Recent Changes in Firm Dynamics and the Nature of Rising Firm-Size Dispersion*

Markus Kondziella†

June 16, 2026

Abstract

In Swedish administrative data, I document and decompose a rise in firm-size dispersion over 1997–2017. Rising dispersion among same-aged firms accounts for most of the increase, concentrated among older firms and absent among the young — pointing to forces that accumulate with age. In a firm-dynamics model with entry, exit, and permanent productivity differences, a widening productivity gap reproduces this age profile; cheaper expansion or entry barriers do not. The wider gap slows long-run growth but raises the level of productivity through reallocation; accounting for the transition, the welfare cost is only 0.025%, far below what balanced-growth-path comparisons suggest.

Keywords: Firm dynamics, Firm-size dispersion, Administrative data

JEL codes: D22, L11, L25, E23, O47

*I am very grateful to Philippe Aghion, Gualtiero Azzalini, Timo Boppart, Maarten De Ridder, Ryan Decker, Niklas Engbom, Sergio Feijoo, Jonas Gathen, Sampreet Goraya, Basile Grassi, Winfried Koeniger, Per Krusell, Kieran Larkin, Kurt Mitman, Alessandra Peter, Michael Peters, Filip Rozsypal, Carolina Villegas-Sanchez, Anthony Savagar, Joshua Weiss, as well as seminar, conference, and workshop participants at various institutions for many useful comments.

†University of St. Gallen and Swiss Finance Institute. E-mail: markus.kondziella@unisg.ch

1 Introduction

Firm-size dispersion has risen substantially across advanced economies over the past decades (Autor et al., 2020; Akcigit and Ates, 2021).[1] Understanding what drives this trend matters, because the underlying mechanisms differ sharply depending on whether dispersion rises through the aging of the firm population or through the divergence of same-aged firms. This paper decomposes the rise in firm-size dispersion into three observable margins of firm dynamics and shows that the overwhelming share of its increase is accounted for by rising heterogeneity among firms of the same age, consistent with widening productivity dispersion.

In comprehensive administrative data covering the universe of Swedish firms between 1997 and 2017, cross-sectional firm-size dispersion rose substantially: the interquartile range of log sales widened by 0.5 log points, raising the ratio of the 75th to the 25th size percentile from 7.4 to 12.2. To identify what underlies this rise, I apply the law of total variance, decomposing dispersion in log firm size into three margins of firm dynamics: differences in mean size across firm ages, the firm-age distribution, and size dispersion conditional on age. For the US, Hopenhayn et al. (2022) and Karahan et al. (2024) document two facts about the first two margins: the age distribution has shifted toward older firms, and mean size conditional on age has remained stable. I confirm both in Swedish data — the share of firms younger than two fell from 14.9% to 10.4%, the share aged ten or older rose from 44.9% to 56.4%, and mean size conditional on age barely changed. The third margin is this paper's new fact: size dispersion conditional on age rose substantially among older firms — the interquartile range of log sales widened by about 0.3 log points — while staying flat among the young. These patterns are robust across sectors and alternative size measures including employment. Through the lens of the decomposition, which is model-free and holds exactly in the data, this rising within-age dispersion accounts for 89.2% of the increase in cross-sectional dispersion, while the shift in the age distribution accounts for the remainder and changes in mean size by age contribute negligibly. Despite dominating the rise, size dispersion conditional on age has received comparatively little attention.

What drives the rise in within-age dispersion? Its concentration among older firms is the key discriminating fact: it points to forces whose effect accumulates with firm age, such as productivity divergence (Aghion et al., 2023), and is hard to reconcile with three alternatives. Two of them — changes in the composition of entrants, and a uniform rise in measurement error — would affect dispersion among the young as

[1] See also Grullon et al. (2019); Decker et al. (2016*a*, 2020); Gourio et al. (2014); Chen et al. (2023). De Loecker et al. (2020) further document a reallocation of market shares from low- to high-markup firms, consistent with rising firm-size dispersion.

well; the flatness of dispersion among the young is hence evidence against both. A third, weakening selection over the life cycle as low interest rates keep unproductive incumbents alive, would inflate dispersion most in the capital-intensive sectors that depend on debt financing; instead the rise is weakest precisely there.

Motivated by these facts, I ask what force can generate them. Aghion et al. (2023) develop a model of creative destruction with permanent heterogeneity in firm productivity. Their model abstracts from firm entry and exit, leaving mean size by age, the firm-age distribution, and size dispersion conditional on age undefined. I embed this mechanism — permanent productivity heterogeneity — into a quality-ladder model of firm dynamics with entry and exit in the spirit of Klette and Kortum (2004), in which these objects are well defined. I calibrate the model to match firm dynamics around the turn of the millennium, which fixes a baseline balanced growth path with no productivity gap between firms. I then ask how much of the subsequent changes a single force — a widening productivity gap — can explain: holding all other parameters at their baseline values, I raise the gap to match the 4.5 percentage-point fall in the share of young firms and the rise in size dispersion conditional on age among older firms. A modest rise matches both targets quantitatively, despite the fit being overidentified. The mechanism is differential expansion: more productive firms expand faster, so size differences cumulate with tenure, raising within-age dispersion among older firms while leaving young-firm dispersion flat — the age signature in the data. Without being targeted, it also reproduces the roughly 12 percentage-point rise in the share of old firms and slows long-run productivity growth by 0.062 percentage points relative to its baseline of 1.5%.

The fit is specific to the productivity gap. As a robustness check, I instead fit the same two targets by lowering the cost of firm expansion or by raising the cost of entry — natural alternatives from the literature on declining business dynamism. Neither reproduces the age signature: both raise dispersion among firms of all ages and steepen the mean size–age profile, counter to the data. A cheaper expansion technology also accelerates aggregate growth, whereas a widening gap slows it, consistent with the productivity slowdown that has accompanied rising dispersion.

This growth slowdown is central to the model's welfare implications, which I evaluate by solving the full transition between balanced growth paths. The wider gap reallocates market shares toward more productive firms, raising the level of aggregate productivity, but slower long-run growth is costly to consumers. Accounting for the transition, the level gains offset most of the growth loss, leaving a welfare cost of only 0.025% in consumption-equivalent terms — far below the 1.24% implied by comparing balanced growth paths alone, which misses the productivity gains realized along the way. This also contrasts with the 3.3% loss in Aghion et al. (2023), who discipline the rise in the productivity gap to the entire long-run growth slowdown rather than, as here, to the rise in firm-size dispersion.

Further literature. This paper contributes to a literature on the drivers of recent macroeconomic trends — rising firm-size dispersion, declining business dynamism, and slowing productivity growth (Aghion et al., 2023; De Ridder, 2024; Akcigit and Ates, 2023; Liu et al., 2022; Peters and Walsh, 2021). Before asking which force is responsible, one must establish which margins of firm dynamics actually move; the decomposition supplies these disciplining moments, requiring any candidate mechanism to generate rising within-age dispersion among incumbents, concentrated at older ages.

A separate literature uses models of firm dynamics to explain the firm-size distribution itself, and in particular its heavy, approximately Pareto tail, with dispersion arising from idiosyncratic growth, selection, and entry (Klette and Kortum, 2004; Luttmer, 2007; Gabaix, 2009). That work concerns the *level* of firm-size dispersion — the shape of the cross-sectional distribution at a point in time. This paper concerns instead its *change*: how dispersion has risen over time, and which margins of firm dynamics account for that rise.

The findings also relate to Decker et al. (2016*b*), who document a decline in the dispersion of firm *growth rates*. Rising dispersion in size *levels* conditional on age is consistent with their finding if growth has become more persistent and type-driven — the same firms increasingly experiencing high or low growth. Decker et al. (2020) provide direct evidence that firm growth has become less responsive to shocks, supporting this interpretation.[2]

Finally, the model's mechanism — a widening dispersion in firm productivity — is corroborated by direct empirical evidence. Andrews et al. (2016) show that both labor productivity and multifactor productivity diverged between the most productive firms, the global frontier, and the rest throughout the 2000s, a gap robust to controls for markups. Berlingieri et al. (2017) document the same divergence in labor and multifactor productivity within many OECD countries, and link the growing gap between high- and low-productivity firms to digitalization.

The remainder of the paper is organized as follows. Section 2 introduces a model-free decomposition of firm-size dispersion based on the law of total variance. Section 3 de-

[2] In the model in this paper, rising dispersion in permanent firm productivity raises the dispersion of the type-specific, persistent component of firm growth, but at the same time reduces dispersion in the idiosyncratic component: creative destruction falls, so product loss becomes rarer, and the economy is increasingly populated by large firms operating many products, for which the loss of a single product is proportionally smaller and growth correspondingly less random. Other forces, such as rising labor adjustment costs (Decker et al., 2020), could also account for the decline in the dispersion of growth rates.

scribes the administrative data and measurement choices. Section 4 documents trends in firm-size dispersion, decomposes its rise, and rules out alternative drivers. Section 5 develops a model of firm dynamics with permanent productivity heterogeneity, and Section 6 embeds the rising productivity gap, runs a horse race against alternative shifters, and computes the transition and welfare effects. Section 7 concludes.

2 A decomposition of firm-size dispersion

To structure the empirical analysis, I begin with a model-free decomposition of firm-size dispersion. Consider the law of total variance, $\text{Var}(Y) = \text{Var}(\mathbb{E}[Y|X]) + \mathbb{E}[\text{Var}(Y|X)]$, which decomposes the total dispersion in Y into (i) variation across groups defined by X and (ii) variation within those groups. Setting $Y \equiv \ln s_f$ and $X \equiv a_f$, where s_f and a_f denote firm size and age, respectively, this decomposition partitions cross-sectional dispersion in log firm size into two components: differences in average size across firm ages and size dispersion conditional on age. Specifically,

$$\text{Var}\left(\ln s_f\right) = \sum_{a_f} p(a_f)\left[\mathbb{E}[\ln s_f|a_f] - \sum_{a_f} p(a_f)\mathbb{E}[\ln s_f|a_f]\right]^2 + \sum_{a_f} p(a_f)\text{Var}(\ln s_f|a_f), \tag{1}$$

where $p(a_f)$ denotes the share of firms of age a_f. The definition of firm age can be either discrete (e.g., integer years) or based on broader age bins; the decomposition applies in either case.

Although it is stated in terms of the variance, the decomposition more generally reveals that cross-sectional firm-size dispersion is shaped by three margins: (i) differences in mean size across firm ages, $\mathbb{E}[\ln s_f|a_f]$; (ii) dispersion in size within age groups, $\text{Var}(\ln s_f|a_f)$; and (iii) the age distribution, $p(a_f)$. Accordingly, total dispersion can increase for three distinct reasons: greater divergence in mean size across ages, increased within-age heterogeneity, or shifts in the age distribution toward older firms, which tend to be larger on average and exhibit greater within-age dispersion.

How have these components evolved over time? The next sections examine each in turn. I first describe the data, then turn to the empirical trends.

3 Data

The data come from Statistics Sweden's main firm-level dataset, Företagens Ekonomi, which contains annual balance sheet and income statement information for the universe of Swedish firms over the period 1997–2017. It includes firm-level variables such as sales, assets, intermediate inputs, and employment, as well as information on industry and legal form. I restrict the sample to firms in the private sector that eventually employ at least one worker. For a detailed description of the data, see Section A in the Supplemental Appendix.

Firm age is defined as the number of years since the first worker — either employee or self-employed — joins the firm. This information comes from the auxiliary dataset Registerbaserad Arbetsmarknadsstatistik, which covers the universe of employer-employee matches and extends back to 1992. This allows me to observe untruncated firm age for all firms founded from 1993 onward.

All nominal variables are deflated to 2017 Swedish kronor (SEK) using the GDP deflator. The results are virtually unchanged when using sector-specific deflators; moreover, sectoral deflation does not affect measures of within-sector dispersion in log firm size.

Table 1: Summary statistics

	25th Pct.	50th Pct.	75th Pct.	Mean	SD
*Sales**	0.4	1.4	4.3	10.9	62.1
*Intermediate Inputs**	0.2	0.6	1.8	4.9	34.8
*Capital stock**	0.0	0.2	0.8	4.6	96.2
Employment	0	1	3	4.9	29.3
*Wage bill**	0.0	0.2	0.9	1.8	11.0

Notes: variables marked with * are in units of million 2017-SEK (1 SEK ≈ 0.1 US dollars). The capital stock is defined as fixed assets minus depreciation. Statistics are computed on the pooled sample of firm-years over 1997–2017, comprising 7,590,605 observations.

The final dataset comprises 7,590,605 firm-year observations. Summary statistics for the pooled sample of firm-years are reported in Table 1. The median firm has sales of approximately 1.4 million SEK (1 SEK ≈ 0.1 USD) and employs one full-time equivalent worker. The distributions of sales and inputs are highly right-skewed: mean sales (10.9 million SEK) and employment (4.9 workers) substantially exceed their respective 75th percentiles. A nontrivial fraction of firms report zero capital stock (fixed assets net of depreciation), employment, or wage bill, as reflected in zero-valued 25th percentiles for these variables. Zero employment is common among the self-employed, which is typical in the early years of a firm's life, before any additional workers are hired. This feature has implications for the choice of the firm-size measure.

I use firm sales as the baseline measure of size and provide robustness checks using alternative proxies.[3] Sales offer two advantages. First, the incidence of zero sales is negligible, avoiding the log-of-zero problem that arises with employment. Second, sales provide a more continuous measure of firm size, whereas employment is highly discrete. For example, the 25th, 50th, and 75th percentiles of employment are 0, 1, and 3, respectively, limiting the informativeness of dispersion measures such as the interquartile range (IQR). While the IQR is the primary dispersion measure used in the paper, I also report results using the standard deviation. To mitigate the influence of outliers, sales are winsorized at the 0.1 and 99.9 percent levels.[4]

4 Size dispersion and firm dynamics in the data

This section examines cross-sectional size dispersion and the three margins of the decomposition in eq. (1) over time. All margins are documented in the same data, allowing for an exact decomposition of firm-size dispersion.

4.1 Firm-size dispersion

Figure 1 shows cross-sectional firm-size dispersion over time, measured by the interquartile range (IQR) of log sales. The IQR rises smoothly from 2.0 to 2.5 log points, implying that the 75th-to-25th percentile size ratio grew from 7.4 to 12.2.

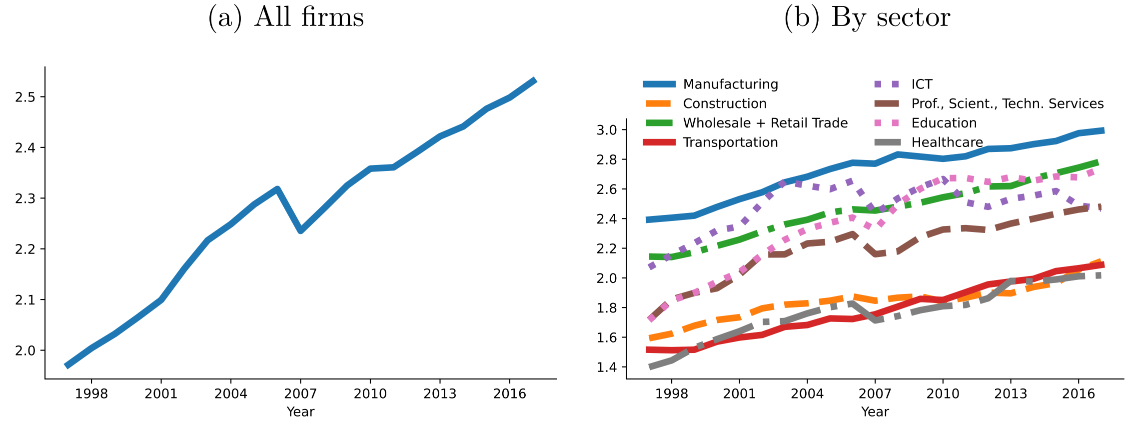
The right panel shows the same dispersion measure by sector. The increase in dispersion is shared across all sectors and is of similar magnitude in each, with most displaying the same 0.5 log point rise shown in the left panel. Hence, the rise in aggregate dispersion is not driven by differential trends across sectors.

The advantage of the IQR is that it does not require dropping firms with zero sales, since the 25th percentile of the sales distribution is strictly positive. Other dispersion measures such as the standard deviation would require dropping observations with zero sales. Beyond this practical advantage, the IQR is robust to the extreme right-skewness of the sales distribution and captures dispersion in the bulk of the firm population rather than being driven by the tails. I show that results are unchanged when computing size dispersion using the standard deviation after dropping firms with zero sales. This is because changes in the tails are not systematically different from changes in the bulk of the distribution, and dropping firms with zero sales introduces negligible selection given its rarity. The picture is virtually unchanged when

[3] Throughout, sales are interpreted as a proxy for firm size rather than market share.

[4] The filter is applied on the pooled sample after deflating.

Figure 1: Cross-sectional firm-size dispersion



Notes: The figure shows the interquartile range (IQR) defined as the difference between the 75th and 25th percentile of the log sales distribution. The left panel shows the IQR for all firms; the right panel shows the IQR by sector.

dispersion is measured by the standard deviation instead (Figure A-1, Supplemental Appendix): the standard deviation increases by 0.4 log points, close to the 0.5 log point increase in the IQR.

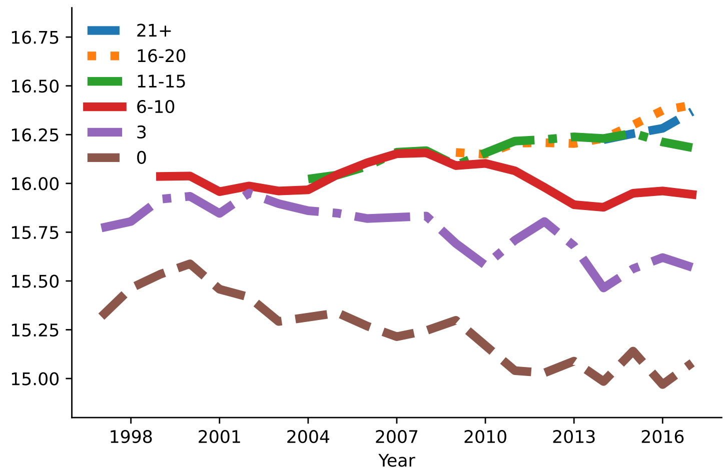
4.2 Firm dynamics

Firm size conditional on age Figure 2 shows the log of the average firm size by firm age.[5] Two patterns are noteworthy. First, there are stark level differences across ages: older firms are, on average, much larger than younger firms. At the start of the sample, the average firm size among firms aged 6–10 is 0.501 log points higher than among entrants (age zero). Second, mean size conditional on age is stable over time. Young firms, represented in the graph by firms aged zero or three, show a relatively stable or slightly negative trend in average size. Firms of older ages show no systematic trend over time. Similar patterns were first documented by Hopenhayn et al. (2022) and Karahan et al. (2024) for the US and by Engbom (2023) for Sweden.

Firm-age distribution Figure 3 shows two margins of the firm-age distribution: the young-firm share in the left panel and the share of mature firms in the right panel. The young-firm share, defined as the share of firms aged zero (entrants) or one, declined from 14.9% in 1997 to 10.4% in 2017. At the same time, the share of

[5] Note that the decomposition in eq. (1) contains the average of the log firm size rather than the log of the average firm size. I show the latter to avoid dropping firms with zero sales.

Figure 2: Firm size conditional on age



Notes: The figure shows the log of the average firm size (sales) by firm age indicated in the legend. Sales are in 2017 SEK.

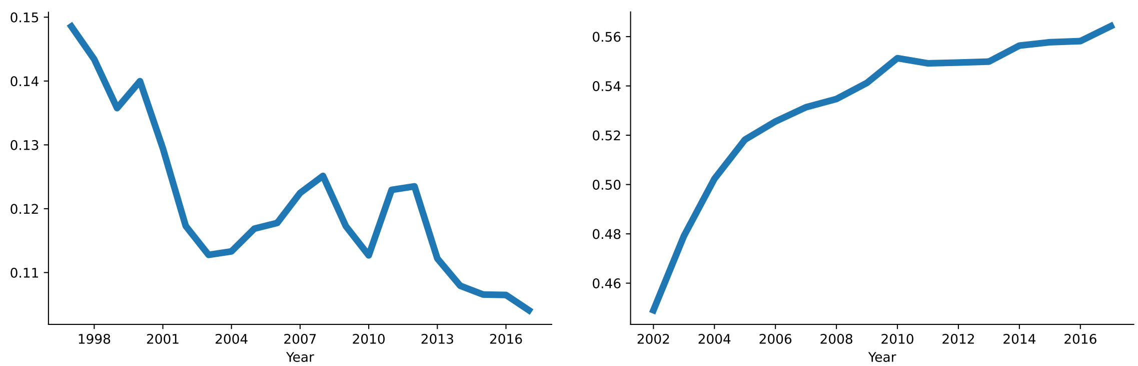
mature firms, defined as firms at least ten years old, increased. I report the series from 2002 onward, as firm age (10+ years) is untruncated only from that year. The share of mature firms increased from 44.9% in 2002 to 56.4% in 2017, an increase of about 12 percentage points over 15 years. Hopenhayn et al. (2022) and Karahan et al. (2024) previously documented a quantitatively similar shift in the firm-age distribution for the US.

Firm-size dispersion conditional on age The left panel of Figure 4 shows the third and last element of the dispersion decomposition in eq. (1), namely size dispersion conditional on age, again measured using the IQR of log sales. Two points are noteworthy. First, there are stark differences across ages: among older firms, sizes are much more dispersed than among younger firms. Second, dispersion has risen systematically over time, but the increase is concentrated among older firms. For ages 0 and 3, size dispersion was stable throughout the sample period. For age groups 6–10, 11–15, and 16–20, within-group size dispersion increased by about 0.3 log points.

The right panel of Figure 4 shows size dispersion within the age group 11–15 by sector, with all sectors normalized to zero at the start. All sectors show an increase in size dispersion conditional on age. The increase is quantitatively similar across sectors, with most displaying a rise of about 0.3 log points consistent with the rise shown in the left panel for this age group. Among the sectors, Manufacturing displays the smallest increase whereas Healthcare shows the largest. Hence, the rise in firm-size

Figure 3: Firm-age distribution

(a) Young-firm share (b) Share mature firms



Notes: The left panel shows the young-firm share, defined as the share of firms aged zero or one. The right panel shows the share of firms that are at least ten years old.

dispersion conditional on age is a within-sector phenomenon.

These trends are robust to a range of specification choices, as shown in Supplemental Appendix B, including using the standard deviation of log sales conditional on age as the dispersion measure (excluding firms with zero sales), using integer ages as the conditioning variable without any age groupings, using sector-level GDP deflators (instead of the aggregate GDP deflator) to deflate firm sales, excluding mergers and acquisitions, and using alternative measures of firm size such as intermediate inputs, capital, or employment.

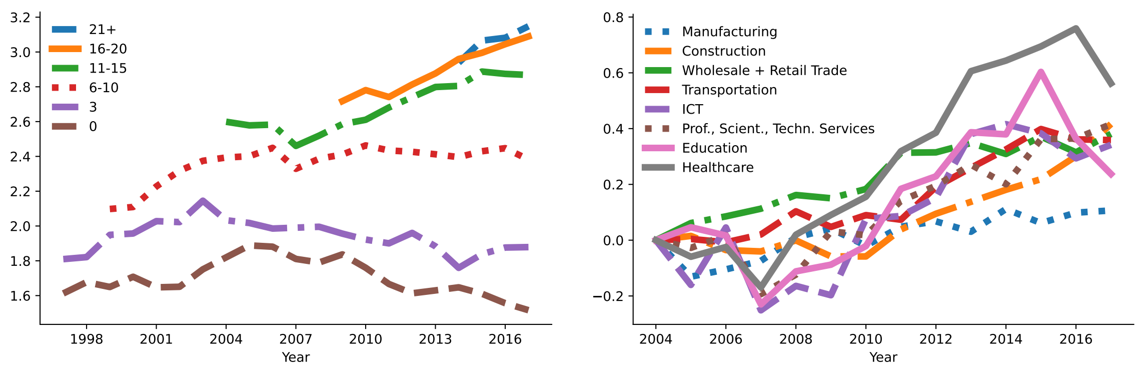
4.3 Decomposing the rise in firm-size dispersion

The graphical analysis in the previous section documented a shift in the firm-age distribution toward older firms and a rise in size dispersion conditional on age. How much does each margin contribute quantitatively to the rise in cross-sectional size dispersion?

In this section, I decompose the rise in firm-size dispersion into its components. To this end, I fix two points in time, compute the cross-sectional dispersion of firm size using eq. (1) and then separately change either the average size conditional on age, the firm-age distribution, or size dispersion conditional on age. I do so twice; once starting from the initial point and changing each component to its values in the end period and vice versa. I report averages across both. Because the decomposition is nonlinear, the contributions of the three margins do not necessarily sum to the total change, but in practice they do so up to the fourth decimal. For this decomposition,

Figure 4: Firm-size dispersion conditional on age

(a) All firms (b) By sector (ages 11-15)



Notes: The figure shows the interquartile range (IQR) defined as the difference between the 75th and 25th percentile of the log sales distribution. The left panel shows the IQR for all firms conditional on age as indicated in the legend; the right panel shows the IQR by sector for firms aged 11-15.

I restrict to firms with positive sales so that I can implement the decomposition in eq. (1) for the standard deviation exactly. As shown before, dropping observations with zero values is not consequential when using firm sales as the size measure.

The end point is naturally the last year in the data, 2017. The start point is less obvious. The further back in time, the longer the period one can decompose; however, firm age is untruncated only for firms born 1993 or later, so an earlier starting date requires a coarser age grouping for older firms. I choose 2002 as the starting point, which allows me to decompose a period of 15 years, from 2002 to 2017, and to differentiate firm ages up to age ten, i.e., in the decomposition in eq. (1), $a_f \in \{0, 1, \ldots, 8, 9, 10+\}$. Note that the decomposition in eq. (1) holds for any age grouping and that the age grouping does not systematically favor any component when decomposing changes in total firm-size dispersion.

Between 2002 and 2017, dispersion of firm sales increased by 0.202 log points, as shown for the standard deviation in Figure A-1 (Supplemental Appendix). Table 2 decomposes this rise into its components. Changes in mean size conditional on age contributed negatively to the total change by -0.006 log points (or -2.8%). This is consistent with average size conditional on age declining slightly in Figure 2. The shift in the age distribution contributed positively but modestly, at 0.027 log points (13.6% of the total). Lastly, the increase in size dispersion conditional on age contributed 0.18 log points, accounting for the largest part (89.2%). Rising within-age heterogeneity thus accounts for the overwhelming majority of the increase in size dispersion. Table 2 further shows that this result is robust across sectors. In most sectors, changes in

Table 2: Decomposition of the change in $\sqrt{\text{Var}(\ln s_f)}$ (2002–2017)

	$\mathbb{E}[\ln s_f \mid a_f]$	$p(a_f)$	$\text{Var}(\ln s_f \mid a_f)$	Total
Total economy	-0.006	0.027	0.180	0.202
Manufacturing	-0.016	0.052	0.226	0.262
Construction	-0.004	-0.006	0.275	0.265
Wholesale + Retail Trade	-0.009	0.031	0.271	0.293
Transportation	-0.007	0.028	0.276	0.297
ICT	-0.008	0.096	0.060	0.149
Prof., Scient., Techn. Services	0.008	0.023	0.158	0.189
Education	0.001	0.050	0.215	0.266
Healthcare	0.009	-0.008	0.187	0.187

Notes: The table decomposes the change in the standard deviation of log firm sales, $\sqrt{\text{Var}(\ln s_f)}$, between 2002 and 2017. Each component is computed by varying one element at a time — the conditional mean $\mathbb{E}[\ln s_f \mid a_f]$, the age distribution $p(a_f)$, and within-age dispersion $\text{Var}(\ln s_f \mid a_f)$ — while holding the others fixed. Each element is varied relative to the start and end period and the table reports the average of both.

mean size conditional on age make a small negative contribution, while changes in the age distribution contribute modestly and positively. By contrast, changes in size dispersion conditional on age are large and positive.[6]

4.4 Discussion

The dominant role of within-age heterogeneity disciplines theories of the rise in firm-size dispersion. Any such theory must answer a single question: what drives rising within-age dispersion?

The age profile of the rise is the key discriminating fact: size dispersion conditional on age increased substantially among older firms while remaining stable among young firms throughout the sample period (Figure 4). This pattern points to forces that accumulate with firm tenure. Three alternative explanations are inconsistent with the data.

First, weakening selection among unproductive firms — for instance, due to falling interest rates keeping otherwise non-viable firms alive — predicts that the effect should be strongest in capital-intensive sectors reliant on debt financing, and that it

[6] The ICT sector, where the largest part is accounted for by the aging of firms, is the only exception. This is specific to the decomposed time period: the starting year of the decomposition (2002) falls into the dotcom boom, which saw an increase in entrant activity in the ICT sector. The ICT sector was therefore characterized by a relatively young firm population at that time, which experienced substantial aging over the subsequent period.

should primarily inflate the left tail of the size distribution by keeping small firms alive. Both predictions fail. The rise in size dispersion conditional on age is weakest precisely in Manufacturing (Figure 4), the most capital-intensive sector in the data. Moreover, as shown separately for each tail in Figure A-8 (Supplemental Appendix), the 75th percentile of log sales gradually rises while the 25th percentile steadily falls over the observed time period. The simultaneous expansion of both tails is more consistent with genuine productivity divergence than with zombie-firm survival.

Second, rising size dispersion conditional on age could reflect changes in the composition of entrants over time rather than post-entry divergence among incumbents. Figure 4 rules this out: size dispersion among young firms remained stable relative to dispersion among older firms throughout the sample period, implying that divergence accumulates after entry rather than being inherited from entry conditions.[7]

Third, the same age profile rules out explanations based on noise common to firms of all ages, such as a rise in measurement error. Because such noise is realized independently of firm tenure, it would inflate within-age dispersion uniformly across ages — raising it among entrants and young firms as much as among older firms. The data show the opposite: the rise is concentrated among older firms and absent among the young, the signature of persistent differences that accumulate over a firm's life rather than of age-invariant noise.[8]

5 Model

As an illustration that rising productivity dispersion is consistent with the documented trends, I study its effects on the firm size–age relationship, the firm-age distribution, and size dispersion conditional on age in a model of firm dynamics. To this end, the following section embeds ex-ante productivity heterogeneity across firms into a model of firm dynamics with entry and exit in the spirit of Klette and Kortum (2004).[9]

[7] One might worry that mean entrant quality has declined over time, even if dispersion among entrants is stable. Average sales among entrants did decline slightly (Figure 2), but average employment among entrants shows no systematic trend (Supplemental Appendix B.2), suggesting that the real size of entrants did not meaningfully change.

[8] The pervasiveness of the rise in size dispersion conditional on age across service sectors (Figure 4) — whose output is largely non-tradable — further rules out foreign demand as a driver.

[9] Rising dispersion of ex-post shocks to firm size, or growing differences in the rate at which firms accumulate demand, would also be consistent with rising size dispersion among old firms. Sterk et al. (2021), however, show that size heterogeneity conditional on age is overwhelmingly accounted for by ex-ante differences across firms rather than by the accumulation of ex-post shocks, which motivates the ex-ante formulation here.

5.1 Preferences and aggregate economy

Time is continuous and indexed by t. The economy consists of a representative household that chooses the path of consumption C_t and wealth A_t to maximize lifetime utility

$$U = \int_0^\infty \exp(-\rho t) \ln C_t dt,$$

subject to the budget constraint $\dot{A}_t = r_t A_t + w_t L_t - C_t$. ρ denotes the discount rate, r_t the interest rate and w_t the real wage. The household supplies one unit of labor inelastically, i.e., $L_t = 1$. The optimality condition (Euler equation) for the household problem reads

$$\frac{\dot{C}_t}{C_t} = r_t - \rho.$$

Aggregate output is produced competitively using a Cobb-Douglas technology over a continuum of different products indexed by i (t subscripts are suppressed when convenient)

$$Y = \exp\left(\int_0^1 \ln\left[q_i y_i\right] di\right),$$

where y_i and q_i denote the quantity and quality of product i. Output is consumed entirely such that $Y = C$. Expenditure minimization leads to the standard demand function

$$y_i = \frac{YP}{p_i}. \tag{2}$$

P is defined as the aggregate price index, which I normalize to 1.

5.2 Production

Firms, indexed by f, produce in product market i with a linear technology

$$y_{ift} = \varphi_f l_{ift},$$

where y_{ift} denotes output, l_{ift} labor hired, and φ_f (time-invariant) firm productivity. Importantly, φ_f differs across firms, which captures the notion that some firms are systematically more efficient at producing than others in all of their product lines, e.g., due to a better business plan. As in Aghion et al. (2023), I assume two productivity types, i.e., $\varphi_f \in \{\varphi^h, \varphi^\ell\}$ where $\varphi^h/\varphi^\ell > 1$, which I refer to as high- and low-type firms.[10]

5.3 Static allocation

Taking the joint distribution of product qualities and firm productivity as exogenous in this section, I characterize the static allocations at the product, firm and aggregate levels.

5.3.1 Product level

Firms in product market i compete in prices (Bertrand competition). In equilibrium, only the firm with the highest quality-adjusted productivity $q_{if}\varphi_f$ produces product i (henceforth, incumbent). Under Bertrand competition, the incumbent firm engages in limit pricing and sets its price equal to the quality-adjusted marginal costs of the follower (the firm with the second highest quality-adjusted productivity)

$$p_{if} = \frac{q_{if}}{q_{if'}} \frac{w}{\varphi_{f'}}, \tag{3}$$

where f' indexes the follower in product market i. The price that the incumbent sets is increasing in the quality gap between the incumbent and the follower, as eq. (3) shows. Defining the product markup as the output price over marginal costs, it follows

$$\mu_{if} \equiv \frac{p_{if}}{w/\varphi_f} = \frac{q_{if}}{q_{if'}} \frac{\varphi_f}{\varphi_{f'}}. \tag{4}$$

The incumbent's markup for product i is increasing in its quality and productivity gap. The price setting of the incumbent gives rise to the following profits for product i

$$\pi_{if} = p_{if} y_{if} - w l_{if} = Y \left(1 - \frac{1}{\mu_{if}}\right),$$

[10] Alternative mechanisms work similarly; in De Ridder (2024), for instance, ex-ante heterogeneity in intangible capital adoption, which lowers the marginal costs of some firms, is akin to ex-ante heterogeneity in productivity.

with labor demand for product i

$$l_{if} = \frac{Y}{w}\mu_{if}^{-1}. \tag{5}$$

Employment in product line i is decreasing in the markup.

5.3.2 Firm level

Firm employment is the sum of employment across the firm's product lines

$$l_f = \sum_{i \in N_f} l_{if} = \frac{Y}{w}\left(\sum_{i \in N_f} \mu_{if}^{-1}\right),$$

where N_f denotes the set of product lines in which firm f is the incumbent producer. Firm employment decreases in the markups within each product line but increases in the number of product lines. Hence, holding product markups constant, firms that produce in more product lines feature higher employment. Vice versa, holding the number of product lines constant, firms with higher product markups employ less labor. As sales are equalized across product lines, firm sales are given by $\left|N_f\right|Y \equiv n_f Y$, where n_f denotes the number of products firm f is producing. Hence, firms that produce in more product lines feature higher sales. Lastly, I define the firm markup μ_f as total firm sales over the wage bill wl_f.

5.3.3 Aggregate level

Integrating employment across firms or products yields the total workforce in production:

$$L_P = \int_f l_f df = \frac{Y}{w}\int_0^1 \mu_{if}^{-1} di. \tag{6}$$

Taking logs and integrating eq. (4), one obtains an expression for the wage

$$w = \exp\left(\int_0^1 \ln q_{if} di\right) \times \exp\left(\int_0^1 \ln \varphi_{f(i)} di\right) \times \exp\left(\int_0^1 \ln \mu_{if}^{-1} di\right). \tag{7}$$

To find an expression for aggregate output, insert eq. (7) into eq. (6) to obtain

$$Y = Q\Phi\mathcal{M}L_P, \tag{8}$$

where

$$Q = \exp\left(\int_0^1 \ln q_i di\right), \quad \Phi = \exp\left(\int_0^1 \ln \varphi_{f(i)} di\right), \quad \mathcal{M} = \frac{\exp\left(\int_0^1 \ln \mu_i^{-1} di\right)}{\int_0^1 \mu_i^{-1} di}.$$

Aggregate output Y depends on geometric averages of quality Q and productivity Φ as well as on misallocation $\mathcal{M}$ and production labor L_P. Misallocation arises from markup dispersion ($\mathcal{M}$ is bounded by unity from above) that is due to quality and productivity heterogeneity. The product of Q, Φ and $\mathcal{M}$ captures aggregate Total Factor Productivity (TFP).

5.4 Dynamic firm problem

Firms compete for product markets through innovation (R&D). A firm improves the quality of a randomly selected product operated by a competitor through expansion R&D, thereby becoming the highest-quality producer in a product line that is new to the firm.[11] Product quality is improved step-wise such that every innovation increases q_i by a factor of λ.[12] Denoting by $[\mu_i]$ a firm's set of markups across its product lines, firm profits follow

$$\pi_f(n, [\mu_i]) \equiv \sum_{k=1}^{n} \pi_i(\mu_k) = \sum_{k=1}^{n} Y \left(1 - \frac{1}{\mu_k}\right) = \sum_{k=1}^{n} Y \left(1 - \frac{1}{\lambda} \frac{1}{\frac{\varphi_{f(k)}}{\varphi_{f'(k)}}}\right).$$

Incumbent firms choose the rate of expansion R&D, x_{it}, for each of their product lines, i. When choosing x_{it}, firms take aggregate output Y_t, the real wage w_t, the share of lines operated by high-productivity firms S_t, the interest rate r_t and the rate of creative destruction τ_t as given. Note that S_t is predetermined in t and its evolution is characterized shortly. Denoting the time derivative by $\dot{V}_t^h()$, the value function of a high-productivity type firm (indexed by h) satisfies the following HJB equation:

[11] In the model, firms differ ex-ante in productivity and innovate on product quality. Since the productivity and quality gaps are interchangeable in the markup equation (4), innovation on product productivity paired with ex-ante heterogeneity in firm quality would deliver the same predictions.

[12] As in Aghion et al. (2023), I assume that the step size of quality improvements exceeds the productivity differential, $\lambda > \varphi^h/\varphi^\ell$. This assumption ensures that the firm with the highest quality version in a product line is the incumbent producer. Relaxing this assumption would give room for a race for incumbency between low-productivity entrants and high-productivity incumbents, from which I abstract. The parameter assumption is fulfilled in all model estimations.

$$
\begin{aligned}
&r_t V_t^h(n,[\mu_i],S_t) - \dot{V}_t^h(n,[\mu_i],S_t) = \\
&\quad \sum_{k=1}^{n} \underbrace{\pi_i(\mu_k)}_{\text{Flow profits}} + \sum_{k=1}^{n} \tau_t \underbrace{\Big[V_t^h\left(n-1,[\mu_i]_{i\neq k},S_t\right) - V_t^h(n,[\mu_i],S_t)\Big]}_{\text{Creative destruction}} \\
&\quad + \max_{[x_{it}]} \Bigg\{ \sum_{k=1}^{n} \underbrace{x_{kt}\Big[S_t V_t^h\left(n+1,\left[[\mu_i],\lambda\right],S_t\right) + (1-S_t)V_t^h\left(n+1,\left[[\mu_i],\lambda\cdot\frac{\varphi^h}{\varphi^\ell}\right],S_t\right) - V_t^h(n,[\mu_i],S_t)\Big]}_{\text{Expansion R\&D}} \\
&\quad - \underbrace{w_t \Gamma\left([x_{it}];n,[\mu_i]\right)}_{\text{R\&D costs}} \Bigg\}
\end{aligned}
$$

The value of a firm consists of flow profits, research costs, and two terms related to creative destruction and expansion R&D. At the rate of creative destruction τ_t (determined in equilibrium), the firm loses one of its n products, in which case it remains with $n-1$ products. At the optimally chosen rate x_{it}, expansion R&D is successful (third row), and the firm acquires a new product (n increases by one).

Productivity-type heterogeneity introduces novel elements into the value function of the firm. The share of product lines operated by high-productivity firms, S_t, is a state variable that firms keep track of to build markup expectations. When taking over a new product line through expansion R&D (third row), the probability of replacing a high-type incumbent is S_t, in which case the high-type entrant charges a markup of λ. With probability $1-S_t$, the replaced incumbent is of the low type, and the high-type entrant charges a markup of $\lambda \cdot \varphi^h/\varphi^\ell$. Firms take S_t as given; however, they affect it through their expansion R&D efforts x_{it} in equilibrium. The HJB equation for a low-productivity firm follows the same structure and is listed in the Appendix, Section C.1. The term related to expansion R&D (third row) is distinct since markup expectations differ for low-productivity firms.

$\Gamma\left([x_{it}];n,[\mu_i]\right)$ denote the R&D costs. For their R&D activities, firms pay a cost of

$$
\Gamma\left([x_{it}];n,[\mu_i]\right) = \sum_{k=1}^{n} c\left(x_{kt}\right) = \sum_{k=1}^{n} \frac{1}{\psi_x}(x_{kt})^\zeta
$$

in terms of labor. $\zeta > 1$ ensures convexity of the cost function. R&D costs are additively separable to render a closed-form solution of the value function along the balanced growth path. ψ_x scales research costs and captures the R&D efficiency.

The rate of creative destruction τ_t that firms take as given in their optimization problem is, in equilibrium, equal to the sum of expansion R&D rates and the rate of entry z_t

$$\tau_t = \int_0^1 x_{it} di + z_t. \tag{9}$$

Firm entry is determined as follows. Potential entrants produce a flow rate of entry z_t using a technology that is linear in labor: $z_t = \psi_z L_{Et}$, where ψ_z denotes the entry efficiency and L_{Et} research labor of entrants. Entrants improve the quality of a randomly selected product line. The productivity type is realized after entry and assigned according to the exogenous probabilities p^h and $1-p^h$, respectively. Entrants start with a one-step quality gap. When $z_t > 0$, the free entry condition requires that the expected value of firm entry equals the entry costs

$$p^h E[V_t^h(1, \mu_i)] + (1 - p^h) E[V_t^\ell(1, \mu_i)] = \frac{1}{\psi_z} w_t, \tag{10}$$

where the expected value of entering as a high- or low-type firm is

$$\begin{aligned} E[V_t^h(1, \mu_i)] =& S_t V_t^h(1, \lambda) + (1 - S_t) V_t^h(1, \lambda \times \varphi^h / \varphi^\ell) \\ E[V_t^\ell(1, \mu_i)] =& S_t V_t^\ell(1, \lambda \times \varphi^\ell / \varphi^h) + (1 - S_t) V_t^\ell(1, \lambda). \end{aligned}$$

Firms' expansion R&D and entry generate a distribution of productivity gaps across product lines, all carrying a one-step quality gap λ, denoted by $\nu_t(\lambda, \frac{\varphi_f}{\varphi_{f'}})$. These gaps determine product markups, so the distribution ν_t characterizes the overall markup distribution in the economy. I derive ν_t as a function of firm policies in Appendix C.3. Importantly, ν_t characterizes S_t, the share of product lines operated by high-productivity firms, which firms take as given in their optimization problem. The law-of-motion for S_t follows from ν_t:

$$\dot{S}_t = \dot{\nu}_t \left(\lambda, \frac{\varphi^h}{\varphi^h} \right) + \dot{\nu}_t \left(\lambda, \frac{\varphi^h}{\varphi^\ell} \right) = S_t x_t^h + z_t p^h - S_t \tau_t. \tag{11}$$

Hence, S_t increases through the expansion of high-type incumbents into new product lines, $S_t x_t^h$, and the entry of new high-type firms, $z_t p^h$, and decreases due to creative destruction, $S_t \tau_t$.

Lastly, labor market clearing requires that production labor L_{Pt} and total research labor L_{Rt} add up to one, the aggregate labor endowment

$$1 = L_{Pt} + L_{Rt} = \frac{Y_t}{w_t} \int_0^1 \mu_i^{-1} di + \int_0^1 \frac{x_{it}^{\zeta}}{\psi_x} di + \frac{z_t}{\psi_z}. \quad (12)$$

5.5 Balanced growth path characterization

I define a balanced growth path of the economy as follows.

Definition 1. *A balanced growth path (BGP) is a set of allocations* $[y_{it}, l_{it}, x_{it}, z_t, S_t, C_t]_t$ *and prices* $[r_t, w_t, p_{it}]_t$ *such that firms choose* $[x_{it}, p_{it}]_t$ *optimally, the representative household maximizes utility choosing* $[y_{it}, C_t]_t$*, the growth rate of aggregate variables is constant, the free-entry condition holds, all markets clear and the distribution of quality and productivity gaps* ν_t *is stationary.*

Along the balanced growth path, the economy can be characterized in closed form.

Proposition 1. *In the above setup, along a balanced growth path:*

1. *The value of a product line* i *for firm-productivity type* $d \in \{h, \ell\}$ *is given by*

$$V_i^d(1, \mu_i, S) = \frac{1}{\rho + \tau}\left[Y_t\left(1 - \frac{1}{\mu_i}\right) + \frac{\zeta - 1}{\psi_x} x_i^{\zeta} w_t\right]. \quad (13)$$

   The firm value is the sum of its product values $V^d(n, [\mu_i], S) = \sum_{k=1}^{n} V_i^d(1, \mu_k, S)$*.*

2. *Expansion R&D rates are firm productivity-type specific, i.e.,* $x_i \in \{x^h, x^\ell\}$*. High-productivity firms expand faster than low-productivity ones,* $x^h > x^\ell$*.*

3. *The share of product lines operated by high-productivity type firms* S *is determined by*

$$S\tau = Sx^h + zp^h. \quad (14)$$

4. *The growth rate of aggregate variables is given by*

$$g = \frac{\dot{Q}_t}{Q_t} = \tau \times \ln(\lambda) = \Big(\underbrace{Sx^h + (1-S)x^\ell}_{\textit{Incumbent expansion R\&D}} + \underbrace{z}_{\textit{Entry}} \Big) \times \ln(\lambda). \quad (15)$$

Proof. Sections C.3 and C.2 and C.1 contain the proofs. □

The value of a product line in eq. (13) consists of two terms: profits for a given markup and the continuation value of expansion R&D. The sum of the two terms is discounted by the discount rate and the rate of creative destruction. The more impatient the household or the higher the risk of replacement, the lower the value of a product line. Importantly, expansion R&D rates are productivity-type specific. More productive incumbents charge higher markups, which increases the expected value of a product line. The optimality condition for expansion R&D stated in Supplemental Appendix C.1 equates the expected value of a product line with the marginal cost of expansion R&D. Hence, in equilibrium, more productive firms pay a higher marginal cost of expansion R&D and choose $x^h > x^\ell$. The productivity heterogeneity renders expansion R&D rates and firm value functions type specific.

Proposition 1 further shows that S_t, the share of product lines operated by high-productivity type firms, is constant along the balanced growth path. The expression in eq. (14) directly follows by setting the law-of-motion for S_t in eq. (11) equal to zero. Eq. (14) captures that, along the balanced growth path, the share of product lines operated by high-productivity firms that fall victim to creative destruction is exactly equal to the share they gain through incumbent expansion into new product lines and through firm entry. Eq. (14) can further be rearranged to

$$S = \frac{zp^h}{(1-S)(x^\ell - x^h) + z}, \tag{16}$$

which shows that S depends on the difference in the expansion R&D rates between firm types, $x^\ell - x^h$. Holding firm entry z fixed, an increase in the expansion rate of high-productivity incumbents must be matched by an equal rise in the expansion rate of less-productive firms for S to remain constant. Note that positive firm entry is necessary for both firm types to co-exist in equilibrium where $x^\ell \neq x^h$ and S is constant at its interior solution.

Long-run growth results from R&D at the product level. This occurs through successful creative destruction. Multiplying the rate of creative destruction by the log step size of innovation delivers the aggregate growth rate g, as shown in eq. (15) of Proposition 1. Since expansion R&D rates are heterogeneous, i.e., $x^h > x^\ell$, changes in the share of product lines operated by each productivity type, S and $1 - S$, affect the aggregate growth rate. Along the balanced growth path, both τ and g are constant. I further characterize the aggregate labor income share, the TFP misallocation measure $\mathcal{M}$, and the aggregate markup analytically in the Supplemental Appendix C.3.

To find the balanced growth path, I jointly solve the optimality conditions of the firm (derived in Supplemental Appendix C.1), the free entry condition, eq. (10), the labor market clearing condition, eq. (12), and the system of differential equations characterizing the distribution of productivity and quality gaps (Supplemental Appendix

C.3).

5.6 Firm dynamics

Firms lose products according to the same stochastic process as in Klette and Kortum (2004). However, in this model, firms add products at systematically different rates as optimally chosen expansion R&D rates vary with the firm's productivity type.[13] The following sections derive type-specific firm dynamics.

5.6.1 Firm size dynamics

Firm sales are proportional to the number of products a firm produces. As such, successful expansion R&D increases firm sales. Since optimal expansion R&D rates are productivity-type specific, so is the sales-age relationship. For productivity type φ_f, the average log sales of surviving firms at age a_f relative to the average log sales of entrants is

$$E\left[\ln n_f Y | a_f, \varphi_f\right] - E\left[\ln n_f Y | 0, \varphi_f\right] = \underbrace{E\left[\ln n_f | a_f, \varphi_f\right]}_{\text{Avg. product count}},$$

where n_f is a firm's number of products. The probability of producing n products at age a conditional on survival is $\left(1-\gamma^j(a)\right)\left(\gamma^j(a)\right)^{n-1}$, where $\gamma^j(a) = x^j \frac{1-e^{-(\tau - x^j)a}}{\tau - x^j e^{-(\tau - x^j)a}}$ and $j \in \{h, l\}$. Therefore, the type-specific sales-age relationship can be reformulated as

$$E\left[\ln n_f Y | a_f, \varphi_f\right] - E\left[\ln n_f Y | 0, \varphi_f\right] = \underbrace{\left(1-\gamma^f(a_f)\right) \sum_{n=1}^{\infty} \ln n \times \left(\gamma^f(a_f)\right)^{n-1}}_{\text{Avg. product count}}. \quad (17)$$

Since firm markups are constant, employment is proportional to sales, and employment differences by productivity type across ages are similarly captured by eq. (17).

5.6.2 Firm survival

Firms that lose their last product exit the economy. Since firm expansion is type-dependent, so is firm survival. The survival function in Klette and Kortum (2004)

[13]Therefore, the properties related to firm-size growth and survival in Klette and Kortum (2004) hold conditional on the firm type. In particular, conditional on the type, firm size and growth are unrelated, as in Lentz and Mortensen (2008).

holds conditional on the firm type, i.e., the share of high and low type firms surviving until age a_f is

$$\chi^h(a_f) = 1 - \tau \frac{1 - e^{-(\tau - x^h)a_f}}{\tau - x^h e^{-(\tau - x^h)a_f}} \quad \text{and} \quad \chi^\ell(a_f) = 1 - \tau \frac{1 - e^{-(\tau - x^\ell)a_f}}{\tau - x^\ell e^{-(\tau - x^\ell)a_f}}. \tag{18}$$

Hence, the share of high-type firms among surviving firms at age a_f is

$$s^h(a_f) = \frac{p^h \chi^h(a_f)}{p^h \chi^h(a_f) + (1 - p^h)\chi^\ell(a_f)}, \tag{19}$$

which corresponds to the mass of high-type survivors relative to the total mass of survivors. Since $x^h > x^\ell$, it is easy to show that size differences across ages are larger for high-productivity firms and that their share among surviving firms $s^h(a_f)$ increases in a_f.

Firm size dynamics and survival by productivity type characterize the firm-age distribution and size dispersion conditional on age. I derive these objects in Supplemental Appendix C.4.

6 Model application

This section confronts the model with the data. I first compare balanced growth paths and then solve for the transition between them.

6.1 Balanced growth paths

I begin by estimating the model along an initial balanced growth path, calibrated to firm dynamics and macroeconomic conditions in Sweden around the turn of the millennium. I then trace out comparative statics in the productivity differential and, finally, re-estimate the model on a new balanced growth path to gauge how far rising productivity dispersion goes in accounting for the observed trends in firm dynamics.

Initial balanced growth path The model has seven parameters: the expansion R&D efficiency ψ_x, the innovation-cost curvature ζ, the entry efficiency ψ_z, the step size of quality improvements λ, the productivity differential φ^h/φ^ℓ, the share of high-productivity firms among entrants p^h, and the discount rate ρ. I estimate three of these and set the remaining four outside the model. Two of the latter are standard: the discount rate ρ is fixed at 0.05 and, following Acemoglu et al. (2018), the cost curvature ζ is set to two, in line with evidence from the microeconometric innovation literature (Blundell et al., 2002; Hall and Ziedonis, 2001). The productivity differential φ^h/φ^ℓ and the entrant composition p^h are discussed below.

The three estimated parameters — ψ_x, ψ_z, and λ — are pinned down jointly by three targets: the firm size–age relationship, the share of young firms, and aggregate TFP growth. Although the parameters are identified jointly, each maps tightly to one target, as I now explain.

The size–age relationship identifies the expansion efficiency ψ_x. Successful expansion R&D adds product lines and thereby raises a firm's sales, so the speed at which firms grow with age pins down how cheaply they expand: a higher ψ_x lowers the cost of expansion R&D and steepens the size–age profile. I target the average log size of firms aged six to ten relative to entrants of age zero, a value of 0.501 log points at the start of the sample (Figure 2).

The entry efficiency ψ_z is identified by the share of young firms. Since ψ_z governs the labor cost of sustaining a given flow of entrants, it maps directly into the mass of recently born firms. I target the share of firms aged zero or one (Figure 3), which stood at 14.9% in 1997.

Finally, the step size of quality improvements λ is identified by aggregate TFP growth, on which it bears directly through the growth equation (15). I target a growth rate of 1.5%, the Swedish average over 1995–2005 (FRED).

Two parameters remain: the productivity gap φ^h/φ^ℓ and the share of high-type entrants p^h. The natural strategy would be to estimate the productivity gap by targeting size dispersion conditional on age, which is substantial in the data. It is well understood that canonical models of firm dynamics built on creative destruction struggle to produce realistic size dispersion (Luttmer, 2011; Berlingieri et al., 2024), and the present model is no exception.[14] The paper's interest, however, lies in changes in dispersion rather than in its level. I therefore set φ^h/φ^ℓ equal to unity in the initial balanced growth path, so that the baseline is a standard Klette and Kortum (2004) type economy in which all firms are ex-ante identical. The changes induced by the rising productivity gaps that the following sections explore can then be read as deviations from this homogeneous benchmark.[15]

[14]Permanent productivity heterogeneity raises size dispersion, as illustrated shortly. For sufficiently large gaps, however, entry converges to zero and the steady state collapses before empirical dispersion levels are reached. This failure persists even when all model parameters are included in the estimation.

[15]Adding features such as transitions between firm types or heterogeneous product counts at entry would help match the *level* of size dispersion conditional on age, but would not alter the mechanism through which rising productivity gaps operate.

When the two types are equally productive, as in the initial balanced growth path, p^h is immaterial. For the subsequent exercises, in which the gap exceeds unity, I set $p^h = 0.5$, so that entrants are equally likely to be born of either type. Table 3 summarizes the targets and the resulting estimates.

Table 3: Initial balanced growth path. Moments and parameters

	Data	Model
Moments		
Size-age: 6–10 vs. entrants in logs	0.501	0.501
Young-firm share in %	14.9	14.9
TFP growth g in %	1.5	1.5
Parameters		
ψ_x *Expansion R&D efficiency*		0.618
ψ_z *Entry R&D efficiency*		2.009
λ *Step size of quality improvements*		1.084
Set exogenously		
ρ *Discount rate*		0.05
ζ *R&D cost curvature*		2
φ^h/φ^ℓ *Productivity differential*		1
p^h *Share of high-type firms among entrants*		0.5

Notes: the size-age relationship measures the average log size of surviving firms aged six to ten relative to that of entrants (age zero) (Figure 2). The young-firm share measures the share of firms aged zero or one. The targets are computed using Swedish registry data, except for aggregate productivity (TFP) growth (FRED).

Estimation proceeds in two steps. A global step evaluates the loss — the sum of squared percentage deviations from the targeted moments, weighting each equally — over a large Sobol sequence of parameter vectors. A local step then refines the best candidates by minimizing the same loss. The procedure is robust: every local search converges to the same parameter vector.[16]

Table 3 reports the estimates. The model matches all three targeted moments exactly. One estimated parameter admits a direct interpretation: a successful innovation raises product quality by 8.4% ($\lambda = 1.084$). Because the initial balanced growth path features no productivity heterogeneity, every product earns the same markup, so the aggregate markup — cost- and sales-weighted measures coincide — equals 1.084 as well. Before turning to the main exercise, I use a set of comparative statics in the productivity gap to illustrate the model's mechanics.

[16] In the global step, I evaluate 12,000 parameter vectors from the Sobol sequence and run the local search on the 50 best candidates.

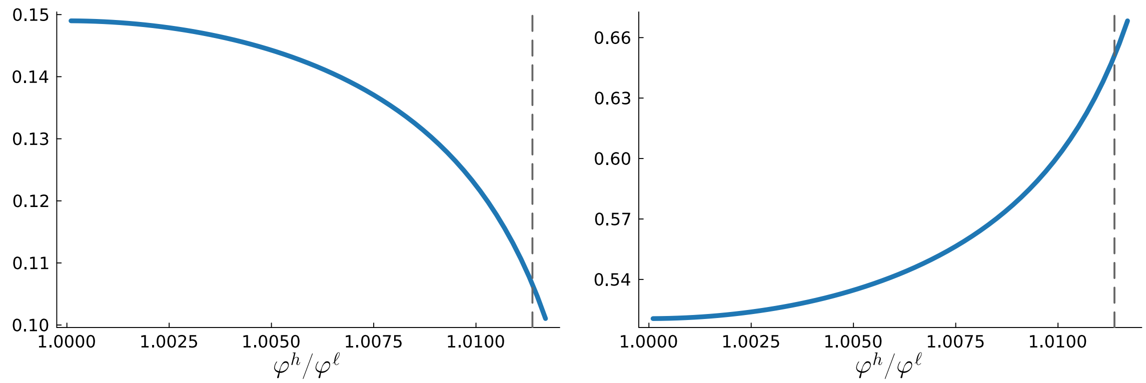
Figure 5 and Figure 6 show the model's implied firm-age distribution and size dispersion conditional on age for a range of productivity gaps, starting from no gap ($\varphi^h/\varphi^\ell = 1$, the initial balanced growth path). Figure 5 shows the firm-age distribution. Without any productivity heterogeneity, the share of young firms, defined as the share of firms aged zero or one, shown in the left panel is exactly matched at 14.9%. The right panel shows the share of firms at least ten years old. This share is untargeted: the model (52%) slightly overstates the data (45%; Figure 3), but the fit is reasonable. The comparative statics then show that as the productivity gap rises, the share of young firms falls and the share of old firms rises. The mechanism runs through firm survival. A firm loses each of its products to creative destruction at rate τ and replaces products by expanding at rate x, so the gap $\tau - x$ governs how quickly it is whittled down to exit. A larger productivity gap raises the value of a high-type's product, and with it the high type's incentive to expand, so x^h rises toward τ and the high-type exit rate $\tau - x^h$ falls toward zero: high-type firms come to replace destroyed products almost as fast as they lose them and survive for longer. Because a fixed share of entrants are high type, these long-lived firms accumulate in the upper tail of the age distribution, raising the share of old firms. At the same time, the lower tail of the age distribution thins. Entry's role in steady state is to replenish a population depleted by exit; once high-type firms scarcely exit, the economy needs far fewer entrants to sustain itself, and the entry rate z falls sharply. With the firm population renewing more slowly, the share of recently born firms falls. The graphs show that the age distribution is sensitive to the productivity differential. For the displayed range of productivity differentials ($\approx 1.2\%$), the share of young firms falls by five percentage points and the share of old firms increases by 15 percentage points.

Figure 6 turns to size dispersion conditional on age, measured by the standard deviation of log sales,[17] over the same range of gaps. As anticipated, the model undershoots the empirical *levels* of within-age dispersion (Figure A-3). Their *gradient* across age groups, however, is captured well, even though it was not targeted. In the initial balanced growth path, the standard deviation of log sales among firms aged 16 to 20 is 22% larger than among firms aged six to ten (0.733/0.601); in the data the corresponding gap is 16% (Figure A-3). The model without permanent heterogeneity thus reproduces how within-age dispersion rises across age groups even as it misses the level — so the cross-sectional age gradient does not, by itself, call for permanent productivity differences in the initial balanced growth path. Raising the gap steepens this gradient further. As the comparative statics in Figure 6 show, size dispersion conditional on age rises with the productivity gap, markedly so among older firms, while dispersion among young firms stays flat. The mechanism is again differential

[17] In product-count models, the IQR is a discrete measure of size dispersion.

Figure 5: Firm-age distribution (model)

(a) Young-firm share (b) Share mature firms



Notes: The left panel shows the young-firm share, defined as the share of firms aged zero or one. The right panel shows the share of firms that are at least ten years old. Both panels start from the initial steady state ($\varphi^h/\varphi^\ell = 1$) and then continuously raise the gap. The vertical dashed line marks the estimated new-balanced-growth-path gap ($\varphi^h/\varphi^\ell = 1.011$).

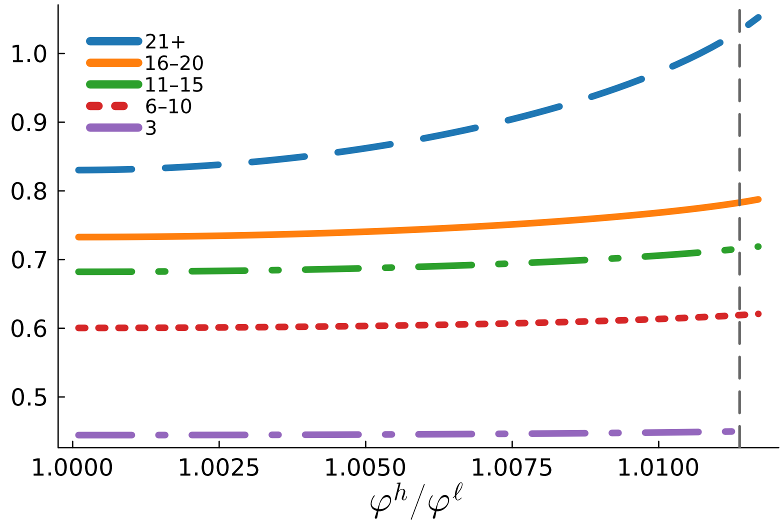
expansion: more productive firms expand faster, and the resulting size differences cumulate with tenure, so dispersion widens among old firms while young-firm dispersion is left almost unchanged. This age-dependent rise directly mirrors the empirical pattern.

The mean size–age relationship, by contrast, is largely insensitive to the gap. Faster expansion by high productivity firms is offset by slower expansion among low productivity firms, so the net effect on the average size–age profile is small. Across the same range of gaps, the size difference between firms aged six to ten and entrants — targeted at 0.501 log points in the initial steady state — rises only to 0.525, consistent with the empirically flat mean size–age profile.

New balanced growth path I now ask how far a rise in the productivity gap between high- and low-type firms can, on its own, account for the observed trends in firm dynamics. I re-estimate the model on a new balanced growth path that reproduces the changes observed in the data relative to the initial one, deliberately allowing only a single parameter to move — the productivity differential — while targeting both the shift in the firm-age distribution and the rise in size dispersion conditional on age among older firms. Holding all other parameters at their initial values makes this a demanding test: one parameter must match two independent margins of the data at once.

The exercise targets the two documented changes in firm dynamics. The first is the

Figure 6: Firm-size dispersion conditional on age (model)



Notes: The figure shows the standard deviation of log firm size (sales) by firm age indicated in the legend. The figure starts from the initial steady state ($\varphi^h/\varphi^\ell = 1$) and then continuously raises the gap. The vertical dashed line marks the estimated new-balanced-growth-path gap ($\varphi^h/\varphi^\ell = 1.011$).

decline in the young-firm share, from 14.9% in 1997 to 10.4% in 2017 — a fall of 4.5 percentage points (Figure 3). The second is the rise in within-age dispersion among older firms: the standard deviation of log sales rose by roughly 0.2 log points for the older age groups (Figure A-3), which I target for firms aged 21 and above.[18]

The productivity-gap column of Table 4 reports the results. The productivity gap rises from unity in the initial steady state to 1.011. Modest as this increase appears, it opens a sizable wedge between the expansion rates of the two types ($x^h = 0.176$, $x^\ell = 0.132$), and the size differences this wedge generates over the life cycle are enough to reproduce the targeted rise in dispersion among old firms.[19] The model matches both targets well. The standard deviation of log size for firms aged 21 and above rises from 0.83 to 1.03 — an increase of 0.2 log points, precisely as targeted — and the share of young firms falls to 10.7%, against a target of 10.4%. The vertical line in Figures 5 and 6 marks this estimated gap, locating the new balanced growth path within the comparative statics shown earlier. The shift in the age distribution, though untargeted, is reproduced as well: the share of firms at least ten years old rises by 13 percentage points, close to the 12-point increase in the data (Figure 3). A single

[18] I target the change in the standard deviation in levels (log points) rather than in proportional terms. This is the natural target here: the variance decomposition in eq. (1) is additive, and the empirical contribution it isolates — the within-age component of the rise in dispersion (Table 2) — is itself measured as a change in log points.

[19] In a similar vein, Aghion et al. (2023) find that a small, 4% increase in the productivity gap more than accounts for the long-run fall in U.S. productivity growth.

parameter thus accounts for both the rise in within-age dispersion and the shift in the age distribution — a nontrivial result, since the productivity gap is overidentified by the two targets.[20] As a further untargeted implication, the rise in the productivity gap slows long-run growth: aggregate TFP growth falls by 0.062 percentage points relative to its baseline of 1.5%. The rise in size dispersion conditional on age is therefore consistent with a mild decline in long-run productivity growth — a connection whose welfare implications I take up in the transition analysis.

Table 4: New balanced growth path: the productivity gap versus alternative shifters

	Data	Gap $\varphi^h/\varphi^\ell\uparrow$	Cheaper exp. $\psi_x\uparrow$	Entry barriers $\psi_z\downarrow$
Targeted				
Young-firm share in %	10.4	10.7	9.8	8.5
$\Delta\sigma$ (ln Size|Age=21+)	0.20	0.20	0.20	0.18
Untargeted				
$\Delta\sigma$ (ln Size|Age=3)	≈ 0	0.01	0.07	0.03
Size-age: 6–10 vs. entrants	0.50	0.52	0.67	0.60
Implied parameter		1.011	0.832	1.779

Notes: Each column re-estimates a single parameter — the productivity gap φ^h/φ^ℓ, the expansion efficiency ψ_x, or the entry efficiency ψ_z — to the two targeted moments (the young-firm share and the rise in old-age dispersion relative to the initial balanced growth path), holding all other parameters at their initial-balanced-growth-path values. Two discriminating moments are untargeted. $\Delta\sigma$ (ln Size|Age=3) is the change in the standard deviation of log sales among firms aged three, flat in the data. Size-age: 6–10 vs. entrants is the average size of firms aged six to ten relative to entrants, in logs; it is flat in the data and in the initial balanced growth path at 0.50. Initial-balanced-growth-path values: young-firm share 14.9%, $\psi_x = 0.618$, $\psi_z = 2.009$.

Distinguishing the productivity gap from alternative shifters That a rising productivity gap *can* reproduce the trends does not establish that it is the only force that could. Within the model, two alternatives are natural candidates: a rise in expansion efficiency ψ_x — a cheaper technology for adding product lines — and a fall in entry efficiency ψ_z — the entry-barriers mechanism emphasized in the declining-business-dynamism literature. I subject each to the same test as the gap. Holding all other parameters at their initial values — and the productivity gap at unity — I re-estimate the single parameter to the same two targets, then ask what it implies for the two margins those targets leave free but that the data discipline (Section 4.4): dispersion among *young* firms and the mean size–age profile, both flat in the data.

[20]The estimation attributes all of the shift in the age distribution to the rising productivity gap. A natural alternative would instead attribute the decline in entry, and the resulting aging of the firm population, to slower labor force growth. This confound is absent here: labor force growth in Sweden did not decline over the sample period (Engbom, 2023).

Table 4 reports the race. All three shifters can be stretched to approximate the two targeted moments — they are estimated to do so — but they diverge sharply on the discriminating margins. Only the productivity gap holds young-firm dispersion flat (a rise of 0.01 log points at age three, against 0.07 under cheaper expansion and 0.03 under entry barriers) and the mean size–age profile flat (firms aged six to ten remain 0.52 log points above entrants, against 0.67 and 0.60, relative to 0.50 in the data). The reason is mechanical: both alternatives raise dispersion by speeding expansion for *all* firms, which widens the product-count distribution at every age and steepens the size–age profile; the productivity gap instead raises dispersion only through the divergence between types, which accrues with tenure and so concentrates among older firms while leaving the average profile flat. The productivity gap also fits the two targets more closely than either alternative, and — unlike cheaper expansion, which *raises* aggregate growth — it slows growth, consistent with the productivity slowdown that accompanies the rise in dispersion. Among the candidate shifters considered, then, only rising productivity dispersion reproduces the full age signature of the data within the model.

6.2 Transition analysis

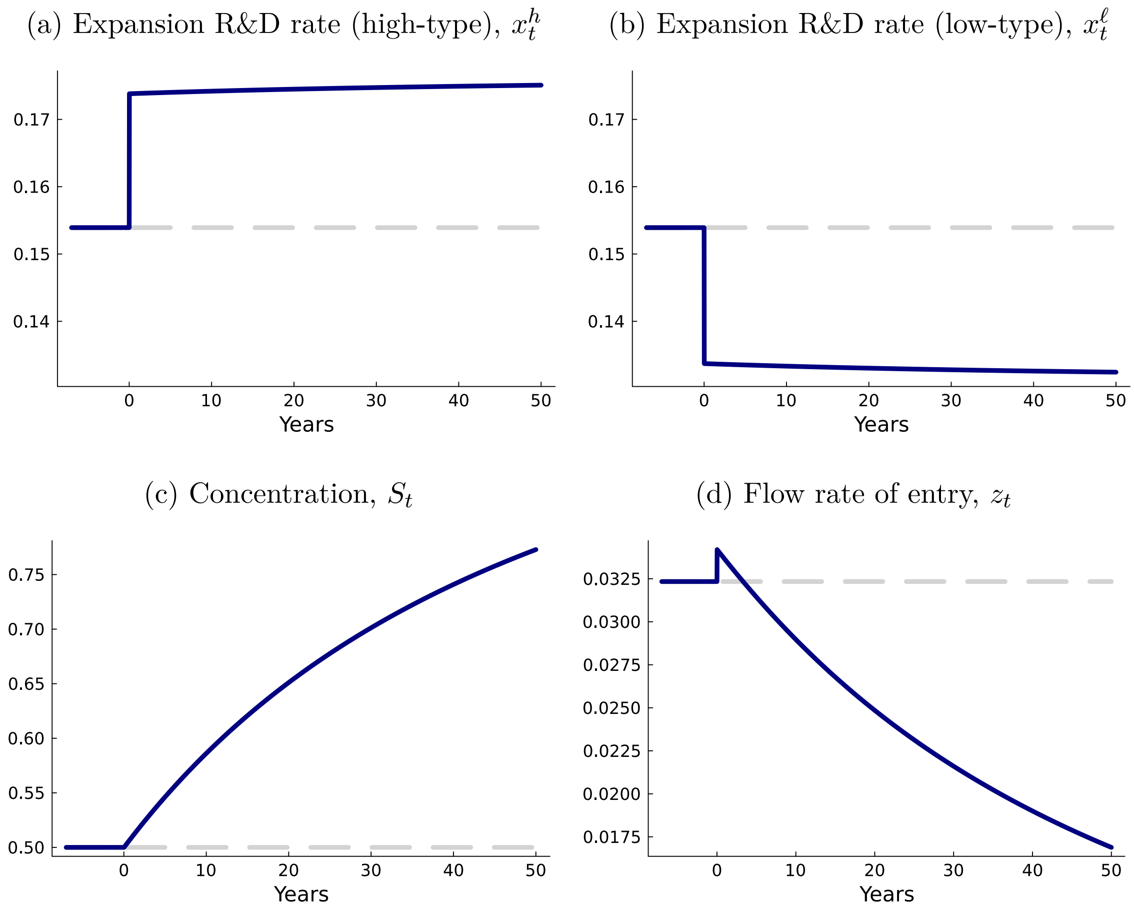
The balanced-growth-path analysis compared two long-run equilibria. I now solve for the full transition path between them, both to trace how the economy adjusts over time and to evaluate the welfare consequences of the rising productivity gap. Starting from the initial balanced growth path, I raise the productivity gap once and for all to its new-balanced-growth-path value as an unanticipated, permanent shock in period zero, holding all other parameters fixed. I model the increase in the gap as a (geometric) mean-preserving spread: as φ^h/φ^ℓ widens, the high type becomes more productive and the low type less productive, holding the geometric mean of the two types' productivities fixed. The wider gap is therefore pure productivity dispersion and does not raise average productivity; the *level* of aggregate productivity then moves only through the reallocation of market share toward the more productive type.[21]

The sign of the net welfare effect is unclear ex ante. On the one hand, long-run growth falls, which is costly to the representative consumer. On the other hand, high-productivity firms expand and gain market share, raising the *level* of aggregate productivity, which is beneficial. The welfare calculation therefore trades off these level gains against the loss from slower long-run growth. Which force dominates can only be settled by solving the transition.

[21] All equilibrium objects depend on the two types' productivities only through their ratio φ^h/φ^ℓ; the mean-preserving-spread assumption therefore leaves the model solution unchanged and bears only on the level of aggregate productivity Φ, which depends on the productivities themselves.

The algorithm to solve for the transition path works as follows. I solve for firms' policy and value functions from the new balanced growth path backward for a guessed sequence of wage growth, interest rates, and the market share of high-productivity firms, S_t. I then use the obtained policy functions over the transition period to simulate the distribution of productivity gaps forward, starting from the initial balanced growth path. Using the evolution of this distribution, together with market clearing and optimality conditions, I back out the implied sequences of wage growth, interest rates, and S_t. The transition path is the fixed point between the guessed and implied sequences. I set the time step to $dt = 0.02$ and the transition horizon to 300 years, and outline the algorithm in detail in Supplemental Appendix D.

Figure 7: Transitional dynamics for equilibrium outcomes



(a) Expansion R&D rate (high-type), x_t^h

(b) Expansion R&D rate (low-type), x_t^ℓ

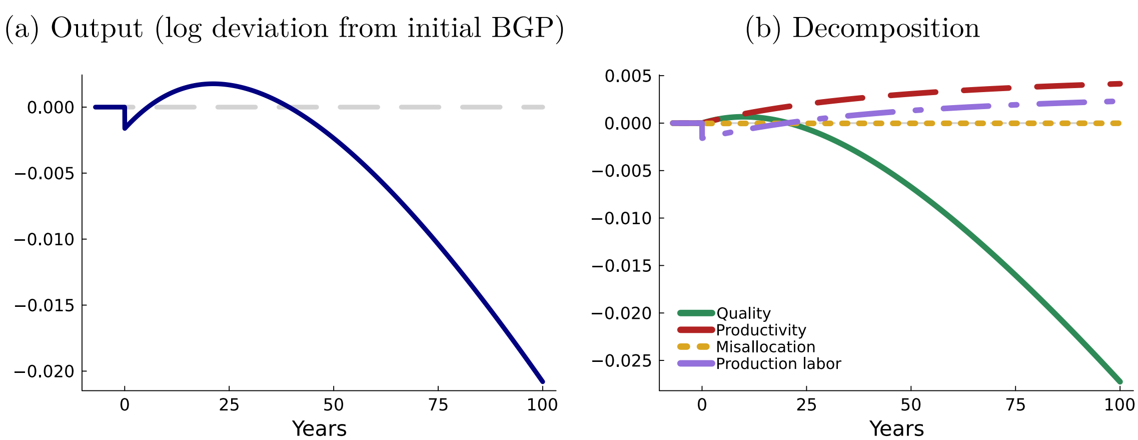
(c) Concentration, S_t

(d) Flow rate of entry, z_t

Notes: the figure shows the response in equilibrium outcomes during the transition toward the new balanced growth path. The productivity gap φ^h/φ^ℓ is raised from 1 to 1.011 (Table 4) as a one-time shock in period zero; all other parameters are held fixed. The dashed line indicates the initial balanced growth path.

Figure 7 shows the response of the main equilibrium outcomes over the transition; the dashed line indicates the initial balanced growth path. Because a larger productivity gap raises the value of a high-type firm's products, high-productivity firms expand more aggressively: x^h jumps from 0.154 to 0.174 on impact and edges up further to 0.176 over the transition. Low-productivity firms move in the opposite direction. Facing competitors that are both more numerous and more productive, their expansion rate x^ℓ falls on impact from the same value of 0.154 to 0.134 and continues to decline to 0.132. The convergence of the choice variables x^h and x^ℓ is relatively quick: they reach their new equilibrium values already after roughly 20 years. The widening gap in expansion rates reallocates market share toward the high type: the share of product lines operated by high-productivity firms, S_t, rises gradually from 0.5. The bottom-right panel shows the flow rate of entry z_t. Entry rises slightly on impact, from 0.032 to 0.034, as the higher gap raises the value of entering as a high-productivity firm. The increase is short-lived: as x^h rises and high-type firms come to dominate, the economy needs fewer entrants to replenish a population that exits ever more slowly, and the higher expansion rate crowds out entry through labor-market clearing. The entry rate therefore declines steadily.

Figure 8: Transitional dynamics for aggregate output



Notes: Panel (a) shows the log deviation of aggregate output along the transition from the counterfactual path on which the economy remains on the initial balanced growth path. Panel (b) decomposes this deviation into the contributions of aggregate quality Q, productivity Φ, misallocation $\mathcal{M}$, and production labor L_P (eq. 8), each in log points relative to the initial balanced growth path. The productivity gap is raised from 1 to 1.011 in period zero.

Figure 8 turns to aggregate output. Panel (a) plots the log deviation of output along the transition from the counterfactual on which the economy would have remained on the initial balanced growth path, $\ln(Y_t/Y_t^{\text{init}})$ where Y_t^{init} denotes output on that path; removing the common trend isolates the effect of the shock. Output dips slightly on impact and a modest, short-lived boom follows: the boom peaks at about 0.002 log points (0.2%) around year twenty before declining, crossing back below the initial-balanced-growth-path trend after roughly forty years and remaining below thereafter. In the long run, output falls further and further below the initial balanced growth path trend, at a constant rate equal to the difference in long-run growth rates, $g^{\text{new}} - g^{\text{initial}} < 0$: once the economy reaches the new balanced growth path, output grows at the constant rate g^{new} while output on the initial balanced growth path grows at the faster rate g^{initial}, so the log ratio between them falls linearly. This steadily widening gap exactly signals that the new balanced growth path has been reached. Panel (b) decomposes this deviation using eq. (8). Since $Y = Q\Phi\mathcal{M}L_P$, in logs the deviation of output is the sum of the deviations of aggregate quality Q, productivity Φ, misallocation $\mathcal{M}$, and production labor L_P from their initial-balanced-growth-path values.

The decomposition reveals two offsetting forces. The level gain in aggregate output comes from aggregate productivity Φ. Φ rises steadily as high-productivity firms gain market share and an increasing share of product lines is operated by the more productive type. This reallocation lifts aggregate output 0.004 log points (0.4%) above its initial balanced growth path value by year one hundred. Working in the opposite direction is aggregate quality Q. Its contribution is in fact slightly positive in the early years of the transition: on impact, entry z and high-type expansion x^h rise, so the aggregate rate of creative destruction — and with it quality growth — briefly exceeds its initial-balanced-growth-path rate. This reverses as the transition proceeds: entry declines, outweighing the higher expansion intensity of the now-dominant high-productivity firms, so creative destruction falls below its initial level and quality grows more slowly thereafter. The resulting quality shortfall is negligible at first but cumulates over time, reaching −0.027 log points (−2.7%) by year one hundred, and is the force that eventually pulls output below trend. Misallocation $\mathcal{M}$ contributes negligibly. Production labor L_P falls on impact — firms shift labor out of production and into expansion R&D — before recovering and turning mildly positive; this fall in production labor is what makes output dip slightly on impact. In sum, the rising productivity gap produces a modest short-run boom through the reallocation of market share toward more productive firms, but a long-run decline through permanently slower growth from reduced creative destruction.

That output rises modestly in the short run but falls in the long run raises the question of the net welfare effect. I compute the permanent percentage change in consumption along the initial balanced growth path that makes the representative consumer indifferent to the consumption stream realized over the transition toward the new balanced growth path. Accounting for the transition, welfare *falls* by only 0.025%: the level gain from reallocating market shares toward more productive firms offsets most — but not quite all — of the discounted cost of slower long-run growth, leaving a small net loss.[22] Ignoring the transition substantially overstates this loss. Comparing

[22] Aghion et al. (2023) instead find a sizable negative welfare effect of −3.3% from rising produc-

the two balanced growth paths directly, the consumption-equivalent change ξ solves $\ln(1+\xi) = (g^{\text{new}} - g^{\text{initial}})/\rho$; with g falling by 0.062 percentage points, this across-balanced-growth-path comparison implies a welfare loss of 1.24% ($\xi = -0.0124$). The contrast is instructive: the across-balanced-growth-path comparison captures only the change in long-run growth and misses the level gain in aggregate productivity from the reallocation of market share, which the transition makes explicit; accounting for it shrinks the welfare loss from 1.24% to 0.025%.

7 Conclusion

This paper documents a sustained rise in firm-size dispersion in Sweden from 1997 to 2017 and decomposes it, using the law of total variance, into three margins of firm dynamics: differences in mean size across firm ages, the firm-age distribution, and size dispersion conditional on age. Over this period the age distribution shifted toward older firms, mean size by age remained stable, and size dispersion conditional on age rose, concentrated among older firms. The decomposition — model-free and exact in the data — attributes the overwhelming share of the rise in dispersion between 2002 and 2017 to rising size dispersion conditional on age, while the aging of the firm population contributes modestly and changes in mean size by age contribute negligibly.

The changes in these three margins — a shift in the age distribution toward older firms, a stable mean size–age profile, and a rise in within-age dispersion concentrated among older firms — provide disciplining moments for theories of rising firm-size dispersion, which any candidate mechanism should be required to reproduce. The concentration of the rise in dispersion among older firms is an additional discriminating fact: it points to forces whose effect cumulates with firm age, such as widening productivity dispersion, and against alternatives — shifts in the composition of entrants or measurement error — that would raise dispersion among young firms as well.

I embed permanent productivity heterogeneity into a quality-ladder model of firm dynamics with entry and exit and ask how much of the observed changes a single widening productivity gap can explain. Holding all other parameters at their calibrated baseline, a modest rise in the productivity gap reproduces both the shift in the

tivity gaps. The difference stems from how the gap is disciplined. They estimate the productivity gap to match the entire long-run fall in U.S. productivity growth, so in their calibration the gap must rise enough to make slower growth the dominant force. Here the gap is instead estimated to match the rise in firm-size dispersion conditional on age; long-run growth falls only mildly and does not account for the entire observed decline in growth, so the growth loss is small and is largely offset by the level gain from reallocation.

firm-age distribution and the rise in within-age dispersion among older firms. The mechanism is differential expansion: more productive firms expand faster, so size differences cumulate with tenure, raising dispersion among older firms while leaving young-firm dispersion flat. The fit is specific to the productivity gap: lowering the cost of expansion or raising the cost of entry — natural alternatives from the literature on declining business dynamism — fits the same targets but raises dispersion at all ages and steepens the mean size–age profile, counter to the data.

The widening gap reallocates market share toward more productive firms, raising the level of aggregate productivity while slowing long-run growth only mildly. These forces offset along the transition between balanced growth paths: accounting for it, the welfare cost is only 0.025% in consumption-equivalent terms, far below the 1.24% implied by comparing balanced growth paths alone — which misses the level gain from reallocation — and below the 3.3% in Aghion et al. (2023), who discipline the gap to the entire long-run growth slowdown rather than, as here, to the rise in firm-size dispersion.

Several directions for future work are worth noting. The first two address limitations of the model analysis. First, like standard product-count models of firm dynamics, the model has difficulty matching the level of firm-size dispersion in the data. While matching the level of dispersion is not itself the aim of this paper, one could seek to make the model consistent with it. Introducing heterogeneity in firm size at entry would be one approach to help on this dimension. Second, the data cover only two decades, and the analysis is necessarily confined to this horizon. Since the model contrasts two balanced growth paths, documenting these trends over a longer period would bring the data closer to this long-run comparison and allow the transition between the two paths to be traced directly.

Third, the trends documented here closely align with those documented for the US by Hopenhayn et al. (2022) and Karahan et al. (2024) on the margins they study, suggesting the dominance of within-age dispersion may be a broader feature of advanced economies; a natural next step is to apply the decomposition to other countries. Fourth, the same model-free decomposition applies to the *level* of firm-size dispersion observed in the data, not only to its change: in ongoing work, I show that almost all firm-size dispersion in the cross-section is accounted for by dispersion conditional on age, with size differences across firm ages contributing negligibly — underscoring the importance of within-age heterogeneity for the level of dispersion as well.

References

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N. and Kerr, W. (2018), 'Innovation, reallocation, and growth', *American Economic Review* **108**(11), 3450–91.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J. and Li, H. (2023), 'A theory of falling growth and rising rents', *Review of Economic Studies* **90**(6), 2675–2702.

Akcigit, U. and Ates, S. T. (2021), 'Ten facts on declining business dynamism and lessons from endogenous growth theory', *American Economic Journal: Macroeconomics* **13**(1), 257–98.

Akcigit, U. and Ates, S. T. (2023), 'What happened to us business dynamism?', *Journal of Political Economy* **131**(8), 2059–2124.

Andrews, D., Criscuolo, C. and Gal, P. N. (2016), 'The best versus the rest: the global productivity slowdown, divergence across firms and the role of public policy'.

Autor, D., Dorn, D., Katz, L. F., Patterson, C. and Van Reenen, J. (2020), 'The fall of the labor share and the rise of superstar firms', *The Quarterly Journal of Economics* **135**(2), 645–709.

Berlingieri, G., Blanchenay, P. and Criscuolo, C. (2017), 'The great divergence (s)'.

Berlingieri, G., De Ridder, M., Lashkari, D. and Rigo, D. (2024), 'Growth through innovation bursts'.

Blundell, R., Griffith, R. and Windmeijer, F. (2002), 'Individual effects and dynamics in count data models', *Journal of econometrics* **108**(1), 113–131.

Chen, Y.-F., Hsu, W.-T. and Peng, S.-K. (2023), 'Innovation, firm size distribution, and gains from trade', *Theoretical Economics* **18**(1), 341–380.

De Loecker, J., Eeckhout, J. and Unger, G. (2020), 'The rise of market power and the macroeconomic implications', *The Quarterly Journal of Economics* **135**(2), 561–644.

De Ridder, M. (2024), 'Market power and innovation in the intangible economy', *American Economic Review* **114**(1), 199–251.

Decker, R. A., Haltiwanger, J., Jarmin, R. S. and Miranda, J. (2016*a*), 'Declining business dynamism: What we know and the way forward', *American Economic Review* **106**(5), 203–07.

Decker, R. A., Haltiwanger, J., Jarmin, R. S. and Miranda, J. (2016*b*), 'Where has all the skewness gone? the decline in high-growth (young) firms in the us', *European Economic Review* **86**, 4–23.

Decker, R. A., Haltiwanger, J., Jarmin, R. S. and Miranda, J. (2020), 'Changing business dynamism and productivity: Shocks versus responsiveness', *American Economic Review* **110**(12), 3952–90.

Engbom, N. (2023), 'Misallocative growth', *Working Paper* .

Gabaix, X. (2009), 'Power laws in economics and finance', *Annu. Rev. Econ.* **1**(1), 255–294.

Gourio, F., Messer, T. and Siemer, M. (2014), A missing generation of firms? aggregate effects of the decline in new business formation, *in* 'CES Ifo Conference'.

Grullon, G., Larkin, Y. and Michaely, R. (2019), 'Are us industries becoming more concentrated?', *Review of Finance* **23**(4), 697–743.

Hall, B. H. and Ziedonis, R. H. (2001), 'The patent paradox revisited: an empirical study of patenting in the us semiconductor industry, 1979-1995', *rand Journal of Economics* pp. 101–128.

Hopenhayn, H., Neira, J. and Singhania, R. (2022), 'From population growth to firm demographics: Implications for concentration, entrepreneurship and the labor share', *Econometrica* **90**(4), 1879–1914.

Karahan, F., Pugsley, B. and Şahin, A. (2024), 'Demographic origins of the start-up deficit', *American Economic Review* **114**(7), 1986–2023.

Klette, T. J. and Kortum, S. (2004), 'Innovating firms and aggregate innovation', *Journal of political economy* **112**(5), 986–1018.

Lentz, R. and Mortensen, D. T. (2008), 'An empirical model of growth through product innovation', *Econometrica* **76**(6), 1317–1373.

Liu, E., Mian, A. and Sufi, A. (2022), 'Low interest rates, market power, and productivity growth', *Econometrica* **90**(1), 193–221.

Luttmer, E. G. (2007), 'Selection, growth, and the size distribution of firms', *The Quarterly Journal of Economics* **122**(3), 1103–1144.

Luttmer, E. G. (2011), 'On the mechanics of firm growth', *The Review of Economic Studies* **78**(3), 1042–1068.

Peters, M. (2020), 'Heterogeneous markups, growth, and endogenous misallocation', *Econometrica* **88**(5), 2037–2073.

Peters, M. and Walsh, C. (2021), Population growth and firm dynamics, Technical report, National Bureau of Economic Research.

Sterk, V., Sedláček, P. and Pugsley, B. (2021), 'The nature of firm growth', *American Economic Review* **111**(2), 547–79.

Supplemental Appendix for

"Recent Changes in Firm Dynamics and the Nature of Rising Firm-Size Dispersion"

[FOR ONLINE PUBLICATION ONLY]

A Data

The main dataset, *Företagens Ekonomi* (FEK), contains information from balance sheets and profit and loss statements for the universe of Swedish firms. From this dataset, I obtain the firm-size measures, including sales, intermediate inputs, the capital stock, employment, and the wage bill. In the FEK code-book by Statistics Sweden (SCB), these variables are defined as follows.[23] SCB defines sales as income from the companies' main business from goods sold and services provided. As the measure of the firm's capital stock I use fixed assets minus depreciation. Employment refers to the average number of employees in full-time units in accordance with the company's annual report. The measure of the firm's wage bill includes salaries and other remuneration, including severance pay. As described in the main text, I focus on firms in the private sector. These firms have a legal type (variable name: *JurForm*) less than 50 or equal to 96.

Starting in 2004, the data include unincorporated self-employed with negative profits. These firms were previously not recorded, leading to an inflow of new firms in that year. Since I cannot differentiate between firms established in 2004 and firms entering the data in 2004 with an earlier birth year, I drop firms that enter the data in

[23] `https://www.scb.se/contentassets/9dd20ce462644cc19f6f04eb2edbbe28/nv0109_kd_slut2022_v1_20240510.pdf`, accessed 26.03.2025.

2004 with negative profits when showing the firm-age distribution. When showing the size-by-age patterns, I leave the 2004 cohort out.[24] For the main exercise — which decomposes the change in size dispersion between 2002 and 2017 — this is inconsequential: firm ages are grouped, so firms born in 2004 fall into the highest age group in 2017 regardless.

To measure size dispersion within sectors, I use the industry classification (SNI codes) provided by SCB. The industry classification changed twice between 1997 and 2017, once in 2002 and once in 2007. The changes in industry classification mainly affect the more detailed codes at the five-digit level, whereas I use the ones at the one-digit level for the main analysis. Nevertheless, to ensure a consistent industry classification over time I proceed as follows. During the year of the change, I observe both the old and the new industry classifications. For the firms present in the data in the year of the classification change, extending the new industry classification further back in time before the change is straightforward. This way, the industry codes of almost all firms are updated. A firm might be in the data before and after the classification change but not for the year of the change. For these firms, the above method does not work. If the firm appears in the data one year after the classification change, I use the observed classification after the change to update the classification before the change. For firms that are absent for several years around the year of change, I use industry mappings provided by Statistics Sweden. These mappings do not always provide a 1:1 mapping between industries before and after the classification change, so I use the most common transitions for the many-to-many mappings.

One concern is that changes in the firm structure, e.g., when firms merge with other firms, change the firm ID. To address this concern, I impute changes in firm IDs using worker flows between firms. The auxiliary data set *Registerbaserad Arbetsmarknadsstatistik* (RAMS) contains the universe of employer-employee matches. I impute changes in the firm ID of firms with at least five employees as follows: if more than 50% of the workforce of firm A in year t makes up more than 50% of the workforce of firm B in year $t+1$, I replace firm B's firm ID with firm A's firm ID from $t+1$ onward. The empirical results remain unchanged when excluding firms for whom the imputed firm ID differs from the observed firm ID.

[24] Leaving the 2004 cohort out when computing the age distribution would bias it in all years following 2004.

Other data sources

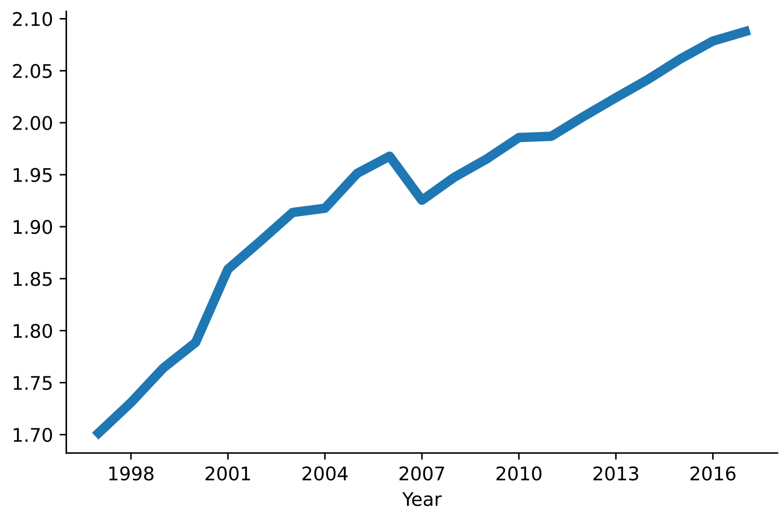
- Federal Reserve Economic Data (FRED): Total Factor Productivity at Constant National Prices for Sweden (series RTFPNASEA632NRUG). Accessed Tuesday, January 30, 2024.

B Size dispersion and firm dynamics in the data

B.1 Firm-size dispersion

Figure A-1 restricts to firms with positive sales and plots the standard deviation of log sales. The standard deviation increases by about 0.4 log points.

Figure A-1: Firm-size dispersion (standard deviation)

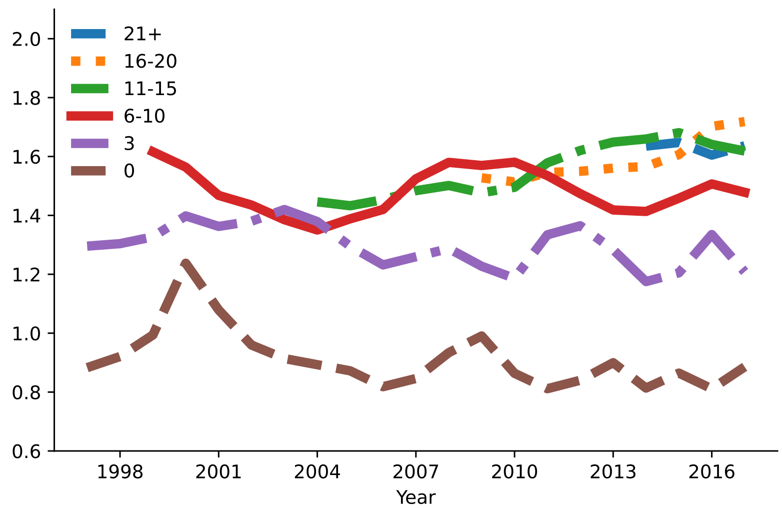


Notes: The figure shows the standard deviation of log sales, restricting to firm-year observations with positive sales.

B.2 Firm size conditional on age

Figure A-2 shows the log of the average firm size by age, where size is measured using employment. There are no systematic trends, including among entrants.

Figure A-2: Firm size conditional on age (employment)



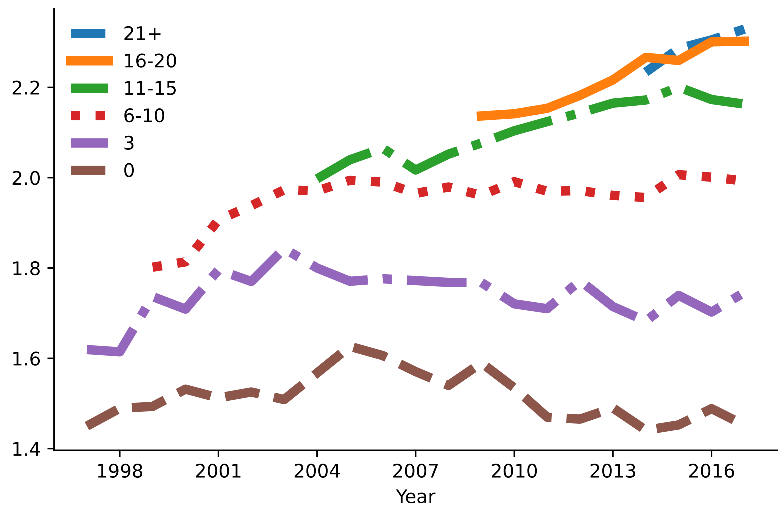
Notes: The figure shows the log of the average firm size (employment) by firm age indicated in the legend.

B.3 Firm-size dispersion conditional on age

As a first robustness check, I use the standard deviation instead of the IQR. This is unproblematic for sales since zero sales are very infrequent, so restricting to positive observations is not consequential. For other size measures such as employment or the wage bill, restricting to positive values is more consequential and introduces stronger selection — self-employed workers, for instance, commonly report a zero wage bill.

As shown in Figure A-3, restricting to firms with positive sales and using the standard deviation of log sales as the measure of dispersion leaves the trends unchanged. Size dispersion conditional on age remains relatively stable for young firms and increases by about 0.2 log points for the age groups 6-10, 11-15 and 16-20.

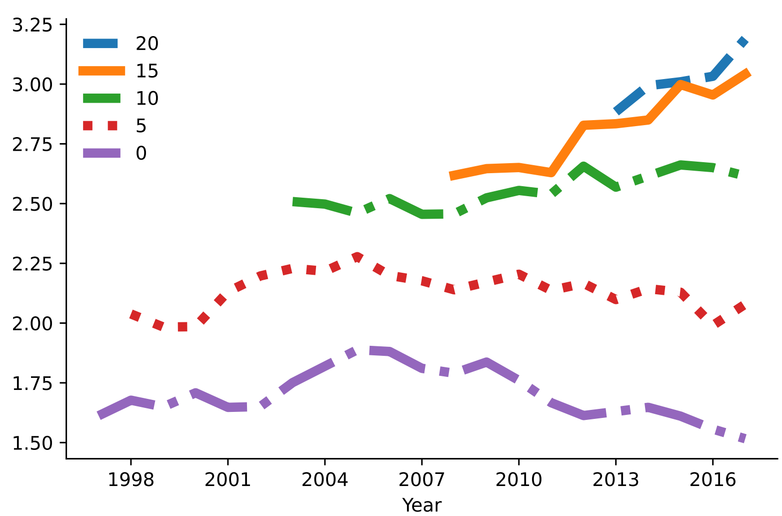
Figure A-3: Size dispersion conditional on age (standard deviation)



Notes: The figure shows the standard deviation of log sales by firm age as indicated in the legend. The figure restricts to firm-year observations with positive sales.

I next show that the age grouping is inconsequential. Figure A-4 shows size dispersion conditional on age for firms aged zero, five, ten, fifteen and twenty. The patterns remain the same. Size dispersion remained stable among young firms, but increased systematically among old firms.

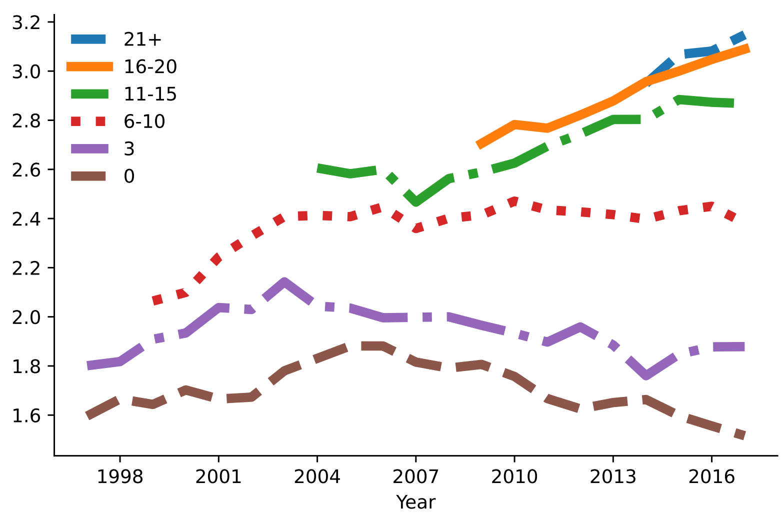
Figure A-4: Size dispersion conditional on age (alternative ages)



Notes: The figure shows the interquartile range (IQR) defined as the difference between the 75th and 25th percentile of the log sales distribution by firm age as indicated in the legend.

The picture also remains unaffected when using alternative (finer) price deflators to deflate firm sales. Figure A-5 uses sector-level GDP deflators to deflate firm sales to 2017 SEK values. Note that these finer deflators have no effect on size dispersion within sectors, but could affect it at the aggregate level. Figure A-5 shows that this is not the case. The figure is virtually unchanged relative to the one in the main text. Differential inflation trends across sectors have no effect on changes in log sales dispersion at the aggregate level.

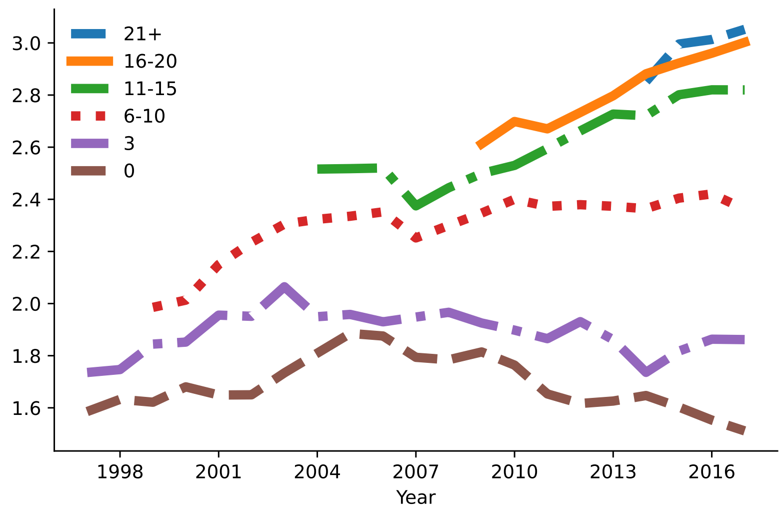
Figure A-5: Size dispersion conditional on age (sector deflators)



Notes: The figure shows the interquartile range (IQR) defined as the difference between the 75th and 25th percentile of the log sales distribution by firm age as indicated in the legend. Sales are deflated using sector-specific GDP deflators.

I also show that mergers and acquisitions have not contributed to the rise in firm-size dispersion conditional on age. In the data, any change in firm structure renders a new firm ID. I identify mergers between firms (and any other changes in firm structure) based on worker flows as described in Supplemental Appendix A and connect the firm IDs for identified cases. Figure A-6 shows size dispersion conditional on age when excluding firms for which the observed firm ID differs from the imputed one. The graph remains virtually unchanged.

Figure A-6: Size dispersion conditional on age (excluding mergers)



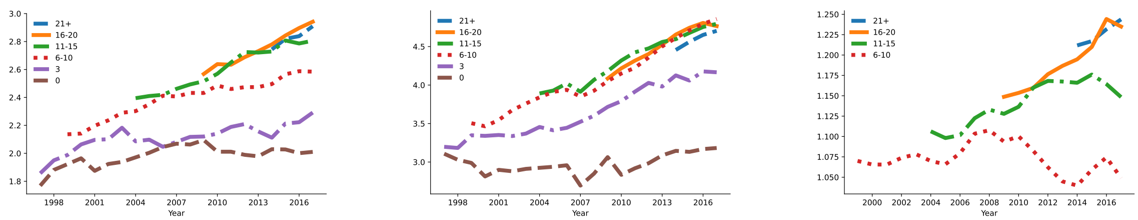
Notes: The figure shows the interquartile range (IQR) defined as the difference between the 75th and 25th percentile of the log sales distribution by firm age as indicated in the legend. The figure excludes firms with an imputed firm ID change.

Next, I assess the robustness of the documented increase in firm-size dispersion conditional on age using alternative size measures. As explained in the main text, the advantage of using firm sales as the main size measure is that sales are a continuous measure of size and the share of firms reporting zero sales is very low, so restricting to positive values is not consequential.

I document trends in size dispersion conditional on age for three alternative size measures: intermediate inputs, capital stock, and employment. As intermediate inputs are a firm's most frequently used input, the share of firms reporting zero intermediate inputs is also relatively low, so the standard dispersion measures can be applied without problems. The capital stock and employment pose more problems. When using the capital stock as the firm-size measure, the 25th percentile of the size distribution commonly features a capital stock of zero, making the IQR of log capital undefined. To circumvent this problem, I instead show the difference in the log capital stock between the 95th and 50th percentile. Hence, this dispersion measure focuses on dispersion in the right tail. For employment, the same problem occurs. The 25th percentile of the employment distribution generally features zero employment as self-employed do not hire any workers. An additional complication is that employment is a very discrete measure of firm size for the majority of firms (recall that the median firm employs one worker). The IQR therefore does not accurately reflect changes in size dispersion. To make progress, when using employment as the size measure I exclude firms with zero employment and use the standard deviation as the dispersion measure.

Figure A-7: Size dispersion conditional on age (alternative size measures)

(a) Intermediate inputs (b) Capital stock (c) Employment

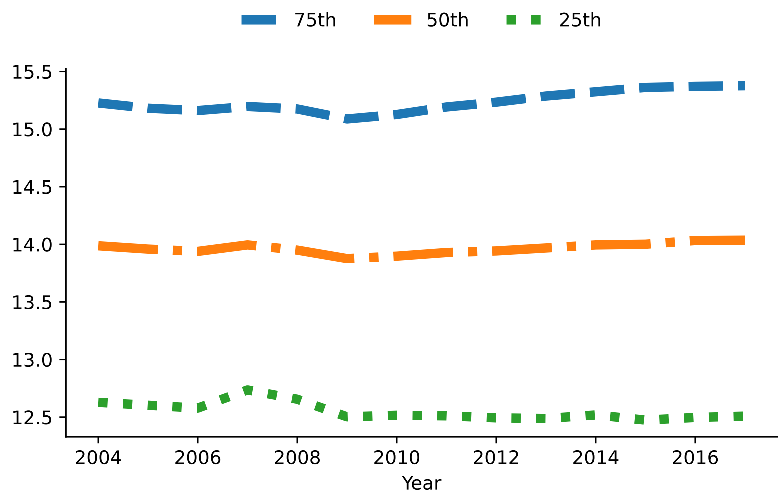


Notes: The figure shows the dispersion of log firm size by firm age as indicated in the legend for alternative size measures. The left panel uses intermediate inputs and applies the interquartile range as the dispersion measure. The middle panel uses the capital stock and shows the difference between the 95th and 50th percentile. The right panel uses employment and applies the standard deviation as the dispersion measure (restricting to positive values).

Figure A-7 shows firm-size dispersion conditional on age for the three alternative size measures. The increase in size dispersion conditional on age is even more visible when using intermediate inputs, and is again particularly pronounced for older firms: within-group size dispersion for the age groups 6–10, 11–15, and 16–20 increases by about 0.4 log points. The increase is even stronger when using the capital stock as the size measure: for the same age groups the increase exceeds 0.5 log points. This should, however, be interpreted with caution, as the measure focuses on dispersion in the right tail of the size distribution. Moreover, the capital stock captures only physical capital, so the increase in dispersion conditional on age is not driven by intangible assets. The last panel shows dispersion in log employment conditional on age over time for firms with at least one worker. Since this restriction heavily selects the sample, particularly among young firms, I only show the trends for older firm ages. For the age groups 11–15 and 16–20, a clear upward trend is noticeable. This increase is slightly smaller than the increase in sales dispersion shown for the standard deviation in Figure A-3, though this should be interpreted with caution given sample selection.

Lastly, I show the size distribution conditional on age to examine how individual percentiles have shifted. Figure A-8 shows the 25th, 50th, and 75th percentile of the log sales distribution for 11–15 year old firms. The figure shows that the IQR increases due to changes in both the 25th and 75th percentile: the former decreased slightly while the latter rose.

Figure A-8: Size distribution (ages 11-15)



Notes: The figure shows percentiles (25th, 50th, 75th) of the log sales distribution for 11-15 year old firms.

C Model

C.1 Solving the dynamic firm problem

The HJB for a high-productivity type firm h reads

$$\begin{aligned}
&r_t V_t^h(n,[\mu_i],S_t) - \dot{V}_t^h(n,[\mu_i],S_t) = \\
&\quad \sum_{k=1}^{n} \pi(\mu_k) + \sum_{k=1}^{n} \tau_t \Big[V_t^h\left(n-1,[\mu_i]_{i\neq k},S_t\right) - V_t^h(n,[\mu_i],S_t) \Big] \\
&\quad + \max_{[x_k]} \Bigg\{ \sum_{k=1}^{n} x_k \Big[S_t V_t^h\left(n+1,\left[[\mu_i],\lambda\right],S_t\right) + (1-S_t) V_t^h\left(n+1,\left[[\mu_i],\lambda\times\varphi^h/\varphi^\ell\right],S_t\right) - V_t^h(n,[\mu_i],S_t) \Big] \\
&\qquad\quad - \frac{w_t}{\psi_x}(x_k)^\zeta \Bigg\}
\end{aligned}$$

The HJB for a low-productivity type firm ℓ reads

$$\begin{aligned}
&r_t V_t^\ell(n,[\mu_i],S_t) - \dot{V}_t^\ell(n,[\mu_i],S_t) = \\
&\quad \sum_{k=1}^{n} \pi(\mu_k) + \sum_{k=1}^{n} \tau_t \Big[V_t^\ell\left(n-1,[\mu_i]_{i\neq k},S_t\right) - V_t^\ell(n,[\mu_i],S_t) \Big] \\
&\quad + \max_{[x_k]} \Bigg\{ \sum_{k=1}^{n} x_k \Big[S_t V_t^\ell\left(n+1,\left[[\mu_i],\lambda\times\varphi^\ell/\varphi^h\right],S_t\right) + (1-S_t) V_t^\ell\left(n+1,\left[[\mu_i],\lambda\right],S_t\right) - V_t^\ell(n,[\mu_i],S_t) \Big] \\
&\qquad\quad - \frac{w_t}{\psi_x}(x_k)^\zeta \Bigg\}
\end{aligned}$$

For clarity, I suppress the dependence of the value function on S_t in the following. Following the guess-and-verify approach in Peters (2020), I obtain the value function of a firm of productivity type d with n products

$$V_t^d(n,[\mu_i]) = n\frac{1}{(\rho+\tau)}\frac{\zeta-1}{\psi_x}(x^d)^\zeta w_t + \sum_{k=1}^{n}\frac{\pi(\mu_k)}{\rho+\tau} \tag{A-1}$$

with the optimality condition for expansion R&D for high-productivity firms given by

$$\frac{\zeta-1}{\psi_x}(x^h)^\zeta + S_t\frac{Y_t}{w_t}\left(1-\frac{1}{\lambda}\right) + (1-S_t)\frac{Y_t}{w_t}\left(1-\frac{\varphi^\ell}{\varphi^h}\frac{1}{\lambda}\right) = (\rho+\tau)\frac{\zeta}{\psi_x}(x^h)^{\zeta-1}$$

and for low-productivity firms by

$$\frac{\zeta-1}{\psi_x}(x^\ell)^\zeta + S_t\frac{Y_t}{w_t}\left(1-\frac{1}{\lambda}\frac{\varphi^h}{\varphi^\ell}\right) + (1-S_t)\frac{Y_t}{w_t}\left(1-\frac{1}{\lambda}\right) = (\rho+\tau)\frac{\zeta}{\psi_x}(x^\ell)^{\zeta-1}.$$

I prove that more productive firms choose higher expansion R&D rates, i.e., $x^h > x^\ell$. Intuitively, the proof shows that an increase in productivity raises the stream of profits in a product line. The continuation value and the marginal cost of expansion R&D have to rise for the optimality condition of the expansion R&D rate to hold, implying that the expansion R&D rate increases. First note that product markups are increasing in firm productivity, as shown in eq. (4). Next, I totally differentiate the optimality condition for the expansion R&D rate and show that the expansion R&D rate is increasing in the markup.

Write the optimality condition for expansion R&D as[25]

$$\frac{\zeta-1}{\psi_x}(x^h)^\zeta + S_t\frac{Y_t}{w_t}\left(1-\frac{1}{\mu^h}\right) + (1-S_t)\frac{Y_t}{w_t}\left(1-\frac{1}{\mu^\ell}\right) = (\rho+\tau)\frac{\zeta}{\psi_x}(x^h)^{\zeta-1},$$

where μ^h and μ^ℓ denote the initial markup charged when facing a high- and low-productivity firm. Totally differentiate with respect to markups and the expansion R&D rate

$$S\frac{1}{(\mu^h)^2}\frac{Y_t}{w_t}d\mu + (1-S)\frac{1}{(\mu^\ell)^2}\frac{Y_t}{w_t}d\mu = \frac{\zeta(\zeta-1)}{\psi_x}\left((\rho+\tau)(x^h)^{\zeta-2} - (x^h)^{\zeta-1}\right)dx^h,$$

where $d\mu^h = d\mu^\ell \equiv d\mu$. Since $x^h > 0$, the above can be rearranged to

$$\frac{1}{(x^h)^{\zeta-2}}\left[S\frac{1}{(\mu^h)^2}\frac{Y_t}{w_t} + (1-S)\frac{1}{(\mu^\ell)^2}\frac{Y_t}{w_t}\right]d\mu = \frac{\zeta(\zeta-1)}{\psi_x}\left(\rho+\tau-x^h\right)dx^h.$$

The left-hand side captures the effect of changes in markups on profits. The right-hand side captures the effect of changes in the expansion R&D rate on the continuation value and marginal costs of expansion R&D. We know that $\frac{Y_t}{w_t} > 0$. For a stationary firm-size distribution, we must further have $\tau > x^h$. From this, it follows that $dx^h/d\mu > 0$, which concludes the proof.

[25] This uses the optimality condition of the high-type firm but the one of the low type works equivalently.

C.2 Steady-state growth rate of aggregate variables

The growth rate of Q_t determines the growth rate of aggregate variables

$$g = \frac{\dot{Q}_t}{Q_t} = \frac{\partial \ln(Q_t)}{\partial t}.$$

Quality of a product in a given product line increases through expansion R&D or firm entry. For the growth rate of Q_t over a discrete time interval Δ, we have

$$\ln(Q_{t+\Delta}) = \int_0^1 \Big[(\Delta S x^h + \Delta(1-S)x^\ell + \Delta z)\ln(\lambda) + \ln(q_{t,i})\Big]\, di$$

so that

$$\frac{\ln(Q_{t+\Delta}) - \ln(Q_t)}{\Delta} = \Big(Sx^h + (1-S)x^\ell + z\Big)\ln(\lambda).$$

For $\Delta \to 0$, $g = \Big(Sx^h + (1-S)x^\ell + z\Big)\ln(\lambda)$ as stated in Proposition 1.

C.3 Joint distribution of quality and productivity gaps

The distribution of quality and productivity gaps between incumbent and laggard firms characterizes economic aggregates in the model. On the one hand, quality and productivity gaps define product markups that determine labor demand. On the other hand, the joint distribution characterizes the share of product lines operated by each productivity type, which is a state variable in the firm's optimization problem. This section derives the joint distribution of quality and productivity gaps as a function of firm policies, which allows the equilibrium distribution to be solved jointly with the policies.

The distribution of quality and productivity gaps across product lines is characterized by four differential equations. For simplicity, I characterize the differential equations for firm-type specific expansion R&D rates, x_t^h and x_t^ℓ, as proven in Proposition 1 for a balanced growth path. The measure $\nu_t\left(\lambda, \frac{\varphi_f}{\varphi_{f'}}\right)$ of product lines in which an incumbent of productivity type φ_f is a λ quality step ahead of a $\varphi_{f'}$ type laggard follows

$$\dot{\nu}_t\left(\lambda, \frac{\varphi^\ell}{\varphi^h}\right) = (1-S_t)x_t^\ell S_t + z_t(1-p^h)S_t - \nu_t\left(\lambda, \frac{\varphi^\ell}{\varphi^h}\right)\tau_t$$

$$\dot{\nu}_t\left(\lambda, \frac{\varphi^\ell}{\varphi^\ell}\right) = (1-S_t)x_t^\ell(1-S_t) + z_t(1-p^h)(1-S_t) - \nu_t\left(\lambda, \frac{\varphi^\ell}{\varphi^\ell}\right)\tau_t$$
$$\dot{\nu}_t\left(\lambda, \frac{\varphi^h}{\varphi^h}\right) = S_t x_t^h S_t + z_t p^h S_t - \nu_t\left(\lambda, \frac{\varphi^h}{\varphi^h}\right)\tau_t$$
$$\dot{\nu}_t\left(\lambda, \frac{\varphi^h}{\varphi^\ell}\right) = S_t x_t^h(1-S_t) + z_t p^h(1-S_t) - \nu_t\left(\lambda, \frac{\varphi^h}{\varphi^\ell}\right)\tau_t. \tag{A-2}$$

Changes in ν_t are due to inflows and outflows arising from creative destruction. For example, in the first line, the measure of products with a low-type incumbent and high-type laggard $\nu_t\left(\lambda, \frac{\varphi^\ell}{\varphi^h}\right)$ increases due to low-type incumbents and entrants replacing high-type incumbents that occupy a share S_t of product lines, captured by $(1-S_t)x_t^\ell S_t + z_t(1-p^h)S_t$. The first term captures creative destruction by low-type incumbents and the second one creative destruction by low-type entrants. At the same time, the measure declines at rate τ_t as these firms themselves are subject to creative destruction.

To solve for $\nu_t\left(\lambda, \frac{\varphi^\ell}{\varphi^h}\right), \nu_t\left(\lambda, \frac{\varphi^\ell}{\varphi^\ell}\right), \nu_t\left(\lambda, \frac{\varphi^h}{\varphi^h}\right), \nu_t\left(\lambda, \frac{\varphi^h}{\varphi^\ell}\right)$ along the balanced growth path, I set the differential equations in (A-2) equal to zero. Denote the solutions to these four equations by $S_{\varphi^\ell,\varphi^h}, S_{\varphi^\ell,\varphi^\ell}, S_{\varphi^h,\varphi^h}, S_{\varphi^h,\varphi^\ell}$.

The following uses the obtained results to solve for the aggregate labor income share, the misallocation measure and the aggregate markup. The aggregate labor income share equals

$$\Lambda = \frac{wL_P}{Y} = \int_0^1 \mu_i^{-1} di = \frac{1}{\lambda}\sum_{k\in\{h,\ell\}}\sum_{n\in\{h,\ell\}}\frac{1}{\varphi^k/\varphi^n}S_{\varphi^k,\varphi^n}. \tag{A-3}$$

The TFP misallocation statistic $\mathcal{M}$ is given by

$$\mathcal{M} = \frac{\exp\left(\int_0^1 \ln \mu_i^{-1} di\right)}{\int_0^1 \mu_i^{-1} di} = \frac{1}{\lambda}\frac{\left(\frac{\varphi^h}{\varphi^\ell}\right)^{S_{\varphi^\ell,\varphi^h}-S_{\varphi^h,\varphi^\ell}}}{\Lambda}.$$

The aggregate markup (sales weighted across products) is given by

$$E[\mu] = \int_0^1 \mu_i di = \lambda \sum_{k\in\{h,\ell\}}\sum_{n\in\{h,\ell\}}\frac{\varphi^k}{\varphi^n}S_{\varphi^k,\varphi^n}. \tag{A-4}$$

The cost weighted aggregate markup equals $1/\Lambda$.

C.4 Firm dynamics

The following subsection derives the components of the size dispersion decomposition, in particular the age distribution and size dispersion conditional on age.

Mean size by age Mean log size conditional on age pools both productivity types with the type-share weights from eq. (19):

$$\mu(a_f) \equiv E\left[\ln \text{Size}_{f,t} \mid a_f\right] = s^h(a_f) \cdot \mu^h(a_f) + \left(1 - s^h(a_f)\right) \cdot \mu^\ell(a_f), \tag{A-5}$$

where $\mu^d(a_f) = E[\ln \text{Size}_{f,t} | a_f, \varphi_f = \varphi^d]$ denotes the type-specific size conditional on age.

Age distribution Since each entrant independently draws a high-type productivity with probability p^h and a low-type productivity with probability $1 - p^h$, the unconditional survival probability to age a_f is

$$\chi(a_f) = p^h \chi^h(a_f) + (1 - p^h)\chi^\ell(a_f), \tag{A-6}$$

where χ^h and χ^ℓ are defined in eq. (18). In the numerical implementation of the decomposition, continuous age is approximated on a fine grid with step size Δa and midpoints $\{a_k\}$. The discrete probability mass assigned to each age bin is

$$\omega(a_k) = \frac{p^h \chi^h(a_k) + (1 - p^h)\chi^\ell(a_k)}{\sum_j \left[p^h \chi^h(a_j) + (1 - p^h)\chi^\ell(a_j)\right]}.$$

Size dispersion conditional on age Within any age cohort, log-size dispersion arises from two sources: the allocation of firms across productivity types and the stochastic accumulation of product lines. The number of product lines of a surviving firm of type d at age a_f follows a geometric distribution with parameter $\gamma^d(a_f)$ (see eq. (17)):

$$P(n_f = n \mid a_f, \varphi_f = \varphi^d) = \left(1 - \gamma^d(a_f)\right) \left(\gamma^d(a_f)\right)^{n-1}, \quad n = 1, 2, \ldots$$

Pooling both types with weight $s^h(a_f)$ from eq. (19), the mixture distribution of product count conditional on age is

$$P(n_f = n \mid a_f) = s^h(a_f) \cdot \Big(1 - \gamma^h(a_f)\Big)\Big(\gamma^h(a_f)\Big)^{n-1} + \Big(1 - s^h(a_f)\Big) \cdot \Big(1 - \gamma^\ell(a_f)\Big)\Big(\gamma^\ell(a_f)\Big)^{n-1}.$$

The variance of log sales conditional on age a_f therefore equals the variance of log product count:

$$\mathrm{Var}(\ln \mathrm{Size}_{f,t} \mid a_f) = \sum_{n=1}^{\infty} P(n_f = n \mid a_f) \cdot (\ln n - \bar{\mu}(a_f))^2\,, \tag{A-7}$$

where $\bar{\mu}(a_f) = \sum_{n=1}^{\infty} \ln(n) \cdot P(n_f = n \mid a_f)$ is the mean log product count.

Implementation with integer age groupings Empirically, firm age is observed at annual frequency. For firms of integer age a (i.e., $a_f \in [a, a+1)$), the law of total variance applied across exact continuous ages within the bin gives within-bin size variance

$$V_a = \underbrace{\sum_{k\colon a_k \in [a,a+1)} \frac{\omega(a_k)}{p_a} \cdot \mathrm{Var}(\ln n_f \mid a_k)}_{\text{avg. within-exact-age variance}} + \underbrace{\sum_{k\colon a_k \in [a,a+1)} \frac{\omega(a_k)}{p_a} \cdot (\mu(a_k) - \bar{\mu}_a)^2}_{\text{variance of mean size across exact ages}}, \tag{A-8}$$

where $p_a = \sum_{k\colon a_k \in [a,a+1)} \omega(a_k)$ is the mass of firms in age bin a, $\mu(a_k)$ is the mean log size at fine-grid midpoint a_k from eq. (A-5), and $\bar{\mu}_a = \sum_{k\colon a_k \in [a,a+1)} (\omega(a_k)/p_a) \cdot \mu(a_k)$ is the within-bin conditional mean. The total cross-sectional variance then decomposes as

$$\mathrm{Var}(\ln \mathrm{Size}) = \underbrace{\sum_a p_a \cdot V_a}_{\text{within-age}} + \underbrace{\sum_a p_a \cdot (\bar{\mu}_a - \bar{\mu})^2}_{\text{between-age}},$$

where $\bar{\mu} = \sum_a p_a \bar{\mu}_a$ is the grand mean log size.

C.5 Firm-size distribution

The theory makes precise predictions about the firm-size distribution. Complementary to the size dispersion decomposition in the previous section, I derive the cross-sectional size distribution from firms' product counts. The mass of high- and low-productivity type firms with $n \geq 2$ products follows the differential equations

$$\begin{aligned}\dot{M}_t^h(n) &= (n-1)x_t^h M_t^h(n-1) + (n+1)\tau_t M_t^h(n+1) - n(x_t^h + \tau_t)M_t^h(n)\\ \dot{M}_t^\ell(n) &= (n-1)x_t^\ell M_t^\ell(n-1) + (n+1)\tau_t M_t^\ell(n+1) - n(x_t^\ell + \tau_t)M_t^\ell(n),\end{aligned} \tag{A-9}$$

whereas the mass of firms with one product evolves according to

$$\begin{aligned}\dot{M}_t^h(1) &= z_t p^h + 2\tau_t M_t^h(2) - (x_t^h + \tau_t)M_t^h(1)\\ \dot{M}_t^\ell(1) &= z_t(1-p^h) + 2\tau_t M_t^\ell(2) - (x_t^\ell + \tau_t)M_t^\ell(1).\end{aligned} \tag{A-10}$$

The mass of firms with n products increases through firms with $n-1$ products expanding to size n at rate x_t^h or x_t^ℓ per product or through firms with $n+1$ products losing a product at the rate of aggregate creative destruction τ_t. The mass of firms with n products decreases through firms with n products either gaining or losing a product through expansion or creative destruction. The mass of firms with one product additionally increases through firm entry.

Proposition A-1. *The stationary firm-size distribution along the balanced growth path is characterized as follows.*

1. *The mass of high and low productivity firms with n products is*

$$\begin{aligned}M^h(n) &= \frac{(x^h)^{n-1} z p^h}{n\tau^n} = \frac{zp^h}{x^h}\frac{1}{n}\left(\frac{x^h}{\tau}\right)^n\\ M^\ell(n) &= \frac{(x^\ell)^{n-1} z(1-p^h)}{n\tau^n} = \frac{z(1-p^h)}{x^\ell}\frac{1}{n}\left(\frac{x^\ell}{\tau}\right)^n.\end{aligned}$$

2. *The total mass of firms with n products is*

$$M(n) = M^h(n) + M^\ell(n) = \frac{(x^h)^{n-1} z p^h + (x^\ell)^{n-1} z(1-p^h)}{n\tau^n}.$$

3. The mass of firms of each productivity type is

$$M^h = \sum_{n=1}^{\infty} M^h(n) = \frac{zp^h}{x^h} \sum_{n=1}^{\infty} \frac{1}{n} \left(\frac{x^h}{\tau}\right)^n = \frac{zp^h}{x^h} \ln\left(\frac{\tau}{\tau - x^h}\right)$$
$$M^\ell = \sum_{n=1}^{\infty} M^\ell(n) = \frac{z(1-p^h)}{x^\ell} \sum_{n=1}^{\infty} \frac{1}{n} \left(\frac{x^\ell}{\tau}\right)^n = \frac{z(1-p^h)}{x^\ell} \ln\left(\frac{\tau}{\tau - x^\ell}\right)$$

4. The total mass of firms is

$$M = M^h + M^\ell.$$

Proof. These results follow from setting the time derivatives in equations (A-9) and (A-10) equal to zero and solving the system of equations. □

For each firm type, the share of firms with n products, $M^h(n)/M^h$ and $M^\ell(n)/M^\ell$, follows the PDF of a logarithmic distribution with parameter x^h/τ and x^ℓ/τ as in Lentz and Mortensen (2008). The firm-size distribution is highly skewed to the right.

Since there is a continuum of products of mass one and each product is mapped to one firm, $\sum_{n=1}^{\infty} M(n) \times n = 1$. Further, the mass of high-productivity type firms producing n products is related to the share of product lines operated by high-type firms, S, as follows

$$S = \sum_{n=1}^{\infty} M^h(n) \times n = \frac{zp^h}{\tau - x^h}.$$

The share of high-productivity type firms in the cross section is

$$S_{M^h} = \frac{M^h}{M}, \tag{A-11}$$

and the firm entry rate equals

$$\text{Firm entry rate} = \frac{z}{M}. \tag{A-12}$$

D Computation of transition dynamics

In this section, I lay out the numerical procedure to solve for the transition path. Since time is continuous, I solve a discretized version of the model where the solution converges to the one in continuous time for small enough time intervals. As shown in Supplemental Appendix C.1, value functions are additive across product lines. Therefore, I solve the problem of two representative one-product firms: one of the high productivity type and one of the low productivity type.

I normalize the value function by the wage w_t to obtain a stationary problem. The value function for the high-type firm (in discrete time) reads

$$
\begin{aligned}
\frac{V_t^h(1,\mu_i,S_t)}{w_t} &= \frac{Y_t}{w_t}\left(1-\frac{1}{\mu_i}\right)dt \\
&- \tau_t \exp(-r_{t+dt}dt)\frac{V_{t+dt}^h(1,\mu_i,S_{t+dt})}{w_{t+dt}}\frac{w_{t+dt}}{w_t}dt \\
&+ \max_{x_t^h}\left\{x_t^h \exp(-r_{t+dt}dt)\left(S_{t+dt}\frac{V_{t+dt}^h(1,\lambda,S_{t+dt})}{w_{t+dt}} + (1-S_{t+dt})\frac{V_{t+dt}^h(1,\lambda\frac{\varphi^h}{\varphi^\ell},S_{t+dt})}{w_{t+dt}}\right)\frac{w_{t+dt}}{w_t}dt - \frac{1}{\psi_x}(x_t^h)^\zeta dt\right\} \\
&+ \exp(-r_{t+dt}dt)\frac{V_{t+dt}^h(1,\mu_i,S_{t+dt})}{w_{t+dt}}\frac{w_{t+dt}}{w_t}.
\end{aligned}
$$

The value function for the low-type firm reads

$$
\begin{aligned}
\frac{V_t^\ell(1,\mu_i,S_t)}{w_t} &= \frac{Y_t}{w_t}\left(1-\frac{1}{\mu_i}\right)dt \\
&- \tau_t \exp(-r_{t+dt}dt)\frac{V_{t+dt}^\ell(1,\mu_i,S_{t+dt})}{w_{t+dt}}\frac{w_{t+dt}}{w_t}dt \\
&+ \max_{x_t^\ell}\left\{x_t^\ell \exp(-r_{t+dt}dt)\left(S_{t+dt}\frac{V_{t+dt}^\ell(1,\lambda\frac{\varphi^\ell}{\varphi^h},S_{t+dt})}{w_{t+dt}} + (1-S_{t+dt})\frac{V_{t+dt}^\ell(1,\lambda,S_{t+dt})}{w_{t+dt}}\right)\frac{w_{t+dt}}{w_t}dt - \frac{1}{\psi_x}(x_t^\ell)^\zeta dt\right\} \\
&+ \exp(-r_{t+dt}dt)\frac{V_{t+dt}^\ell(1,\mu_i,S_{t+dt})}{w_{t+dt}}\frac{w_{t+dt}}{w_t}.
\end{aligned}
$$

From this, one obtains the first order conditions for the policy functions. For the optimal expansion R&D rate of the high type firm x_t^h (again suppressing the dependence of the value function on S_t):

$$
\exp(-r_{t+dt}dt)\left(S_{t+dt}\frac{V_{t+dt}^h(1,\lambda)}{w_{t+dt}} + (1-S_{t+dt})\frac{V_{t+dt}^h(1,\lambda\frac{\varphi^h}{\varphi^\ell})}{w_{t+dt}}\right)\frac{w_{t+dt}}{w_t} = \frac{\zeta}{\psi_x}(x_t^h)^{\zeta-1} \tag{A-13}
$$

and for the low type firm x_t^ℓ:

$$\exp(-r_{t+dt}dt)\left(S_{t+dt}\frac{V_{t+dt}^\ell(1,\lambda\frac{\varphi^\ell}{\varphi^h})}{w_{t+dt}}+(1-S_{t+dt})\frac{V_{t+dt}^\ell(1,\lambda)}{w_{t+dt}}\right)\frac{w_{t+dt}}{w_t}=\frac{\zeta}{\psi_x}(x_t^\ell)^{\zeta-1}. \tag{A-14}$$

Both are independent of the markup μ_i.

Equations (D) to (A-14) characterize the firm problem in discrete time. These equations are supplemented by the law of motion for the distribution of productivity gaps

$$\nu_{t+dt}\left(\lambda,\frac{\varphi^\ell}{\varphi^h}\right)-\nu_t\left(\lambda,\frac{\varphi^\ell}{\varphi^h}\right)=dt\left[(1-S_t)x_t^\ell S_t+z_t(1-p^h)S_t-\nu_t\left(\lambda,\frac{\varphi^\ell}{\varphi^h}\right)\tau_t\right]$$
$$\nu_{t+dt}\left(\lambda,\frac{\varphi^\ell}{\varphi^\ell}\right)-\nu_t\left(\lambda,\frac{\varphi^\ell}{\varphi^\ell}\right)=dt\left[(1-S_t)x_t^\ell(1-S_t)+z_t(1-p^h)(1-S_t)-\nu_t\left(\lambda,\frac{\varphi^\ell}{\varphi^\ell}\right)\tau_t\right]$$
$$\nu_{t+dt}\left(\lambda,\frac{\varphi^h}{\varphi^h}\right)-\nu_t\left(\lambda,\frac{\varphi^h}{\varphi^h}\right)=dt\left[S_tx_t^hS_t+z_tp^hS_t-\nu_t\left(\lambda,\frac{\varphi^h}{\varphi^h}\right)\tau_t\right]$$
$$\nu_{t+dt}\left(\lambda,\frac{\varphi^h}{\varphi^\ell}\right)-\nu_t\left(\lambda,\frac{\varphi^h}{\varphi^\ell}\right)=dt\left[S_tx_t^h(1-S_t)+z_tp^h(1-S_t)-\nu_t\left(\lambda,\frac{\varphi^h}{\varphi^\ell}\right)\tau_t\right]$$

and a standard Euler equation

$$\frac{C_{t+dt}}{C_t}=\exp(-\rho dt)(1+r_{t+dt}dt). \tag{A-15}$$

Further, the (static) free entry and labor market clearing conditions remain unchanged and are characterized in the main text by equations (10) and (12).

The algorithm to compute the transition path assumes that the initial and final balanced growth paths have been solved for, including the (stationary) distribution of productivity gaps. I choose $dt = 0.02$ and set the transition period to 300 years (T), after which I assume the economy has reached its new balanced growth path. I then compute the transition path as follows:

1. Guess a path of wage growth $\frac{w_{t+dt}}{w_t}$ and income to wage ratios Y_t/w_t over the transition (equal to their values in the final balanced growth path)

    (a) Guess a path for S_t over the transition (equal to its value in the final balanced growth path).

i. Starting backwards in period T, solve for optimal policy functions in $T - dt$ using equations (A-13) and (A-14).

ii. Solve for τ_{T-dt} that ensures that the free entry condition (10) holds.

iii. Compute the value function in $T - dt$ using equations (D) and (D).

iv. Iterate backwards until the first time period.

v. Starting from the initial balanced growth path, simulate S_t forward using[26]

$$S_{t+dt} = S_t + dt\left[S_t x_t^h(1-S_t) - (1-S_t)x_t^\ell S_t + z_t(p^h(1-S_t) - (1-p^h)S_t)\right],$$

where z_t can be substituted out by equation (9).

(b) Update the guess for S_t from step v using bisection and go back to step i. Iterate until the guessed path for S_t converges to the implied one.

2. Starting from the initial balanced growth path, simulate the distribution of productivity gaps forward using equation D.

3. Solve for the implied sequence of $\frac{Y_t}{w_t}$ from the labor market clearing condition.

4. Compute the sequence of quality growth using

$$\frac{Q_{t+dt}}{Q_t} = \exp\left(\left[S_{t+dt}x_{t+dt}^h + (1-S_{t+dt})x_{t+dt}^\ell + z_{t+dt}\right] dt \ln(\lambda)\right).$$

5. Compute the sequence of aggregate productivity growth using

$$\frac{\Phi_{t+dt}}{\Phi_t} = \left(\frac{\varphi^h}{\varphi^\ell}\right)^{S_{t+dt}-S_t}.$$

6. Using the distribution of productivity gaps, compute the sequence of $\mathcal{M}_t$ defined in equation (8).

7. Compute the sequence of production labor L_{Pt} using equation (6).

[26] One could already simulate the entire distribution of productivity gaps forward here. However, for the inner loop, it is sufficient to iterate on S_t.

8. Compute the sequence of aggregate output growth $\frac{Y_{t+dt}}{Y_t}$ using equation (8).
9. With the paths of aggregate output growth and $\frac{Y_t}{w_t}$, obtain the implied path of wage growth $\frac{w_{t+dt}}{w_t}$.
10. Update the guesses for $\frac{w_{t+dt}}{w_t}$ and $\frac{Y_t}{w_t}$ using bisection and go back to step (a). Iterate until the guessed and implied paths converge.

I declare convergence when the difference between the three guessed and implied sequences is less than $\epsilon = 10^{-5}$ in absolute terms. I annualize wage growth before applying the convergence threshold.