percentage-point rise in the Herfindahl index of sales concentration is associated with about a 3-percentage-point lower growth rate over a five-year period. Second, when the resource allocation is distorted, differences between sales- and cost-based concentration should further magnify this slowdown. A similar increase in the gap between the Herfindahl indices of sales and costs predicts a 13-percentage-point lower growth rate. I further corroborate these findings using industry-level data from the United States and a broad set of European countries. 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.

The calibrated model generates a rich set of patterns across all levels of aggregation. Even though all firms face identical random growth shocks, the model generates firm-level volatility and skewness profiles for sales growth that decline with current firm size, consistent with the data. At the sector level, an increase in concentration leads to a contemporaneous rise in productivity, followed by a prolonged slowdown in growth. 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 even long-term horizons.

An important takeaway from these results concerns how we interpret recent macroeconomic trends through the lens of endogenous growth theory. A large body of this literature has linked rising market concentration to subsequent lackluster productivity growth because dominant firms face weaker incentives to innovate. The findings in this paper suggest that such interpretations should be made with caution. These patterns do not rule out innovation-based explanations, but show that concentration alone can reduce expected growth in an economy in which (i) the

allocation of resources is efficient and (ii) the productivity process of individual firms remains unchanged. In such an economy, the slowdown in productivity growth following an increase in concentration arises endogenously from the granular drag mechanism emphasized in this paper, which operates through reallocation rather than through innovation incentives.

The granular drag is quantitatively relevant for two empirical reasons. First, firm-size distributions exhibit Pareto tails, with a few large firms accounting for a disproportionate share of sectoral output (Axtell, 2001; Chen, 2023). Second, firm-level productivity growth is highly dispersed, with large and persistent differences in performance across firms even within the same narrowly defined industries (Bartelsman and Doms, 2000; Foster et al., 2001). Many models of firm dynamics and growth acknowledge these two facts, but often rely on a continuum of infinitesimal firms for tractability, thus abstracting from the finite nature of granular firms.¹ In such settings, reallocation is maximized, as for every "unlucky" firm, there is a similarly sized "lucky" firm to reallocate to.

With a finite number of firms, however, granularity shapes how effectively production 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 negative shocks to large firms, and 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 by misallocating resources 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 them. These sectoral effects naturally propagate to the aggregate economy: as more sectors become concentrated, expected aggregate growth slows.

Related Literature This paper contributes to three main strands of literature. First, it adds to the literature on the propagation of microeconomic shocks to aggregate outcomes. Propagation of shocks at the micro-level requires discreteness: in a continuum of infinitesimal agents, idiosyncratic shocks average out and have no aggregate effect. In the presence of a large but discrete number of agents, propagation further requires a heavy-tailed size distribution, such that a few large agents carry sufficient weight to affect aggregates. Gabaix (2011) introduces this idea as the *granular* hypothesis, showing that, given the empirically observed firm size distribution, idiosyncratic shocks

¹The 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 accounts for a sizable share of the aggregates. See, for example, Luttmer (2007, 2011); Acemoglu and Cao (2015) for growth models with a continuum of firms.

to large firms can generate aggregate fluctuations of significant magnitude. Subsequent work extends this idea to trade (di Giovanni et al., 2014, 2024; Gaubert and Itskhoki, 2021), propagation through input-output linkages (Acemoglu et al., 2012; Carvalho and Gabaix, 2013; Grassi, 2018), firm dynamics and aggregate volatility (Carvalho and Grassi, 2019), and markup fluctuations (Burstein et al., 2025). My paper differs from these contributions by focusing on how granularity shapes *expected* productivity growth rather than short-term fluctuations. As Carvalho and Grassi (2019) show, for empirically realistic firm size distributions, the distribution of firm sizes is highly persistent, such that the conditional expectations derived in this paper are quantitatively relevant over medium and long-term horizons. The closest related paper is Gaubert and Itskhoki (2021), which examines how firm granularity predicts reversals in comparative advantage and trade flows but does not examine productivity growth. A closely related line of research studies more broadly the aggregation and propagation of microeconomic shocks in general equilibrium. Building on Hulten (1978), Baqaee and Farhi (2019) and Baqaee and Farhi (2020) develop frameworks to trace how micro-level shocks both in efficient and distorted economies aggregate.² This paper is related in that it also studies how idiosyncratic shocks aggregate to shape aggregate outcomes. However, while these papers focus on how realized shocks propagate, I take the probability distribution of shocks as given and study how market structure shapes *expected* productivity growth, conditional on the current allocation of resources.

Second, recent macroeconomic trends have spurred research on market concentration and productivity growth, to which this paper also contributes. Several studies report rising market concentration and markups in the US and other developed countries (Autor et al., 2020; De Loecker et al., 2020; Kwon et al., 2024; Ma et al., 2025). Furthermore, productivity growth has been lackluster, with a burst in the late 1990s that coincided with an increase in concentration, followed by a prolonged slowdown (Aghion et al., 2023). Building on the seminal work of Aghion and Howitt (1992), several studies have linked this increase in concentration to a decline in productivity growth, driven by reduced innovation incentives. For example, in Aghion et al. (2023) and De Ridder (2024), an increase in concentration is contemporaneously associated with a burst in productivity growth, followed by a persistent slowdown, as larger incumbents have weaker incentives to innovate. Several papers in this literature also study the role of large granular firms for innovation incentives and productivity growth (Aghion et al., 2001, 2005; Olmstead-Rumsey, 2019; Weiss, 2020; Akcigit and Ates, 2023; Cavenaile et al., 2025). In these papers, there is dispersion

²See Baqaee and Rubbo (2023) for a comprehensive review of this framework.

in firm productivity growth for three main reasons: (i) innovation intensity varies across firms due to equilibrium forces, (ii) there is randomness in the arrival of innovations, and (iii) because computational constraints induce coarse discretizations of the firm productivity distribution. This paper shows that this dispersion can interact with market structure and lead to dispersion in aggregate productivity growth, beyond innovation incentives. My approach is related to the Real Business Cycle literature (Kydland and Prescott, 1982; Prescott, 1986), which emphasizes that business cycles can be *efficient* and arise endogenously from economic agents' responses to technology shocks. Here, I emphasize how concentration can endogenously reduce expected productivity growth through *efficient* resource reallocation, even when the productivity process of individual firms is identically distributed and unaffected by market structure.

Finally, the paper also 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 (Haltiwanger et al., 2013), but it fails for higher moments. It is well documented that firm-growth volatility decreases slowly with size (Stanley et al., 1996; Sutton, 1997; Yeh, 2025). I further document that firm-growth skewness decreases with size. On the theoretical side, matching these patterns has proven challenging, see Moran et al. (2024) for a recent technical discussion. Standard models, such as Klette and Kortum (2004), imply that volatility should shrink rapidly with size. I show that granularity naturally generates size-dependent volatility and skewness profiles consistent with the data. 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) highlights how long-term contracting frictions in buyer-supplier networks can give rise to persistent deviations from Gibrat's law.

Outline The remainder of the paper is organized as follows. Section 2 presents the static equilibrium, which holds at any point in time. In Section 3, I introduce the stochastic productivity process for firms and derive the dynamics at the sectoral and aggregate levels. Section 4 presents the data, tests the model's predictions, and estimates the model using the simulated method of moments. Section 5 illustrates the quantified version of the model. Finally, Section 6 concludes.

2 Model

This section presents how production in the economy is organized at any point in time. The representative household derives utility from consuming a discrete set of differentiated goods, each produced by a single firm. Since the equilibrium holds at a point in time, I refer to this setting as the *static* equilibrium. The next section introduces 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 consuming 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 Y_j :

$$Y = \prod_{j=1}^N Y_j^{\omega_j} \quad (1)$$

where ω_j for $j = 1, \dots, N$ 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:

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

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. For simplicity, I use Cobb–Douglas preferences, but the analysis extends to more general CES preferences across sectors, provided that the elasticity of substitution across sectors is lower than within sectors.

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

$$Y_{ij} = A_{ij}L_{ij}. \quad (3)$$

Here, A_{ij} is firm-specific productivity and labor L_{ij} is the only input. 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 across 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. 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 distribution of firm shocks. For empirically plausible distributions and parameters, 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 on each variety ($P_{ij}Y_{ij}$) not exceeding the 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 :

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

where $P_j \equiv \left(\sum_{i=1}^{N_j} P_{ij}^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}}$ is the price index for sector j , and $P \equiv \prod_{j=1}^N P_j^{\omega_j}$ 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:

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

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:

$$\mu_{ij} = \frac{\zeta_{ij}}{\zeta_{ij} - 1} \frac{1}{1 - \tau_{ij}} \quad (5)$$

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 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 firm granularity, 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.9 for details.

2.3 Equilibrium Definition and Efficient Allocation

I normalize the labor wage to $W = 1$. The equilibrium is defined as follows. Given a choice of market structure for the perceived elasticity of demand ζ_{ij} , and a sequence of firm productivity vectors $\{\{A_{ij}\}_{i=1}^{N_j}\}_{j=1}^N$, a *static equilibrium* is (i) vectors of prices and quantities $\{\{P_{ij}, Y_{ij}, L_{ij}\}_{i=1}^{N_j}\}_{j=1}^N$, (ii) vectors of sectoral prices and quantities $\{P_j, Y_j, L_j\}_{j=1}^N$, and (iii) aggregate prices and quantities $\{P, Y, L\}$ such that: firms set prices and quantities to maximize profits given the demand curves (4) and ζ_{ij} ; household demand (4) holds; and the labor market and the government budget clear.

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.

³I abstract from the possibility of firms internalizing their impact on the aggregate price index P . See Appendix A.8 in [Burstein et al. \(2025\)](#) for a case in which this assumption is relaxed.

2.4 Firm-Level Outcomes

I denote the sales share of firm i in sector j as s_{ij} , and its cost share as \tilde{s}_{ij} . This notation follows [Baqae and Farhi \(2020\)](#), who show that, in the presence of distortions, cost-based shares play a central role in determining how shocks propagate through the economy. Given vectors of firm productivities and markups, we can express sales and cost shares as composites of markup-adjusted productivities in the sector:

$$s_{ij} := \frac{P_{ij}Y_{ij}}{P_jY_j} = \frac{(A_{ij}/\mu_{ij})^{\varepsilon-1}}{\sum_{k=1}^{N_j} (A_{kj}/\mu_{kj})^{\varepsilon-1}}, \quad \text{and} \quad \tilde{s}_{ij} := \frac{WL_{ij}}{WL_j} = \frac{s_{ij}/\mu_{ij}}{\sum_{k=1}^{N_j} s_{kj}/\mu_{kj}}. \quad (6)$$

Note that, under homogeneous markups, sales and cost shares coincide. A gap between both reflects markup dispersion, and thus misallocation of resources within the sector. Holding markups and aggregate expenditure constant, the productivity elasticity of sales and cost shares are given by:

$$\frac{\partial \ln s_{ij}}{\partial \ln A_{kj}} = \begin{cases} (\varepsilon - 1)(1 - s_{ij}) & \text{if } k = i, \\ -(\varepsilon - 1)s_{kj} & \text{if } k \neq i, \end{cases} \quad \text{and} \quad \frac{\partial \ln \tilde{s}_{ij}}{\partial \ln A_{kj}} = \begin{cases} (\varepsilon - 1)(1 - \tilde{s}_{ij}) & \text{if } k = i, \\ -(\varepsilon - 1)\tilde{s}_{kj} & \text{if } k \neq i. \end{cases} \quad (7)$$

When the number of firms in the sector is large and each firm is small ($s_{ij} \rightarrow 0$), the elasticity is constant and equal to $\varepsilon - 1$, as in the standard framework with a continuum of firms ([Lucas, 1978](#)). However, when firms are granular, the elasticity decreases with firm size, as larger firms eventually saturate their markets of operation. Another difference with the standard framework with infinitesimal firms is that shocks to one firm affect the sales and cost shares of all other firms in the sector.

2.5 Aggregation

Sector- and aggregate-level productivity are defined as labor productivity at the respective levels of aggregation: $A_j := Y_j / \sum_{i=1}^{N_j} L_{ij}$ and $A := Y / L$. Sectoral and aggregate markups are defined as the ratio of price to marginal cost at the respective levels of aggregation: $\mu_j := P_j A_j$ and $\mu := P A$, and can be expressed as cost-share-weighted averages of firm- and sector-level markups, respectively:

$$\mu_j = \sum_{i=1}^{N_j} \tilde{s}_{ij} \mu_{ij}, \quad \text{and} \quad \mu = \sum_{j=1}^N \tilde{\omega}_j \mu_j \quad (8)$$

where $\tilde{\omega}_j := \frac{\omega_j/\mu_j}{\sum_{k=1}^N \omega_k/\mu_k}$ is the aggregate cost share of sector j .⁴ Using these expressions, we can write sector- and aggregate-level productivity as functions of productivities and markups at lower levels of aggregation:

$$A_j = \left(\sum_{i=1}^{N_j} \left(\frac{\mu_{ij}}{\mu_j} \right)^{-\varepsilon} A_{ij}^{\varepsilon-1} \right)^{\frac{1}{\varepsilon-1}}, \quad \text{and} \quad A = \prod_{j=1}^N \left(\frac{\mu}{\mu_j} A_j \right)^{\omega_j}. \quad (9)$$

The presence of markup dispersion within and across sectors leads to an inefficient allocation of resources and, ultimately, lower sectoral and aggregate productivity. From equation (6), if a firm (sector) charges a higher markup than the sector (aggregate) markup, i.e., $\mu_{ij} > \mu_j$ ($\mu_j > \mu$), its sales share s_{ij} (ω_j) exceeds its cost share \tilde{s}_{ij} ($\tilde{\omega}_j$). From an efficiency perspective, that firm (sector) is smaller than socially optimal, and it would be beneficial to reallocate labor toward it. By the same logic, if instead $\mu_{ij} < \mu_j$ ($\mu_j < \mu$), the firm (sector) is larger than socially optimal, such that reallocating labor away from it would be beneficial.

Since there is a finite number of firms and sectors, firms are *granular* at both levels of aggregation. Section 3 introduces firm productivity dynamics and characterizes how granularity affects sectoral and aggregate productivity growth in both efficient and misallocated economies.

3 Productivity Dynamics

Having characterized the static allocation, this section shows how idiosyncratic firm-level shocks propagate to sectoral and aggregate productivity, and how the concentration of within-sector sales and costs affects expected growth.

3.1 Firm-Level Productivity Dynamics

Following the tradition in the firm dynamics literature, I assume that firm-level productivity follows a proportional random growth process, often referred to as Gibrat's law. Specifically, firm productivity features: (i) a trend component g which is common across all firms, (ii) frequent, thin-tailed continuous shocks, captured by an i.i.d. Brownian motion W_{ijt} with diffusion coefficient σ , and (iii) an i.i.d. jump component driven by a Poisson process Q_{ijt} with intensity λ and i.i.d.

⁴Markups can also be expressed as harmonic means of sales-share-weighted markups, or as the inverse sector and aggregate labor shares respectively.

jump size $J_{ijt} \sim F_J$, capturing rare and potentially asymmetric large shocks.⁵ Formally, firm productivity evolves according to the following stochastic differential equation:

$$\frac{dA_{ijt}}{A_{ijt}} = gdt + \sigma dW_{ijt} + (e^{J_{ijt}} - 1)dQ_{ijt}. \quad (10)$$

It is 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:

$$\mathbb{E}_t \left[\frac{1}{dt} d \ln A_{ijt} \right] = g - \frac{\sigma^2}{2} + \lambda \mathbb{E}[J], \quad (11)$$

that is, expected average firm productivity is the common drift minus the concavity correction $\sigma^2/2$ due to Jensen's inequality, plus the expected jump contribution $\lambda \mathbb{E}[J]$.⁶

In contrast to the case with infinitesimal firms, random growth in firm productivity does not imply that firm sales satisfy Gibrat's law, since changes in firm sales depend on the market share of the firm. One contribution of this paper is to show how granularity generates size-dependent firm dynamics in line with the data even with random growth in individual productivity.

Entry and Exit Incumbent firms exit randomly at a Poisson rate $\delta \geq 0$, while new firms enter at rate $\nu \geq 0$. Entrants draw their initial productivity from a time-shifting distribution $F_{e,t}$ that moves rightward at the trend growth rate η , so that $F_{e,t}(A) = F_e(Ae^{-\eta t})$. The parameter η captures the economy's underlying trend in technological progress. A higher η means new entrants start with increasingly higher productivities, reflecting overall technological improvement. For economically plausible values of η , there will be a stationary firm productivity distribution across sectors with a Pareto tail ($\mathbb{P}(A_{ij} < a) \sim a^{-\alpha_{\text{tail}}}$ for $\alpha_{\text{tail}} > 0$), which is uniquely determined by the model parameters. See Gabaix (2009); Beare and Toda (2022) and Appendix A.4 for details.

Continuous Time Working in continuous time makes the analysis tractable. Over a discrete time period, many firms can experience shocks at once, so tracking how simultaneous shocks reallocate

⁵Proportional random growth is the canonical baseline because: (i) firm growth rates are roughly size-independent for medium and large firms, e.g., Haltiwanger et al. (2013), and (ii) with a stabilizing force like entry and exit (Gabaix, 2009), it yields a stationary Pareto-tailed size distribution consistent with the data. 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, $\Pr(\Delta Q_{ijt} = 1) = \lambda \Delta t + o(\Delta t)$.

⁶Throughout the paper, I use $\mathbb{E}_t[d \ln X_t / dt]$ as shorthand for $\lim_{\Delta t \rightarrow 0} \Delta t^{-1} \mathbb{E}_t[\ln X_{t+\Delta t} - \ln X_t]$.

demand across a finite set of producers becomes intractable. In continuous time, Brownian motions have continuous paths and, over an infinitesimal interval dt , at most one Poisson jump can occur. These properties make it possible to study sectoral and aggregate expected productivity growth analytically, which is the focus of the remainder of this section.

3.2 The Granular Drag in Efficient Economies

This subsection analyzes how firm-level productivity dynamics interact with granularity to shape expected sectoral productivity growth in efficient economies. For expositional simplicity, I begin with the case without jumps ($\lambda = 0$) or entry and exit ($\delta = \nu = 0$), such that (10) reduces to $dA_{ijt} / A_{ijt} = gdt + \sigma dW_{ijt}$. The contributions of jumps, entry, and exit appear *additively* in expected sectoral productivity growth, such that the intuition built in this baseline case extends naturally to the more general setting. The following proposition characterizes expected sectoral productivity growth in this baseline setting.

Proposition 1. *Let $\mathcal{H}_{jt} := \sum_i s_{ijt}^2$ denote the sectoral sales Herfindahl–Hirschman index, where s_{ijt} are the firms' sales shares. Consider an efficient allocation in which firm productivities evolve according to $dA_{ijt} / A_{ijt} = g dt + \sigma dW_{ijt}$, where $\{W_{ijt}\}$ are i.i.d. standard Brownian motions. Then, the expected sectoral productivity growth rate $\gamma_{jt} := \mathbb{E}_t[\frac{1}{dt} d \ln A_{jt}]$ is*

$$\gamma_{jt} = \underbrace{g - \frac{\sigma^2}{2}}_{\text{Average-firm}} + \underbrace{(\varepsilon - 1) \frac{\sigma^2}{2} (1 - \mathcal{H}_{jt})}_{\text{Reallocation}}, \quad (12)$$

where $g - \frac{\sigma^2}{2}$ is the expected average-firm productivity growth, and $(\varepsilon - 1) \frac{\sigma^2}{2} (1 - \mathcal{H}_{jt})$ is a positive reallocation residual.

In an efficient economy with idiosyncratic diffusion shocks, the Herfindahl–Hirschman index (HHI) measure of sales concentration $\mathcal{H}_{jt} := \sum_i s_{ijt}^2$ is a sufficient statistic for the role of granularity in shaping expected sectoral productivity growth. The HHI lies in $[1/N_j, 1]$: it equals $1/N_j$ when all firms have equal sales shares and 1 when a single firm accounts for the entire sector. Thus, more concentrated sectors (higher \mathcal{H}_{jt}) exhibit relatively lower expected productivity growth due to reduced reallocation gains. I refer to the proof in Appendix A.3 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:

$$\gamma^1 := \lim_{s_{1jt} \rightarrow 1} \gamma_{jt} = \underbrace{g - \frac{\sigma^2}{2}}_{\text{Average-firm}}. \quad (13)$$

I refer to this term as the *average-firm* contribution to growth, $\gamma^1 = \mathbb{E}_t[d \ln A_{ijt}/dt]$. Second, consider the polar opposite case of a sector with a continuum of infinitesimal firms. I refer to this setting as the *fully diversified* case since the 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:

$$\gamma^\infty := \lim_{N_j \rightarrow \infty} \gamma_{jt} = \underbrace{g - \frac{\sigma^2}{2}}_{\text{Average-firm}} + \underbrace{(\varepsilon - 1) \frac{\sigma^2}{2}}_{\text{Reallocation}}. \quad (14)$$

Where does the reallocation term in (14) 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:

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

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 are 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:

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

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 goods are gross substitutes ($\varepsilon > 1$), granularity reduces expected reallocation gains, leading to lower expected productivity growth. In the extreme case of a monopolist, there are no reallocation gains at all. I refer to this phenomenon as the *granular drag* on expected productivity growth.

Jumps I extend the previous analysis to include jumps ($\lambda > 0$) in firm productivity. To keep the algebra light, I assume that there are no common trend or diffusion components ($g = 0, \sigma = 0$), so that firm productivity evolves purely through jumps: $dA_{ijt}/A_{ijt} = (e^{J_{ijt}} - 1)dQ_{ijt}$.⁷ The expected growth rate of sectoral productivity is now:

Proposition 2 (Jumps). *Consider an efficient allocation in which firms' productivities evolve according to $dA_{ijt}/A_{ijt} = (e^{J_{ijt}} - 1)dQ_{ijt}$ with Q_{ijt} a Poisson process of intensity λ and $J_{ijt} \sim F_J$ such that $\mathbb{E}[e^{(\varepsilon-1)J}] < \infty$. Then the expected sectoral productivity growth rate $\gamma_{jt} := \mathbb{E}_t[\frac{1}{dt}d \ln A_{jt}]$ is*

$$\gamma_{jt} = \frac{\lambda}{\varepsilon - 1} \sum_{i=1}^{N_j} \mathbb{E} \left[\ln \left(1 + s_{ijt} (e^{(\varepsilon-1)J} - 1) \right) \right]. \quad (15)$$

See Appendix A.3 for the proof. 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 ($N_j \rightarrow \infty$) yield

⁷The general case with both diffusion and jumps is just the sum of the two components, as they are independent. See Appendix A.3 for the general case.

respectively the following expected growth rates:

$$\begin{aligned}\gamma^1 &= \underbrace{\lambda \mathbb{E}[J]}_{\text{Average-firm}}, \\ \gamma^\infty &= \underbrace{\lambda \mathbb{E}[J]}_{\text{Average-firm}} + \underbrace{\lambda \frac{\mathbb{E}[e^{(\varepsilon-1)J} - 1] - (\varepsilon - 1)\mathbb{E}[J]}{\varepsilon - 1}}_{\text{Reallocation}}.\end{aligned}$$

The expected growth rate in the fully-diversified case can again be decomposed into an average-firm term plus a reallocation term. As Proposition 6 shows, 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 are 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 cross-sectional distribution of incumbent productivities does not affect expected 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:

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

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 $0 \leq 3\mathcal{H}_{jt} - 2\mathcal{H}_{3,jt} \leq 1$. Without jumps, it is obvious that expected sectoral productivity growth is bounded below by the monopolist ($\mathcal{H}_{jt} = 1$) case and above by the fully diversified ($\mathcal{H}_{jt} = 0$) case. Proposition 6 in Appendix A.3 shows that this result continues to hold with jumps, implying that the reallocation term is always non-negative for any well-behaved jump distribution.

Entry and Exit Entry and exit also contribute to expected sectoral productivity growth additively. To isolate their role, I assume that incumbent firms' productivity is constant over time ($g = 0$, $\sigma = 0$, $\lambda = 0$), so that $dA_{ijt}/A_{ijt} = 0$, but firms exit at rate δ and new firms enter at rate ν , drawing their initial productivity from a time-shifting distribution $F_{e,t}$ that grows at rate η . The following proposition shows how entry and exit contribute to expected sectoral productivity growth.

Proposition 3 (Entry and Exit). *Consider an efficient economy where incumbent firms have constant productivity, but exit at rate δ and new firms enter at rate ν , drawing their initial productivity from a distribution $F_{e,t}$. Then, the expected sectoral productivity growth rate $\gamma_{jt} := \mathbb{E}_t[\frac{1}{dt}d \ln A_{jt}]$ is given by:*

$$\gamma_{jt} = \underbrace{\frac{\nu}{\varepsilon - 1} \mathbb{E}_t \left[\ln \left(1 + \left(\frac{A_{et}}{A_{jt}} \right)^{\varepsilon - 1} \right) \right]}_{\text{Entry}} + \underbrace{\frac{\delta}{\varepsilon - 1} \sum_{i=1}^{N_j} \ln(1 - s_{ijt})}_{\text{Exit}},$$

where $A_{et} \sim F_{e,t}$ is the productivity of entrants.

The expected contribution of entry is larger when entrants are more productive relative to incumbents. Exit contributes negatively to expected sectoral productivity growth, with larger contributions when large firms exit. Applying a second-order approximation for small sales shares $s_{ijt} \approx 0$ makes the role of granularity explicit:

$$\text{Exit}_{jt} \approx -\frac{\delta}{\varepsilon - 1} \left(1 + \frac{1}{2} \mathcal{H}_{jt} \right).$$

Intuitively, exogenous exit is similar to an extremely left-skewed jump distribution. The higher the

elasticity of substitution ε , the smaller the negative impact of exit on expected sectoral productivity growth, as workers are reallocated toward surviving firms. However, concentration amplifies the negative impact of exit, as losing a large firm has a bigger effect on sector productivity than losing a small firm. Exogenous exit might appear extreme, but as I show in Appendix D.2, in the data, large establishments do exit, and when they do, sectoral output contracts significantly.

Percentage Growth The focus so far has been on expected *log* growth rates $\mathbb{E}_t[d \ln A_{jt} / dt]$. Do these results carry over to expected *percentage* growth rates $\mathbb{E}_t[\frac{1}{dt} dA_{jt} / A_{jt}]$? The answer is yes, subject to the elasticity of substitution being large enough. The following corollary shows that when $\varepsilon \geq 2$, the expected percentage growth rate can also be decomposed into an average-firm term plus a reallocation term that scales with one minus the sales HHI.

Corollary 1 (Percentage Growth). *In an efficient economy with firm productivity dynamics given by $dA_{ijt} / A_{ijt} = gdt + \sigma dW_{ijt}$, expected sectoral percentage productivity growth is given by:*

$$\mathbb{E}_t \left[\frac{1}{dt} \frac{dA_{jt}}{A_{jt}} \right] = \underbrace{g}_{\text{Avg. Firm}} + \underbrace{(\varepsilon - 2) \frac{\sigma^2}{2} (1 - \mathcal{H}_{jt})}_{\text{Reallocation}}. \quad (16)$$

where g is the average firm percentage growth rate $\mathbb{E}_t[\frac{1}{dt} dA_{ijt} / A_{ijt}]$, and $(\varepsilon - 2) \frac{\sigma^2}{2} (1 - \mathcal{H}_{jt})$ is the reallocation residual, with $\mathcal{H}_{jt} := \sum_{i=1}^{N_j} s_{ijt}^2$ the sales HHI of sector j .

The reallocation term is non-negative if and only if $\varepsilon \geq 2$. Because firms are granular, sectoral productivity is volatile. Taking the logarithm of a volatile variable induces a negative concavity correction due to Jensen's inequality. In this case, the correction term is $-\frac{\sigma^2}{2} \mathcal{H}_{jt}$. Combining this correction with the reallocation term from Proposition 1 gives the result. Reassuringly, both log and percentage growth coincide as the number of firms grows large and the law of large numbers holds such that productivity growth becomes deterministic. This result extends to the general case with jumps as well; see Appendix A.3 for details.

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 $A_t = \prod_{j=1}^N A_{jt}^{\omega_j}$, where ω_j is sector j 's expenditure share which is fixed over time. The following corollary to Proposition 1 characterizes aggregate productivity growth.

Corollary 2 (Aggregate Productivity Growth). *In an efficient economy with sectoral sales shares ω_j and 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:*

$$\gamma_t = g - \frac{\sigma^2}{2} + (\varepsilon - 1) \frac{\sigma^2}{2} \left(1 - \sum_{j=1}^N \omega_j \mathcal{H}_{jt} \right) \quad (17)$$

where $\sum_{j=1}^N \omega_j \mathcal{H}_{jt}$ is the sales-weighted aggregate HHI.

The proof follows directly from Proposition 1 and the Cobb-Douglas aggregation across sectors. The relevant measure of granularity at the aggregate level is the sales-weighted average of sectoral HHIs. There is an aggregate *granular drag* on productivity growth when (i) a few concentrated sectors dominate the economy (high ω_j for sectors with high \mathcal{H}_{jt}) or (ii) many sectors contribute significantly to the economy, but each sector is itself concentrated (high \mathcal{H}_{jt} for many j). Even if the number of sectors is large, the sales-weighted average HHI does not converge to zero unless the number of firms per sector also grows to infinity. If, for example, there is a secular trend of increasing sectoral concentration, aggregate productivity growth is dragged down as well.

3.3 The Granular Drag in Economies with Distortions

The preceding analysis assumes that the allocation of labor resources is efficient. 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 has a higher sales share relative to its cost share (i.e., $s_{ijt} > \tilde{s}_{ijt}$). This firm employs fewer workers than in the efficient allocation. Reallocating labor toward this firm increases 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 4 (Sectoral Productivity Growth under Misallocation). *Let $\mathcal{H}_{jt} := \sum_{i=1}^{N_j} s_{ijt}^2$ and $\tilde{\mathcal{H}}_{jt} := \sum_{i=1}^{N_j} (\tilde{s}_{ijt})^2$ denote the sectoral sales- and cost-based Herfindahl–Hirschman indices, where s_{ijt} and \tilde{s}_{ijt}*

are, respectively, the firms' sales and cost shares. Consider a sector where firm productivity follows $dA_{ijt}/A_{ijt} = g dt + \sigma dW_{ijt}$ and firms have fixed markups μ_{ij} . Then the expected sectoral productivity growth rate $\gamma_{jt} = \mathbb{E}_t[d \ln A_{jt}/dt]$ is given by:

$$\gamma_{jt} = \underbrace{g - \frac{\sigma^2}{2}}_{\text{Average-firm}} + \underbrace{(\varepsilon - 1) \frac{\sigma^2}{2} [1 - \mathcal{H}_{jt} - (\varepsilon - 1)(\mathcal{H}_{jt} - \tilde{\mathcal{H}}_{jt})]}_{\text{Reallocation}}. \quad (18)$$

When $\mathcal{H}_{jt} = \tilde{\mathcal{H}}_{jt}$ (homogeneous markups), this reduces to the efficient-economy case in Proposition 1.

The proof is in Appendix A.5. In the efficient allocation, sales and cost shares coincide, and the sales HHI \mathcal{H}_{jt} is, up to second order, a sufficient statistic for how granularity affects growth. Under misallocation, however, the difference between sales- and cost-based shares matters as well. If we compare equation (18) to the efficient case in equation (12), for a fixed sales concentration \mathcal{H}_{jt} , misallocation increases or decreases the expected growth rate by $(\varepsilon - 1)^2 \frac{\sigma^2}{2} (\mathcal{H}_{jt} - \tilde{\mathcal{H}}_{jt})$. Intuitively, if more productive firms have high markups (relative to small firms), sales concentration \mathcal{H}_{jt} is high relative to cost concentration $\tilde{\mathcal{H}}_{jt}$. In this case, the granular drag on growth is amplified; as resources are misallocated away from the most productive firms, it is 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 is 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 ($\mathcal{H}_{jt}, \tilde{\mathcal{H}}_{jt} \rightarrow 1$), there is a single firm, so there is no misallocation within the sector. In the fully diversified case ($\mathcal{H}_{jt}, \tilde{\mathcal{H}}_{jt} \rightarrow 0$), 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 interacts dynamically with granularity in a way that is not present in models with infinitesimal firms. When firms are discrete rather than infinitesimal, the joint distribution of productivity and markups shapes the *granular drag* on growth. With diffusion shocks, the sales

and cost HHIs are sufficient statistics for this drag.

Sectoral Markup Growth Even with firm markups fixed over time, sectoral markups μ_{jt} evolve endogenously as the sales and costs shares of firms change. Using that $\mu_{jt} = P_{jt}A_{jt}$, a simple Corollary to Proposition 4 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:*

$$\mathbb{E}_t \left[\frac{1}{dt} d \ln \mu_{jt} \right] = (\varepsilon - 1)^2 \frac{\sigma^2}{2} \left(\tilde{\mathcal{H}}_{jt} - \mathcal{H}_{jt} \right). \quad (19)$$

The proof is in Appendix A.5. Comparing (19) to the misallocation growth expression in (18), 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 is high relative to cost concentration, and sectoral markups tend to decline on average. Intuitively, a firm with an above average markup responds more with its cost share than with its sales share to a positive productivity shock (see Equation (7)), leading to a decline in the sectoral markup. Conversely, if more productive firms have low markups, sectoral markups tend to rise on average. Thus, the joint distribution of productivity and markups across firms shapes the endogenous dynamics of sectoral markups.

Aggregate Growth For efficient economies, aggregate productivity growth inherits 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:

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

Proposition 5 (Aggregate Productivity Growth under Misallocation). *Let \mathcal{H}_{jt} and $\tilde{\mathcal{H}}_{jt}$ denote sectoral sales- and cost-based HHIs, ω_j and $\tilde{\omega}_j$ sectoral sales- and cost-based shares, and $\mathcal{V}_{jt} := \sum_{i=1}^{N_j} (s_{ijt} - \tilde{s}_{ijt})^2$ the within-sector dispersion between sales and cost shares. If $dA_{ijt}/A_{ijt} = g dt + \sigma dW_{ijt}$ and firm markups μ_{ij}*

are fixed, then expected aggregate markup growth and aggregate productivity growth $\gamma_t := \mathbb{E}_t[\frac{1}{dt}d \ln A_t]$ are:

$$\mathbb{E}_t \left[\frac{1}{dt} d \ln \mu_t \right] = (\varepsilon - 1)^2 \frac{\sigma^2}{2} \sum_{j=1}^N \tilde{\omega}_j (\tilde{\mathcal{H}}_{jt} - \mathcal{H}_{jt}) - (\varepsilon - 1)^2 \frac{\sigma^2}{2} \sum_{j=1}^N \tilde{\omega}_j (1 - \tilde{\omega}_j) \mathcal{V}_{jt}, \quad (20)$$

$$\gamma_t = g - \frac{\sigma^2}{2} + (\varepsilon - 1) \frac{\sigma^2}{2} \sum_{j=1}^N \omega_j (1 - \mathcal{H}_{jt}) + \mathbb{E}_t \left[\frac{1}{dt} d \ln \mu_t \right]. \quad (21)$$

The proof is in Appendix A.5 and follows directly from Proposition 4 and Corollary 3. 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 second term captures the dependence of sectoral cost shares $\tilde{\omega}_j := \omega_j \mu_{jt}^{-1} / \sum_{k=1}^N \omega_k \mu_{kt}^{-1}$ on sectoral markups μ_{jt} which evolve endogenously according to Corollary 3. Again, aggregate productivity growth, is the sum of two components: average-firm growth plus a reallocation term that depends on the sales-weighted aggregate HHI, plus expected aggregate markup growth.

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.⁸ The exact construction of the final sample is detailed in Appendix D.1.

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 D.3.

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 D.4.

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 D, 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.

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,

⁸See Weidenman (2016); data retrieved 15/10/2023.

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 (22) and (23). 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 $\beta_1 < 0$. 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 (23). The theory predicts that both coefficients will be negative and that the coefficient on the gap will be larger in magnitude $\beta_2 < \beta_1 < 0$.

$$\ln A_{jt+\Delta t} - \ln A_{jt} = \alpha + \beta_1 \mathcal{H}_{jt} + \beta_A \ln A_{jt} + \epsilon_{jt+\Delta t} \quad (22)$$

$$\ln A_{jt+\Delta t} - \ln A_{jt} = \alpha + \beta_1 \mathcal{H}_{jt} + \beta_2 (\mathcal{H}_{jt} - \tilde{\mathcal{H}}_{jt}) + \beta_A \ln A_{jt} + \epsilon_{jt+\Delta t} \quad (23)$$

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: Concentration and Productivity Growth: Swedish 5-digit industries, 1998–2022

	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*** (0.062)	-0.215** (0.062)	-0.046 (0.066)	-0.013 (0.065)
HHI _t sales – HHI _t costs			-1.349*** (0.199)	-1.164*** (0.198)
$\ln(\text{Prod}_t)$		-0.106*** (0.017)		-0.070*** (0.016)
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

Notes: Unit of observation is 5-digit SNI in Sweden, 1998–2022. $\ln(\text{Prod})$ is log-labor productivity, HHI (sales/costs) based on firm sales and personnel cost shares within the industry. All regressions include 2-digit \times year fixed effects. SEs clustered by 5-digit industry and year. “Efficient” refers to the specification motivated by the predictions of the model without distortions, while “With Distortions” includes the additional control for the HHI gap (sales - costs) as motivated by the model with markup dispersion.

I run similar regressions for U.S. and CompNet data in Appendix D.3 and D.4, 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.⁹ Panel (A) shows the relationship between sales HHI and five-year ahead productivity growth, controlling for current productivity and fixed effects. While the relationship is negative, linearity is not apparent. Panel (B) shows the relationship productivity growth and the difference between sales and cost HHIs. As the theory predicts, a negative linear relationship is clearly visible, supporting the model predictions that granularity affects sectoral dynamics through misallocation when markups differ across firms.

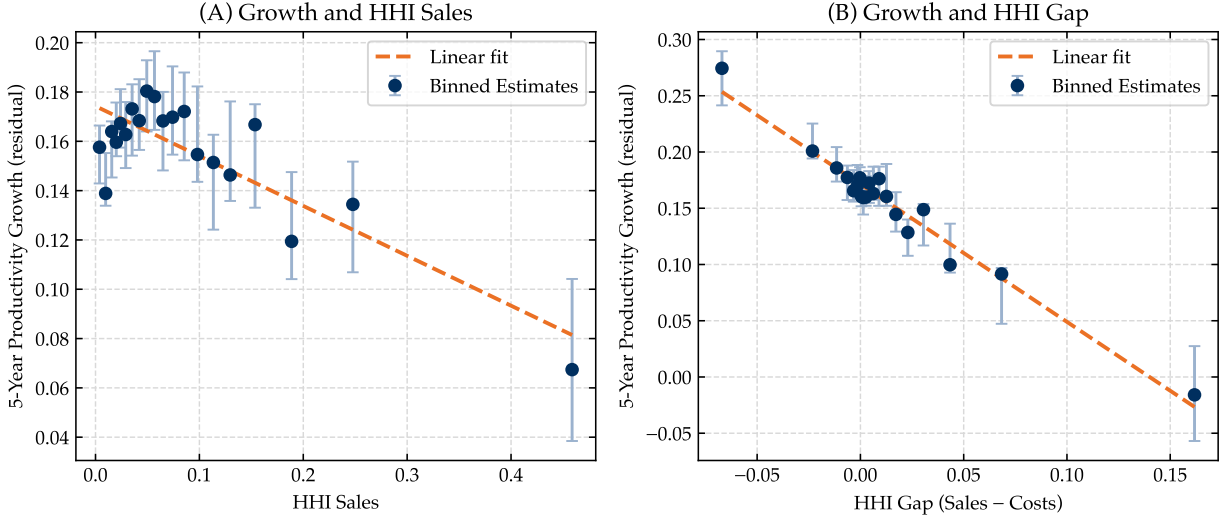


Figure 1: Binscatters of five-year labor-productivity growth on concentration: Swedish 5-digit industries, 1998–2022. Panel (A) plots $\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$ against the sales HHI; Panel (B) plots the same outcome against the HHI gap (sales – costs). Bins are ventiles (20). Each panel partials out current $\ln(\text{Prod}_t)$ and 2-digit \times year fixed effects; Panel (B) additionally partials out the sales HHI. Partialing out and 95% confidence bands are computed using the semiparametric binscatter procedure of Cattaneo et al. (2024). Sample: 7,218 industry–year observations.

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 (19), 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

⁹I use the methodology developed by (Cattaneo et al., 2024) to residualize and estimate the confidence intervals for the binned scatter plots.

industries may not correspond to the relevant market boundaries for competition.¹⁰ If, for example, 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.¹¹

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 (10) includes a common deterministic drift (g), a diffusion coefficient (σ) that reflects the standard deviation of thin-tailed shocks, and a jump component that captures the frequency and size of large shocks. For the jump distribution, I use an asymmetric Laplace distribution:¹²

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

¹⁰See [Berry et al. \(2019\)](#) for a comprehensive discussion of the role of market definition in empirical industrial organization.

¹¹In 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.

¹²An 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. \(1996\)](#).

with mean $\frac{1}{\mu_+} - \frac{1}{\mu_-}$ and variance $\frac{1}{\mu_-^2} + \frac{1}{\mu_+^2}$. As we saw in Section 3, higher-order moments like skewness and kurtosis might interact with granular concentration in a non-trivial way. Allowing for asymmetry in the jumps allows matching skewness, while kurtosis is controlled by the jump intensity λ .¹³ 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 $\frac{P97.5-P2.5}{P75-P25} - 2.91$.¹⁴ 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 set to 5, 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. The number of firms per industry is Poisson distributed with mean 140, corresponding to the median number of firms per 5-digit industry in the data.

Table 2 summarizes the calibration results. The tail index and the exit rate are estimated separately, while the productivity process parameters are estimated jointly given α_{tail} and δ . The estimated productivity process features a jump roughly every three years, with left skewed jumps that are larger on average than right-skewed jumps. The diffusion component is relatively small compared to the jump component, indicating that large shocks play a significant role in firm productivity dynamics. The tail index α_{tail} is estimated at 3.96, implying Zipf's law for sales shares within industries.¹⁵

¹³A low rate λ makes jumps rare, leading to excess kurtosis.

¹⁴I 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.

¹⁵Since $\varepsilon = 5$, Zipf's law for sales shares implies a tail index of $\alpha_{\text{tail}} = \varepsilon - 1 = 4$.

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	$\frac{P97.5-P2.5}{P75-P25} - 2.91$ 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

Notes: “Data” are empirical targets; “Model” are simulated moments at the estimated parameters. g is firm trend; σ is the diffusion coeff.; λ is jump arrival rate; (μ_+, μ_-) govern right/left jump tails; α_{tail} is the Pareto tail index; δ is the exit rate. P_x denotes the x th percentile. CR4 is the four-firm sales concentration ratio. Baseline elasticity $\varepsilon = 5$. Sample: Swedish firms by 5-digit SNI, 1998–2022.

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.

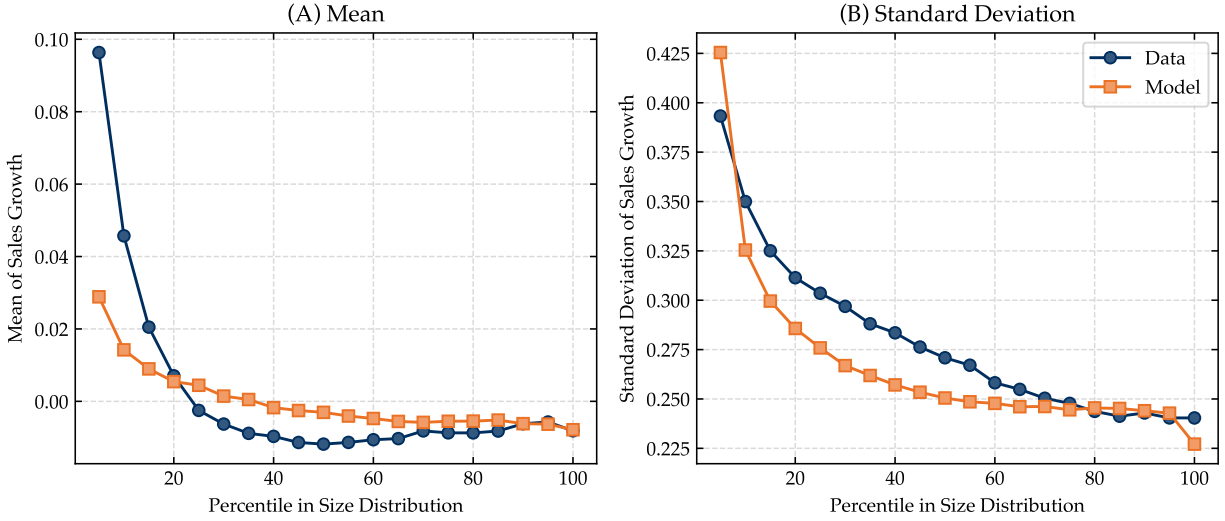


Figure 2: Mean and standard deviation of one-year sales growth across current sales bins. Data correspond to Swedish firms, 1998–2022, detrended with industry-year fixed effects. Model correspond to the calibrated model using the parameters in Table 2.

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.¹⁶ The right panel shows the volatility profile, 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.

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.

¹⁶See appendix D.2 for evidence on the exit hazard.

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.

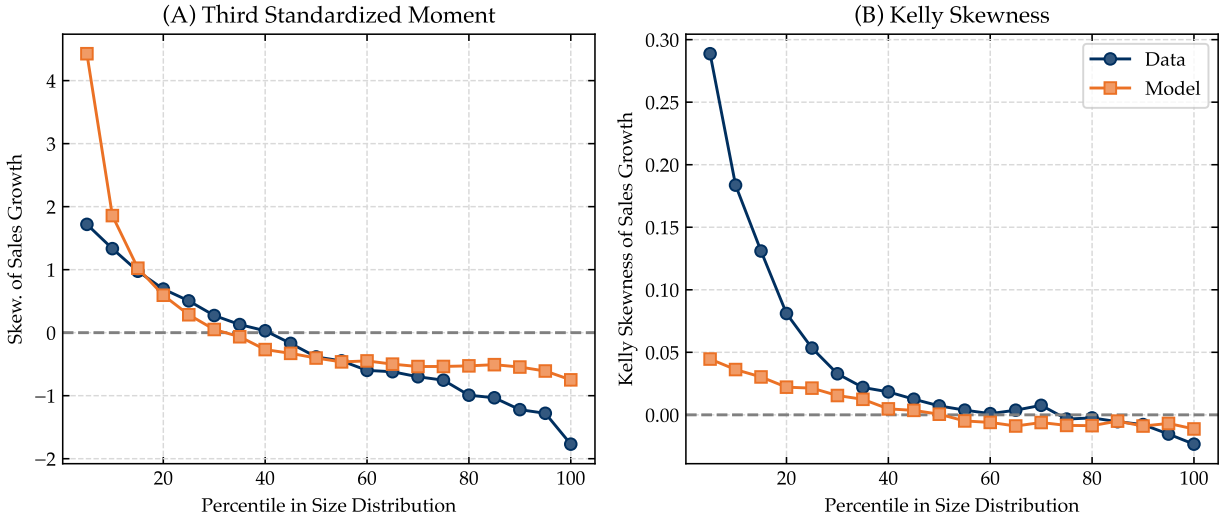


Figure 3: Standardized skewness and Kelly skewness of one-year sales growth across current sales bins. Data correspond to Swedish firms, 1998–2022, detrended with industry-year fixed effects. Model correspond to the calibrated model using the parameters in Table 2.

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 4.¹⁷ 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,

¹⁷Formally, 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.

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.

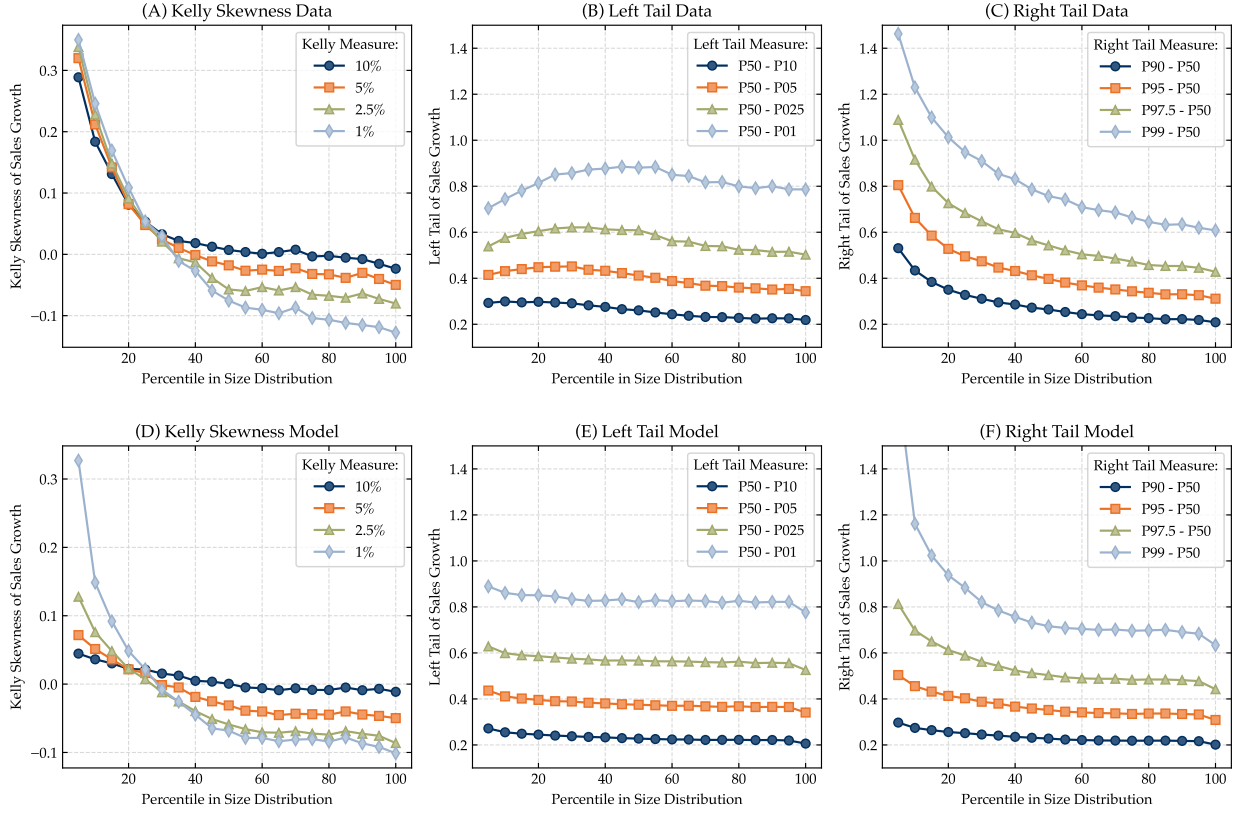


Figure 4: Kelly skewness and left/right-tail decomposition of one-year sales growth across current sales bins. Kelly skewness with tail threshold τ is defined as $(P_{1-\tau} + P_{\tau} - 2P_{50}) / (P_{1-\tau} - P_{\tau})$, where P_x is the x th percentile of the distribution. Panels (A)–(C) correspond to data; panels (D)–(F) to the model. Data correspond to Swedish firms, 1998–2022, detrended with industry-year fixed effects. Model correspond to the calibrated model using the parameters in Table 2.

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.

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 allow for i.i.d. taxes and subsidies on firm sales of $\pm 20\%$ in the spirit of Restuccia and Rogerson (2008). 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: Concentration and Productivity Growth: Swedish 5-digit Industries vs. Model

	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** (0.062)	-0.013 (0.065)	-0.059*** (0.006)	-0.060*** (0.005)
HHI _t sales – HHI _t costs		-1.164*** (0.198)		-0.215*** (0.026)
$\ln(\text{Prod}_t)$	-0.106*** (0.017)	-0.070*** (0.016)	-0.066*** (0.008)	-0.068*** (0.007)
Observations	7218	7218	200000	200000
R^2	0.318	0.352	0.064	0.068
R^2 Within	0.062	0.109	-	-

Notes: Data corresponds to 5-digit SNI industries in Sweden, 1998-2022. Model corresponds to a simulated cross-section of CES nests. $\ln(\text{Prod})$ is log-labor productivity, HHI (sales/costs) based on firm sales and personnel cost shares within the industry. All data regressions include 2-digit \times year fixed effects and SEs clustered by 5-digit industry and year.

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 $h = 0, 5, 10, \dots, 140$:

$$\Delta_h \ln A_{j,t} = \beta_h \Delta_5 \mathcal{H}_{j,t} + \theta_h \Delta_5 (\mathcal{H}_{j,t} - \tilde{\mathcal{H}}_{j,t}) + \beta_A \ln A_{j,t-1} + \alpha_j + \tau_t + \epsilon_{j,t+h}, \quad (24)$$

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.

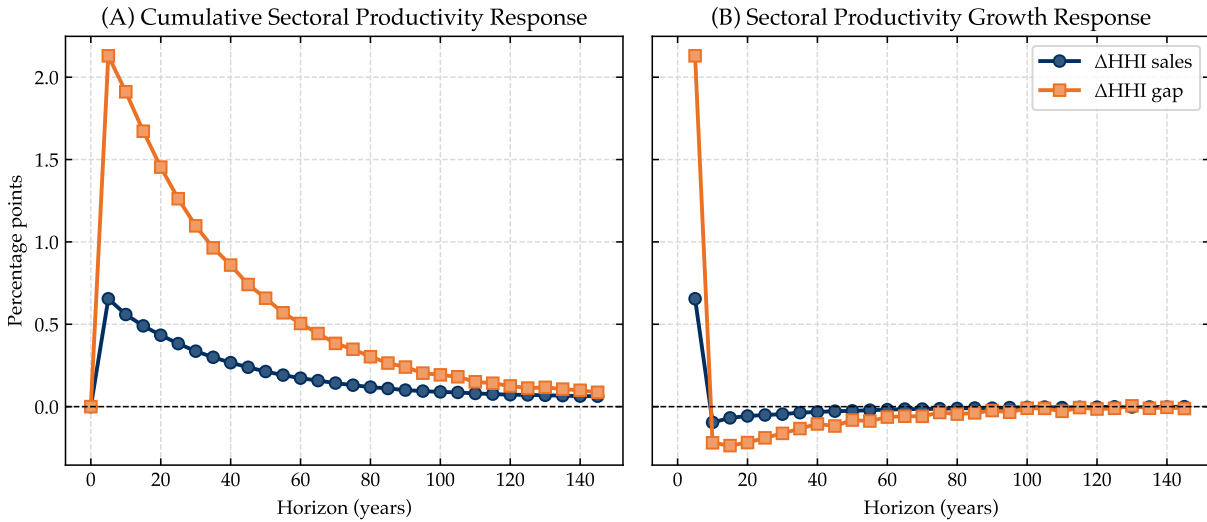


Figure 5: Model IRFs to One-Percentage-Point Increases in Sales HHI (Blue) and HHI Gap (Orange). (A) Cumulative Productivity Response. (B) 5-Year Productivity Growth Response. HHI Gap refers to the difference between sales- and cost-based HHIs.

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 (17), 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.¹⁸ 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

¹⁸In 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.

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 growth by about 0.5 percentage points.

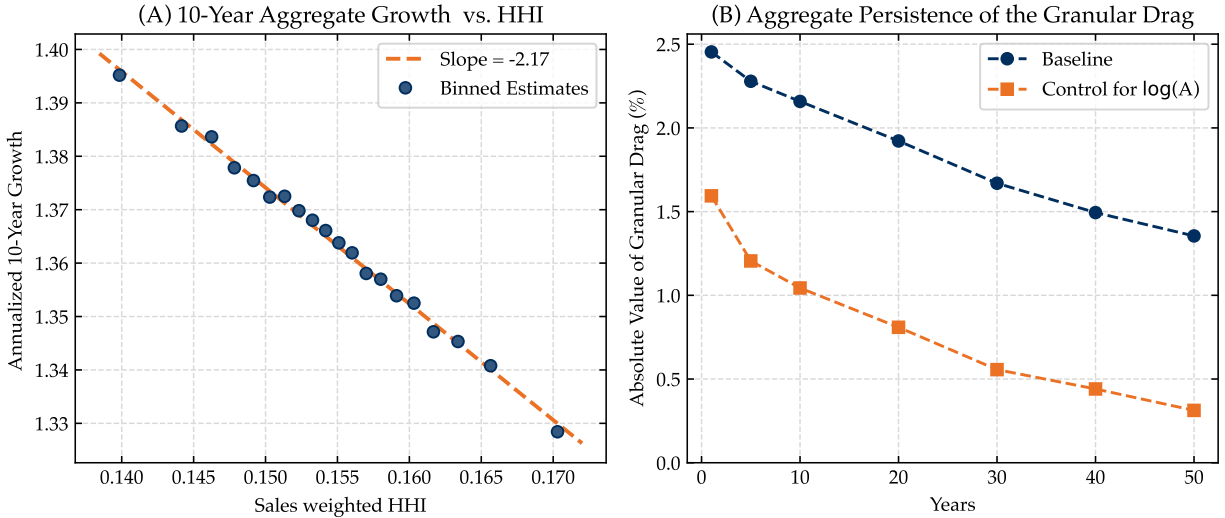


Figure 6: Model-Simulated Long-Run Effects of Aggregate Concentration on Productivity Growth. (A) Binscatter of annualized 10-year aggregate productivity growth vs. sales-weighted aggregate HHI. (B) Absolute value of the annualized effect of sales-weighted aggregate HHI on aggregate productivity growth across 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 micro-founded, exogenous growth model providing to study how firm granularity shapes 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, the reallocation of resources across producers becomes imperfect. As a result, higher concentration reduces the expected gains from micro-level reallocation, generating a *granular drag* on productivity growth.

The model generates substantial and persistent cross-sectional growth heterogeneity across firms and aggregates, in line with the empirical evidence. 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 charge higher markups. Industries that become more concentrated, or where sales are more concentrated than costs, subsequently experience slower productivity growth. These predictions further aggregate up to the macro level, implying that economies with more concentrated dominant sectors grow more slowly on average.

Consistent with these predictions, I find strong support for these mechanisms using firm- and industry-level data from Sweden, the United States, and Europe. Across Swedish 5-digit industries, a 10-percentage-point rise in the Herfindahl index of sales concentration is associated with about a 3-percentage-point decline in five-year productivity growth. In the presence of distortions, the theory predicts that the granular drag is amplified when sales concentration exceeds cost concentration. Consistent with this, I find that a 10-percentage-point increase in the gap between the Herfindahl indices of sales and costs predicts a decline of roughly 13 percentage points in five-year productivity growth.

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 *granular drag* channel through which market concentration influences medium- to 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

- Acemoglu, Daron and Dan Cao**, “Innovation by entrants and incumbents,” *Journal of Economic Theory*, 2015, 157, 255–294.
- , **Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi**, “The Network Origins of Aggregate Fluctuations,” *Econometrica*, 2012, 80 (5), 1977–2016.
- 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.
- , **Christopher Harris, Peter Howitt, and John Vickers**, “Competition, Imitation and Growth with Step-by-Step Innovation,” *The Review of Economic Studies*, 07 2001, 68 (3), 467–492.
- , **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.
- Akcigit, Ufuk and Sina T. Ates**, “What Happened to US Business Dynamism?,” *Journal of Political Economy*, 2023, 131 (8), 2059–2124.
- 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.
- Baqae, David and Elisa Rubbo**, “Micro Propagation and Macro Aggregation,” *Annual Review of Economics*, September 2023, 15 (1), 91–123.
- Baqae, 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 2020, 135 (1), 105–163.
- Bartelsman, Eric J. and Mark Doms**, “Understanding Productivity: Lessons from Longitudinal Microdata,” *Journal of Economic Literature*, 2000, 38 (3), 569–594.
- Beare, Brendan K. and Alexis Akira Toda**, “Determination of Pareto exponents in economic models driven by Markov multiplicative processes,” *Econometrica*, 2022, 90 (4), 1811–1833.
- 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.
- **and Xavier Gabaix**, “The Great Diversification and Its Undoing,” *American Economic Review*, August 2013, 103 (5), 1697–1727.
- 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.
- Chen, Zhang**, “Economic Growth and the Rise of Large Firms,” STEG Working Paper WP078, Structural Transformation and Economic Growth (STEG), Centre for Economic Policy Research (CEPR) August 2023.
- Chodorow-Reich, Gabriel, Xavier Gabaix, and Davide Viviano**, “Propagation of Shocks in Networks: Identification and Applications,” Working Paper, SSRN March 2025.
- 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.
- Foster, Lucia, John C. Haltiwanger, and C. J. Krizan**, “Aggregate Productivity Growth: Lessons from Microeconomic Evidence,” in Charles R. Hulten, Edwin R. Dean, and Michael J. Harper, eds., *New Developments in Productivity Analysis*, National Bureau of Economic Research Studies in Income and Wealth, University of Chicago Press, January 2001, pp. 303–372.
- 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.

- Grassi, Basile**, "IO in I-O: Size, Industrial Organization, and the Input-Output Network Make a Firm Structurally Important," Working Papers 619, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University 2018.
- 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.
- Hanoch, Giora**, "CRESH Production Functions," *Econometrica*, September 1971, 39 (5), 695–712.
- 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.
- Jordà, Òscar**, "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, March 2005, 95 (1), 161–182.
- Klette, Tor Jakob and Samuel Kortum**, "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 2004, 112 (5), 986–1018.
- Kwon, Spencer Y., Yueran Ma, and Kaspar Zimmermann**, "100 Years of Rising Corporate Concentration," *American Economic Review*, July 2024, 114 (7), 2111–40.
- Kydland, Finn E and Edward C Prescott**, "Time to Build and Aggregate Fluctuations," *Econometrica*, November 1982, 50 (6), 1345–1370.
- Loecker, Jan De, Jan Eeckhout, and Gabriel Unger**, "The Rise of Market Power and the Macroeconomic Implications," *The Quarterly Journal of Economics*, None 2020, 135 (2), 561–644.
- Lucas, Robert E.**, "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 1978, 9 (2), 508–523.
- Luttmer, Erzo G. J.**, "Selection, Growth, and the Size Distribution of Firms," *The Quarterly Journal of Economics*, 2007, 122 (3), 1103–1144.
- , "On the Mechanics of Firm Growth," *The Review of Economic Studies*, 2011, 78 (3), 1042–1068.
- 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.
- Moran, José, Angelo Secchi, and Jean-Philippe Bouchaud**, "Revisiting Granular Models of Firm Growth," 2024.
- Olmstead-Rumsey, Jane**, "Market Concentration and the Productivity Slowdown," Technical Report 107000, Munich Personal RePEc Archive (MPRA) 2019.

- Prescott, Edward C.**, "Theory ahead of business cycle measurement," Staff Report 102, Federal Reserve Bank of Minneapolis 1986.
- Restuccia, Diego and Richard Rogerson**, "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic Dynamics*, 2008, 11 (4), 707–720.
- Ridder, Maarten De**, "Market Power and Innovation in the Intangible Economy," *American Economic Review*, January 2024, 114 (1), 199–251.
- 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 Literature*, 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.
- Weiss, Joshua**, "Intangible Investment and Market Concentration," 2020. Working Paper.
- Yeh, Chen**, "Revisiting the Origins of Business Cycles With the Size-Variance Relationship," *The Review of Economics and Statistics*, 2025, 107 (3), 864–871.

A Theory Appendix

A.1 Derivation of Sectoral and Aggregate Productivity Indices

In this subsection, I derive the sectoral and aggregate productivity indices (9). Recall that the sectoral productivity index is defined as labor productivity $A_j = Y_j/L_j$, where Y_j is sectoral output and $L_j = \sum_{i=1}^{N_j} L_{ij}$ is sectoral labor input. The sectoral markup is defined as $\mu_j = P_j A_j$, where P_j is the sectoral price index and the wage has been normalized to 1. Using the definition of sectoral productivity we have:

$$\begin{aligned}\mu_j &= \frac{P_j Y_j}{L_j} \\ &= \sum_{i=1}^{N_j} \frac{P_{ij} Y_{ij}}{L_j} \\ &= \sum_{i=1}^{N_j} \frac{\mu_{ij} L_{ij}}{L_j}\end{aligned}$$

such that the sectoral markup is the arithmetic mean of firm-level markups μ_{ij} weighted by their cost shares $\tilde{s}_{ij} = L_{ij}/L_j$. Using that $L_j = \sum_{i=1}^{N_j} L_{ij}$ and $L_{ij} = \mu_{ij}^{-\varepsilon} A_{ij}^{\varepsilon-1} C_j$ where C_j is a sector-level term. We can then write:

$$\begin{aligned}\mu_j &= \frac{\sum_{i=1}^{N_j} \mu_{ij}^{1-\varepsilon} A_{ij}^{\varepsilon-1}}{\sum_{i=1}^{N_j} \mu_{ij}^{-\varepsilon} A_{ij}^{\varepsilon-1}} \\ &= \frac{\sum_{i=1}^{N_j} P_{ij}^{1-\varepsilon}}{\sum_{i=1}^{N_j} \mu_{ij}^{-\varepsilon} A_{ij}^{\varepsilon-1}} \\ &= \frac{(\mu_j / A_j)^{1-\varepsilon}}{\sum_{i=1}^{N_j} \mu_{ij}^{-\varepsilon} A_{ij}^{\varepsilon-1}}\end{aligned}$$

where in the last line we used the definition of the sectoral price index $P_j^{1-\varepsilon} = \sum_{i=1}^{N_j} P_{ij}^{1-\varepsilon} = \sum_{i=1}^{N_j} (\mu_{ij} / A_{ij})^{1-\varepsilon}$. Rearranging yields the expression for sectoral productivity in (9):

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

A.2 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, changes in a transformation of that variable depend not only on the changes in the original variable, but also on how randomness propagates through the transformation. For example, if firm productivity follows a geometric Brownian motion, then the growth rate of its logarithm must correct for curvature of the concave transformation (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 me to map micro-level stochastic processes for firms into laws of motion for aggregates such as sectoral or economy-wide productivity indices. Itô's lemma does not give an approximation, but an *exact* characterization of the dynamics of transformed stochastic processes.

Itô's Lemma (with Jumps) Let X_t follow

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

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

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

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

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

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

A.3 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. 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 standard Brownian motion 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:

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

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 $s_{ijt} = A_{ijt}^{\varepsilon-1} / \left(\sum_{k=1}^{N_j} A_{ikt}^{\varepsilon-1}\right)$. It will be useful to note that $\sum_{i=1}^N s_{ijt} dW_{ijt} \stackrel{d}{=} \sqrt{\mathcal{H}_{jt}} dW_{jt}$, 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_{jt} :

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

Using Ito's lemma for the log transformation again, we get:

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

We are now ready to prove Proposition 1, and Proposition 2.

Proof of Proposition 1 and Proposition 2. The focus of the paper is on expected productivity growth $\gamma_{jt} := \mathbb{E}_t[\frac{1}{dt} d \ln A_{jt}]$. We have $\mathbb{E}_t[dW_{jt}] = 0$ and $\mathbb{E}_t[dQ_{ijt}] = \lambda dt$, so taking expectations and using independence of the process and of the jump arrival and jump size yields:

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

To recover the case without jumps, we set $\lambda = 0$, which yields equation (12). To recover the case with jumps only, we set $g = \sigma = 0$, which yields equation (15). ■

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 6. *Consider an efficient economy where firm productivity follows the process in (10). 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 :*

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

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

Proof. The proof for the case without jumps is immediate from (12). For the case with jumps, the result follows from the elementary inequality $e^x \geq 1 + x$ for all $x \in \mathbb{R}$. ■

From Proposition 6, it follows that in efficient economies, the reallocation residual is always positive.

A.3.1 Concentration and Growth over Finite Horizons

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

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

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 $\vec{s}_{jt} > \vec{s}_{kt}$, if

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

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.¹⁹ It implies higher values for standard measures of concentration, including the HHI and top- m concentration ratios. With this notion of concentration in hand, we can establish a negative relationship between concentration and growth even over finite horizons.

Proposition 7. Consider an efficient economy where firm productivity follows a random growth process (e.g., (10)) and there is no entry or exit. 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

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

Proof. A symmetric function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is Schur-concave if for any two vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ such that \vec{x} is more Lorenz-concentrated than \vec{y} (i.e., $\vec{x} > \vec{y}$), it holds that $f(\vec{x}) \leq f(\vec{y})$. The expected growth rate $\Gamma_{jt}(\Delta t)$ is a symmetric function of the share vector \vec{s}_{jt} , which is also strictly (Schur-)concave for $\varepsilon > 1$, such that if sector j is more Lorenz-concentrated than sector k , then $\Gamma_{jt}(\Delta t) < \Gamma_{kt}(\Delta t)$. In the case of percentage growth over finite horizons, the elasticity of substitution ε needs to be greater than 2 for Schur-concavity to hold. ■

¹⁹Mathematically, Lorenz concentration is referred to as the majorization order. See (Marshall et al., 2011) for a textbook treatment.

A.4 Entry, Exit, and Stationary Firm Productivity Distribution

Without a "stabilizing force" (Gabaix, 2009), random growth does not admit a stationary distribution: firm productivities fan out over time.²⁰ 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 (10). 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):

$$0 = -\mu_x(\eta) \phi'(x) + \frac{\sigma^2}{2} \phi''(x) + \lambda \mathbb{E}[\phi(x - J) - \phi(x)] - \delta \phi(x), \quad x \in \mathbb{R} \setminus \{x_e\}, \quad (\text{A.2})$$

with an inflow of mass at $x_e := \ln A_e$ at rate ν . A standard implication of (A.2) 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 (A.2) with jumps. Guessing $\phi(x) \propto e^{-\alpha x}$ away from x_e and substituting into (A.2) yields a mapping between the traveling-wave speed η and the tail index α

$$\eta = g + \frac{\alpha - 1}{2} \sigma^2 + \lambda \frac{\mathbb{E}[e^{\alpha J}] - 1}{\alpha} - \frac{\delta}{\alpha}. \quad (\text{A.3})$$

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 . If the entry distribution is not degenerate, a further requirement is that the entry tail is thinner than the stationary tail.

With finitely many firms, the empirical sectoral distribution fluctuates around the stationary density.²¹ As the number of firms N_j increases, the empirical distribution converges to the

²⁰For 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.

²¹One necessary modification with finitely many firms is that, to have on average \bar{N}_j incumbents, the rate of entry $\nu = \delta \bar{N}_j$.

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

Proof of Proposition 3. Define

$$S_{jt} \equiv \sum_{i=1}^{N_j} A_{ijt}^{\varepsilon-1}, \quad A_{jt} = S_{jt}^{1/(\varepsilon-1)}, \quad s_{ijt} = \frac{A_{ijt}^{\varepsilon-1}}{S_{jt}}.$$

I model entry and exit as Poisson jumps of the sectoral aggregator:

- *Entry:* A marked Poisson random measure $Q^\nu(dt, dA_e)$ with intensity ν and jump distribution $F_{e,t}(dA_e) dt$ adds a mass $A_e^{\varepsilon-1}$ to S_{jt} at each jump.
- *Exit:* For each incumbent i , a Poisson process $Q_i^\delta(dt)$ with intensity δdt removes the mass $A_{ijt}^{\varepsilon-1}$ when firm i exits.

Hence the law of motion for S_{jt} is

$$dS_{jt} = \int A_e^{\varepsilon-1} Q^\nu(dt, dA_e) - \sum_{i=1}^{N_j} A_{ijt}^{\varepsilon-1} Q_i^\delta(dt). \quad (\text{A.4})$$

Applying Itô's lemma (A.1) to (A.4) gives

$$\begin{aligned} d \ln S_{jt} &= \int \ln\left(1 + \frac{A_e^{\varepsilon-1}}{S_{jt}^-}\right) Q^\nu(dt, dA_e) + \sum_{i=1}^{N_j} \ln\left(1 - \frac{A_{ijt}^{\varepsilon-1}}{S_{jt}^-}\right) Q_i^\delta(dt) \\ &= \int \ln\left(1 + \left(\frac{A_{et}}{A_{jt}}\right)^{\varepsilon-1}\right) Q^\nu(dt, dA_e) + \sum_{i=1}^{N_j} \ln(1 - s_{ijt}) Q_i^\delta(dt), \end{aligned} \quad (\text{A.5})$$

where I used $A_{jt} = S_{jt}^{1/(\varepsilon-1)}$ and $s_{ijt} = A_{ijt}^{\varepsilon-1}/S_{jt}$. Since $\ln A_{jt} = \frac{1}{\varepsilon-1} \ln S_{jt}$, taking expectations and using $\mathbb{E}_t[Q^\nu(dt, dA_e)] = \nu F_{e,t}(dA_e) dt$ and $\mathbb{E}_t[Q_i^\delta(dt)] = \delta dt$, and the independence of jumps and arrival times, the instantaneous expected growth rate $\gamma_{jt} \equiv \mathbb{E}_t[\frac{1}{dt} d \ln A_{jt}]$ follows from (A.5):

$$\gamma_{jt} = \underbrace{\frac{\nu}{\varepsilon-1} \mathbb{E}_t\left[\ln\left(1 + \left(\frac{A_{et}}{A_{jt}}\right)^{\varepsilon-1}\right)\right]}_{\text{Entry}} + \underbrace{\frac{\delta}{\varepsilon-1} \sum_{i=1}^{N_j} \ln(1 - s_{ijt})}_{\text{Exit}},$$

which proves Proposition 3. ■

Second-order exit approximation. For small shares, $\ln(1-s) = -s - \frac{1}{2}s^2 + O(s^3)$, hence

$$\text{Exit}_{jt} = \frac{\delta}{\varepsilon - 1} \sum_i \ln(1 - s_{ijt}) \approx -\frac{\delta}{\varepsilon - 1} \left(1 + \frac{1}{2}\mathcal{H}_{jt}\right), \quad \mathcal{H}_{jt} = \sum_i s_{ijt}^2,$$

making the role of concentration explicit.

A.5 Sectoral Productivity Growth with Misallocation

Proof of Proposition 4. The sectoral productivity index can be written as:

$$A_{jt} = \frac{\left(\sum_{i=1}^{N_j} A_{ijt}^{\varepsilon-1} \mu_{ij}^{1-\varepsilon}\right)^{\frac{\varepsilon}{\varepsilon-1}}}{\sum_{i=1}^{N_j} A_{ijt}^{\varepsilon-1} \mu_{ij}^{-\varepsilon}}. \quad (\text{A.6})$$

Recall that

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

and let $\mathcal{H}_{jt} := \sum_i s_{ijt}^2$ and $\tilde{\mathcal{H}}_{jt} := \sum_i (\tilde{s}_{ijt})^2$ be their HHI indices. Write

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

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

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

Then we have $s_{ijt} := X_{ijt}^{(N)} / N_{jt}$ and $\tilde{s}_{ijt} := X_{ijt}^{(D)} / D_{jt}$. 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

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

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

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

with $\tilde{\mathcal{H}}_{jt} := \sum_i (\tilde{s}_{ijt})^2$. Combining via $\ln A_{jt} = \frac{\varepsilon}{\varepsilon - 1} \ln N_{jt} - \ln D_{jt}$ and taking expectations gives (18). ■

A.6 Sectoral Markup Dynamics

I derive here the SDE for sectoral markups under the diffusion specification in (10) 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

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

As with sectoral productivity, note that $\sum_{i=1}^{N_j} (s_{ijt} - \tilde{s}_{ijt}) dW_{ijt} \stackrel{d}{=} \sqrt{\mathcal{V}_{jt}} dW_{\mu_{jt}}$, where $\mathcal{V}_{jt} := \sum_{i=1}^{N_j} (s_{ijt} - \tilde{s}_{ijt})^2$ 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

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

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

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

A.8 Aggregate Productivity Dynamics with Misallocation

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

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

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

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

Taking expectations and simplifying gives the expected aggregate productivity growth rate

under misallocation:

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

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

A.9 Endogenous Markups

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

sectoral price index P_j .

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

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

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

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

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

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

$$\frac{1}{\zeta_{ij}} \equiv -\frac{\partial \ln P_{ij}}{\partial \ln Y_{ij}}$$

$$\text{Given } \ln P_{ij} = -\frac{1}{\varepsilon} \ln Y_{ij} + \left(\frac{1}{\varepsilon} - 1 \right) \ln Y_j$$

$$\frac{1}{\zeta_{ij}} = \frac{1}{\varepsilon} - \left(\frac{1}{\varepsilon} - 1 \right) \frac{\partial \ln Y_j}{\partial \ln Y_{ij}}$$

$$\text{Since } \frac{\partial \ln Y_j}{\partial \ln Y_{ij}} = \frac{Y_{ij}}{Y_j} \frac{\partial Y_j}{\partial Y_{ij}} = \left(\frac{Y_{ij}}{Y_j} \right)^{(\varepsilon-1)/\varepsilon} = s_{ij}$$

$$\implies \frac{1}{\zeta_{ij}} = \frac{1}{\varepsilon} - \left(\frac{1}{\varepsilon} - 1 \right) s_{ij} = \frac{1}{\varepsilon} (1 - s_{ij}) + s_{ij}.$$

A.10 The Concentration Drag with Endogenous Markups

We postulate that

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

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

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

A.11 Quantitative Results with Endogenous Markups

I now reestimate the full model under Cournot and Bertrand competition to compare the results with the constant-markup benchmark. I follow the same estimation procedure as in Section 4, but now allowing for endogenous markups. Table A.2 reports the regression results under the three modes of competition. The model replicates the negative effect of concentration on future productivity growth under all modes of competition. However, the magnitude of the effect varies: it is smallest under constant markups, larger under Bertrand competition, and largest under Cournot competition. This ranking is consistent with the strength of the endogenous markup channel across competition modes.

Table A.1: Estimated Parameters under Different Modes of Competition

Parameter	Constant Markups	Cournot	Bertrand
α_{tail}	3.96	3.35	3.6
g	0.019	0.02	0.019
σ	0.025	0.026	0.025
λ	0.36	0.42	0.38
μ_+	19.6	19.8	19.7
μ_-	15.0	15.0	15.0
δ	0.034	0.034	0.034

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

Table A.2: Model Regressions under Different Modes of Competition

	CPED	Cournot	Bertrand
	$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$	$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$	$\ln(\text{Prod}_{t+5}) - \ln(\text{Prod}_t)$
	(1)	(2)	(3)
HHI _t sales	-0.059*** (0.006)	-0.268*** (0.013)	-0.128*** (0.008)
$\ln(\text{Prod}_t)$	-0.066*** (0.008)	-0.018*** (0.005)	-0.040*** (0.006)
Observations	200000	200000	200000
R^2	0.064	0.063	0.062
Adj. R^2	0.064	0.063	0.062

Notes: Constant markups, Cournot, and Bertrand correspond to simulated cross-sections of CES nests under different modes of competition. $\ln(\text{Prod})$ is log-labor productivity, HHI (sales/costs) based on firm sales and labor cost shares within the sector.

A.12 Measurement

I consider now how statistical agencies might (mis-)measure the granular drag. Consider the efficient case with constant markups $\mu_{ij} = \mu$ for all i . Individual firm prices are given by $P_{ij} = \mu / A_{ij}$ and quantities by $Y_{ij} = \mu_j^{-\epsilon} A_{ij}^\epsilon P_j^\epsilon \omega_j Y$. Individual firm productivities follow a geometric Brownian motion $dA_{ijt} / A_{ijt} = gdt + \sigma dW_{ijt}$. Sector level expenditure is $P_j Y_j = \omega_j P Y$

The true sectoral productivity index is $A_j = \left(\sum_{i=1}^{N_j} A_{ij}^{\epsilon-1} \right)^{\frac{1}{\epsilon-1}}$. If a statistical agency happens to know the elasticity of substitution ϵ and the individual firm prices, it can compute the true

productivity growth rate as $d \ln A_{jt} = -d \ln P_{jt}$, where $P_{jt} = \left(\sum_{i=1}^{N_j} P_{ijt}^{1-\varepsilon} \right)^{\frac{1}{1-\varepsilon}}$. We obtain:

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

Laspeyres Price Index and the Itô Integral However, statistical agencies typically do not observe all individual firm prices. Instead, they may construct a Laspeyres price index using initial period sales shares as weights:

$$P_{jt,t+\Delta t}^L := \frac{\sum_{i=1}^{N_j} P_{ijt+\Delta t} Y_{ijt}}{\sum_{i=1}^{N_j} P_{ijt} Y_{ijt}}$$

I define the Laspeyres log-change over the interval $[t, t + \Delta t]$ as

$$\Delta \ln P_{jt}^L := \sum_{i=1}^{N_j} s_{ijt} (\ln P_{ijt+\Delta t} - \ln P_{ijt}) = \sum_{i=1}^{N_j} \int_t^{t+\Delta t} s_{ijt} d \ln P_{iju}$$

Note that this is an Itô stochastic integral: the weights s_{ijt} are evaluated at the beginning of the interval and are therefore \mathcal{F}_t -measurable with respect to the filtration generated by the firm-level shocks. Equivalently, $\Delta \ln P_{jt}^L$ is obtained as the limit of left-endpoint Riemann sums,

$$\sum_{k=0}^{n-1} \sum_{i=1}^{N_j} s_{ij,t_k} (\ln P_{ij,t_{k+1}} - \ln P_{ij,t_k}) = \sum_{i=1}^{N_j} \int s_{ij,t} d \ln P_{ij,t},$$

which corresponds exactly to holding base-period quantities fixed in the Laspeyres price index. Taking the limit $\Delta t \rightarrow 0$ yields the continuous-time Laspeyres price dynamics

$$d \ln P_{jt}^L = \sum_{i=1}^{N_j} s_{ijt} d \ln P_{ijt}.$$

Using Itô's lemma on $\ln P_{ijt} = \ln \mu - \ln A_{ijt}$ gives

$$d \ln P_{jt}^L = - \left(g - \frac{\sigma^2}{2} \right) dt - \sigma \sqrt{\mathcal{H}_{jt}} dW_{jt}$$

where $\mathcal{H}_{jt} = \sum_{i=1}^{N_j} s_{ijt}^2$ and W_{jt} is a standard Wiener process. As a result, deflating by a Laspeyres price index completely eliminates the granular drag term in sectoral productivity growth that would be present if the true price index were used.

Törnqvist Price Index and the Stratonovich Integral The Törnqvist price index uses average expenditure shares over the interval. The Törnqvist log-change over $[t, t + \Delta t]$ is defined as

$$\Delta \ln P_{jt}^T := \sum_{i=1}^{N_j} \frac{1}{2} (s_{ijt} + s_{ij,t+\Delta t}) (\ln P_{ij,t+\Delta t} - \ln P_{ijt}).$$

Taking the continuous-time limit yields the Stratonovich integral. In differential form, we have

$$d \ln P_{jt}^T = \sum_{i=1}^{N_j} s_{ijt} \circ d \ln P_{ijt},$$

where \circ denotes Stratonovich integration, corresponding to midpoint evaluation of the weights.

Using the Stratonovich–Itô conversion,

$$s_{ijt} \circ d \ln P_{ijt} = s_{ijt} d \ln P_{ijt} + \frac{1}{2} d \langle s_{ij}, \ln P_{ij} \rangle_t,$$

we obtain

$$d \ln P_{jt}^T = \sum_{i=1}^{N_j} s_{ijt} d \ln P_{ijt} + \frac{1}{2} \sum_{i=1}^{N_j} d \langle s_{ij}, \ln P_{ij} \rangle_t.$$

We can rewrite the quadratic covariation term, we have

$$\sum_{i=1}^{N_j} d \langle s_{ij}, \ln P_{ij} \rangle_t = -(\varepsilon - 1) \sigma^2 (1 - \mathcal{H}_{jt}) dt, \quad \mathcal{H}_{jt} := \sum_i s_{ijt}^2,$$

so that

$$d \ln P_{jt}^T = -\left(g - \frac{1}{2} \sigma^2\right) dt - \frac{1}{2} (\varepsilon - 1) \sigma^2 (1 - \mathcal{H}_{jt}) dt - \sigma \sqrt{\mathcal{H}_{jt}} dW_{jt}.$$

That is, deflating by a Törnqvist price index recovers the granular drag term in sectoral productivity growth.

Paasche Price Index and Right-Endpoint Weights The Paasche price index holds current-period quantities fixed. The discrete Paasche price index over $[t, t + \Delta t]$ is

$$P_{jt,t+\Delta t}^P := \frac{\sum_{i=1}^{N_j} P_{ij,t+\Delta t} Y_{ij,t+\Delta t}}{\sum_{i=1}^{N_j} P_{ijt} Y_{ij,t+\Delta t}}.$$

I define the Paasche log-change over $[t, t + \Delta t]$ as

$$\Delta \ln P_{jt}^P := \sum_{i=1}^{N_j} s_{ij,t+\Delta t} (\ln P_{ij,t+\Delta t} - \ln P_{ijt}),$$

where the weights are evaluated at the end of the interval. Taking the continuous-time limit (refining the partition) yields a right-endpoint weighted stochastic integral. Rewriting the corresponding Riemann sums gives:

$$\sum_{i=1}^{N_j} s_{ij,t+\Delta t} d \ln P_{ijt} = \sum_{i=1}^{N_j} s_{ijt} d \ln P_{ijt} + \sum_{i=1}^{N_j} d \langle s_{ij}, \ln P_{ij} \rangle_t,$$

so that

$$d \ln P_{jt}^P = \sum_{i=1}^{N_j} s_{ijt} d \ln P_{ijt} + \sum_{i=1}^{N_j} d \langle s_{ij}, \ln P_{ij} \rangle_t.$$

In the efficient CES case,

$$\sum_{i=1}^{N_j} d \langle s_{ij}, \ln P_{ij} \rangle_t = -(\varepsilon - 1) \sigma^2 (1 - \mathcal{H}_{jt}) dt, \quad \mathcal{H}_{jt} := \sum_i s_{ijt}^2,$$

and therefore

$$d \ln P_{jt}^P = -\left(g - \frac{1}{2} \sigma^2\right) dt - (\varepsilon - 1) \sigma^2 (1 - \mathcal{H}_{jt}) dt - \sigma \sqrt{\mathcal{H}_{jt}} dW_{jt}.$$

That is, deflating by a Paasche price index overstates the granular drag term in sectoral productivity growth by a factor of two.

Sectoral Productivity under Different Price Indices Summarizing, the sectoral productivity

dynamics under different price indices are given by

$$\begin{aligned}
d \ln A_{jt}^{true} &= \left(g - \frac{\sigma^2}{2} \right) dt + (\varepsilon - 1) \frac{\sigma^2}{2} (1 - \mathcal{H}_{jt}) dt + \sigma \sqrt{\mathcal{H}_{jt}} dW_{jt} \\
d \ln A_{jt}^L &= \left(g - \frac{\sigma^2}{2} \right) dt + \sigma \sqrt{\mathcal{H}_{jt}} dW_{jt} \\
d \ln A_{jt}^T &= \left(g - \frac{\sigma^2}{2} \right) dt + (\varepsilon - 1) \frac{\sigma^2}{2} (1 - \mathcal{H}_{jt}) dt + \sigma \sqrt{\mathcal{H}_{jt}} dW_{jt} \\
d \ln A_{jt}^P &= \left(g - \frac{\sigma^2}{2} \right) dt + (\varepsilon - 1) \sigma^2 (1 - \mathcal{H}_{jt}) dt + \sigma \sqrt{\mathcal{H}_{jt}} dW_{jt}
\end{aligned}$$

The Törnqvist index recovers the true granular drag, while the Laspeyres index eliminates it and the Paasche index overstates it by a factor of two. This is expected since the Laspeyres index uses base-period weights and overstates inflation as it does not account for substitution (reallocation) effects. The Paasche index, on the other hand, uses future-period weights and understates inflation. The Törnqvist and Fisher indices, are "superlative" indices and exactly capture the reallocation effects, and thus the granular drag.

B Imperfect Labor Reallocation

B.1 Setup with Firm-Specific Fixed Labor

In the main text, I assume that individual firms produce with a linear production function $Y_{ij} = A_{ij}L_{ij}$, where labor inputs L_{ij} are freely reallocated across firms within a sector. In this appendix, I consider the case where labor inputs are fixed. In particular, I follow [Baqae and Farhi \(2019\)](#) and I assume that each firm i in sector j has a firm specific fixed labor endowment \bar{L}_{ij} which is

$$Y_{ij} = A_{ij}L_{ij}^\alpha \bar{L}_{ij}^{1-\alpha} \tag{B.7}$$

with $\alpha \in [0, 1]$.²² When $\alpha = 1$, firms produce with only variable labor input and we recover the main text case. When $\alpha = 0$, labor inputs are completely fixed. The associated cost function is therefore

$$C_{ij}(Y_{ij}) = \left(\frac{Y_{ij}}{A_{ij}\bar{L}_{ij}^{1-\alpha}} \right)^{1/\alpha},$$

²²I thank Basile Grassi for suggesting this exercise.

and marginal cost is given by

$$\begin{aligned} MC_{ij}(Y_{ij}) &= \frac{\partial C_{ij}(Y_{ij})}{\partial Y_{ij}} \\ &= \frac{1}{\alpha} \left(\frac{Y_{ij}}{A_{ij} \bar{L}_{ij}^{1-\alpha}} \right)^{\frac{1-\alpha}{\alpha}} \left(A_{ij} \bar{L}_{ij}^{1-\alpha} \right)^{-1}. \end{aligned} \quad (\text{B.8})$$

When $\alpha = 1$, marginal cost is proportional to A_{ij}^{-1} and independent of output, recovering the linear-production case studied in the main text. Firms face CES demand with elasticity $\varepsilon > 1$ and charge a constant markup over marginal cost,

$$P_{ij} = \mu MC_{ij}, \quad \mu = \frac{\varepsilon}{\varepsilon - 1}.$$

Since firms charge an homogeneous markup over marginal cost, the allocation of flexible labor is efficient. Changes in sectoral productivity are defined as

$$d \ln A_j \equiv d \ln Y_j - \alpha d \ln L_j$$

where $L_j = \sum_{i=1}^{N_j} L_{ij}$ is total flexible labor input in sector j .

B.2 Sectoral Productivity and Its SDE

Define effective firm productivity $\tilde{A}_{ijt} \equiv A_{ijt} \bar{L}_{ij}^{1-\alpha}$ and let

$$A_{jt} \equiv \left(\sum_{i=1}^{N_j} \tilde{A}_{ijt}^\varphi \right)^{1/\varphi}, \quad \varphi \equiv \frac{\varepsilon - 1}{\varepsilon - (\varepsilon - 1)\alpha}. \quad (\text{B.9})$$

Let weights and the associated Herfindahl be

$$s_{ijt} \equiv \frac{\tilde{A}_{ijt}^\varphi}{\sum_{k=1}^{N_j} \tilde{A}_{kjt}^\varphi}, \quad \mathcal{H}_{jt} \equiv \sum_{i=1}^{N_j} s_{ijt}^2. \quad (\text{B.10})$$

Assume firm productivity follows

$$\frac{dA_{ijt}}{A_{ijt}} = g dt + \sigma dW_{ijt}, \quad dW_{ijt} dW_{kjt} = \mathbf{1}\{i = k\} dt, \quad (\text{B.11})$$

and \bar{L}_{ij} is time-invariant, so $d\tilde{A}_{ijt}/\tilde{A}_{ijt} = dA_{ijt}/A_{ijt}$. Applying Itô's lemma to (B.9) yields

$$\begin{aligned} d \log A_{jt} &= \left(g + \frac{\varphi-1}{2} \sigma^2 - \frac{\varphi}{2} \sigma^2 \mathcal{H}_{jt} \right) dt + \sigma \sum_{i=1}^{N_j} s_{ijt} dW_{ijt} \\ &= \left(g - \frac{1}{2} \sigma^2 + \frac{\varphi}{2} \sigma^2 (1 - \mathcal{H}_{jt}) \right) dt + \sigma \sum_{i=1}^{N_j} s_{ijt} dW_{ijt}. \end{aligned} \quad (\text{B.12})$$

Using $\mathbb{E}_t[dW_{ijt}] = 0$, (B.12) implies

$$\gamma_{jt} = g - \frac{1}{2} \sigma^2 + \frac{\varphi}{2} \sigma^2 (1 - \mathcal{H}_{jt}). \quad (\text{B.13})$$

We see that imperfect labor reallocation ($\alpha < 1$) dampens reallocation gains by reducing φ , while concentration still enters only through \mathcal{H}_{jt} . When $\alpha = 0$, we obtain the minimum value $\varphi = (\varepsilon - 1)/\varepsilon$, which is still positive. Even if labor inputs are completely fixed, productivity growth still benefits from reallocation effects, and hence the granular drag remains present.

C Beyond CES

To extend the analysis beyond CES preferences, I consider the demand system introduced by [Hanoch \(1971\)](#) and extended by [Chodorow-Reich et al. \(2025\)](#). For simplicity, I focus on a single-sector economy. The representative consumer aggregates goods $\vec{Y} = (Y_1, \dots, Y_N)$ into a composite consumption index $Y = F(\vec{Y})$, defined implicitly by

$$\sum_{i=1}^N \phi_i \frac{\left(\frac{Y_i}{Y}\right)^{1-\frac{1}{\varepsilon_i}} - 1}{1 - \frac{1}{\varepsilon_i}} + \phi_0 = 0, \quad (\text{C.14})$$

where $\phi_i > 0$ are taste shifters, $\varepsilon_i > 1$ govern substitution elasticities, and ϕ_0 is a normalization constant. CES preferences are nested as the special case $\varepsilon_i = \varepsilon$ for all i . The consumer minimizes nominal expenditure

$$E = \sum_{i=1}^N P_i Y_i$$

subject to achieving utility level Y . Let $q_i := Y_i/Y$ denote quantity shares. Expenditure minimization yields the inverse-demand system

$$P_i = \frac{\lambda}{Y} \phi_i q_i^{-1/\varepsilon_i}, \quad (\text{C.15})$$

where λ is the Lagrange multiplier on the utility constraint. Define the demand index

$$d^{-1} := \sum_{i=1}^N \phi_i q_i^{1-\frac{1}{\varepsilon_i}}. \quad (\text{C.16})$$

Multiplying (C.15) by q_i and summing over i implies $\sum_i P_i q_i = \lambda d$. It is therefore convenient to define the *ideal expenditure deflator*

$$P^Y := \sum_{i=1}^N P_i q_i, \quad (\text{C.17})$$

so that the inverse-demand system can be written compactly as

$$\frac{P_i}{P^Y} = d \phi_i q_i^{-1/\varepsilon_i}. \quad (\text{C.18})$$

Define the *substitution price index*

$$P := \frac{P^Y}{d}, \quad (\text{C.19})$$

which implies

$$\frac{P_i}{P} = \phi_i q_i^{-1/\varepsilon_i}. \quad (\text{C.20})$$

There are two relevant price indices. The index P governs substitution across goods, while P^Y deflates nominal expenditure to recover welfare.

Technology and productivity. Each good is produced using labor according to

$$Y_i = A_i L_i,$$

with the wage normalized to one. Total labor is $L = \sum_i L_i$, and aggregate productivity is defined as

$$A := \frac{Y}{L} = \left(\sum_{i=1}^N \frac{q_i}{A_i} \right)^{-1}. \quad (\text{C.21})$$

Pricing. Firms set prices as a constant markup over marginal cost,

$$P_i = \frac{\mu_i}{A_i}, \quad (\text{C.22})$$

where markups $\mu_i \geq 1$ may be heterogeneous across goods but are constant over time. When $\mu_i = \mu$ for all i , relative prices coincide with planner shadow prices and the allocation is efficient.

Aggregate productivity and markups. Let

$$P^Y := \sum_{i=1}^N P_i q_i \quad (\text{C.23})$$

denote the ideal expenditure deflator and define the *aggregate markup*

$$\mu := \left(\sum_{i=1}^N \frac{s_i}{\mu_i} \right)^{-1}. \quad (\text{C.24})$$

Aggregate productivity can then be written as

$$A = \frac{Y}{L} = \frac{\mu}{P^Y}, \quad d \ln A = d \ln \mu - d \ln P^Y. \quad (\text{C.25})$$

When markups are common, $\mu = \mu_i$ and aggregate productivity log-changes depend only on the expenditure price index. Firm-level productivities follow

$$d \ln A_{it} = \left(g - \frac{1}{2} \sigma^2 \right) dt + \sigma dW_{it}, \quad dW_{it} dW_{jt} = \mathbf{1}\{i = j\} dt. \quad (\text{C.26})$$

C.0.1 Common Markups

I first consider the benchmark case in which all firms charge a common markup,

$$\mu_i = \mu \quad \text{for all } i. \quad (\text{C.27})$$

In this case, relative prices coincide with planner shadow prices and the allocation is efficient. With common markups, revenue and labor shares coincide, $s_i = \tilde{s}_i$ for all i , and the aggregate

markup is constant, $\mu = \mu_i$. Aggregate productivity therefore satisfies

$$d \ln A = -d \ln P^Y, \quad (\text{C.28})$$

where $P^Y = \sum_i P_i q_i$ is the ideal expenditure deflator.

Applying Itô's lemma to (C.28) yields

$$d \ln A = \left[g - \sigma^2 + \frac{\sigma^2}{2} \sum_{i=1}^N s_i \varepsilon_i - \frac{\sigma^2}{2} \frac{1}{\bar{\varepsilon}_s} \sum_{i=1}^N s_i^2 \varepsilon_i^2 + \frac{\sigma^2}{2} \sum_{i=1}^N s_i^2 \right] dt + \sigma \sum_{i=1}^N s_i dW_{it}, \quad (\text{C.29})$$

where $\bar{\varepsilon}_s := \sum_i s_i \varepsilon_i$ is the revenue-share-weighted mean elasticity. With constant elasticity $\varepsilon_i = \varepsilon$ for all i , for the granular drag to operate we required $\varepsilon > 1$. Here, the granular drag operates when $\frac{\sum_i s_i^2 \varepsilon_i^2}{\sum_i s_i^2} > \bar{\varepsilon}_s$. Two extreme cases are useful again. First, when there is a single dominant firm, we again obtain $d \ln A = \left(g - \frac{\sigma^2}{2} \right) dt + \sigma dW_{1t}$. Second, when all firms have equal revenue shares $s_i = 1/N$,

$$d \ln A = \left[g - \frac{\sigma^2}{2} + \frac{\sigma^2}{2} \frac{1}{N} \sum_{i=1}^N (\varepsilon_i - 1) \right] dt + \frac{\sigma}{\sqrt{N}} dW_t,$$

where W_t is a standard Wiener process. As long as the average elasticity exceeds one, expected productivity growth is higher than in the monopolistic case, as there will be reallocation.

The derivation of (C.29) proceeds as follows: Since markups are common, $d \ln A = -d \ln P^Y$ by (C.29). $P^Y(\vec{P}, Y)$ is the expenditure price index associated with HCES preferences. Let $x_i := \ln A_i$. Since $d \ln P_i = -d \ln A_i$, we have

$$d \ln A = \sum_{i=1}^N \frac{\partial \ln A}{\partial x_i} dx_i + \frac{1}{2} \sum_{i=1}^N \frac{\partial^2 \ln A}{\partial x_i^2} (dx_i)^2.$$

By Shephard's lemma,

$$\frac{\partial \ln A}{\partial x_i} = s_i,$$

Furthermore, the Hicksian (compensated) elasticities of demand are given by

$$e_{ij} := \frac{\partial \ln Y_i}{\partial \ln P_j} \Big|_Y = \varepsilon_i \left(\frac{s_j \varepsilon_j}{\bar{\varepsilon}_s} - \mathbf{1}\{i = j\} \right)$$

where $\bar{\varepsilon}_s := \sum_{\ell=1}^N s_\ell \varepsilon_\ell$. As a result, the price elasticity of revenue shares satisfies

$$\frac{\partial s_i}{\partial \ln P_j} = s_i (\mathbf{1}\{i = j\} + e_{ij} - s_j).$$

Hence,

$$\frac{\partial^2 \ln A}{\partial x_i^2} = -\frac{\partial s_i}{\partial \ln P_i} = -s_i (1 + e_{ii} - s_i),$$

and substituting $e_{ii} = \varepsilon_i \left(\frac{s_i \varepsilon_i}{\bar{\varepsilon}_s} - 1 \right)$ gives

$$\frac{\partial^2 \ln A}{\partial x_i^2} = -s_i \left(1 + \varepsilon_i \left(\frac{s_i \varepsilon_i}{\bar{\varepsilon}_s} - 1 \right) - s_i \right).$$

Using the dynamics of x_i ,

$$dx_i = \left(g - \frac{1}{2} \sigma^2 \right) dt + \sigma dW_i,$$

with $(dx_i)^2 = \sigma^2 dt$ and $dx_i dx_j = 0$ for $i \neq j$, we have

$$\begin{aligned} d \ln A &= \sum_{i=1}^N s_i dx_i + \frac{1}{2} \sum_{i=1}^N \left(-s_i \left(1 + \varepsilon_i \left(\frac{s_i \varepsilon_i}{\bar{\varepsilon}_s} - 1 \right) - s_i \right) \right) \sigma^2 dt \\ &= \left(g - \frac{1}{2} \sigma^2 \right) dt + \sigma \sum_{i=1}^N s_i dW_i - \frac{1}{2} \sigma^2 \sum_{i=1}^N s_i \left(1 + \varepsilon_i \left(\frac{s_i \varepsilon_i}{\bar{\varepsilon}_s} - 1 \right) - s_i \right) dt. \end{aligned}$$

Simplifying the drift term yields the expression in (C.29).

Example (two elasticities). Suppose firms belong to two groups G_1 and G_2 with $\varepsilon_i \in \{\varepsilon_1, \varepsilon_2\}$. Let $S_g := \sum_{i \in G_g} s_i$ and $\mathcal{H}_g := \sum_{i \in G_g} (s_i / S_g)^2$ denote within-group concentration. Then $\bar{\varepsilon}_s = \varepsilon_1 S_1 + \varepsilon_2 S_2$ and

$$\sum_i s_i^2 = S_1^2 \mathcal{H}_1 + S_2^2 \mathcal{H}_2, \quad \sum_i s_i^2 \varepsilon_i^2 = \varepsilon_1^2 S_1^2 \mathcal{H}_1 + \varepsilon_2^2 S_2^2 \mathcal{H}_2.$$

Substituting into (C.29) yields

$$\mathbb{E} \left[\frac{1}{dt} d \ln A \right] = g - \sigma^2 + \frac{\sigma^2}{2} (\varepsilon_1 S_1 + \varepsilon_2 S_2) - \frac{\sigma^2}{2} \frac{\varepsilon_1^2 S_1^2 \mathcal{H}_1 + \varepsilon_2^2 S_2^2 \mathcal{H}_2}{\varepsilon_1 S_1 + \varepsilon_2 S_2} + \frac{\sigma^2}{2} (S_1^2 \mathcal{H}_1 + S_2^2 \mathcal{H}_2). \quad (\text{C.30})$$

As the share of either group increases, we recover the CES expression.

D Data

D.1 Swedish Firm Data

D.1.1 Data Construction

This appendix describes the construction of the final Swedish firm-level sample from the Serrano database. The raw data cover most legal forms in the Swedish business community, with full coverage of limited liability companies. To ensure consistency and comparability over time, I apply the following steps: Nominal flows such as sales, value added, personnel costs, production costs, and other operating expenses are deflated using the GDP deflator. I drop observations with missing or zero values for sales, employment, or value added. In addition, I exclude firm-years with imputed accounts flagged in the Serrano database (codes 11, 51, 91, and 99). Firms are required to have positive deflated sales above 500.000 SEK, positive deflated personnel and production costs, positive deflated value added, and positive total assets. To avoid inactive entities, I further exclude firms that never exceed one employee over their lifetime. To mitigate extreme reporting errors, I winsorize key accounting ratios—value added to materials, capital, and labor; materials to capital and labor; and capital to labor—at the 1st and 99th percentiles within each year cell. Observations flagged as outside this range are dropped. Following standard practice, I exclude sectors with large public or financial components. Specifically, I drop two-digit SNI codes 35–39 (utilities), 64–66 (finance and insurance), 68 (real estate), 84–88 (public administration, education, and health), 90–96 (arts and other services), and 97–99 (households and extraterritorial organizations). Within each year–industry, I compute firms’ market shares from nominal sales and costs shares from personnel costs, and construct measures of concentration (Herfindahl–Hirschman Index and maximum market share). Finally, I restrict the sample to 5-digit industries with at least 20 firms in each year. After these restrictions, the final dataset consists of 2.629.458 firm-year observations. Table D.3 presents descriptive statistics for the main variables in the final sample.

Table D.3: Descriptive Statistics of the Swedish Firm-Level Data

	Sales	Value added	Employment	Personnel costs
N	2629458	2629458	2629458	2629458
Mean	24,766.7	7,229.2	11.3	5,513.9
SD	207,764.4	54,013.3	77.0	41,185.1
P25	2,256.2	854.0	2.0	661.0
P50	5,117.6	1,841.0	4.0	1,421.0
P75	13,263.9	4,518.0	8.0	3,492.0
P90	35,913.4	11,362.0	18.0	8,703.0
P95	70,224.7	20,906.0	32.0	16,018.0
P99	303,264.8	78,734.0	108.0	59,913.9
Max	54,028,144.8	9,473,085.0	20,699.0	8,372,000.0

Notes: Sales, value added, and personnel costs in 1000 SEK, deflated using the GDP deflator.

D.2 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 D.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. A concern is that the measured exit of large firms might reflect mergers and acquisitions, or changes in legal form, rather than true market exit. In principle, the Serrano database tracks such events, for which I drop exits due to mergers and acquisitions or registering of a new legal entity. To further test whether exits are

"true" exits, I regress the yearly change in 5-digit industry sales on the sales of large firms exiting the previous year, as reported in Table D.4. The coefficient is large and close to 1, suggesting that exits of large firms correspond to actual market exits rather than mergers or legal form changes.

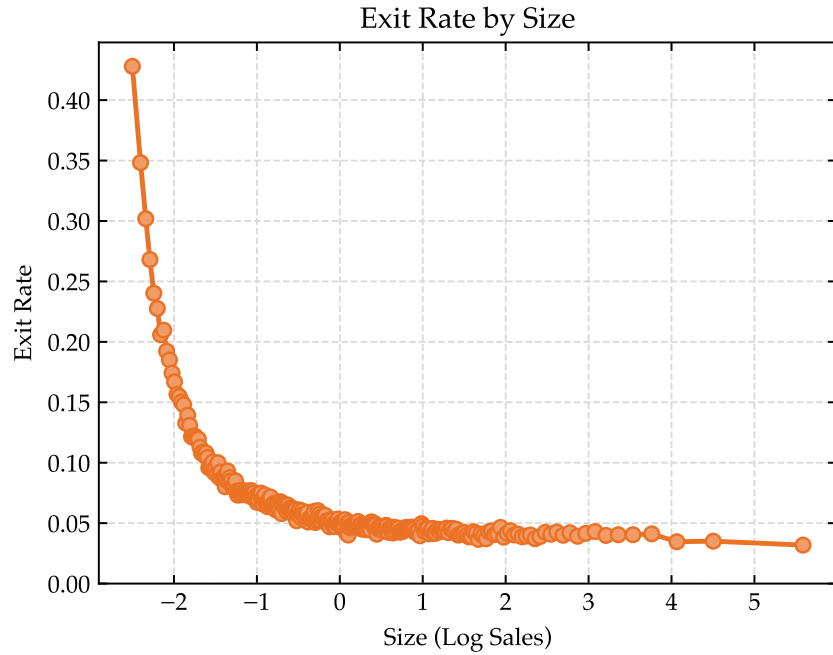


Figure D.1: Exit Hazard by Log-Sales, conditioning on year fixed effects

Table D.4: Exit Hazard Regression Results

	Sales _t – Sales _{t-1}	
	(1)	(2)
Exit Sales _{t-1} ^{Top 2}	-0.852*** (0.076)	
Entry Sales _t ^{Top 2}	0.967*** (0.072)	
Exit Sales _{t-1} ^{Top 4}		-0.832*** (0.077)
Entry Sales _t ^{Top 4}		0.974*** (0.063)
Year FE	x	x
5-digit Industry FE	x	x
Observations	8964	8964
S.E. type	by: industry	by: industry
R ²	0.154	0.157
R ² Within	0.046	0.050

SEs clustered by 5-digit industry and year.

D.3 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. 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 D.5 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.

Table D.5: U.S. Industry-Level Regressions of 5-Year Productivity Growth on Concentration

	$\Delta_5 \ln(\text{TFP})$							
	(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

Unit of observation is 6-digit NAICS in the U.S. 1972-2012 in windows of 5-years from the replication package of [Ganapati \(2021\)](#). CR4 based on firm sales shares within the industry. TFP is 5 factor TFP. Standard errors clustered by industry and year in parentheses.

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.](#)

D.4 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 D.6, 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 (0.126)	0.007 (0.125)	0.134 (0.077)	0.229* (0.099)
HHI _t gap (sales - costs)	-0.762*** (0.180)	-0.493** (0.139)	-0.556*** (0.107)	-0.434** (0.136)
$\ln(\text{Prod}_t)$		-0.198*** (0.026)		-0.042* (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 NACE.

Table D.6: CompNet Industry-Level Regressions of 5-Year Productivity Growth on Concentration

[Click here to go back to Section 4.2.](#)