Firm-Size Dispersion and the Limits of Age

Markus Kondziella*

August 3, 2026

Abstract

Firm age predicts average firm size—in Swedish administrative data, firms older than 25 years are more than a log point larger than entrants—yet age differences explain only 3.6% of the cross-sectional variance of log size; size dispersion *conditional on age* accounts for the residual 96.4%. A standard quality-ladder firm-dynamics model estimated on the mean size–age profile overstates the between-age share eightfold; adding persistent differences in firm growth—following from type heterogeneity in innovation productivity, disciplined by within-age dispersion—brings it close to the data. A firm's size then reflects its type more than its age, and small firms are negatively selected on it, so the size-screened R&D subsidies used in U.S. and EU small-business policy reach the least productive innovators. Screening eligibility on firm age instead delivers 0.02 to 0.36 percentage points more annual growth than the size screen at equal fiscal cost—at every statutory EU size ceiling.

Keywords: Firm dynamics, Firm-size dispersion, R&D subsidies, Administrative data

JEL codes: D22, L11, L25, E23, O38, O47

1 Introduction

Firms differ enormously in size. Among Swedish firms in 2017, the firm at the 75th percentile of the sales distribution is seven times as large as the firm at the 25th percentile. A long tradition organizes the analysis of this heterogeneity around the firm life cycle: firms enter small, and grow as they age. Central mechanisms of firm growth in the literature include gradual learning about one's type (Jovanovic, 1982),

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

expansion into new product markets over time (Klette and Kortum, 2004), growth out of financial constraints (Midrigan and Xu, 2014), or the slow accumulation of demand (Foster et al., 2016). Firm age is indeed a strong predictor of size—Swedish firms older than 25 years are, on average, more than a log point larger than entrants. It is therefore natural to presume that cross-sectional size dispersion reflects firms observed at different stages of their life cycle.[1]

This paper shows that this presumption fails. Using administrative data on the universe of Swedish private-sector firms, I apply the law of total variance to decompose the cross-sectional variance of log firm size into size differences across firm ages (the between-age component) and size dispersion conditional on age (the within-age component). This decomposition is model-free and holds exactly in the data. The main finding is that within-age dispersion accounts for 96.4% of the variance of log sales in 2017; differences in mean size across ages account for only 3.6%. Existing firm-dynamics work emphasizes the mean size-age profile as a central object for disciplining models. But the mean profile explains almost none of the cross-sectional size dispersion. There is no contradiction: the entire size–age profile is dwarfed by dispersion conditional on age. This result does not hinge on measurement choices: the within-age share exceeds 93% in every sector, is 95.7% when size is measured by employment instead of sales, and is robust to how firm age is measured. Nor is it an artifact of measuring dispersion by the variance: removing the entire mean size–age profile shrinks the p90–p10 gap of log size by only 1.9%. Nor is it specific to Sweden: combining public U.S. Census counts of firms and employment by age with the profile of dispersion conditional on age that Sterk et al. (2021) report for U.S. firms under twenty years of age puts the within-age share at 94.8% in the United States. Whatever generates firm-size dispersion must operate almost entirely conditional on age. These are the limits of age.[2]

What must a firm-dynamics model contain to respect this discipline? I take a quality-ladder model of firm dynamics and estimate three versions of it to the Swedish cross-section. The first version is a standard Klette and Kortum (2004) economy: firms enter the economy with a single product line, expand into new lines as they age, but eventually fall victim to creative destruction themselves. Final goods production is Cobb-Douglas so that a firm's total sales are proportional to the number of product lines it holds. The second version replaces the Cobb-Douglas aggregator with a CES

[1] The presumption is quantitatively substantive rather than rhetorical. A standard creative-destruction model of firm dynamics, estimated in Section 5 to the mean size–age profile and the age distribution, attributes about 29% of the cross-sectional size dispersion to differences across ages.

[2] The finding parallels a classic result on wage inequality: most earnings dispersion is residual, arising within rather than between narrowly defined demographic groups (Juhn et al., 1993). Firm-size dispersion is residual in the same sense—the size distribution is not a snapshot of firms moving along a common growth path and observed at different points of it.

technology so that a firm's size reflects the quality of its products and not only its product count. The third version further relaxes the assumption that all firms are ex-ante identical. At birth, firms are assigned a persistent type that governs their innovation productivity, introducing systematic differences in firms' growth rates to the model. Such ex-ante heterogeneity takes two forms in the literature, and I estimate both: types that differ in the size of the quality improvements their innovations achieve, as in Lentz and Mortensen (2008), and types that differ in the cost of innovating at a common step, as in Acemoglu et al. (2018). The step form is the baseline for exposition; I present the main results under it, estimate the cost form as a variant, and run the policy experiments under both, returning to the comparison at the end of this introduction. All three models are estimated to the mean size–age profile, the age distribution, and aggregate growth, which pin down the expansion costs, entry costs, and the average innovation step size, respectively. Targeting the size–age profile and the age distribution is standard practice for disciplining firm-dynamics models; size dispersion conditional on age is commonly left untargeted. The version with ex-ante firm heterogeneity additionally targets it, which directly disciplines the type heterogeneity.

The Cobb–Douglas version attributes 28.8% of size dispersion to the between-age component, against 3.6% in the data, an overstatement of nearly an order of magnitude. The model has no problem fitting the mean size–age profile, as is well known in the literature. Instead, the failure is too little size dispersion conditional on age: with sales proportional to the product count, the model's standard deviation of log sales conditional on age is only about one-seventh of the data's among the youngest firms. Replacing the unitary price elasticity of the Cobb-Douglas aggregator with an elasticity equal to five—allowing for substantial substitution towards firms with high quality items—reduces the between-share to 18.5%. Further adding the ex-ante heterogeneity in the size of firms' innovations brings the between-age share to 5.9%, close to the data.[3]

Ex-post shocks could not have played the types' role: they could fit any dispersion profile while contradicting the evidence that ex-ante differences account for most of size dispersion conditional on age (Sterk et al., 2021). Nor is the estimated type heterogeneity more than the data ask for: it is supported out of sample by the growth of the largest firms of a given age—a dynamic moment no version of the model is estimated to. In the data, these firms hold their growth and sustain a

[3]That the models without type heterogeneity miss the within-age share is not an artifact of leaving size dispersion conditional on age untargeted. In them, the only way to raise within-age dispersion is to raise firms' expansion rates, which simultaneously steepens the mean size–age profile; matching the observed dispersion would therefore require a counterfactually steep mean profile. Firm-type heterogeneity introduces exactly the margin these models lack—one that separates size dispersion from mean size, conditional on age.

population of fast growers that the homogeneous benchmark cannot produce and that the estimated type model reproduces quantitatively; if anything, the largest firms of an age mean-revert even less in the data than in the estimated model. As an untargeted by-product, the type-model generates a Pareto right tail of firm size with an exponent near one, the Zipf regularity of the largest firms.

The decomposition also carries a warning for policy. Heterogeneity within age groups dwarfs the differences between them, so a small firm need not be a young firm that has yet to grow—the left-tail counterpart of Luttmer (2011)'s observation that the largest U.S. firms are too large for their age to be the product of typical life-cycle growth. If firms are small for reasons other than a young age, size is a poor screen for policies whose economic case rests on reaching young firms with room to grow. Small size—measured by both sales and employment—is nonetheless used as a direct eligibility criterion for government subsidies under the Small Business Act in the United States and the Small Business Act for Europe.[4]

The estimated model turns this warning into a quantitative statement about whom a size screen reaches. Because size reflects a firm's type, the firms it selects are negatively selected on innovation productivity rather than young firms waiting to grow: below the EU's "micro-enterprise" turnover ceiling (€2 million)—carried into the model by the share of aggregate sales such firms hold in the Swedish data—eligible firms are 11% high types and take innovation steps only 0.73 times as large as those of the firms left out, and close to half of them are more than ten years old. The youngest firms holding the same share of aggregate sales are instead 22% high types and take steps 1.9 times as large as those they leave out. At equal fiscal outlay, screening eligibility on firm age rather than size therefore delivers 0.31 percentage points more annual growth, and 0.36 and 0.35 points at the "small-enterprise" (€10 million) and "medium-enterprise" (€50 million) ceilings. The asymmetry has a simple root: size is endogenous, so being small is itself evidence of innovating poorly, and a rule written on size inherits that adverse selection; no firm, by contrast, can choose its age, so an age cutoff works like an immutable tag.

In the estimated model, the misdirection is costly enough that the size screen does not merely buy less growth than the age screen—it lowers growth outright. Eligible firms do raise their R&D, but crowding out high-type innovation through the R&D labor market shifts innovation toward firms that improve their products in small steps,

[4]Both criteria are size-based in sales and employment, and bind tightly for some industries. Under the EU SME definition (Commission Recommendation 2003/361/EC), which underpins the Small Business Act for Europe, eligibility can require as few as ten employees and turnover below €2 million. In the United States, the Small Business Administration sets size standards by industry under the Small Business Act, with thresholds as low as roughly $2 million in average annual receipts or 100 employees for some industries.

and the fall in the economy's average innovation step outweighs the extra effort: the micro-ceiling subsidy *reduces* aggregate growth by 0.24 percentage points, about twelve percent of the economy's 2% growth rate. The loss is nearly identical at the small-enterprise ceiling and only modestly smaller at the medium-enterprise one, so no current statutory cutoff escapes it, while the age screen raises growth by 0.07 to 0.16 points. Without the type heterogeneity, the size screens increase growth, by about 0.05 percentage points: a policymaker who read size dispersion as life-cycle would predict a small gain and deliver a sizable loss.

Whether the size screen lowers growth outright depends on the form the heterogeneity takes: because the loss runs through the average innovation step, it requires types that differ in the *size* of their innovations rather than in the *cost* of innovating at a common step. I therefore re-estimate the model with cost heterogeneity in place of the step, on exactly the same moments. What survives the change of form is the verdict on the screen. Small firms are negatively selected on innovation productivity under both, so the size screen reaches the least productive innovators either way, and screening on age dominates screening on size under both, at every statutory ceiling. What the form determines is the price of using the size screen. Under cost heterogeneity, the subsidy no longer lowers growth—it raises it by 0.04 to 0.06 percentage points—and the age screen's advantage narrows to 0.02 to 0.03 percentage points, from 0.31 to 0.36 under the step form: the misdirected subsidy now merely buys less growth than a better-targeted design would, rather than lowering growth outright. The estimation provides suggestive evidence for the step form, though not a final verdict: the cost form matches most targeted profiles about as well, but features no size dispersion across types at entry and so undershoots the dispersion of young firms—the moment that identifies the type heterogeneity in the first place—and, carrying every type difference on the expansion rate alone, its dispersion then rises steeper with age than in the data. I carry both forms through the policy experiments and state the paper's policy conclusion at the level both support: size-based screening is not the natural way to reach high-growth innovators, and how expensive a screen it is depends on the source of the persistent heterogeneity.

Related literature. The paper's argument runs in three steps—the life-cycle view of firm size is empirically incomplete; standard firm-dynamics models cannot match the dominant within-age dispersion without persistent firm heterogeneity; and that heterogeneity overturns the conventional logic of size-based innovation policy—and each step engages a different literature. The first relates the paper to the empirical literature on firm dynamics and the firm life cycle (Haltiwanger et al., 2013; Hsieh and Klenow, 2014; Hopenhayn et al., 2022; Karahan et al., 2024). This literature documents how firms grow with age and how firm demographics have shifted over time; relative to it, I quantify how much of cross-sectional size dispersion the mean size–age profile accounts for, and find that it is negligible.

The paper speaks to theories of firm dynamics in which age is a driver of firm size, whether through learning, selection, gradual expansion, financial constraints, or demand accumulation (Jovanovic, 1982; Hopenhayn, 1992; Klette and Kortum, 2004; Luttmer, 2007; Peters, 2020; Midrigan and Xu, 2014; Foster et al., 2016). The decomposition provides a model-free moment that disciplines this class of models.

The paper relates to the literature on the origin of the heavy tail of the firm-size distribution (Luttmer, 2007; Gabaix, 2009; Luttmer, 2011). Random-growth models generate the cross-sectional distribution by mixing growth histories across ages, with growth common to a cohort and diffusion within it as the two engines of dispersion. The decomposition caps the first engine at 3.6% of the variance—and it caps only growth *common* to firms of the same age: heterogeneity in growth rates within a cohort loads on the within component. Nearly all dispersion must therefore be generated within cohorts. The decomposition thereby disciplines the random-growth class directly, bounding how much dispersion any such model may load on the common age profile. The result is also the cross-sectional counterpart of Luttmer (2011): where he finds the largest U.S. firms too large for their age to have grown at typical life-cycle rates, the decomposition finds firm sizes too dispersed over the whole cross-section to be explained by age differences, extending his observation from the right tail to the entire size distribution.[5]

The paper connects to a literature on the persistent effects of the aggregate conditions at a firm's birth: firms born in recessions start smaller and remain smaller over their life cycle (Moreira, 2016), and cohort-level employment is shaped by aggregate conditions at entry, largely through the share of startups with high growth potential (Sedláček and Sterk, 2017). Because a firm's cohort is its birth year, age and cohort coincide in the cross-section, so any effect of birth conditions common to a cohort is absorbed into the between-age component in its entirety, and the decomposition bounds the dispersion it can generate at 3.6%. Birth conditions that affect firms within a cohort differentially instead load on the within component, and the decomposition places no bound on them. This is not in tension with the role of birth conditions for aggregate dynamics, which concern the *level* of firm outcomes. The within-cohort emphasis of cross-sectional size dispersion is, in fact, consistent with Sedláček and Sterk (2017)'s finding that cohort outcomes are governed by the

[5] In recent work, Berlingieri et al. (2024) generate the heavy tail of the firm-size distribution through innovation bursts—episodes in which a single idea yields applications across many product lines at once—so that the size distribution inherits the power-law tail of the burst-size distribution rather than requiring the transient phase of fast growth of Luttmer (2011). The present paper instead focuses on dispersion over the entire cross-section and its split into size differences across ages and dispersion conditional on age.

composition of firm types present at birth.

The paper complements the literature on ex-ante versus ex-post firm heterogeneity. Sterk et al. (2021) decompose size dispersion conditional on age into an ex-ante component, reflecting differences across firms determined at entry, and an ex-post component that accumulates thereafter, using the autocovariance structure of employment. Their decomposition answers a different question: the ex-ante/ex-post split of size dispersion conditional on age does not determine how much the latter contributes to total size dispersion. This paper shows that within-age dispersion accounts for essentially all of size dispersion.[6][7] Put together, the decomposition here locates nearly all of cross-sectional size dispersion within age groups, and Sterk et al. (2021) attribute most of the dispersion within age groups to ex-ante differences.[8]

The model and policy analysis in the second half of the paper connect it to the literature on innovation-driven growth with heterogeneous firms and the growth policy built on it. In a quality-ladder model of creative destruction, Lentz and Mortensen (2008) model permanent firm types that differ in the quality—the step size—of their innovations, the form my baseline adopts; Akcigit and Kerr (2018) likewise make innovation step sizes heterogeneous, though across an innovation's type—internal versus external—rather than across firms, and study their consequences for firm-level growth rather than policy. Acemoglu et al. (2018) instead let types differ in innovation costs and use the resulting reallocation of innovation across types to study R&D and growth policy—the alternative form estimated here as a variant. It differs from these papers in two respects. First, in identification: I additionally target size dispersion conditional on age, the moment that pins down the type heterogeneity and that estimations of standard models typically leave untargeted. Second, in focus: I ask whether a common feature of small-firm policy—eligibility screened on firm size—misallocates R&D, and show that it does once firm heterogeneity is disciplined to match the cross-section, whereas an age screen delivers more growth at equal fiscal outlay.

Lastly, the policy exercise speaks to the empirical literature evaluating size-targeted

[6] In principle, size dispersion conditional on age could reflect ex-ante heterogeneity as Sterk et al. (2021) show and yet account for little of total dispersion, with cross-sectional size differences dominated by size-age differences instead.

[7] A separate conceptual difference between both decompositions is that the within/between decomposition requires no identifying assumptions because age is observable, whereas separating ex-ante from ex-post heterogeneity requires panel data and a model, as a firm's type is unobservable.

[8] The reduced-form wedges of the misallocation literature (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009) map into the distinction between ex-ante and ex-post heterogeneity. A wedge drawn at entry and persistent thereafter is firm-type heterogeneity in all but name; an idiosyncratic wedge drawn after entry is ex-post heterogeneity, and it is this margin that Sterk et al. (2021) bounds.

innovation support, which finds sizable positive effects on the subsidized firms: Howell (2017) for grants under the U.S. Small Business Innovation Research program, Dechezleprêtre et al. (2023) for the U.K.'s SME R&D tax credit—identified at the very kind of size threshold this paper models—and Criscuolo et al. (2019) for size-targeted investment subsidies. The estimated model agrees with that evidence: eligible firms raise their R&D in every design of Section 6. The aggregate loss arises in general equilibrium, from the crowding out of ineligible firms' innovation—an effect outside the reach of firm-level research designs—which is why evaluating the screen requires a structural model. The paper also relates to the literature on size-dependent policies (Guner et al., 2008; Garicano et al., 2016), which quantifies the cost of size-contingent rules through the incentive they create for firms to stay small; the cost identified here operates instead through selection—whom the eligibility rule reaches—though the stay-small margin is also priced into the solved equilibrium.

The remainder of the paper is organized as follows. Section 2 describes the administrative data and measurement choices. Section 3 introduces the model-free decomposition of firm-size dispersion, documents its ingredients in the 2017 cross-section—mean size conditional on age, the firm-age distribution, and size dispersion conditional on age—and decomposes firm-size dispersion into its between- and within-age components. Section 4 develops a quality-ladder model of firm dynamics with ex-ante firm heterogeneity and Section 5 estimates it, showing that a Cobb–Douglas benchmark overstates the between-age component by an order of magnitude while the model with ex-ante types matches the decomposition. Section 6 uses the estimated model to evaluate small-firm R&D subsidies screened on firm size versus age. Section 7 concludes.

2 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. All nominal variables are deflated to 2017 Swedish kronor (SEK) using the GDP deflator; deflation leaves the within-year decomposition unchanged—a common price level shifts all log sizes equally. To mitigate the influence of outliers, I winsorize sales at the 0.1th and 99.9th percentiles.[9] I further restrict to firms with positive sales and at least one employee since the main exercise decomposes log size dispersion. For a

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

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. The decomposition is robust to how firm age is measured.

The final dataset comprises 4,833,580 firm-year observations. Summary statistics for the pooled sample of firm-years are reported in Table 1. The median firm has sales of approximately 2.8 million SEK (1 SEK ≈ 0.1 USD) and employs two full-time equivalent workers. The distributions of sales and inputs are highly right-skewed: mean sales (16.5 million SEK) and employment (7.7 workers) substantially exceed their respective 75th percentiles. A nontrivial fraction of firms report zero capital stock (fixed assets net of depreciation), as reflected in the zero-valued 25th percentile of that variable.

Table 1: Summary statistics

	25th Pct.	50th Pct.	75th Pct.	Mean	SD
*Sales**	1.3	2.8	7.9	16.5	76.7
*Intermediate Inputs**	0.4	1.1	3.1	7.3	42.6
*Capital stock**	0.0	0.2	1.1	6.0	116.0
Employment	1	2	5	7.7	36.4
*Wage bill**	0.3	0.6	1.6	2.7	13.7

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 4,833,580 observations.

I use firm sales as the baseline measure of size and provide robustness checks using employment.[10] Sales provide a continuous measure of firm size, whereas employment is discrete. For example, the 25th, 50th, and 75th percentiles of employment shown in Table 1 are 1, 2, and 5, respectively.

3 Decomposing firm-size dispersion

How much do the between and within components contribute to cross-sectional size dispersion? This section states the decomposition, measures its ingredients in the 2017 cross-section, computes the two components, and probes the result's robustness.

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

3.1 The decomposition

Consider the law of total variance, $\mathrm{Var}(Y) = \mathrm{Var}(\mathbb{E}[Y|X]) + \mathbb{E}[\mathrm{Var}(Y|X)]$, 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 (the "between" component) and size dispersion conditional on age (the "within" component). Specifically,

$$\mathrm{Var}\left(\ln s_f\right) = \underbrace{\mathrm{Var}(\mathbb{E}[\ln s_f|a_f])}_{\text{Between}} + \underbrace{\mathbb{E}[\mathrm{Var}(\ln s_f|a_f)]}_{\text{Within}}. \tag{1}$$

This decomposition highlights that cross-sectional size dispersion may be large because old firms are on average much larger than young firms (the between component) or because firm sizes conditional on age are dispersed (the within component).[11] Denoting by $p(a_f)$ the share of firms of age a_f, one can write eq. (1) as

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

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.[12]

3.2 The ingredients

I now document, in the cross section, the overall dispersion of firm size and the three ingredients of the decomposition in eq. (1): mean size conditional on age, the firm-age distribution, and size dispersion conditional on age. All four objects are measured in the same data and the same year, so the decomposition holds exactly. Throughout I use 2017, the final year of the sample, because firm age is then least truncated; the

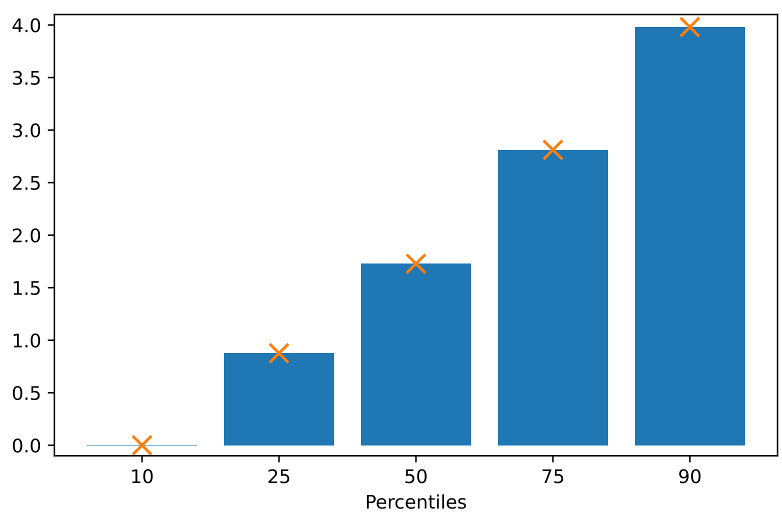
[11] Similarly, the decomposition can be leveraged to decompose *changes* in firm-size dispersion over time, see Kondziella (2026).

[12] Because age is discrete, the between-age share is equivalent to the R^2 of a regression of $\ln s_f$ on a full set of age dummies. Two remarks. First, that regression is saturated, so the share is not one specification's fit but the largest share of the variance of log size that any function of age can account for. Second, the relevant benchmark for this R^2 is the one implied by the models the moment is meant to discipline, which is larger by nearly an order of magnitude (Section 5).

decomposition is repeated for earlier years in Section 3.4.

Overall firm-size dispersion. Figure 1 summarizes the overall dispersion of firm size in the 2017 cross section (288,229 observations). It plots percentiles of the log sales distribution relative to the 10th percentile, which is normalized to zero, so each bar measures the log size gap to the bottom decile. The gaps are large: the median firm is about 1.7 log points larger than the 10th-percentile firm, a ratio of roughly 5.6; the 75th-percentile firm is 2.8 log points (a factor of about 17) larger; and the 90th-percentile firm is 4.0 log points larger—a roughly fifty-fold difference in sales. Equivalently, the interquartile range of log sales is about 1.9 log points, so the firm at the 75th percentile is roughly seven times as large as the firm at the 25th percentile. Firm-size dispersion is thus substantial; the rest of this section asks how much of it is accounted for by firm age.

Figure 1: Cross-sectional firm-size dispersion

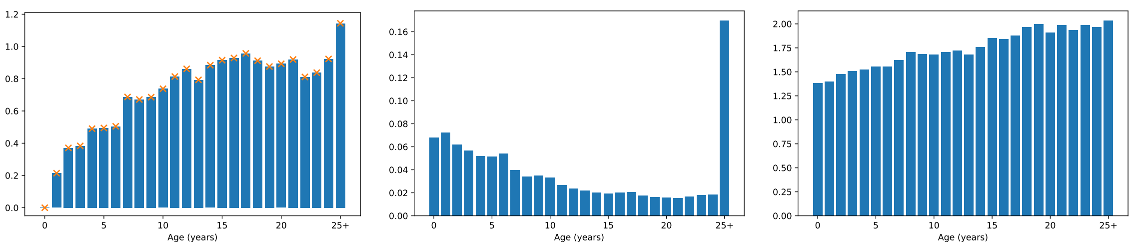


Notes: The figure shows percentiles (10th, 25th, 50th, 75th, 90th) of the log sales distribution in 2017, each expressed relative to the 10th percentile, which is normalized to zero. Sales are in 2017 SEK.

Firm size conditional on age. The first ingredient of the decomposition is the mean size–age profile, the source of variation in the between-age component. Panel (a) of Figure 2 plots the average of log sales by integer firm age in 2017, with age zero (entrants) normalized to zero. Average size rises steeply across the youngest age groups and then flattens out among older ages. Firms aged five are about 0.49 log points (a factor of 1.6) larger than entrants, firms aged ten about 0.74 log points larger, and firms aged twenty about 0.9 log points (a factor of 2.4) larger. The oldest firms, grouped at ages 25 and above, are 1.14 log points—more than a factor of three—larger than entrants. Firm age is thus a strong predictor of average firm size.

Figure 2: The three ingredients of the decomposition

(a) Firm size cond. on age (b) Firm-age distribution (c) Dispersion cond. on age



Notes: All three panels use the 2017 cross-section. *(a)* The average of log sales by integer firm age, normalized to zero at age zero. *(b)* The share of firms at each integer age. *(c)* The standard deviation of log sales within each integer age group. Firms aged 25 and above are grouped; sales are in 2017 SEK.

The firm-age distribution. The second ingredient is the firm-age distribution, which supplies the weights $p(a_f)$ in eq. (1). Panel (b) of Figure 2 shows the share of firms at each integer age in 2017. Entrants make up about 6.8% of all firms, and the share declines with age—to about 5% at age five, 3% at age ten, and under 2% at age twenty. The exception is the oldest group: firms aged 25 and above account for about 17% of firms, since this bin pools all older cohorts. The age distribution is thus heavily skewed toward young firms.

Firm-size dispersion conditional on age. The third ingredient is size dispersion conditional on age, the within-age component of the decomposition. Panel (c) of Figure 2 shows the standard deviation of log sales within each integer age group in 2017. Two features stand out. First, there is already substantial dispersion at entry: the standard deviation of log sales among entrants is about 1.39. This is large: it implies that two same-aged entrants one standard deviation apart in log sales differ in size by a factor of four, and that an entrant one standard deviation above the mean is roughly sixteen times the size of one an equal distance below. Second, dispersion increases gradually with age—to about 1.56 at age five, 1.68 at age ten, 1.91 at age twenty, and 2.03 in the oldest group. Dispersion conditional on age is thus large at every age, which foreshadows its dominant role in the decomposition.

3.3 The result

I decompose the variance of log firm size in 2017 according to eq. (1), partitioning firm ages into integers up to age 24, with older firms grouped, i.e., $a_f \in \{0, 1, \ldots, 24,$

25+}.

Table 2 reports the results. The within component—size dispersion conditional on age—accounts for 96.4% of the variance of log sales Var($\ln s_f$). Because the decomposition is linear in the two components, the between component—differences in mean size across ages—accounts for the residual 3.6%. The overwhelming share of cross-sectional size dispersion thus reflects dispersion conditional on age rather than size differences across ages.

Table 2: Decomposition of Var($\ln s_f$)

	Share within (%)
Total economy	96.4
Manufacturing	93.9
Construction	97.1
Wholesale + Retail Trade	95.5
Transportation	95.0
ICT	96.0
Prof., Scient., Techn. Services	99.5
Education	93.9
Healthcare	97.1

Notes: The table decomposes the variance of log firm sales, Var($\ln s_f$), in 2017 into the within and between component according to eq. (1) and shows the share of the within component.

That the within component dominates is not obvious ex ante, because the firm size–age gradient is itself substantial: Figure 2a showed that firms in the oldest group, aged 25 and above, are on average 1.14 log points larger than entrants. Yet that entire profile is smaller than the standard deviation of log sales within a single age group, which is already 1.39 among entrants and reaches 2.03 among the oldest firms. The between-age share is small not because age fails to predict size, but because what age predicts is dwarfed by the dispersion that remains once age is held fixed. Firm age is thus highly informative about a firm's expected size, yet nearly uninformative about where in the size distribution the firm falls.

What the decomposition constrains deserves equal precision. The 3.6% between-age share caps the contribution of the mean size–age profile—the life-cycle path common to firms of the same age—to cross-sectional size dispersion; it does not rule out the mechanisms behind that profile per se. Learning, gradual expansion, the escape from financial constraints, and demand accumulation all generate a positive size–age relationship, but firms experience each of them heterogeneously, so the same mechanisms also generate size dispersion conditional on age, and that part of their contribution loads on the within component. The result is thus a bound on the common, mean-profile part of any growth mechanism—and of all of them jointly. Section 5 makes the distinction concrete for a creative-destruction model of firm dynamics: the same expansion process generates both the mean size–age profile and dispersion around it, and the decomposition disciplines the split between the two.

3.4 Robustness

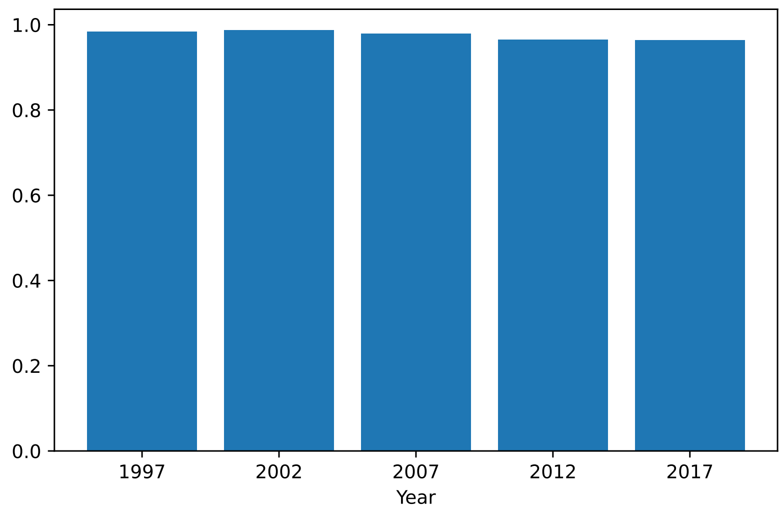
The dominance of the within component is robust to the measure of dispersion and across sectors, base years, size measures, and the grouping and measurement of firm age, and it is not specific to Sweden.

Other dispersion measures. The decomposition in eq. (1) measures dispersion by the variance, which weights deviations from the mean quadratically. The result does not depend on that choice. Consider instead the p90–p10 gap of log sales, which is insensitive to the tails of the distribution. In the 2017 cross section, the raw gap is 3.98 log points (Figure 1). Netting out mean log sales within each age group—that is, removing the entire mean size–age profile, using the same age grouping as in the baseline—leaves a gap of 3.91 log points, a reduction of only 1.9%. Differences in mean size across ages thus account for essentially none of the p90–p10 gap. The two measures agree closely once they are expressed in the same units: because the residualized variable has mean zero within every age group, its variance is exactly the within component of eq. (1), so the same residualization scales the standard deviation of log sales by $\sqrt{0.964}$, a reduction of 1.8%, against the 1.9% reduction in the p90–p10 gap.

Sectors. The high within-share also obtains when decomposing the variance of size dispersion within sectors. As shown in Table 2, across one-digit sectors, the within share ranges from 93.9% (Manufacturing) to 99.5% (Professional, Scientific, and Technical Services), so the result is not an artifact of pooling across sectors.

Other years. The baseline uses 2017 because firm age is least truncated in the final year of the sample. The employer-employee records that date firm births begin in 1992, so firm age is observed without truncation, starting from age zero, for firms founded in 1993 or later. The oldest untruncated integer age is therefore 24 in 2017, and fewer integer ages are untruncated in earlier base years—up to age 19 in 2012, age 14 in 2007, and so on. A base year with fewer untruncated ages amounts to a coarser age grouping at the top, which mechanically raises the within share. Figure 3 reports the within share for base years at five-year intervals back to 1997. The within share is stable, between about 96% and 99%, and edges toward 100% in earlier (more

Figure 3: Within-age share by base year



Notes: The figure shows the share of the within component in the decomposition of $\text{Var}(\ln s_f)$ in eq. (1), computed for base years at five-year intervals. Each base year uses integer ages up to the oldest untruncated age in that year, with older firms grouped.

truncated) years, exactly as this logic implies. The 2017 baseline is thus the most conservative choice.

Employment. The high within share is not specific to sales. Table 3 repeats the decomposition using employment. The within share is 95.7% for the total economy and ranges from 91.9% (Manufacturing) to 97.7% (Professional, Scientific, and Technical Services) across sectors. That the result survives the switch to employment indicates that it reflects genuine size heterogeneity among same-aged firms, not dispersion in markups that would show up in sales. The two ingredients of the decomposition are similar to their sales counterparts, only flatter (Figure 4): the oldest firms employ about 0.76 log points more than entrants, against 1.14 log points for sales, and dispersion conditional on age is lower at every age, rising from about 0.69 at entry to 1.31 in the oldest group, against 1.39 and 2.03 for sales.

Age grouping and measurement. As the comparison across base years already showed, the split between the two components depends on the age grouping: the coarser the grouping, the larger the within component. The baseline uses the finest feasible grouping, integer ages up to 24, so any coarser grouping would only raise the within share further. The result is also robust to how firm age is measured: dating firm birth by the year the firm first files tax returns and appears in the balance-sheet data, rather than by the year the first worker joins the firm, leaves the split at 94.2%

Table 3: Decomposition of Var($\ln s_f$), size measured by employment

	Share within (%)
Total economy	95.7
Manufacturing	91.9
Construction	95.9
Wholesale + Retail Trade	93.4
Transportation	95.0
ICT	94.4
Prof., Scient., Techn. Services	97.7
Education	92.7
Healthcare	96.2

Notes: The table decomposes the variance of log firm size in 2017 into the within and between component according to eq. (1), with firm size s_f measured by employment instead of sales, and shows the share of the within component.

within and 5.8% between, and excluding firms for which a restructuring—a merger or reorganization imputed from worker flows—renders a new firm ID gives 96.6% and 3.4%. Appendix A.2 details both exercises.[13]

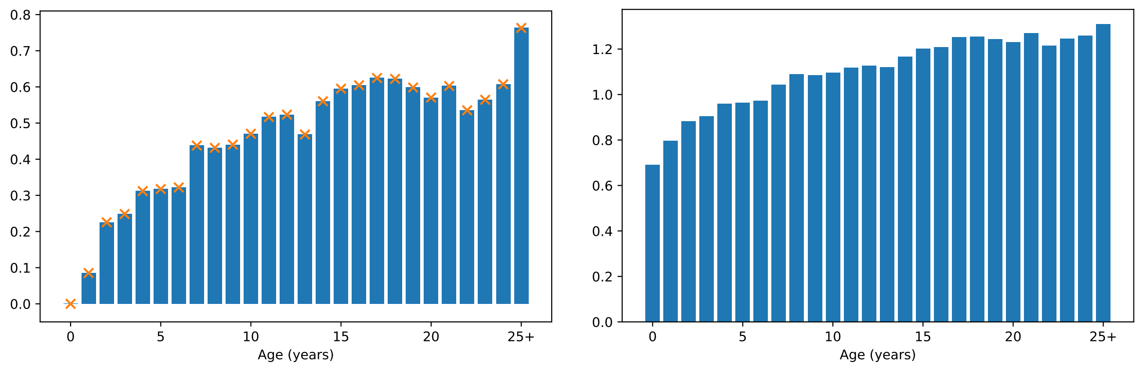
The United States. The decomposition can be carried out for the United States, in employment, by assembling its three margins from published sources. The Business Dynamics Statistics (BDS) report the number of firms and total employment by firm age, which give the age distribution and mean size by age. Dispersion conditional on age is not published in the BDS and is taken from Sterk et al. (2021), who use U.S. Census data on firms under twenty years of age and report that the standard deviation of log employment rises approximately linearly with age, from about 0.9 at entry to about 1.3 by age nineteen; I interpolate linearly between the two to obtain the intermediate ages. Their sample sets the age range of the comparison, which therefore covers firms aged nineteen and below—70% of U.S. and 75% of Swedish firms.

Two steps are needed to bring the BDS to the integer-age grid used above. First, the BDS group ages above five into five-year cells. Since the imported dispersion profile is defined at integer ages, I spread each BDS age groups over integer ages by linear interpolation, preserving the published firm counts and employment exactly. Second, the BDS report totals, so mean log employment has to be recovered from mean em-

[13] Entrants are not necessarily observed for a full calendar year, so age-zero sales may not cover a full year. Excluding age zero raises the within share to 97.3%: it removes the steepest step of the concave size–age profile, shrinking the between component, and drops the least dispersed age group, raising the within component. Retaining age zero is the conservative choice.

Figure 4: Firm size and dispersion conditional on age (employment)

(a) Firm size conditional on age (b) Dispersion conditional on age



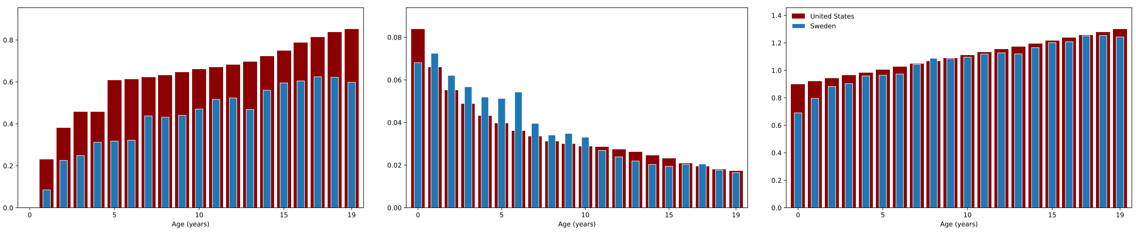
Notes: The left panel shows the average of log employment by integer firm age in 2017, normalized to zero at age zero. The right panel shows the standard deviation of log employment within each integer age group. Firms aged 25 and above are grouped.

ployment. Under within-age lognormality $\mathbb{E}[\ln s_f|a_f] = \ln \mathbb{E}[s_f|a_f] - \mathrm{Var}(\ln s_f|a_f)/2$, so the imported dispersion profile also supplies the correction. Because the firm-size distribution has a thicker right tail than the lognormal, this is a lower bound on the correct adjustment, so the size–age profile it implies is if anything too steep and the between-age component it delivers is an upper bound.

Figure 5 shows the three margins in both countries. The age distributions are close, and dispersion conditional on age differs mainly at entry, where the U.S. standard deviation is 0.90 against Sweden's 0.69; from age two onward the two are within 0.06 log points at every age. The mean size–age profile is steeper in the U.S.: mean log employment rises 0.85 log points from entry to age nineteen, against 0.60 in Sweden. That difference matters less than it appears. The steeper profile does raise the level of the between-age component relative to Sweden, but the U.S. between-age component remains tiny next to the U.S. within-age component. In levels, the between-age component of the variance of log employment is 0.034 in Sweden and 0.062 in the United States, while the within-age component is 1.00 and 1.11. Sweden therefore splits 96.7% within and 3.3% between; the United States, 94.8% and 5.2%. The within share is insensitive to the size–age profile precisely because age accounts for so little to begin with. The Swedish evidence thus does not appear special.

Figure 5: The three ingredients in the United States and Sweden

(a) Firm size cond. on age (b) Firm-age distribution (c) Dispersion cond. on age



Notes: All three panels use the 2017 cross-section of firms aged 0–19 in the United States (wide dark bars) and Sweden (narrow light bars), with firm size measured by employment. *(a)* The average of log employment by integer firm age, normalized to zero at age zero. *(b)* The share of all firms in the country at each integer age; the bars do not sum to one because firms aged 20 and above are excluded, and the decomposition weights each age by its share within the sample. *(c)* The standard deviation of log employment within each integer age group; the legend shown there applies to all three panels. U.S. firm counts and employment are from the BDS, with ages six and above interpolated within the published five-year cells; U.S. dispersion is the linearly interpolated profile of Sterk et al. (2021).

4 A model of firm-size dispersion

The decomposition disciplines quantitative models of firm dynamics before any structure is imposed: whatever generates firm-size dispersion must do so almost entirely within age groups. This section develops a model to ask what a standard firm-dynamics model must contain to meet that discipline, and Section 5 estimates it. The model is a quality-ladder model of firm dynamics in the Klette–Kortum tradition (Klette and Kortum, 2004) with two departures. First, the final good aggregates intermediate varieties with a constant elasticity of substitution rather than Cobb–Douglas, so that a firm's size reflects the *quality* of its product lines and not only their number. Second, the model relaxes the assumption that all firms are identical at birth: at entry, each firm draws a persistent type that governs its innovation productivity—in the baseline developed here, the size of its innovation steps; the variant in which the type enters innovation costs instead is taken up in Section 6 and Appendix I.

4.1 Environment

Time is continuous. A representative household with preferences $\int_0^\infty e^{-\rho t} \ln C(t)\, dt$ supplies one unit of labor inelastically and owns all firms.

Final good. A competitive final good—the numeraire—is produced from a unit continuum of intermediate varieties $j \in [0, 1]$ with the quality-augmented CES technology

$$Y=\left(\int_0^1 \left(q_j\, y_j\right)^{\frac{\varepsilon-1}{\varepsilon}} dj\right)^{\frac{\varepsilon}{\varepsilon-1}}, \qquad \varepsilon>1, \tag{3}$$

where q_j is the quality and y_j the quantity of variety j. Cost minimization gives the inverse demand $y_j = q_j^{\varepsilon-1} p_j^{-\varepsilon} Y$, and the numeraire normalization is the price-index condition $\int_0^1 q_j^{\varepsilon-1} p_j^{1-\varepsilon} dj = 1$.

Intermediate producers, prices, and sales. Each line j is served by the firm holding its leading-edge technology, producing with the linear technology $y_j = \ell_j$ in labor at wage w. To rule out limit pricing by displaced incumbents I adopt the standard two-stage price-bidding device: a firm wishing to supply line j first pays a vanishing entry fee, then all suppliers set prices simultaneously; only the quality leader can profitably pay the fee, so it prices as an unconstrained monopolist regardless of the quality gap to the next-best version (Akcigit and Kerr, 2018; Acemoglu et al., 2018). Facing the isoelastic demand, every leader charges the same price and markup,

$$p_j = \frac{\varepsilon}{\varepsilon-1}\, w \equiv p \quad \forall j, \qquad \text{markup } \frac{\varepsilon}{\varepsilon-1}, \qquad \text{profit share } \frac{1}{\varepsilon}. \tag{4}$$

The aggregate markup identifies ε directly, outside the size and age moments. Let $\bar{q} \equiv \int_0^1 q_j\, dj$ be average quality, $\hat{q}_j \equiv q_j/\bar{q}$ relative quality, and $M_k \equiv \int_0^1 \hat{q}_j^{\,k}\, dj$ (so $M_0 = M_1 = 1$). With common prices the numeraire condition gives $p = \bar{q}\, M_{\varepsilon-1}^{1/(\varepsilon-1)}$, and line revenue is a power function of relative quality,

$$s_j = \frac{Y}{M_{\varepsilon-1}} \hat{q}_j^{\varepsilon-1}, \qquad \text{so a firm } f \text{ holding lines } \mathcal{J}_f \text{ has sales} \quad S_f = \frac{Y}{M_{\varepsilon-1}} \sum_{j\in\mathcal{J}_f} \hat{q}_j^{\varepsilon-1}. \tag{5}$$

Line profits are s_j/ε and line production labor is s_j/p, so production employment is proportional to sales and the model does not distinguish sales from employment as a size measure.[14] A firm is large either because it holds many lines or because its lines are of high relative quality. Under Cobb–Douglas ($\varepsilon \to 1$) the quality term $\hat{q}^{\varepsilon-1}$ vanishes, sales become proportional to the product count, and quality drops out of firm size—one source, as Section 5 shows, of that model's too-small conditional-on-age dispersion. Writing L_P for aggregate production labor, summing line revenue and

[14]The constant markup is deliberate. Empirically, the within/between decomposition is nearly identical for sales and employment (Table 3); systematic markup dispersion would open a wedge between the two, so its near-absence is consistent with a common markup. Conceptually, a common markup leaves the allocation of production labor across lines undistorted, so the model contains no static misallocation—the size-screened subsidy of Section 6 therefore operates entirely through the dynamic allocation of R&D across types.

labor demand gives output, the wage, and the wage-to-output ratio

$$Y = \bar{q}\, M_{\varepsilon-1}^{1/(\varepsilon-1)} L_P, \qquad w = \frac{\varepsilon-1}{\varepsilon}\,\bar{q}\, M_{\varepsilon-1}^{1/(\varepsilon-1)}, \qquad \omega \equiv \frac{w}{Y} = \frac{\varepsilon-1}{\varepsilon L_P}. \tag{6}$$

Firm types and innovation. Firms differ in their innovation technologies following Lentz and Mortensen (2008). Firms are of a type $\theta \in \{H, L\}$ that governs the *step size* of the quality improvements they create, $\lambda_H > \lambda_L > 0$. An entrant draws H with probability p_h and L with probability $1 - p_h$. Further, a high type reverts to low at an exogenous rate $\nu > 0$, and the low type is absorbing. The type is value-side heterogeneity, and this is the point of contrast with Acemoglu et al. (2018), whose types instead scale the innovation *arrival* technology. The distinction matters for policy: with a common step, growth is the arrival rate times a constant, so policy can move growth only by changing how much innovation occurs. With step heterogeneity the mean step is endogenous to the type composition, so policy moves growth by changing *who* innovates as well (Proposition 1). That composition channel turns out to be important: in Section 6 the size screen *raises* the rate of innovation yet lowers growth, because it lowers the mean step.

A type-θ incumbent holding n lines chooses a per-line innovation intensity x; successes arrive at the firm rate nx and land on a line drawn uniformly at random in the economy (undirected R&D), displacing its incumbent. Producing per-line intensity x requires

$$c_x(x) = \frac{x^\zeta}{\psi_x}, \qquad \zeta > 1, \tag{7}$$

units of labor per line; this per-line specification is the Klette–Kortum scaling that keeps firm value additive across lines. There is no internal (own-product) R&D.[15] A successful innovation on line j by a type-θ firm replaces its quality with

$$q_j \;\longmapsto\; q_j + \lambda_\theta\, \bar{q}, \tag{8}$$

and transfers the line to the innovator: the step is additive and scaled by average quality, so the innovator builds on the line's existing quality and on the aggregate knowledge stock $\bar{q}$, as in Akcigit and Kerr (2018) and Acemoglu et al. (2018).

[15] The baseline excludes internal R&D deliberately: with the mean size–age relationship directly targeted, internal R&D is redundant, and when included, the estimation drives it to zero. In Section 6, subsidized firms could respond on the internal margin—which Garcia-Macia et al. (2019) find accounts for most U.S. growth—and internal innovation displaces no high-type incumbent; but the growth loss there is a composition effect rather than a displacement effect: a low type's innovation carries the small step, and bids R&D labor away from high types, whether it improves its own line or captures a rival's. The internal margin could attenuate the loss, not remove its source.

Entry, exit, and creative destruction. New firms arrive at aggregate flow z at linear labor cost z/ψ_z; a successful entrant captures a random line, learns its type, and starts with one line of quality $q_j + \lambda_\theta \bar{q}$. A firm that loses its last line to creative destruction exits. Writing μ_H for the measure of lines held by high types and $\mu_L = 1 - \mu_H$ for low types, every line is captured at the share-weighted aggregate creative-destruction rate

$$\tau = \mu_H x_H + \mu_L x_L + z. \tag{9}$$

The firm's problem. A type-θ firm holding lines $\mathcal{J}_f$ chooses a per-line R&D intensity x to maximize the expected present discounted value of its lines' profits net of R&D costs, taking as given the constant aggregates of a balanced growth path (Section 4.2)—the creative-destruction rate τ, the growth rate g, and the wage. Because the R&D cost (7) is incurred per line and successes land on lines drawn at random elsewhere in the economy, the problem separates across lines: firm value is additive, $V_\theta = Y \sum_{j \in \mathcal{J}_f} v_\theta(\hat{q}_j)$, and the normalized per-line value $v_\theta(\hat{q})$ solves the Hamilton–Jacobi–Bellman equation

$$\rho\, v_\theta(\hat{q}) = \frac{\hat{q}^{\varepsilon-1}}{\varepsilon M_{\varepsilon-1}} - g\hat{q} v_\theta'(\hat{q}) - \tau v_\theta(\hat{q}) + \max_{x \geq 0} \Big\{ x\, G_\theta - \omega\, c_x(x) \Big\} + \nu\, \mathbf{1}\{\theta = H\} \Big[v_L(\hat{q}) - v_H(\hat{q}) \Big]. \tag{10}$$

The left side is $r - g$ times the value ($r - g = \rho$ under log utility, from the Euler equation below); the terms on the right are, in order, the normalized profit flow $s_j/(\varepsilon Y)$ from (5), the erosion of the line's relative quality as the frontier $\bar{q}$ advances at rate g, the loss of the line to creative destruction at rate τ, the per-line expansion surplus, and—for high types—the switch to low type at rate ν. The expansion surplus rewards R&D through $G_\theta = \mathbb{E}_F[v_\theta(\hat{q}' + \lambda_\theta)]$, the value to a type-$\theta$ firm of capturing one additional line whose relative quality $\hat{q}'$ is drawn from the stationary distribution F characterized in Section 4.3. The optimal intensity x_θ and the closed-form solution for v_θ are derived there.

Free entry and market clearing. Entry is free, so the entry cost equals the expected value of a successful entrant, who captures one line and is high type with probability p_h:

$$\frac{\omega}{\psi_z} = p_h\, G_H + (1 - p_h)\, G_L. \tag{11}$$

Production, per-line R&D, and entry exhaust the unit labor endowment:

$$1 = L_P + \mu_H\, c_x(x_H) + \mu_L\, c_x(x_L) + \frac{z}{\psi_z} = L_P + \frac{\mu_H x_H^2 + \mu_L x_L^2}{\psi_x} + \frac{z}{\psi_z}, \tag{12}$$

where x_H, x_L are the firms' optimal per-line intensities, determined in Section 4.3. R&D and entry are paid in labor rather than in final output. The choice is deliberate here, because the crowding-out channel of the policy experiment operates through the R&D resource constraint.[16] The household's Euler equation on a balanced growth path with growth rate g gives the interest rate $r = \rho + g$.

4.2 Balanced growth path

The balanced growth path of this economy is defined as follows.

Definition 1 (Balanced growth path)**.** *A balanced growth path is a set of constant aggregates* $(\tau, h, \mu_H, g, L_P, \omega, \{M_k\})$*, constant policies* (x_H, x_L, z)*, value coefficients* (A, B_H, B_L)*, and a stationary relative-quality distribution* F *such that: (i)* x_H, x_L *and* z *solve the incumbents' and entrants' problems given* (τ, g, ω)*; (ii) free entry holds; (iii) the labor market clears; (iv) the type composition* μ_H *is stationary; (v)* $\{M_k\}$ *are the moments of* F*; and (vi)* $r = \rho + g$*. Along a balanced growth path, output* Y*, consumption* $C = Y$*, the wage* w*, and average quality* $\bar{q}$ *grow at the common rate* g*, while production labor* L_P *and the relative-quality moments* $\{M_k\}$ *are constant.*

4.3 Characterization

Constant aggregates and the quality distribution. Along the balanced growth path, the relative-quality moments $\{M_k\}$ and production labor L_P in (6) are constant, so output Y, consumption $C = Y$, the wage w, and average quality $\bar{q}$ all grow at the common rate g, as Definition 1 requires. At the line level, between innovations relative quality decays, $\dot{\hat{q}} = -g\hat{q}$ (own quality fixed while $\bar{q}$ grows), and at rate τ it jumps to $\hat{q} + \lambda$ with $\lambda = \lambda_H$ with probability h and λ_L with probability $1 - h$, where

$$h = \frac{\mu_H x_H + p_h z}{\tau}, \qquad m_k \equiv h\,\lambda_H^k + (1-h)\,\lambda_L^k, \qquad \bar{\lambda} \equiv m_1 \tag{13}$$

is the share of innovations made by high types. This mean-reverting shot-noise process has a unique stationary distribution F on $(0, \infty)$ with all moments finite; applying its generator to $\hat{q}^k$ yields the triangular recursion

$$M_k = \frac{\tau}{k\,g} \sum_{i=0}^{k-1} \binom{k}{i} M_i\, m_{k-i}, \tag{14}$$

[16] Klenow and Li (2025) provide direct evidence for this choice: average firm size in the United States does not fall with labor productivity, either over time or across states, which through the free-entry condition requires the cost of creating a firm to rise with growth—as it does when entry is labor intensive; they read the same evidence as corroborating the assumption, common in endogenous growth models, that the cost of innovation rises with the level of technology attained.

whose $k = 1$ case, $M_1 = \tau\bar{\lambda}/g = 1$, is the growth equation below. For integer $\varepsilon - 1$, the moment $M_{\varepsilon-1}$ and the capture moments used below are closed form, which is why I take ε integer.

Type composition. High-type captures convert lines to H-held; low-type captures and type switching convert them to L-held, so the stationary type composition satisfies

$$(\mu_H x_H + p_h z)(1 - \mu_H) = \Big(\mu_L x_L + (1 - p_h)z\Big)\mu_H + \nu\,\mu_H. \tag{15}$$

Because high types revert to low at rate ν, no cohort ever becomes fully high-type ($\mu_H < 1$ always): the economy never approaches the corner at which the faster-expanding type holds every line, the boundary that confines models with *permanent* types to small type gaps. Switching also mean-reverts the type composition of old cohorts, which prevents selection from steepening the mean size–age profile—the property that lets the estimated model respect the data's small between-age component.

Growth. Each innovation adds $\lambda_\theta \bar{q}$ to the quality stock, which pins the growth rate.

Proposition 1 (Growth)**.** *Along a balanced growth path,*

$$g = \tau\,\bar{\lambda},$$

where $\bar{\lambda} \equiv h\lambda_H + (1 - h)\lambda_L$ is the flow-weighted mean innovation step.

Growth depends both on *how much* innovation occurs, τ, and on *who* does it. Reallocating innovation across types moves both: it shifts the aggregate rate τ—because the two types innovate at different rates, reallocating toward the faster-innovating high types raises it, the reallocation-of-innovation channel of Acemoglu et al. (2018)—and, because the types here also differ in the *size* of their innovations, it shifts the mean step $\bar{\lambda}$ through the composition h, the margin whose growth role Akcigit and Kerr (2018) emphasizes.

Firm values and R&D policy. Solving the firm's problem (10) yields the per-line value in closed form and, with it, the optimal expansion rate.

Proposition 2 (Firm value and expansion)**.** *Along a balanced growth path, the per-line value solving* (10) *is*

$$v_\theta(\hat{q}) = A\,\hat{q}^{\,\varepsilon-1} + B_\theta, \qquad A = \frac{1}{\varepsilon\,M_{\varepsilon-1}\Big(\rho + \tau + (\varepsilon - 1)g\Big)}, \tag{16}$$

so the value of capturing a line drawn at random, $G_\theta = \mathbb{E}_F[v_\theta(\hat{q}' + \lambda_\theta)]$, is

$$G_\theta = A\,\Gamma_\theta + B_\theta, \qquad \Gamma_\theta \equiv \mathbb{E}_F\Big[(\hat{q}' + \lambda_\theta)^{\varepsilon-1}\Big] = \sum_{i=0}^{\varepsilon-1} \binom{\varepsilon-1}{i} M_i\,\lambda_\theta^{\varepsilon-1-i}. \tag{17}$$

The first-order condition of the expansion surplus in (10) gives the optimal per-line intensity and the surplus it earns,

$$x_\theta = \left(\frac{\psi_x\,G_\theta}{\zeta\,\omega}\right)^{\frac{1}{\zeta-1}} \overset{\zeta=2}{=} \frac{\psi_x\,G_\theta}{2\omega}, \qquad \Phi_\theta \equiv \max_{x\ge 0}\Big\{x\,G_\theta - \omega\,c_x(x)\Big\} \overset{\zeta=2}{=} \frac{\omega\,x_\theta^2}{\psi_x}, \tag{18}$$

and the type-specific constants close the value with $(\rho+\tau)B_L = \Phi_L$ and $(\rho+\tau+\nu)B_H = \Phi_H + \nu B_L$. Because $\Gamma_H > \Gamma_L$ and $B_H \ge B_L$, the capture value satisfies $G_H > G_L$, so $x_H > x_L$: high types expand faster.

The slope A is common to both types, and its effective discount rate $\rho + \tau + (\varepsilon-1)g$ is the interest rate net of growth, plus the loss of the line to creative destruction, plus the erosion of its relative quality. The constant B_θ capitalizes the option to expand from the line, discounted by $\rho + \tau$ and, for high types, the switching hazard ν. The capture value (17) shows that an innovator appropriates the captured line's inherited quality—through the moments M_i—together with its own step λ_θ, so $\Gamma_H > \Gamma_L$; high types thus expand faster both because their captures are worth more and because their option value is higher ($B_H \ge B_L$), and as $\nu \to \infty$ the gap collapses to the capture-premium component. The proof is in Appendix B.

Equilibrium system. Collecting these results, with $\zeta = 2$ the balanced growth path is the solution of eight equations in $(x_H, x_L, z, \mu_H, g, L_P, B_H, B_L)$: the type composition (15), growth (Proposition 1), the two R&D policies (18) and two value constants of Proposition 2, free entry (11), and labor-market clearing (12), where $\tau, h, \bar{\lambda}, \{M_k\}, A, \Gamma_\theta, G_\theta, \omega$ are the explicit functions of the unknowns given above. Appendix B collects the derivation and the solution algorithm.

4.4 Firm-size moments

The moments that enter the decomposition of Section 3 follow from the firm's state $(\theta; \hat{q}_1, \dots, \hat{q}_n)$ and the sales equation (5). An entrant holds a single line of relative quality $\hat{q}' + \lambda_\theta$, so there is no count dispersion at entry and

$$\operatorname{Var}\Big(\ln S \mid a = 0\Big) = (\varepsilon-1)^2 \operatorname{Var}\Big(\ln(\hat{q}' + \lambda_\theta)\Big), \qquad \hat{q}' \sim F,\ \theta \sim (p_h, 1-p_h). \tag{19}$$

Entry dispersion mixes the common line-quality draw $\hat{q}'$ with the type step; the between-type component enters with weight $p_h(1 - p_h)$, so p_h and the step ratio λ_H/λ_L are partially identified off entry-level dispersion. Mean log size by age combines count growth—the per-line birth rate switches from x_H to x_L at the type-switching time, so the count is a two-phase birth–death process—with the slow erosion of each held line's relative quality at rate $(\varepsilon - 1)g$. Dispersion conditional on age combines count dispersion, the vintage composition of a firm's lines, the type mixture, and the heterogeneity of switching times. Because sales are no longer proportional to counts, these moments are computed by simulating a balanced-growth cohort panel (Appendix C).

5 Estimation and the within/between split

I estimate three versions of the model to the 2017 Swedish cross-section. *Model B* is the two-type model of Section 4. *Model A* is its single-type restriction, $\lambda_H = \lambda_L = \lambda$ (so the switching rate ν and the entrant mix p_h play no role); it is the homogeneous benchmark, in which all firms are ex-ante identical and all size dispersion is ex post. I estimate Model A at two elasticities: $\varepsilon = 1$, the Cobb–Douglas model that the literature commonly calibrates, and $\varepsilon = 5$; and Model B at $\varepsilon = 5$. The ladder is deliberate: the first stands in for—and reproduces the failure of—the Cobb–Douglas benchmark, the second adds the CES quality margin, and the third adds ex-ante types. Reading the between-age share down the ladder shows what each ingredient contributes.

5.1 Targets and identification

Targets. Three parameters are set outside the estimation: $\rho = 0.05$, $\zeta = 2$, and $\varepsilon = 5$, pinned by an aggregate markup of 25% ($\varepsilon/(\varepsilon - 1) = 1.25$, a 20% profit share) and hence identified outside the size moments. A value of $\varepsilon = 5$ is further consistent with Boppart et al. (2023), who estimate elasticities of substitution within product markets for the Swedish economy and report median estimates of 4.4 and 5.5 for the goods and services sectors, respectively.[17] The remaining parameters are chosen to match, with equal block weights, the mean size–age profile and the firm-age distribution—at integer ages up to 24 with firms aged 25 and above pooled, exactly the targets and grouping of Section 3.2—and aggregate growth ($g = 2\%$). Targeting the size–age profile and the age distribution is the standard practice for disciplining this class of models, and it leaves size dispersion conditional on age untargeted; Model

[17] A high ε lowers the degree of type heterogeneity required in the model to explain size dispersion conditional on age. It is therefore a conservative choice for the policy counterfactuals in Section 6.

A follows exactly this practice to answer what such estimation approach implies for the latter. Model B *additionally* targets the standard deviation of log size by age, the moment that disciplines its type heterogeneity—without it the type spread λ_H/λ_L and the switching rate ν would be unidentified. The between/within decomposition itself is not separately targeted by any version. Appendix E details the loss function and the search.

Identification. Identification is transparent. The markup fixes ε; aggregate growth fixes the mean step $\bar{\lambda}$; the mean size–age profile fixes the expansion efficiency ψ_x; the age distribution fixes the entry efficiency ψ_z (the split of τ between incumbent expansion and entry). In Model B, the *level* of dispersion conditional on age at young ages—entry dispersion (19) above all—fixes the type spread λ_H/λ_L and the entrant mix p_h, and the *slope* of dispersion at old ages fixes the switching rate ν: without switching, surviving high types would keep pulling old-age dispersion up. In the Cobb–Douglas model ($\varepsilon = 1$), firm sales equal the product count and the markup is unbounded, so the pricing block degenerates: the wage-to-output ratio $\omega = (\varepsilon - 1)/(\varepsilon L_P) \to 0$, and with the wage collapsing relative to the value of a product line, expansion R&D becomes unboundedly attractive at any fixed efficiency, so the map from the efficiencies (ψ_x, ψ_z) to finite expansion and entry rates has no interior solution. This is a property of the pricing limit, not of the firm dynamics. The count process—firms expand at rate x, lose lines at rate τ, and entry arrives at rate z—is well defined for any finite rates, and the size and age moments depend on those rates alone, not on ω or the efficiencies. I therefore estimate this version directly in its primitive rates (x, z), with λ pinned residually by the growth equation $g = \tau\lambda$, while the CES versions report the estimated efficiencies and the rates they imply.[18]

The estimation leverages cross-sectional moments only: the three margins of the dispersion decomposition—the mean size–age relationship, the age distribution and size dispersion conditional on age—map into the expansion efficiency, the entry efficiency and type heterogeneity, respectively. A dynamic moment—growth and survival conditional on *both* size and age, in the spirit of Haltiwanger et al. (2013)—provides further information about the ex-ante heterogeneity: conditional on age, the homogeneous model's largest firms are large by luck alone and must revert to the mean, whereas Model B's are disproportionately high types in mid-expansion, and keep growing. I

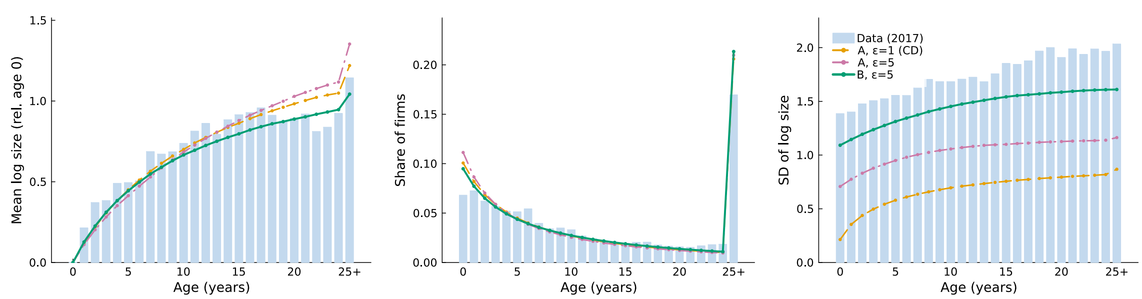
[18] The finite rates recovered this way are precisely the equilibrium rates of the Cobb–Douglas model closed with limit pricing—the standard pricing assumption of Klette and Kortum (2004) and Peters (2020), under which the markup is finite, $\omega > 0$, and the efficiencies do map into finite rates. I estimate the Peters (2020) model, a Cobb–Douglas quality-ladder model with endogenous markups in Supplemental Appendix H to the size–age relationship and the age distribution, and find that it fails on exactly the same dimensions as Model A with $\varepsilon = 1$.

compute exactly this moment in the panel and use it as the out-of-sample test of the estimated type heterogeneity, reported in Section 5.3 (Figure 7 and Table 5).

5.2 Fit and the split

Figure 6 shows the fit to the three age profiles. All three versions reproduce the mean size–age profile (left panel) and the age distribution (middle panel) reasonably well, so, by eq. (1), they agree on the two ingredients of the between-age component. They differ entirely in size dispersion conditional on age (right panel), and hence in the split. The Cobb–Douglas count model undershoots the profile badly: with sales proportional to the product count, its standard deviation of log size is 0.21 among firms aged zero against 1.39 in the data—one-seventh of it—and 0.87 against 2.03 in the oldest group. Model A at $\varepsilon = 5$ improves markedly, to 0.71 among entrants and 1.16 in the oldest group, roughly half the data at every age. Model B, which adds the ex-ante types and is the only version to target the profile, nearly closes the gap: 1.09 among entrants and 1.61 in the oldest group, about four-fifths of the data at every age.

Figure 6: Fit of the three models: size, age distribution, and dispersion



Notes: Each panel overlays the 2017 Swedish cross-section (bars) and the three estimated models (lines): Model A at $\varepsilon = 1$ (the Cobb–Douglas count model), Model A at $\varepsilon = 5$, and Model B at $\varepsilon = 5$. *Left:* mean log size by age, relative to age 0. *Middle:* the firm-age distribution. *Right:* the standard deviation of log size by age; the legend shown there applies to all three panels. All three models target the mean size–age profile and the age distribution; the standard deviation of log size by age is targeted only by Model B; the within/between decomposition is untargeted throughout. Ages 0–24 with 25 and above pooled.

Table 4 reports the estimates, the implied balanced growth paths, and the fit. The Cobb–Douglas model attributes 28.8% of the variance of log size to the between-age component, against 3.6% in the data—an overstatement by a factor of eight.[19] Introducing the CES quality margin (Model A, $\varepsilon = 5$) roughly halves the overstatement,

[19]This is a failure other firm-dynamics models that also feature a Cobb-Douglas aggregator share: they load nearly an order of magnitude too much of size dispersion onto age. Estimating a Peters (2020) model, for instance, to the same cross-sectional moments likewise attributes 28.5% of dispersion to the between-age component (Supplemental Appendix H).

to 18.5%: quality dispersion across lines raises dispersion conditional on age at every age. Adding ex-ante types (Model B) brings the between-age share to 5.9%, close to the data's 3.6% and far from the Cobb–Douglas 28.8%. Model B raises dispersion conditional on age to near the data at every age, so the overall standard deviation of log size, 1.45, approaches the data's 1.75, against 0.78 in the Cobb–Douglas model. The estimated model is thus consistent with size dispersion being overwhelmingly driven by size dispersion conditional on age.

Table 4: Three quality-ladder models: estimates and fit

	A, ε=1	A, ε=5	B, ε=5	Data
Estimated parameters				
Expansion efficiency ψ_x	n.i.	0.600	0.455	
Entry efficiency ψ_z	n.i.	1.059	0.837	
Step size λ (λ_L in B)	0.081	0.061	0.025	
Step size, high type λ_H	—	—	0.175	
Switching rate ν	—	—	0.176	
High-type entrant share p_h	—	—	0.351	
Implied balanced growth path				
Expansion rate x (x_H/x_L in B)	0.207	0.283	0.409 / 0.197	
Entry rate z	0.039	0.035	0.014	
Creative destruction τ	0.246	0.318	0.256	
TFP growth g	0.020	0.020	0.020	0.020
Size dispersion (SD of $\ln S$*)*				
Overall	0.78	1.10	1.45	1.75
Entrants (a=0)	0.21	0.71	1.09	1.39
Oldest (a=25+)	0.87	1.16	1.61	2.03
Decomposition of Var($\ln S$)				
Between-age share	**28.8%**	**18.5%**	**5.9%**	**3.6%**
Within-age share	71.2%	81.5%	94.1%	96.4%

Notes: The first three columns estimate Model A at $\varepsilon = 1$ (the Cobb–Douglas count model), Model A at $\varepsilon = 5$, and Model B at $\varepsilon = 5$ to the 2017 Swedish cross-section (final column), matching the mean size–age profile, the firm-age distribution, and aggregate growth ($g = 2\%$) at integer ages 0–24 with 25+ pooled; Model B additionally targets the standard deviation of log size by age. "n.i.": at $\varepsilon = 1$ the aggregate markup is unbounded and the efficiencies (ψ_x, ψ_z) are not separately identified from the flow rates, so the Cobb–Douglas model is estimated directly in (x, z) (reported below), with λ pinned residually by aggregate growth ($\lambda = g/\tau$). The high-type step size, the switching rate, and the entrant share apply only to Model B. The size-dispersion rows are untargeted diagnostics in the two Model A columns; in Model B the standard deviation by age is targeted, but the between/within share is not separately targeted.

The estimated types describe a small population of high-potential entrants. About 35% of entrants are high types ($p_h = 0.35$); they take innovation steps roughly seven times as large as low types ($\lambda_H/\lambda_L = 6.9$), expand about twice as fast ($x_H/x_L = 2.1$), and revert to low type at rate $\nu = 0.18$—an expected high-type phase of about five and a half years. During that phase, the high type's expansion outruns creative destruction, generating a Pareto right tail of firm size with an exponent near 1.15, an untargeted by-product close to the Zipf regularity of the largest firms.[20] The picture is one of rare, fast-growing entrants—"gazelles"—that grow quickly for a few years and then settle down, which is what supplies the dispersion conditional on age that the homogeneous model lacks.

5.3 Untargeted moments

The estimation ties Model B's type heterogeneity to a single targeted moment, size dispersion conditional on age, and a natural objection is that ex-ante types are not needed to match it: idiosyncratic shocks added to the homogeneous model—including a lumpy process of product creation unrelated to any persistent firm characteristic—could do so equally well, and the policy analysis of Section 6 would then have no bite. But such shocks are *ex-post* heterogeneity, whereas Sterk et al. (2021) find that the greater part of size dispersion among same-aged firms reflects *ex-ante* heterogeneity.[21] Matching the profile with noise would thus fit the moment by moving model A in exactly the wrong direction.[22] What that argument settles is the *kind* of heterogeneity that size dispersion conditional on age calls for. It leaves open, however, the question of *how much*—whether the type margin the estimation assigns is larger than the data themselves warrant. The remainder of this section addresses that question with moments in the Swedish data that no version of the model is estimated to: the growth

[20] The mechanism is that of Luttmer (2011), Proposition 4: a transient phase of fast growth that ends at a constant hazard generates a Pareto tail whose index is the ratio of that hazard to the net growth rate of firms in the phase, here $\nu/(x_H - \tau) = 1.15$. In his economy, the index is the sum of the hazard and the population growth rate, divided by that same net growth rate. He notes that with the fixed set of product lines of Klette and Kortum (2004) incumbents cannot grow on average, which makes it impossible for large firms to arise, and that relaxing Gibrat's law in this way restores a thick tail even without growth in the number of products. The present model is that case: with no population growth and a fixed measure of lines, the tail is generated by type reversion alone.

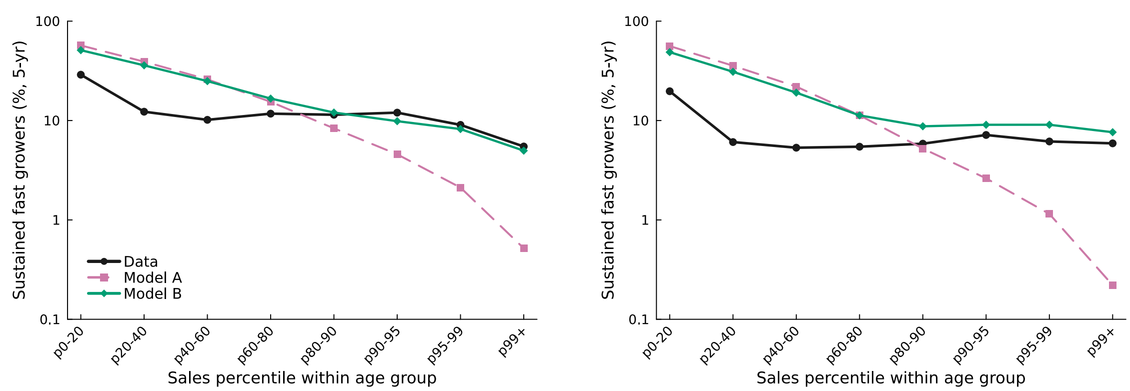
[21] Sterk et al. (2021) classify ex-ante/ex-post heterogeneity by the moment of firm inception; in the model here, it is natural to identify ex-ante heterogeneity with dispersion between types and ex-post heterogeneity with dispersion within them (Appendix D compares them). Where that boundary belongs is not this paper's question, and the conclusion in the text holds under either reading: even under theirs, which would count an entrant's first quality draw as ex-ante, fewer than two percent of the firms alive at age twenty in Model A still hold their initial product line—so the ex-ante component decays to essentially nothing precisely where Sterk et al. (2021) find it converging to roughly 40%.

[22] Berlingieri et al. (2024) generate the heavy tail of the firm-size distribution from such a lumpy process. It would not substitute for ex-ante heterogeneity here, for the reason just given, and alongside it they assign firms permanent types differing in the quality improvements their innovations achieve—the step form of Lentz and Mortensen (2008) adopted in this paper—to account for the persistent, firm-specific component in the initial revenue of new products.

of firms conditional on both their size and their age. I further show how Model B splits size dispersion conditional on age into its between-type and within-type components in Appendix D.

The growth of the largest firms of each age. The panel dimension of the data delivers dynamic firm moments, none of which are targeted by any estimation. Firms in a base-year cross-section are grouped by age and, within each age bin, by sales percentile, and followed for one year (2016 to 2017) and for five (2012 to 2017); the age- and size-cells are similarly computed in simulated balanced-growth panels of Models A and B at their estimates (both for $\varepsilon = 5$ to isolate the effect of type heterogeneity). The two models agree on firm-dynamics among small firms, where shared up-or-out mechanics dominate—small-given-age firms exit more, and conditional on survival grow fast—and they disagree at the top, for the reason anticipated in the identification discussion: at ages 1–19, more than half of Model B's largest-five-percent firms within each age bin are high types (56–69%), and about 80% of its top percentile. Model B's validation therefore rests on the growth of the largest firms at each age. Figure 7 shows the statistic that separates the two models most sharply—the share of firms sustaining fast growth over a five-year horizon (log growth of at least $5 \ln 1.2$, i.e. twenty percent per year, among firms present at both dates)—across the size distribution for firms aged 1–2 (left panel) and aged 3–6 (right panel); Table 5 reports the mean and median one-year survivor growth of the largest five percent of firms of each age. Growth is measured relative to the economy-wide average in both the model and the data.

Figure 7: Sustained fast growers by size, within age groups (untargeted)



Notes: Among firms aged 1–2 (left panel) and 3–6 (right panel) in the 2012 registry cross-section, grouped by within-age sales percentile, the share still present in 2017 that grew at least 20% per year over the five years (survivors only). Growth is measured relative to the economy-wide average. The data (2012–2017) are overlaid on simulated balanced-growth panels of Model A (no ex-ante types) and Model B (ex-ante types) at their $\varepsilon = 5$ estimates. The vertical axes are logarithmic. Over the lower percentiles both models overstate the share, a shared feature of product-count models; over the top percentile groups, the homogeneous model falls toward zero while the data track the type model.

The data side with model B. In both panels of Figure 7, the share of sustained fast growers falls with firm size. Over the lower percentiles, the two models overlap and both overstate the data—a survivor-selection artifact common to product-count models. The models separate only at the top, exactly where type heterogeneity operates. Sustaining twenty-percent annual growth for five years is all but impossible for a large firm in the homogeneous Model A—a firm with many products cannot do it on luck—and its fast-grower share among the largest firms of a given age collapses toward zero: 1.0% at ages 3–6, and just 0.2% in the top percentile. The data do not fall this way; at the higher percentiles, the data and the type model instead converge, staying elevated together in the high single digits near ten percent.

Table 5: Mean and median growth of the largest firms (p95+) of each age (untargeted)

Age bin	Model A	Model B	Data
Mean one-year growth of survivors (percent)			
3–6	−19.0	−10.1	−6.3
7–10	−14.1	−2.8	−4.6
11–19	−11.1	−1.8	−5.4
20+	−8.0	−6.7	−4.1
Median one-year growth of survivors (percent)			
3–6	−12.5	−6.8	−0.9
7–10	−10.7	−2.4	−1.5
11–19	−9.3	−2.4	−1.7
20+	−7.0	−7.3	−1.9

Notes: Firms in the 2016 registry cross-section (outcomes in 2017) are grouped by age and, within each age bin, by sales percentile; the table reports firms above the 95th percentile of size within their age bin (p95+). The statistics are the mean and the median change in log sales, measured relative to the economy-wide average, among firms present at both dates. Model cells apply the identical grouping to simulated balanced-growth panels of Models A and B ($\varepsilon = 5$) at their estimates. Age bins end at 11–19 and 20+ for consistency with the five-year panel of Figure 7, whose 2012 base year pools ages 20 and above. Ages 0–2 are omitted: the youngest firms are large or small mainly because of a lucky entry draw, which then reverts as they capture typical-quality lines; one-year growth in both models reflects this common reversion rather than type, so those cells do not discriminate—unlike the five-year fast-grower share, which no entry draw can mimic. Data cells contain between roughly 1,900 and 3,400 firms.

One-year survivor growth points the same way (Table 5): among the top five percent in each age bin, mean growth in the data is mildly negative, between −4 and −6.5 percent, and the medians are essentially flat (−0.9 to −1.9); i.e., firms that are large given their age shrink slightly relative to the economy-wide average. Model A predicts mean growth of −8 to −19 percent for these firms—far too much mean reversion. Model B's range of −1.8 to −10 covers the data well and its medians are within a percentage point of the data at ages 7–19. The relatively fast growth of high-type firms slows down the mean reversion of firms that are large given their age, bringing the model much closer to the data.

This pattern connects the exercise to Haltiwanger et al. (2013), whose central finding is that the inverse relationship between firm size and growth disappears once firm age is controlled for. The Swedish data behave the same way: regressing one-year survivor growth on log size within each age bin gives a slope of -0.02 at ages 3–6, and essentially zero thereafter (-0.01 at 7–10, -0.02 at 11–19, and $+0.00$ at 20+). Read through a model without ex-ante heterogeneity, that flatness is not a null result but a failure. Conditional on age, size is accumulated luck, so the largest firms of an age are those that drew favorable histories and must revert to the mean: Model A produces slopes of -0.18, -0.15 and -0.12 over ages 3–6, 7–10 and 11–19, roughly an order of magnitude steeper than the data. Ex-ante types flatten the profile, because they make the largest firms of an age large for a reason that persists; Model B cuts the slope by 37 to 40% over the same bins. The absence of a size–growth relationship conditional on age is thus, read through the model, itself a signature of ex-ante heterogeneity—and the data, flatter still than Model B, ask for more of it rather than less.

6 Small-firm subsidies and the size screen

Support for small firms is a pillar of growth policy in both the United States and the European Union. In the United States, the Small Business Act establishes the Small Business Administration and a broad agenda of small-business support: relief from regulatory and administrative burdens, government-guaranteed loans, and dedicated funding for small-firm innovation. The European Union's Small Business Act for Europe is the analogous framework, built explicitly on a "think small first" principle and spanning the same margins. These programs are broad, but the exercise below isolates one instrument from the bundle and asks what it does to aggregate growth: an R&D subsidy for the smallest firms, financed by a fixed share of GDP.[23]

The decomposition licenses a sharp statement about such a policy. Because a firm's age is nearly uninformative about where in the size distribution it falls, size is a poor proxy for age—a small firm is not necessarily a young firm that has yet to grow. It is, however, informative about *type*: in the model, firm type affects its size—high types take larger innovation steps and expand faster, and so grow large—which makes small firms disproportionately low types. A size screen that makes the smallest firms eligible therefore selects low types, precisely the firms whose innovation adds least

[23] In the United States, the Small Business Innovation Research and Small Business Technology Transfer programs reserve fixed shares of the largest federal agencies' external R&D budgets for small firms. In the European Union, the SME Instrument and its successor, the European Innovation Council Accelerator, fund small-firm innovation. Eligibility throughout is defined by firm size: the Small Business Administration's industry size standards (in sales or employment) in the United States, and the EU SME definition (Commission Recommendation 2003/361/EC) in Europe.

to growth. This section asks, within the estimated model, what such a size-screened R&D subsidy does to growth, and how it compares with screening on age instead. The homogeneous Model A is the reference; the policy is evaluated in Model B, the version consistent with the within/between decomposition, and then re-run under the cost form of the type heterogeneity at the close of the section.

Scope. Statutes motivate small-firm R&D support in three ways: the knowledge-spillover externality of innovation, the premise that small firms are a particularly innovative pool, and frictions in small firms' access to finance. The model prices the first exactly: by (8), every innovation raises $\bar{q}$, the base on which all future innovators build, while the innovator appropriates only the profits on the captured line until it is destroyed, so the spillover is proportional to the step λ_θ—a wedge attached to a firm's type, not to its size. In Model A, the wedge is the same for every firm, so the screen has no composition to select on: who receives the subsidy does not matter for the average step. In Model B, it is type-dependent, so which firms the subsidy reaches determines whether it pushes innovation toward large steps or small ones. The second premise the model delimits rather than disputes, and the qualification is age. Small *young* firms are still innovative: high types are common even among the small and young. Small *old* firms are not: their high types have already grown large, so among old firms, small size now reflects a low type. This is the innovation counterpart of Haltiwanger et al. (2013), who show that the job-creation advantage of small U.S. firms is attributable to firm age rather than size and who conclude that "policies targeting firms based on size without taking account of the role of firm age are unlikely to have the desired impact on overall job creation."[24] The third rationale—financial frictions—is absent from the model, and the exercise takes no stand on supporting constrained firms: whatever gains a subsidy delivers by relaxing such constraints lie outside the model, to be weighed alongside the effects computed below.[25]

[24] Their analysis stops at measurement, and deliberately so: effective policy design, they note, additionally requires a model of the equilibrium responses involved. The exercise below supplies that step for R&D policy, and sharpens their conclusion in one respect: once the reallocation of R&D across firms is priced in general equilibrium, the size screen does not merely fail to deliver the desired impact—it delivers the opposite one.

[25] The omission is contained in three respects. First, a friction whose incidence varies with age but not with type shifts the mean size–age profile and is absorbed into the estimated expansion efficiency ψ_x, which is agnostic about whether young firms expand slowly because R&D is costly or because finance is tight. Second, a friction that binds heterogeneously across types does enter size dispersion conditional on age, the moment the type heterogeneity is estimated from, so the estimate would attribute to types some dispersion that frictions generate. Overturning the policy conclusion of this section takes more than shaving the estimated spread, however: that conclusion runs on the eligible pool holding more low types than the ineligible one (Table 7), and the type–size correlation behind the gap is the direct footprint of the innovation advantage that defines a high type, so frictions would have to offset that advantage entirely to eliminate it—and even then the

Instrument and design. The policy is an ad valorem subsidy at rate ς to the expansion-R&D costs of eligible firms, financed by a lump-sum tax on the household. Eligibility follows the EU definitions directly: a firm is eligible if its sales fall below a statutory turnover ceiling, $S_f \leq \bar{S}$—€2 million (micro-enterprise), €10 million (small enterprise), or €50 million (medium; Commission Recommendation 2003/361/EC). Each ceiling is carried into the model by matching the share of aggregate sales that eligible firms hold: in the 2017 Swedish data, firms below the three ceilings (converted at 9.6 SEK/EUR) account for 19.6%, 42.2%, and 71.0% of total sales, and $\bar{S}$ is set so that eligible model firms hold the same shares.[26] Following the fixed-outlay, endogenous-rate convention of the growth-policy literature (Acemoglu et al., 2018), I hold the total fiscal outlay fixed at 1% of aggregate output and let the statutory rate ς adjust to exhaust the budget, so all designs are compared at equal generosity.[27] Because a size cutoff makes a firm's R&D depend on its position relative to the threshold, firm value is no longer additive across lines; the subsidized equilibrium is solved with the firm-level value functions and simulation algorithm of Appendix F.

The growth channel. Differentiating $g = \tau\bar{\lambda}$ (Proposition 1),

$$dg = \underbrace{\bar{\lambda}\, d\tau}_{\text{more innovation}} + \underbrace{\tau\, d\bar{\lambda}}_{\text{who innovates}} . \tag{20}$$

In Model A, only the first term ("effort") is present. In Model B, the second "composition" term can turn negative: subsidizing low-type expansion, and crowding out unsubsidized high-type expansion through the R&D labor market, lowers the mean step $\bar{\lambda}$. Whether growth rises or falls is quantitative, and the estimated model answers it.

ranking of the two screens would survive, because the age screen's advantage rests on the type–age correlation, which frictions would have to undo separately. Third, the untargeted moments push the other way. A constraint only holds a firm below its unconstrained size, so it cannot produce the sustained fast growers among the *largest* firms of each age: fitting Figure 7 and Table 5 requires fast-growing types at the top whether or not frictions are present.

[26] Sales shares rather than firm counts are matched because the eligible sales share is what governs the fiscal footprint of the screen. The mapping is nonetheless validated on the count margin: the model's implied eligible firm shares are 80%, 96%, and 99.8% across the three ceilings, against 88.1%, 97.1%, and 99.5% in the data—close agreement, untargeted, reflecting the model's Pareto size tail.

[27] Fixing the outlay is a device for holding generosity fixed across designs, set at the level used by Acemoglu et al. (2018) so that the growth effects are comparable to theirs. The realism check belongs on the rate rather than the budget: statutes set eligibility ceilings and aid intensities and leave the outlay to take-up, so the rate an eligible firm faces is the quantity with a legislated counterpart—and the budget-exhausting rates of 8–18% reported below sit inside those intensities.

Results. Table 6 reports the effect of the subsidy—eligibility below an EU ceiling, an outlay of 1% of output—on growth under the two screens in the two models, with the exact decomposition (20); the B^{cost} rows, which repeat the exercise under the alternative form of the type heterogeneity, are discussed below. The result is a sign flip at every ceiling. In Model A, the size-screened subsidy raises growth by about 0.05 percentage points regardless of the ceiling[28]: where firms are homogeneous, no screen makes the subsidy harmful, and the differences among them are second-order. In Model B, the identical policies *lower* growth—by 0.24 percentage points at the micro and small ceilings, about twelve percent of the 2% growth rate, and by 0.19 percentage points at the medium ceiling—entirely through the composition term (the effort term is if anything larger than in A). A policymaker who used the homogeneous model to evaluate the subsidy would predict a small gain and deliver a sizable loss. Two features of the statutory design are worth noting. First, the budget-exhausting subsidy rates are 8–18%, squarely in the range of actual R&D support programs; "1% of output" is a plausible program, not an extreme one. Second, the growth-loss is largest at the *tightest* ceiling (0.24 at the micro against 0.19 percentage points at the medium ceiling).[29] European state-aid rules run the gradient the other way, granting the highest R&D aid intensities to the smallest size classes.[30]

The sign flip is not in tension with the empirical evidence that size-targeted R&D support works at the firm level. Evaluations of the Small Business Innovation Research program (Howell, 2017) and of the U.K.'s SME R&D tax credit (Dechezleprêtre et al., 2023) find sizable positive effects on the R&D and growth of the subsidized firms, and the model reproduces exactly that: eligible firms expand their R&D, and the effort term of (20) is positive in every row of Table 6. The loss sits in the composition term—the crowding out of ineligible high types through the R&D labor market in general equilibrium—which event-study designs at the firm-level miss. A firm-level evaluation run inside the model economy would return a positive treatment effect even as aggregate growth falls. Consistent with the aggregate relevance of this mar-

[28]The entries are rounded; the size-screen effect in fact rises slightly with the ceiling, from +0.047 at the micro to +0.051 at the medium. Because R&D costs are convex, spreading the fixed outlay over a broader pool at a lower rate is weakly more cost-effective than concentrating it at a high rate on a narrow one.

[29]The loss survives even where eligibility is nearly universal, because coverage measured in firm counts is not the margin that matters. At the medium ceiling, the 0.2% of firms above the threshold perform a third of aggregate expansion R&D—the Pareto tail of the size distribution—and 40% of them are high types, against 13% in the cross-section. The matched age pool excludes a *larger* share of R&D, 43%, performed by the 9.5% of firms older than 41 years, and still raises growth, because at that age virtually none of them are high types. What matters is not how much R&D a screen leaves out, but whose.

[30]The EU state-aid rules add a rate gradient on top of eligibility: under the General Block Exemption Regulation (Article 25), the ceilings on R&D aid intensity rise by 10 percentage points for medium-sized enterprises and by 20 for small ones.

Table 6: Effect of statutory small-firm R&D subsidies on growth (outlay: 1% of output)

Ceiling	Screen and model	Rate ς^*	Δg (pp)	decomposition (pp): effort $\bar{\lambda}\, d\tau$	decomposition (pp): who $\tau\, d\bar{\lambda}$	CE welfare
Micro (€2m)	Size, Model A	17%	+0.05	+0.05	0	−0.12%
	Size, Model B	14%	**−0.24**	+0.10	**−0.34**	−6.53%
	Size, Model B^{cost}	12%	+0.06	+0.06	0	−0.14%
	Age, Model A	18%	+0.06	+0.06	0	−0.20%
	Age, Model B	14%	+0.07	+0.07	+0.00	−0.03%
	Age, Model B^{cost}	11%	+0.08	+0.08	0	+0.09%
Small (€10m)	Size, Model A	11%	+0.05	+0.05	0	−0.05%
	Size, Model B	9%	**−0.24**	+0.10	**−0.34**	−6.45%
	Size, Model B^{cost}	8%	+0.05	+0.05	0	−0.16%
	Age, Model A	11%	+0.06	+0.06	0	−0.07%
	Age, Model B	10%	+0.12	+0.06	+0.06	+1.12%
	Age, Model B^{cost}	8%	+0.08	+0.08	0	+0.18%
Medium (€50m)	Size, Model A	8%	+0.05	+0.05	0	−0.01%
	Size, Model B	8%	**−0.19**	+0.09	**−0.28**	−5.38%
	Size, Model B^{cost}	8%	+0.04	+0.04	0	−0.15%
	Age, Model A	8%	+0.06	+0.06	0	−0.02%
	Age, Model B	8%	+0.16	+0.05	+0.11	+2.21%
	Age, Model B^{cost}	7%	+0.07	+0.07	0	+0.22%

Notes: Each row applies the same statutory design—eligibility below the EU turnover ceiling shown, mapped into the model by the eligible pool's share of aggregate sales in the 2017 Swedish data (19.6%, 42.2%, and 71.0%), or below the age cutoff holding the same sales share (9, 22, and 41 years)—with the fiscal outlay fixed at 1% of output and the subsidy rate ς^* adjusting to exhaust the budget. Δg is the change in the annual growth rate in percentage points, decomposed as in eq. (20) into the effort term ($\bar{\lambda}\, d\tau$, "more innovation") and the composition term ($\tau\, d\bar{\lambda}$, "who innovates"), with $\bar{\lambda}$ and τ each evaluated at the midpoint of the baseline and subsidized balanced growth paths so that the two terms sum to Δg exactly; in Model A, there are no types, so the composition term is zero. Because τ aggregates type-specific innovation intensities over the line stock, $d\tau$ contains a composition term alongside the intensity responses, so "effort" is a close label rather than an exact one. CE is the consumption-equivalent welfare change comparing balanced growth paths; Table 8 recomputes it along the transition path for the four micro-ceiling designs. The Model B^{cost} rows repeat the two designs in the cost-heterogeneity variant of Section 6, in which the types differ in the cost of innovating at a common step, so the composition term is identically zero; its ceilings and matched age cutoffs (14, 29, and 48 years) are mapped by the same eligible sales shares on that model's own cross-section, and Appendix I details the variant. Baseline growth is 2%.

gin, Lehr (2025) documents that R&D resources have become more misallocated over time, accounting for roughly a quarter of the U.S. growth slowdown, without taking a stand on the driver. An R&D subsidy targeted at small firms is a concrete, policy-made source of exactly this misallocation: it lowers the cost of R&D for the firms that innovate in the smallest steps and, through the R&D labor market, crowds out the firms that innovate in the largest.

Screening on age instead. The failure of the size screen has a simple root, and the root points to the alternative. Size is endogenous—the cumulated outcome of the very innovation process the subsidy targets—so smallness itself signals not having grown, and an eligibility rule written on size inherits that adverse selection. The problem is not that size carries no information about type: high types grow faster, which is exactly why they are missing from the eligible pool. It is that a rule written on the outcome of the subsidized process reads that information backwards. Age, by contrast, is exogenous to the firm—no firm can choose its age—so an age cutoff works like an immutable tag in the sense of Akerlof (1978): it cannot be manipulated, and it selects only through the composition of cohorts.[31]

Screening on age is not a hypothetical alternative to screening on size: France's *Jeune Entreprise Innovante* status grants R&D-related tax and social-contribution relief on precisely this basis, to firms younger than eight years. The question is how much such a design would improve on the statutory size test, and I answer it by re-running each experiment with the turnover ceiling replaced by an age cutoff. The cutoff is set so that the eligible pool holds the same share of aggregate sales—henceforth, the same *footprint*—as the corresponding size pool, so that the two screens are compared at matched economic weight and at the same 1%-of-output outlay. The matched cutoffs are 9, 22, and 41 years of age, the tightest of them close to the French statute. Because eligibility now turns on a firm characteristic that is independent of the firm's portfolio, the age design preserves the additivity of firm value across lines and is solved without the firm-level value functions the size screen requires.

Screening on age overturns the result within Model B. At matched economic weight, the same-budget age subsidy is growth-*positive* at every footprint: +0.07 percentage points at the micro footprint, +0.12 at the small, and +0.16 at the medium, against −0.24, −0.24, and −0.19 for the size screens. Replacing the size test with an age

[31] Endogeneity carries a second cost that tagging avoids: a size ceiling gives firms near the threshold a reason to stay below it, a margin no age cutoff can create. That margin is not assumed away—the value functions of Appendix F price eligibility into expansion incentives, so firms curbing their expansion to preserve eligibility is part of the solved equilibrium—but it is not what carries the results. In Table 6, it is selection, the composition term, and not the effort term, that produces the sign flip.

test at equal outlay thus swings the growth effect by 0.31 to 0.36 percentage points, roughly a sixth of the baseline growth rate.

The gain rises with the breadth of the age pool, which runs against what the model's type dynamics might be expected to deliver: types are drawn at entry and revert in one direction only. Two things account for it. The first is that what matters for growth is not the share of high types among the firms a screen reaches but the share of the economy's product lines those high types hold, and high types accumulate their lines after entry, over the expansion phase that follows. The second is subtler and concerns entry. A tight young-firm subsidy operates in part as a subsidy to entry—entry rises by 128% at the micro footprint—and the additional creative destruction that entry brings falls on high-type incumbents too. The broadest of the three cutoffs, by contrast, subsidizes every firm except the stagnant old low types: entry is essentially undistorted, and the composition term itself turns powerful, +0.11 (Table 6). Both forces cut against the young pool; holding the budget at the same 1% of output and tightening the age cutoff, the growth effect in Model B falls monotonically and turns *negative* at a cutoff of about eighteen months, while at nearly identical subsidy rates it stays positive in Model A throughout. An age test does not beat a size test by construction here; the ranking is a statement about screens at statutory weight, not a mechanical consequence of drawing types at entry.[32]

Table 7 shows what the size and age screen select in Model B at the micro footprint. Firms below the micro ceiling are only 11% high types—below the cross-sectional high-type share of 13%—so the screen selects innovators of below-average step ($\mathbb{E}[\lambda \mid \text{eligible}]$ is 0.73 times $\mathbb{E}[\lambda \mid \text{ineligible}]$), and 42% of them are at least ten years old: small firms are not necessarily young, exactly as the decomposition in Section 3 warns. The youngest firms at the same economic weight, by contrast, are 22% high types, with an average step 1.9 times the ineligible mean, and by construction none are old. Selection is sharpest when age and size are combined in the model. Among young firms (under three years), it is the *largest* that carry the high types: the young-and-large—the top size quintile of firms under three years, 5% of all firms—are 49% high types, with an average step 2.3 times the ineligible mean, against 22% under the age screen and 11% under the size screen alone. Neither margin isolates the high types

[32] Mechanisms that revealed types later in life would not overturn the comparison between the two screens; they would push the age design further in the direction just described, toward pools reaching further up the age distribution. What limits how far the data allow that is the fit. Model B already generates *more* sustained fast growth among the largest firms aged 7–19 than the data show (Section 5.3), and late-blooming high types would widen that gap. Nor can types be permanent: the Pareto right tail comes from high types outgrowing creative destruction ($x_H > \tau$), which no firm can do indefinitely in a stationary firm-size distribution, and permanent types would concentrate the surviving high types among the old, pulling dispersion at old ages above what the data show—the moment that identifies the switching rate—and steepening the mean size–age profile, increasing the between-age share.

as sharply as the two together.

Table 7: What the eligibility screen selects (Model B)

Eligible pool	Pr(high type)	$\frac{\mathbb{E}[\lambda \mid \text{elig.}]}{\mathbb{E}[\lambda \mid \text{inelig.}]}$	aged ≥ 10
Below the micro ceiling (19.6% of sales)	0.11	0.73	42%
Youngest at matched weight (≤ 9 yr)	0.22	1.92	0%
Young (< 3 yr) $\times$ largest 20%	0.49	2.34	0%

Notes: Composition of the eligible pool under three screens in the estimated Model B. Pr(high type) is the high-type share of the pool ($p_h = 0.35$ among entrants, 0.13 in the firm cross-section); the middle column is the ratio of the mean innovation step of eligible to ineligible firms; the last column is the share of the pool older than ten years. The first two rows are the micro-footprint pools of Table 6; the last is the top size quintile among firms under three years old.

Welfare. Household welfare on a balanced growth path is the present discounted value of log consumption. With consumption growing at rate g from an initial level C_0, so that $C(t) = C_0 e^{gt}$, and discount rate ρ,

$$\mathcal{W} = \int_0^\infty e^{-\rho t} \ln C(t)\, dt = \frac{\ln C_0}{\rho} + \frac{g}{\rho^2}. \tag{21}$$

Welfare depends on the consumption level at the policy date, C_0, and on the growth rate g. I summarize a reform by its consumption-equivalent (CE) welfare change ξ, the permanent percentage change in baseline consumption that leaves the household indifferent to it, defined by

$$\ln(1+\xi) = \rho\,(\mathcal{W}_1 - \mathcal{W}_0) = \underbrace{\ln \tfrac{C_0'}{C_0}}_{\text{level}} + \underbrace{\frac{\Delta g}{\rho}}_{\text{growth}}\,, \tag{22}$$

where $\mathcal{W}_0, \mathcal{W}_1$ and C_0, C_0' are welfare and initial consumption on the baseline and reformed paths; the CE thus combines the jump in the consumption level with the discounted change in growth. Initial consumption is initial output: by eq. (6), $C_0 = \bar{q}_0\, M_{\varepsilon-1}^{1/(\varepsilon-1)} L_P$ at the policy date. The comparison holds the initial average quality $\bar{q}_0$ fixed across the two paths so the level term reflects production labor L_P and the stationary quality moment $M_{\varepsilon-1}$.

By this measure, the size-screened subsidy costs 5.4–6.5% of consumption in Model B (Table 6), against 0.0–0.2% in Model A, where the capitalized growth gain, $\Delta g/\rho \approx 1\%$, almost exactly repays the level loss from drawing labor out of production into R&D and entry. The age screen in Model B is essentially neutral at the micro footprint and delivers consumption-equivalent *gains* of 1.1% and 2.2% at the small and medium footprints, where the growth term outweighs the level loss. The contrast across models and screens is carried by the growth term.

The transition. These figures compare balanced growth paths: they credit the reformed economy with its new growth rate from the reform date onwards. That is not innocuous here, because the two terms of (20) move at very different speeds. The effort term is a flow—R&D intensities jump the moment the subsidy arrives. The composition term works through a *stock*: the share of the economy's product lines held by high types, μ_H, which is a state variable, governed off the balanced growth path by the law of motion whose stationary version is (15). I therefore solve the transition for the micro-ceiling subsidies. The economy starts on the baseline balanced growth path; the subsidy arrives unexpectedly at $t = 0$ at the budget-exhausting rate ς^* of Table 6 and is held there; agents have perfect foresight; and the economy converges to the subsidized balanced growth path. Welfare is then evaluated along the realized path, so that (22) is replaced by $\ln(1+\xi) = \rho \int_0^\infty e^{-\rho t}\Big[\ln C_1(t) - \ln C_0(t)\Big]dt$, the discounted log-consumption gap between the reformed and the baseline economy. The algorithm is outlined in Appendix G.

Table 8 and Figure 8 report the result. In Model A, the transition is instantaneous: there are no types and hence no composition term to adjust, and Δg jumps to its long-run value on impact as shown in the left panel of Figure 8, so the balanced-growth welfare number is exact (−0.12% for the size screen; −0.20% for the age screen). Model A also makes plain that more growth need not mean more welfare: there, the age screen buys the larger growth gain (+0.06 against +0.05 percentage points) and is still the more costly design (−0.20% against −0.12%). Eligibility for the young is in part an entry subsidy—entry runs 55% above baseline under the age screen against 44% under the size screen—and the labor it draws into entry is withdrawn from production today against growth delivered later, so the age design pays the larger consumption-level cost on impact (1.4% against 1.1%). The same wedge is present in Model B, and it is why every design in Table 8 carries a welfare cost even where it raises growth.

In Model B, the transition is not instantaneous. The size-screened subsidy destroys growth *slowly*: on impact, Δg is only −0.008 percentage points, the effort term (+0.07) and a composition term that has only begun to build (−0.08) very nearly cancelling. The composition term then compounds as the subsidy tilts the line stock toward the low types: μ_H erodes from 0.212 to 0.164 as shown in the right panel, and Δg reaches half of its long-run −0.24 only after 10.4 years. The middle panel shows why the two terms separate in time. The subsidy raises R&D intensities immediately, and the effort term is already two-thirds of the way to its long-run value on the day

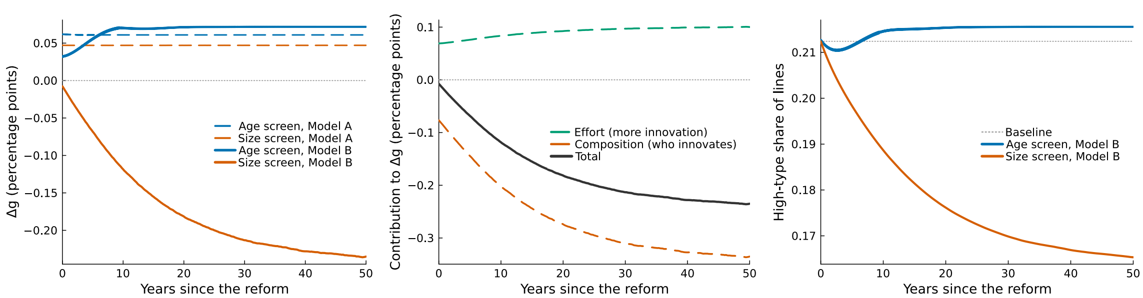
Table 8: Welfare along the transition (micro ceiling, outlay: 1% of output)

	Growth effect Δg			CE welfare	
Screen and model	on impact	long run	half-life	transition	across BGPs
Size, Model A	$+0.047$	$+0.047$	—	-0.12%	-0.12%
Size, Model B	-0.008	$-\mathbf{0.241}$	10.4 yr	$-\mathbf{4.05}\%$	-6.53%
Age, Model A	$+0.062$	$+0.061$	—	-0.20%	-0.20%
Age, Model B	$+0.032$	$+0.072$	1.2 yr	-0.22%	-0.03%

Notes: The four micro-ceiling designs of Table 6—eligibility below the €2m turnover ceiling, or below the age cutoff of 9 years that holds the same share of aggregate sales—with the subsidy introduced unexpectedly at $t = 0$ at the budget-exhausting rate ς^* of Table 6 and held fixed thereafter. The growth columns report the deviation of the annual growth rate from baseline, in percentage points, immediately after the reform and on the new balanced growth path, and the time at which it reaches half of its long-run value ("—": in Model A, the growth effect jumps immediately). "Transition" evaluates welfare along the realized consumption path, "across BGPs" compares balanced growth paths as in Table 6. Baseline growth is 2%.

of the reform (+0.07 of an eventual +0.10); the composition term, which the same intensities can move only by slowly reallocating the stock of lines, is less than a quarter of the way there (−0.08 of an eventual −0.34). Discounting a loss that arrives this late lowers its cost from 6.5% of consumption to 4.1%: two-fifths of the balanced-growth figure is an artifact of crediting the reformed economy immediately with a growth loss it takes a decade to inflict. The damage nonetheless survives, and remains an order of magnitude larger than anything the homogeneous model produces at the same budget.

Figure 8: The transition after the reform (micro ceiling)



Notes: The transition after the unexpected introduction, at $t = 0$, of the micro-ceiling R&D subsidy of Table 6 (outlay 1% of output; the rate ς^* is held fixed thereafter, and the economy converges to the subsidized balanced growth path). *Left:* the deviation of the aggregate growth rate from its baseline value, in percentage points, under the size and the age screen in each model. *Middle:* that deviation for the size screen in Model B, split into the effort and composition terms of (20); each change is weighted at the midpoint of the two paths, so the two terms sum to the total exactly. *Right:* the share of the economy's product lines held by high-type firms in Model B—the stock through which the composition term operates. The size screen erodes that stock, and its growth loss compounds over a decade; the age screen raises it. In Model A, there are no types, the composition term is zero, and the growth effect is a jump.

The age screen moves the same stock in the opposite direction, and the transition therefore works against it. Its long-run growth effect is positive, so delay defers a gain rather than a loss: Δg starts at $+0.03$ percentage points and builds to $+0.07$ over about a decade, as the entry surge that a young-firm subsidy triggers on impact subsides and the subsidized high types accumulate lines (μ_H rising from 0.212 to 0.216). Set against the immediate 1.4% drop in consumption as labor moves into R&D and entry, the delayed gain no longer nets out, and the micro-footprint age screen turns from balanced-growth neutrality (-0.03%) into a small but clear welfare loss (-0.22%). It remains far cheaper than the size screen at the same budget (-4.05%), and it remains the only one of the two that raises growth at all. The general lesson is that, in Model B, the composition effect of any screen has to work through the same slowly-moving stock of lines, so the balanced-growth comparison mis-states welfare in a direction set by the sign of the long-run growth effect: it overstates the losses of a growth-reducing reform and overstates the gains of a growth-raising one. The age-screen gains at the broader ceilings in Table 6 should accordingly be read as upper bounds, and the size-screen losses there as overstatements of the same order as at the micro ceiling. What is not affected is the ranking: replacing the size screen with an age screen at equal outlay raises growth and costs an order of magnitude less.

The timing compounds the inference problem the screen already poses. For its first decade, the size-screened subsidy is nearly invisible in the aggregate: growth is 0.008 percentage points lower on impact and 0.07 lower after five years, against 0.24 in the long run. An evaluation conducted at the horizon at which such programs are actually assessed would find eligible firms raising their R&D—the effort term is positive from the first instant—and aggregate growth all but unchanged, and would conclude that the screen is harmless. The damage is accruing to a stock, the share of the economy's product lines in the hands of the firms that innovate in large steps, which the subsidy erodes with a half-life of a decade. The size screen therefore hides twice. In the cross-section, the firms it harms are not the firms it treats; in time, the harm accrues long after the treatment, so even an aggregate evaluation returns almost nothing for years.

The cost-heterogeneity variant B^{cost}

The sign flip of Δg runs on the composition term $\tau\, d\bar{\lambda}$ of (20)—subsidizing small firms shifts innovation toward low types and lowers the mean step—and so exists only because the types enter the *step*. Had they instead differed in the *cost* of innovating at a common step, as in Acemoglu et al. (2018), $\bar{\lambda}$ would be constant and the composition term, with the sign flip, would vanish identically. The two forms are close to observationally equivalent on the moments firm-dynamics models usually target—both deliver $x_H > x_L$, low types selected into the small-given-age pool, and a

Pareto size tail—so the choice cannot be settled by assertion. I therefore re-estimate the model with cost heterogeneity in place of the step and repeat the policy experiments in it. In the variant B^{cost}, every innovator takes the same step λ and the type instead scales the expansion-R&D cost, $x^2/\psi_{x,\theta}$ with $\psi_{x,H} > \psi_{x,L}$—at the model's quadratic cost exactly the arrival-capacity type of Acemoglu et al. (2018)—leaving entry, type switching, demand, and the labor market unchanged. Appendix I develops the variant and reports the estimation and its fit in full; three findings matter here.

First, the estimation is informative about which form the data prefer, but not definite—and none of what follows rests on it. Estimated on exactly Model B's moments by the same routine, the cost form matches most targets about as well. But with a common step, type heterogeneity does not affect entry size, and the estimated variant undershoots the dispersion of young firms (0.83 in the age-zero bin, against Model B's 1.09 and the data's 1.39). Forced to carry every type difference on the expansion margin, it stretches the expansion gap to $x_H/x_L = 3.0$ (Model B: 2.1); from its too-low starting point, its dispersion schedule then rises faster with age than in the data. The same expansion gap spills into the untargeted growth moments as too many sustained fast growers among large firms (Appendix I, Figures A-3 and A-4). The estimated configuration says the same thing from the other side: with the type invisible at entry, the estimator inverts the entrant composition, $p_h = 0.88$ against Model B's 0.35, so that nearly nine in ten entrants are "high types" whose share then decays to 20% by age ten. A trait carried by almost every entrant and shed with age is not a rare ex-ante difference between firms but a phase that nearly all firms pass through—ex-ante heterogeneity in name, a life-cycle stage in substance.[33]

Second, the sign flip is indeed conditional on the step form. The Model B^{cost} rows of Table 6 repeat the statutory experiments in the variant—the same ceilings, mapped by the same eligible sales shares on its own cross-section, at the same 1%-of-output outlay. With the composition term identically zero, the size screens that cost 0.19–0.24 percentage points of growth in Model B now *raise* growth, by 0.04–0.06, as (20) predicts.

Third, nothing else about the size screen improves. It still selects against innovation productivity: firms below the statutory ceilings remain disproportionately low types—firms that now innovate *slowly* rather than in small steps—carrying 0.86–0.89 times the ineligible pool's mean innovation rate, with roughly half of them at least ten years old. And it is still dominated by the age screen: at matched footprints and

[33] Acemoglu et al. (2018) obtain very similar values—a high-type entrant share of 0.93 and a switching rate of 0.21—despite identifying it off dynamic moments, firm growth rates and transitions across the size distribution by age and size. Identification here rests on cross-sectional moments instead, with the dynamic moments used as untargeted comparisons (Section 5.3).

the same outlay, the age design delivers +0.07 to +0.08 percentage points of growth at every ceiling (Table 6), against the size screen's +0.04 to +0.06—and it is the only one of the two whose growth gain repays its consumption-level cost (consumption equivalents of +0.1 to +0.2% against −0.1 to −0.2%; both designs pay a similar level cost, about 1.4% at the micro ceiling, as labor moves into R&D and entry).[34] The comparison this paper rests on is accordingly the growth ranking, which the transition leaves intact. The mechanism behind the surviving ranking is worth stating. Per subsidy dollar, the low-type pool is the *cheap* source of innovation—R&D costs are convex, and the high types already operate high on their marginal cost curves—so the size screen's failure is not that its dollars buy too little innovation; at equal pool size, they buy *more* innovation events per dollar than the age screen's. What the age screen buys instead is a stock: reaching the high types during their expansion phase raises the share of the economy's product lines they own—by 2.1 percentage points at the micro footprint, where the size screen *lowers* it by 0.6—and those lines keep generating innovation at the high types' privately chosen rates long after, the reallocation channel of Acemoglu et al. (2018) operating through the ownership of lines. That stock accounts for roughly two-thirds of the age design's growth effect at the micro footprint.[35] What depends on the type form, then, is only the strong claim—that the misdirected subsidy lowers growth outright. What does not: small firms are negatively selected on innovation productivity under either form, so the size screen reaches the least productive innovators either way, and screening on age dominates screening on size under both, at every statutory ceiling. The robust conclusion is thus about the screen rather than about the growth accounting—size-based eligibility is not the natural way to reach the firms that innovate most productively—while the price of using it, from growth merely forgone to growth lost outright, depends on where the persistent heterogeneity comes from.

7 Conclusion

This paper decomposes cross-sectional firm-size dispersion in Sweden into size differences across firm ages and size dispersion conditional on age. Although firms older

[34] Those consumption equivalents compare balanced growth paths, and the age design's growth gain here accrues through the same slowly-moving line stock as in Model B, so they should be read as upper bounds in the sense of Table 8, which computes the transition for the micro-ceiling designs of Models A and B.

[35] This resolves an apparent tension in Table 6: Model B's *effort* term ranks the screens the other way (+0.10 size, +0.07 age). The trade-off is the same in both models—the size screen subsidizes cheap low-type innovation, the age screen builds the high-type line stock—but the weights differ: low types hold 79% of lines in Model B, so the cost advantage of subsidizing them prevails; in B^{cost} high types hold 36% of lines (twice Model B's share, the footprint of its larger expansion gap), and the stock term prevails.

than 25 years are, on average, more than a log point larger than entrants, size dispersion conditional on age accounts for 96.4% of the variance of log size in 2017. Firm age predicts average firm size but does not account for its dispersion.

The decomposition is an accounting identity, so it holds exactly in the data without any model. The 3.6% between-age share is therefore an upper bound on how much learning, life-cycle expansion, growth out of financial constraints, selection among survivors, and cohort effects *jointly* contribute to cross-sectional size dispersion *through the mean size–age profile*. It rules out none of these mechanisms per se: each also generates size dispersion conditional on age, which loads on the within component, so the bound constrains only their common, mean-profile part. Firm-dynamics models calibrated to the average life cycle and firm age distribution can nonetheless violate the bound substantially: a Cobb–Douglas quality-ladder model attributes about 29% of size dispersion to the between-age component—nearly an order of magnitude too much—with both the mean size–age profile and the age distribution matched well.

The model in this paper shows what closes the gap. Two departures from the workhorse model—a constant-elasticity aggregator, so that firm size reflects the quality of a firm's products and not only their count, and persistent ex-ante heterogeneity in firms' innovation productivity—bring the between-age share to 5.9%, close to the data. This heterogeneity can enter through the size of firms' innovation steps or through the cost of innovating at a common step—the two forms the literature considers—and both, estimated on the same moments, fit most targeted profiles about as well. The evidence tilts toward the step form—it fits size dispersion among young firms better, because the cost form generates none across types at entry—but it is suggestive rather than definite, and the policy conclusions are stated so as not to depend on it.

That ex-ante heterogeneity carries a policy lesson, one that follows directly from the decomposition. Because size dispersion is overwhelmingly within age groups, a firm's size reveals its type more than its age; and small firms are negatively selected on it—disproportionately the economy's least productive innovators rather than young firms poised to grow. An R&D subsidy screened on firm size—eligibility below the EU statutory turnover ceilings for "micro", "small", or "medium enterprises"—therefore reaches the wrong firms under either form of the heterogeneity, and screening on firm age instead delivers 0.02 to 0.36 percentage points more growth at equal fiscal cost at every ceiling. What is robust is this verdict on the screen; what the source of the heterogeneity determines is the price of using it anyway. Under the step form the misdirection is costly outright—the size-screened subsidy lowers aggregate growth by 0.19 to 0.24 percentage points, whereas the same subsidies raise growth in a homogeneous benchmark—while under the cost form it still raises growth, only by less than a better-targeted design would.

The verdict on the screen is more general than the model that delivers it. In this model, a firm's sales are the same power function of the quality of its products as they would be of the efficiency with which it makes them, so nothing here distinguishes innovations that make a firm's products better from innovations that make them cheaper. Nor does the verdict turn on where the heterogeneity sits, as the two variants estimated above show. All it requires is that firms grow at systematically different rates because of a persistent firm characteristic. Wherever that holds, the firms a size cutoff selects are disproportionately those with least of it, and size is not a usable proxy for the characteristic the policy is trying to reach.

This paper treats a firm's type as an immutable draw at entry. What determines that type, and whether policy can affect it rather than only reallocate resources across fixed types, is the natural next question, and the one on which any rationale for firm support ultimately rests.

References

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

Akcigit, U. and Kerr, W. R. (2018), 'Growth through heterogeneous innovations', *Journal of Political Economy* **126**(4), 1374–1443.

Akerlof, G. A. (1978), 'The economics of "tagging" as applied to the optimal income tax, welfare programs, and manpower planning', *American Economic Review* **68**(1), 8–19.

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

Boppart, T., Carlsson, M., Kondziella, M. and Peters, M. (2023), Micro ppi-based real output forensics, Technical report, Tech. rep.

Criscuolo, C., Martin, R., Overman, H. G. and Van Reenen, J. (2019), 'Some causal effects of an industrial policy', *American Economic Review* **109**(1), 48–85.

Dechezleprêtre, A., Einiö, E., Martin, R., Nguyen, K.-T. and Van Reenen, J. (2023), 'Do tax incentives increase firm innovation? an rd design for r&d, patents, and spillovers', *American Economic Journal: Economic Policy* **15**(4), 486–521.

Foster, L., Haltiwanger, J. and Syverson, C. (2016), 'The slow growth of new plants: Learning about demand?', *Economica* **83**(329), 91–129.

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

Garcia-Macia, D., Hsieh, C.-T. and Klenow, P. J. (2019), ‘How destructive is innovation?’, *Econometrica* **87**(5), 1507–1541.

Garicano, L., Lelarge, C. and Van Reenen, J. (2016), ‘Firm size distortions and the productivity distribution: Evidence from france’, *American Economic Review* **106**(11), 3439–3479.

Guner, N., Ventura, G. and Xu, Y. (2008), ‘Macroeconomic implications of size-dependent policies’, *Review of economic Dynamics* **11**(4), 721–744.

Haltiwanger, J., Jarmin, R. S. and Miranda, J. (2013), ‘Who creates jobs? small versus large versus young’, *Review of Economics and Statistics* **95**(2), 347–361.

Hopenhayn, H. A. (1992), ‘Entry, exit, and firm dynamics in long run equilibrium’, *Econometrica: Journal of the Econometric Society* pp. 1127–1150.

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.

Howell, S. T. (2017), ‘Financing innovation: Evidence from r&d grants’, *American economic review* **107**(4), 1136–1164.

Hsieh, C.-T. and Klenow, P. J. (2009), ‘Misallocation and manufacturing tfp in china and india’, *The Quarterly journal of economics* **124**(4), 1403–1448.

Hsieh, C.-T. and Klenow, P. J. (2014), ‘The life cycle of plants in india and mexico’, *The Quarterly Journal of Economics* **129**(3), 1035–1084.

Jovanovic, B. (1982), ‘Selection and the evolution of industry’, *Econometrica: Journal of the econometric society* pp. 649–670.

Juhn, C., Murphy, K. M. and Pierce, B. (1993), ‘Wage inequality and the rise in returns to skill’, *Journal of political Economy* **101**(3), 410–442.

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

Klenow, P. J. and Li, H. (2025), ‘Entry costs rise with growth’, *Journal of Political Economy Macroeconomics* **3**(1), 43–74.

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

Kondziella, M. (2026), ‘Recent changes in firm dynamics and the nature of rising firm-size dispersion’.

Lehr, N. H. (2025), *Did R&D Misallocation Contribute to Slower Growth?*, International Monetary Fund.

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

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.

Midrigan, V. and Xu, D. Y. (2014), 'Finance and misallocation: Evidence from plant-level data', *American economic review* **104**(2), 422–458.

Moreira, S. (2016), 'Firm dynamics, persistent effects of entry conditions, and business cycles', *Persistent Effects of Entry Conditions, and Business Cycles (October 1, 2016)* .

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

Restuccia, D. and Rogerson, R. (2008), 'Policy distortions and aggregate productivity with heterogeneous establishments', *Review of Economic dynamics* **11**(4), 707–720.

Sedláček, P. and Sterk, V. (2017), 'The growth potential of startups over the business cycle', *American Economic Review* **107**(10), 3182–3210.

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

Supplemental Appendix for

“Firm-Size Dispersion and the Limits of Age”

[FOR ONLINE PUBLICATION ONLY]

A Empirics

A.1 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.[36] SCB defines sales as income from the companies' main business for goods sold and provided services. Employment refers to the average number of employees in full-time units in accordance with the company's annual report. 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 2004 with negative profits.

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

The U.S. firm counts and employment by firm age used in Section 3.4 and Figure 5 are from the Business Dynamics Statistics. Source: U.S. Census Bureau – Center for Economic Studies – Business Dynamics Statistics (2023).

A.2 Further robustness

One concern is that changes in the firm structure, e.g., when firms merge with other firms resulting in a new firm ID, affect the measured firm age. 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 split of the cross-sectional size dispersion into the within- and between-age component remains virtually unchanged when excluding firms for whom the imputed firm ID differs from the observed firm ID (96.6% and 3.4%).

A related concern is the definition of firm age itself. In the main text, I measure firm age as the number of years since the first worker—employee or self-employed—joins the firm, using the employer-employee records in RAMS (see Section 2). As an alternative, I measure firm age as the number of years since the firm first files tax returns and appears in the balance-sheet data (FEK). The decomposition is robust to this alternative age measure: the within- and between-age components account for 94.2% and 5.8% of the cross-sectional variance of log size in 2017, respectively, close to the 96.4% and 3.6% obtained under the baseline age measure.

B The model: equilibrium and the subsidized balanced growth path

This appendix proves Propositions 1–2, derives the stationary quality distribution and its moment recursion (14), collects the balanced-growth system, and states the solution algorithm. Section F derives the value functions and algorithm for the eligibility-contingent subsidies of Section 6.

B.1 Additivity of firm value

A firm's state is $(\theta; \hat{q}_1, \ldots, \hat{q}_n)$. Let $V_\theta(\hat{q}_1, \ldots, \hat{q}_n; t)$ denote its value and normalize $V_\theta = Y(t)\, v_\theta(\hat{q}_1, \ldots, \hat{q}_n)$; since Y grows at g and $r - g = \rho$ under log utility, the normalized Bellman equation reads

$$\begin{aligned}\rho\, v_\theta(\hat{q}_1, \ldots, \hat{q}_n) = \max_{x \geq 0} \Bigg\{ &\sum_{j=1}^{n} \left[\frac{\hat{q}_j^{\,\varepsilon-1}}{\varepsilon M_{\varepsilon-1}} - g\, \hat{q}_j\, \partial_{\hat{q}_j} v_\theta + \tau \Big(v_\theta(\{\hat{q}_i\}_{i \neq j}) - v_\theta \Big) \right] \\ &+ n\, x\, \mathbb{E}_{\hat{q}' \sim F} \Big[v_\theta(\{\hat{q}_i\} \cup \{\hat{q}' + \lambda_\theta\}) - v_\theta \Big] - n\, \omega\, \frac{x^\zeta}{\psi_x} \Bigg\} \\ &+ \nu\, \mathbf{1}\{\theta = H\} \Big[v_L(\hat{q}_1, \ldots, \hat{q}_n) - v_H(\hat{q}_1, \ldots, \hat{q}_n) \Big],\end{aligned} \tag{A-1}$$

with $v_\theta(\varnothing) = 0$ (exit). The terms are the normalized profit flows, the relative-quality decay of each held line, the loss of each line to creative destruction at rate τ, the expansion surplus at firm rate nx with per-line cost $\omega x^\zeta / \psi_x$, and the type switch. Conjecture $v_\theta(\hat{q}_1, \ldots, \hat{q}_n) = \sum_j v_\theta(\hat{q}_j)$. Every term of (A-1) then separates across lines: profits, decay, and loss are line-by-line; the capture value $\mathbb{E}_F[v_\theta(\hat{q}' + \lambda_\theta)]$ is independent of the capturer's portfolio, and cost and capacity are per line, so the maximization is the same for each line; and switching maps the sum into the sum. Dividing by n term by term yields the per-line equation (10), verifying the conjecture.

B.2 Proof of Proposition 2

Substitute the conjecture $v_\theta(\hat{q}) = A_\theta\, \hat{q}^{\,\varepsilon-1} + B_\theta$, with type-specific coefficients not imposed equal, into (10), using $\hat{q}\, \partial_{\hat{q}} \hat{q}^{\,\varepsilon-1} = (\varepsilon - 1) \hat{q}^{\,\varepsilon-1}$. Matching the coefficients on

$\hat{q}^{\varepsilon-1}$,

$$\rho\, A_L = \frac{1}{\varepsilon M_{\varepsilon-1}} - (\varepsilon-1) g\, A_L - \tau A_L, \quad \rho\, A_H = \frac{1}{\varepsilon M_{\varepsilon-1}} - (\varepsilon-1) g\, A_H - \tau A_H + \nu\,(A_L - A_H). \tag{A-2}$$

The first equation gives $A_L = \Big[\varepsilon M_{\varepsilon-1}(\rho + \tau + (\varepsilon-1)g)\Big]^{-1}$; substituting it into the second,

$$\Big(\rho+\tau+(\varepsilon-1)g+\nu\Big) A_H = \frac{1}{\varepsilon M_{\varepsilon-1}} + \nu A_L = \Big(\rho+\tau+(\varepsilon-1)g+\nu\Big) A_L \implies A_H = A_L \equiv A,$$

which is (16): the switching term contributes nothing to the slope, and the common slope is not an assumption but an implication. Matching the $\hat{q}$-free terms,

$$(\rho+\tau)\, B_L = \Phi_L, \qquad (\rho+\tau+\nu)\, B_H = \Phi_H + \nu\, B_L, \tag{A-3}$$

with Φ_θ the maximized expansion surplus of (18). Under the conjecture the capture value is

$$G_\theta = \mathbb{E}_F\Big[v_\theta(\hat{q}' + \lambda_\theta)\Big] = A\,\mathbb{E}_F\Big[(\hat{q}' + \lambda_\theta)^{\varepsilon-1}\Big] + B_\theta = A\,\Gamma_\theta + B_\theta,$$

and expanding $(\hat{q}' + \lambda_\theta)^{\varepsilon-1}$ binomially for integer $\varepsilon - 1$ gives the closed form (17). The first-order condition of $\Phi_\theta = \max_x\{xG_\theta - \omega x^\zeta/\psi_x\}$ is $\zeta\omega x^{\zeta-1}/\psi_x = G_\theta$, which yields x_θ and, substituting back, $\Phi_\theta = (\zeta-1)\,\omega\, x_\theta^\zeta/\psi_x$ as in (18). The ordering claims are Lemma A-1. □

Lemma A-1 (Type ordering)**.** $\Gamma_H > \Gamma_L$. *Moreover, at $\zeta = 2$, on any balanced growth path with $\max\{x_H, x_L\} < \rho+\tau+\nu$ (equivalent, once the lemma delivers $x_H > x_L$, to the transversality condition of Remark A-2; satisfied with a wide margin at the estimates),*

$$B_H > B_L, \qquad G_H > G_L, \qquad x_H > x_L.$$

Proof. $\Gamma_\theta = \sum_{i=0}^{\varepsilon-1}\binom{\varepsilon-1}{i} M_i\,\lambda_\theta^{\varepsilon-1-i}$ is strictly increasing in λ_θ ($M_i > 0$), so $\lambda_H > \lambda_L$ gives $\Gamma_H > \Gamma_L$. Let $\Delta \equiv B_H - B_L$ and $\Delta\Gamma \equiv \Gamma_H - \Gamma_L > 0$. Subtracting the two equations in (A-3),

$$(\rho+\tau+\nu)\,\Delta = \Phi_H - \Phi_L.$$

At $\zeta = 2$, $\Phi_\theta = \psi_x G_\theta^2/(4\omega)$ and $x_\theta = \psi_x G_\theta/(2\omega)$, so

$$\Phi_H - \Phi_L = \frac{\psi_x}{4\omega}\,(G_H + G_L)(G_H - G_L) = \frac{x_H + x_L}{2}\,\Big(A\,\Delta\Gamma + \Delta\Big),$$

using $G_H - G_L = A\,\Delta\Gamma + \Delta$. Collecting terms,

$$\left(\rho + \tau + \nu - \frac{x_H + x_L}{2}\right)\Delta = \frac{x_H + x_L}{2}\,A\,\Delta\Gamma > 0,$$

and the bracket is positive by hypothesis, so $\Delta > 0$. Then $G_H - G_L = A\Delta\Gamma + \Delta > 0$, and x_θ strictly increasing in G_θ gives $x_H > x_L$. □

B.3 Proof of Proposition 1 and the moment recursion

Over an interval Δt, a measure $\tau\Delta t$ of lines receives an innovation; a fraction h of these innovations is made by high types, each adding $\lambda_H \bar{q}(t)$ to the quality stock $\bar{q}(t) = \int_0^1 q_j\, dj$, and a fraction $1-h$ by low types, each adding $\lambda_L \bar{q}(t)$. Hence

$$\begin{aligned} \bar{q}(t + \Delta t) &= \bar{q}(t) + \tau\Delta t\,\Big[h\lambda_H + (1-h)\lambda_L\Big]\,\bar{q}(t) + o(\Delta t) \\ \implies\quad g &= \lim_{\Delta t \to 0} \frac{\bar{q}(t+\Delta t) - \bar{q}(t)}{\Delta t\,\bar{q}(t)} = \tau\,\bar{\lambda}. \end{aligned}$$ □

At the line level, $\hat{q}$ decays deterministically at rate g between innovations and, at rate τ, jumps to $\hat{q} + \lambda$ with $\lambda = \lambda_H$ with probability h and λ_L with probability $1-h$, independently of $\hat{q}$ (undirected R&D). The generator of this process applied to $\phi(\hat{q}) = \hat{q}^k$ is

$$(\mathcal{A}\phi)(\hat{q}) = -k\,g\,\hat{q}^{\,k} + \tau\,\Big[\mathbb{E}_\lambda(\hat{q} + \lambda)^k - \hat{q}^{\,k}\Big].$$

Stationarity of F requires $\int \mathcal{A}\phi\, dF = 0$; with $m_i = \mathbb{E}_\lambda[\lambda^i] = h\lambda_H^i + (1-h)\lambda_L^i$ and $\mathbb{E}_\lambda\mathbb{E}_F[(\hat{q}+\lambda)^k] = \sum_{i=0}^k \binom{k}{i} M_i\, m_{k-i}$, the $i = k$ term ($m_0 = 1$) cancels against $-\tau M_k$ and

$$0 = -k\,g\,M_k + \tau \sum_{i=0}^{k-1} \binom{k}{i} M_i\, m_{k-i},$$

which is the triangular recursion (14). At $k = 1$ it reads $g\,M_1 = \tau\bar{\lambda}$ with $M_1 = 1$—the growth equation again, an internal consistency check.

Remark A-1 (The additive step is required under CES)**.** *The multiplicative-step* *convention of Klette and Kortum (2004), $q_j \mapsto \kappa q_j$ with $\kappa > 1$, admits no stationary relative-quality distribution under CES. Let $u_j = \ln \hat{q}_j$. With multiplicative steps, average quality grows at rate $g_{\bar{q}} = \tau\, \mathbb{E}[\kappa - 1]$ (captures are uniform, so the captured line's expected quality is $\bar{q}$), and u_j drifts down at rate $g_{\bar{q}}$ between innovations and jumps up by $\ln \kappa$ at rate τ with no state dependence: u_j is a Lévy process, and its drift is strictly negative,*

$$\tau\, \mathbb{E}[\ln \kappa] - g_{\bar{q}} = \tau\, \mathbb{E}\Big[\ln \kappa - (\kappa - 1)\Big] < 0,$$

since $\ln \kappa < \kappa - 1$ for all $\kappa \neq 1$. A Lévy process with nonzero drift has no stationary distribution: the cross-section of $\ln \hat{q}$ spreads without bound, the level mean $\mathbb{E}[\hat{q}] = 1$ being preserved by an ever-thinner set of ever-larger lines. Under Cobb–Douglas the non-stationarity is invisible because sales do not depend on q; under CES it is fatal. The additive step (8), anchored to $\bar{q}$, makes $\hat{q}$ mean-reverting (proportional decay, additive jumps) and is the stationarity device of Akcigit and Kerr (2018) and Acemoglu et al. (2018).

B.4 The equilibrium system and its computation

With $\zeta = 2$, the balanced growth path of Definition 1 is a solution of eight equations in the eight unknowns $(x_H, x_L, z, \mu_H, g, L_P, B_H, B_L)$:

$$\begin{aligned}
&\text{(i) type composition (15);} \qquad \text{(ii) growth } g = \tau\bar{\lambda} \text{ (Proposition 1);}\\
&\text{(iii)–(iv) R\&D policies } x_\theta = \frac{\psi_x G_\theta}{2\omega},\ \theta \in \{H, L\} \text{ (18);}\\
&\text{(v) } (\rho + \tau)B_L = \omega\,\frac{x_L^2}{\psi_x}; \qquad \text{(vi) } (\rho + \tau + \nu)B_H = \omega\,\frac{x_H^2}{\psi_x} + \nu B_L;\\
&\text{(vii) free entry (11);} \qquad \text{(viii) labor-market clearing (12),}
\end{aligned} \tag{A-4}$$

where τ, h, $\bar{\lambda}$, m_k, $\{M_k\}_{k \le \varepsilon - 1}$, A, Γ_θ, G_θ, and ω are the explicit functions (9), (13), (14), (16), (17), and (6) of the unknowns. Model A is the restriction $\lambda_H = \lambda_L$, under which ν and p_h drop out and (A-4) collapses to five equations in (x, z, g, L_P, B). The system is solved by a standard nonlinear solver (residuals below 10^{-12}) with continuation in the parameters, checking $z > 0$ ex post; Model B is warm-started from the single-type limit with the type spread opened stepwise. The cross-sectional moments of Section 4.4 are computed by exact event-time simulation of a large balanced-growth cohort panel, on the same integer-age grid (25+ pooled) as the data, and validated against (14), (15), and the Klette–Kortum count distribution absent types (Appendix C). The additive per-line value $v_\theta = A\hat{q}^{\varepsilon-1} + B_\theta$ with a common slope A, the effective discount rate $\rho + \tau + (\varepsilon - 1)g$, and the growth equation $g = \tau\bar{\lambda}$ take the same form as the corresponding expressions in Akcigit and Kerr (2018) and Acemoglu et al. (2018).

C Firm-size moments in the model

This appendix describes how the three margins of the decomposition in eq. (1)—the mean size–age profile, the firm-age distribution, and size dispersion conditional on age—are computed in the model of Section 4. A type-θ firm holding lines with relative qualities $\{\hat{q}_j\}$ has log sales $\ln S_f = \ln(Y/M_{\varepsilon-1}) + \ln\sum_j \hat{q}_j^{\varepsilon-1}$ by eq. (5); in the cross section the constant $\ln(Y/M_{\varepsilon-1})$ is common to all firms, so every conditional moment below derives from the joint distribution of a firm's type, its product count, and the relative qualities of its lines. Under CES ($\varepsilon > 1$) sales are *not* proportional to the product count, so this distribution has no closed form and the moments are computed by simulation; the Cobb–Douglas case ($\varepsilon \to 1$, sales = count) is closed form and is used both for Model A at $\varepsilon = 1$ and to validate the simulation.

Entry dispersion and its ex-ante component. An entrant holds one line of relative quality $\hat{q}'+\lambda_\theta$ with $\hat{q}' \sim F$ independent of θ, so $\ln S = \text{const}+(\varepsilon-1)\ln(\hat{q}'+\lambda_\theta)$ and, by the law of total variance over the type,

$$\mathrm{Var}\Big(\ln S \mid a=0\Big) = (\varepsilon-1)^2\Big[\underbrace{\mathbb{E}_\theta \mathrm{Var}\Big(\ln(\hat{q}'+\lambda_\theta) \mid \theta\Big)}_{\text{ex post: the quality draw}} + \underbrace{p_h(1-p_h)\Big(m_H - m_L\Big)^2}_{\text{ex ante: the type}}\Big], \tag{A-5}$$

where $m_\theta \equiv \mathbb{E}\Big[\ln(\hat{q}'+\lambda_\theta)\Big]$. The second term is the between-type component of (19): it is the variance of the type-conditional means over the two-point type distribution, enters with weight $p_h(1-p_h)$, and vanishes in Model A ($\lambda_H = \lambda_L$) and in the cost variant of Appendix I (common step).

The ex-ante component at higher ages. The same decomposition extends to every age. A firm's entry type $\theta_0 \in \{H, L\}$ is fixed at birth—the switch to low type at rate ν is a post-entry outcome—so it carries the model's ex-ante heterogeneity. Applying the law of total variance over the entry type within each age a gives

$$\begin{aligned}\mathrm{Var}\Big(\ln S \mid a\Big) = &\underbrace{\pi_a \mathrm{Var}\Big(\ln S \mid a, H_0\Big) + (1-\pi_a)\mathrm{Var}\Big(\ln S \mid a, L_0\Big)}_{\text{ex post: within entry type}} \\ &+ \underbrace{\pi_a(1-\pi_a)\Big(\bar{s}_H(a) - \bar{s}_L(a)\Big)^2}_{\text{ex ante: the entry type}},\end{aligned} \tag{A-6}$$

where H_0, L_0 index the entry type, π_a is the share of age-a survivors that entered as high types, and $\bar{s}_\theta(a) \equiv \mathbb{E}[\ln S \mid a, \theta_0 = \theta]$ is their mean log size. Being a two-group partition of the age-a cross-section, this is an exact identity, and at $a = 0$—where every firm holds a single line—it reduces to the entry split (A-5). Its ratio to the total, the ex-ante share of size dispersion at age a, is computed on the cohort panel described next and reported in Appendix D.

Simulation of a balanced-growth cohort panel. I simulate the histories of a large number of entering firms in event time, under the balanced-growth aggregates $(\tau, g, x_H, x_L, z, \{M_k\})$ of Definition 1. An entrant is high type with probability p_h and starts with a single line of relative quality $\hat{q}' + \lambda_\theta$, where $\hat{q}'$ is drawn from the stationary line-quality distribution F (sampled exactly from its shot-noise representation; its moments match the recursion in eq. 14). Each firm then evolves through four events: an expansion success, at firm rate $n\, x_\theta$, adds a line of relative quality $\hat{q}' + \lambda_\theta$ with $\hat{q}' \sim F$; a creative-destruction hit, at firm rate $n\,\tau$, removes a line drawn uniformly at random, and the firm exits when it loses its last line; a high type switches to low at rate ν; and between events every held line's relative quality decays deterministically at rate g (own quality fixed while the frontier $\bar{q}$ advances). Recording each firm at every integer age—jittered within the year for unbiased $[a, a+1)$ draws—and forming $\ln S_f$ from eq. (5), I compute, over the same integer-age grouping as the data (integer ages 0 to 24 with firms aged 25 and above pooled): the mean of $\ln S_f$ by age (the size–age profile), the standard deviation of $\ln S_f$ by age (dispersion conditional on age), and the share of firms in each age group (the age distribution). The variance of log size then decomposes as in eq. (1). The simulation is validated against Propositions 1–2, the moment recursion (eq. 14), the type composition (eq. 15), and—in the single-type case—the closed-form count distribution below.

The count process and the Cobb–Douglas case. When all firms share a single expansion rate x (Model A), the product count is a birth–death process and, following Klette and Kortum (2004), the count of a surviving firm of age a_f is geometric,

$$P(n_f = n \mid a_f) = (1 - \gamma(a_f))\, \gamma(a_f)^{n-1}, \quad \gamma(a_f) = x\, \frac{1 - e^{-(\tau - x)a_f}}{\tau - x\, e^{-(\tau - x)a_f}}, \tag{A-7}$$

and the firm survives to age a_f with probability

$$\chi(a_f) = 1 - \tau \, \frac{1 - e^{-(\tau - x)a_f}}{\tau - x\, e^{-(\tau - x)a_f}}. \tag{A-8}$$

The age distribution is proportional to $\chi(a_f)$; on a fine grid with midpoints $\{a_k\}$ the mass in each cell is $\varpi(a_k) = \chi(a_k)/\sum_j \chi(a_j)$ (the weight ϖ is not to be confused with the wage-to-output ratio ω). In the Cobb–Douglas limit ($\varepsilon \to 1$) size equals the product count up to the common factor Y, so mean size and dispersion conditional on age are the mean and variance of the log count under eq. (A-7):

$$\mu(a_f) \equiv E[\ln n_f \mid a_f] = (1 - \gamma(a_f)) \sum_{n=1}^{\infty} \ln(n)\, \gamma(a_f)^{n-1}, \tag{A-9}$$

$$\mathrm{Var}(\ln s_f \mid a_f) = (1 - \gamma(a_f)) \sum_{n=1}^{\infty} (\ln n - \mu(a_f))^2\, \gamma(a_f)^{n-1}. \tag{A-10}$$

For integer age groups $a_f \in [a, a+1)$, the law of total variance across exact continuous ages within the bin gives the within-bin size variance

$$V_a = \underbrace{\sum_{k:\, a_k \in [a,a+1)} \frac{\varpi(a_k)}{p_a} \mathrm{Var}(\ln s_f \mid a_k)}_{\text{avg. within-exact-age variance}} + \underbrace{\sum_{k:\, a_k \in [a,a+1)} \frac{\varpi(a_k)}{p_a} (\mu(a_k) - \bar{\mu}_a)^2}_{\text{variance of mean size across exact ages}}, \tag{A-11}$$

where $p_a = \sum_{k:\, a_k \in [a,a+1)} \varpi(a_k)$ is the mass in bin a and $\bar{\mu}_a$ is the within-bin mean, and the cross-sectional variance decomposes as in eq. (1), $\mathrm{Var}(\ln s_f) = \sum_a p_a V_a + \sum_a p_a (\bar{\mu}_a - \bar{\mu})^2$, with the grouped 25+ bin. This closed-form count case is what Model A at $\varepsilon = 1$—and the Peters (2020) benchmark of Appendix H—uses, and it is the limit the simulator reproduces when types are switched off.

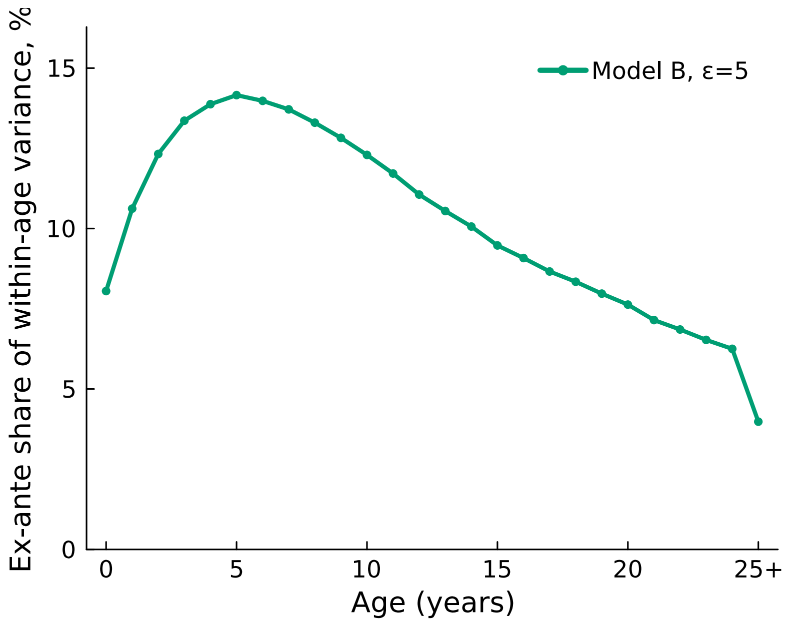
D The ex-ante share of size dispersion by age

Appendix C splits the variance of log size conditional on age into an ex-ante (entry-type) and an ex-post (quality-draw) component, eq. (A-6). This appendix reports that split across the age distribution.

The profile. Figure A-1 plots the ex-ante share by age at the Model B estimates. It is hump-shaped. The type accounts for 6.7% of the variance of log size exactly at entry (eq. A-5); the first age group, $[0, 1)$, which already contains up to a year of divergence, stands at about 8%. The share then rises to a peak near 14% around age five before receding to 4% among firms aged 25 and above. The shape follows from the two ways the types differ. At entry, only the step $\lambda_H > \lambda_L$ has marked size, so the ex-ante share is modest; as firms age, high types expand faster ($x_H > x_L$), the gap between the entry-type means $\bar{s}_H(a) - \bar{s}_L(a)$ widens, and the share climbs over the first several years. Two forces then reverse it: switching at rate ν turns surviving high types into low types and blurs the distinction between the entry groups, while within-type dispersion keeps growing—the surviving high types develop a heavy, near-Zipf size tail—so the ex-post term outpaces the between-type term. The type's imprint on size dispersion is thus largest among firms aged roughly three to seven.

Comparison with Sterk et al. (2021). Sterk et al. (2021) report a much larger ex-ante share of size dispersion conditional on age among age-zero U.S. firms, about 85%. The two numbers are not measurements of the same object, for two reasons. First, the decompositions draw the ex-ante/ex-post boundary in different places. Sterk et al. (2021) classify by the moment of inception: their ex-ante component collects every firm-specific constant drawn before startup—the firm's initial size included—and the ex-post component is only what accumulates thereafter. The split here classifies by type instead: the entry quality draw, though realized at entry, is booked as ex-post, because a level draw carries no information about a firm's subsequent growth (the identification argument of Section 5). Under their timing convention, all of Model B's entry-size variance would count as ex ante, since every determinant of size at age zero is drawn at entry; the model's 6.7% is therefore not the counterpart of their 85%. Second, the ex-ante object itself differs. Their ex-ante heterogeneity lies in levels: firm size converges to a firm-specific steady state ($\theta_i/(1-\rho_u)$ in their notation), and their data reject the unit-root variant in which the ex-ante

Figure A-1: The ex-ante (type) share of size dispersion conditional on age



Notes: The ex-ante component of the variance of log size conditional on age, as a percentage share of that variance (eq. A-6), in a simulated balanced-growth panel of Model B at its $\varepsilon = 5$ estimates. Ages 0–24 with 25 and above pooled.

terms would instead represent heterogeneous growth rates (their footnote 10). Model B's type is precisely such growth-rate heterogeneity, and size in a model of creative destruction has no firm-specific steady state—a firm accumulates product lines until it dies.[37] The two age profiles differ accordingly: theirs is largest at entry and recedes as ex-post shocks accumulate against a bounded ex-ante variance, whereas Model B's is small at entry and hump-shaped, since the growth-rate difference needs time to compound before type switching erodes it (Figure A-1).

Even though the two decompositions differ in whether ex-ante heterogeneity is classified by inception—counting all entry dispersion as ex ante—or by type, as here, this does not affect the conclusion that explaining size dispersion conditional on age with additional ex-post shocks in place of type heterogeneity moves the homogeneous model A in the wrong direction. Grant that model the most generous reading and count all of its entry dispersion as ex ante: creative destruction erases that dispersion.

[37]Sterk et al. (2021) flag (their footnote 16) a potential trade-off between fitting early life-cycle dynamics and reproducing the power-law tail of the ergodic size distribution. Model B's growth-rate form supplies the tail directly, as an untargeted by-product: the transient supercritical high-type phase generates a Pareto right tail with exponent near 1.15 (Section 5).

Fewer than two percent of the firms alive at age twenty still hold the line they entered with, so its ex-ante share falls to essentially zero among older firms—precisely where Sterk et al. (2021) find that share settling near 40%. Even under that reading—which classifies the first random quality draw a firm takes as ex-ante and every later draw from the same distribution as ex-post—the homogeneous model already sits at the wrong end of their evidence before any shocks are added to it.

E The estimation routine

This appendix describes the estimation behind Table 4. The estimated parameters are (x, z, λ) in Model A at $\varepsilon = 1$ (the flow rates directly, for the identification reason given in Section 5), $(\psi_x, \psi_z, \lambda)$ in Model A at $\varepsilon = 5$, and $(\psi_x, \psi_z, \lambda_L, \lambda_H/\lambda_L, \nu, p_h)$ in Model B, with $(\psi_{x,L}, \psi_{x,H}, \psi_z, \lambda, \nu, p_h)$ in the cost variant B^{cost} of Appendix I; $\rho = 0.05$, $\zeta = 2$, and ε are fixed throughout (Section 5).

Loss function. The targets are organized in blocks $b = 1, \ldots, K$: the mean size–age profile (mean log sales relative to age zero, at integer ages 1–24 and the pooled 25+ group), the firm-age distribution (26 groups), aggregate TFP growth ($g = 2\%$), and—in Model B only—the standard deviation of log size by age (26 groups), so $K = 3$ for Model A and $K = 4$ for Model B. Writing ϑ for the parameter vector and $\mu_m(\vartheta)$ for the model counterpart of data moment μ_m^{data}, the loss is the equally block-weighted mean squared proportional deviation

$$\mathcal{L}(\vartheta) = \frac{1}{K} \sum_{b=1}^{K} \frac{1}{|b|} \sum_{m \in b} \left(\frac{\mu_m(\vartheta) - \mu_m^{\text{data}}}{\mu_m^{\text{data}}} \right)^2, \tag{A-12}$$

where $|b|$ is the number of moments in block b. In the Cobb–Douglas model, growth does not enter (A-12): firm size there depends only on the flow rates (x, z), so the step λ is set to match the growth target exactly through $g = \tau\lambda$.

Evaluating the loss. Each evaluation first solves the balanced-growth system of Appendix B exactly (residuals below 10^{-12}); Model B is solved by continuation, warm-starting from the single-type limit and opening the type spread stepwise. The cross-sectional moments are then computed from the simulated balanced-growth cohort panel of Appendix C under common random numbers: all seeds are fixed and shared across parameter vectors, so the loss is a deterministic function of the parameters and comparisons between candidate vectors are free of resampling noise. Parameter bounds ensure a well-defined balanced growth path: entry must remain active, and in Model B, the transient supercritical high-type phase ($x_H > \tau$)—the source of the Pareto size tail, whose exponent is $\nu/(x_H - \tau)$—must satisfy $x_H - \tau < \nu$, imposed with a small margin in the search ($x_H - \tau \leq 0.95\nu$), so that the tail exponent exceeds one and mean firm size is finite; value finiteness, $x_H - \tau < \rho + \nu$, is verified in Remark

A-2 of Appendix F. The cost variant B^{cost} of Appendix I is estimated under Model B's routine unchanged—targets, weights, bounds, and search protocol—so that the two type forms are compared like for like. The reported estimates of Models A and B are interior to all bounds; the cost variant's estimate comes within one percent of the $x_H - \tau \leq 0.95\nu$ fence, which Appendix I reads as part of that variant's diagnosis.

Search. The loss is minimized in two stages: a global quasi-random (Sobol) search over wide parameter boxes, evaluated on a moderate simulated panel (20,000 firms), followed by Nelder–Mead refinement started from the best global points on a larger panel (120,000 firms). The leading starts converge to a common point, which attains the lowest loss and is the reported estimate. All moments, figures, and decompositions in the paper are recomputed at the final estimates on a larger panel (800,000 firms) than used inside the search.

F The size-eligibility subsidy: value functions and algorithm

An ad valorem subsidy $\varsigma \in [0,1)$ to the expansion-R&D costs of eligible firms leaves the static block (3)–(6) and labor-market clearing (12) untouched—the subsidy is a transfer, not a resource cost, so the resource cost of intensity x remains x^2/ψ_x units of labor per line. It changes only the R&D problem, and only for eligible firms. For a uniformly eligible firm the subsidy scales the R&D cost wherever it appears:

$$\Phi_\theta = \max_{x\geq 0}\left\{x\,G_\theta - (1-\varsigma)\,\omega\,\frac{x^2}{\psi_x}\right\} = \frac{\psi_x\,G_\theta^2}{4\omega(1-\varsigma)}, \qquad x_\theta = \frac{\psi_x\,G_\theta}{2\omega(1-\varsigma)}, \tag{A-13}$$

with the constants (A-3) unchanged in form, $(\rho+\tau)B_L = \Phi_L$ and $(\rho+\tau+\nu)B_H = \Phi_H + \nu B_L$. This uniform case is exactly the incumbent-R&D subsidy of Acemoglu et al. (2018), Table 7, whose fixed-outlay, endogenous-rate convention—ς set so the fiscal outlay equals a fixed share of output—is adopted for all comparisons. The three eligibility designs are derived in turn: the age cutoff (exactly solvable), the count cutoff (exactly solvable; not a paper exhibit but the closure-free benchmark against which the size-cutoff solver is validated), and the headline size cutoff.

F.1 Age eligibility ($a \leq \bar{a}$)

Age is a firm-level state common to all lines and independent of the portfolio, so additivity in lines is fully preserved:

$$V_\theta(\hat{q}_1,\ldots,\hat{q}_n;a;t) = Y(t)\Big[A\sum_{j=1}^{n}\hat{q}_j^{\varepsilon-1} + n\,B_\theta(a)\Big],$$

with the same slope A of (16) and *per-line* constants $B_\theta(a)$ that depend on firm age. For $a \geq \bar{a}$ the firm is permanently ineligible and $B_\theta(a) = b_\theta$, the baseline constants solving (A-3); for $a < \bar{a}$ the constants solve the backward system

$$(\rho+\tau)\,B_L(a) = \Phi_L(a) + B_L'(a), \qquad (\rho+\tau+\nu)\,B_H(a) = \Phi_H(a) + \nu B_L(a) + B_H'(a), \tag{A-14}$$

with terminal condition $B_\theta(\bar{a}) = b_\theta$ and

$$\Phi_\theta(a) = \frac{\psi_x\, G_\theta(a)^2}{4\omega(1-\varsigma)}, \qquad x_\theta(a) = \frac{\psi_x\, G_\theta(a)}{2\omega(1-\varsigma)}, \qquad G_\theta(a) = A\,\Gamma_\theta + B_\theta(a). \tag{A-15}$$

Eligibility expires *deterministically*, so (A-14) is a backward ODE rather than a difference system, and integrating it backward from $\bar{a}$ is a contraction. Aggregation is exact because expansion and loss are both per line: the line mass $\ell_\theta(a)$ held by type-θ firms of age a follows the cohort ODEs

$$\ell'_H(a) = \Big(x_H(a) - \tau - \nu\Big)\ell_H(a), \qquad \ell'_L(a) = \Big(x_L(a) - \tau\Big)\ell_L(a) + \nu\,\ell_H(a), \tag{A-16}$$

with $\ell_H(0) = p_h z$ and $\ell_L(0) = (1-p_h)z$; beyond $\bar{a}$ the coefficients are constant at the unsubsidized far-field intensities $x_\theta^\infty \equiv \psi_x(A\Gamma_\theta + b_\theta)/(2\omega)$ and the tail integrals are closed form, requiring $x_H^\infty < \tau + \nu$ for a finite high-type line stock (satisfied at the estimates; cf. Remark A-2). The subsidized balanced growth path is then the solution of the four-equation outer system in (τ, h, L_P, z),

$$\begin{aligned}
&\text{(i)}\;\; \mu_H + \mu_L = 1, \qquad \text{where } \mu_\theta = \int_0^\infty \ell_\theta(a)\,da;\\
&\text{(ii)}\;\; h\,\tau = \mu_H(\tau+\nu) \qquad (h\text{-consistency});\\
&\text{(iii)}\;\; \frac{\omega}{\psi_z} = p_h\Big[A\Gamma_H + B_H(0)\Big] + (1-p_h)\Big[A\Gamma_L + B_L(0)\Big] \qquad \text{(free entry)};\\
&\text{(iv)}\;\; 1 = L_P + \int_0^\infty \frac{\ell_H(a)\,x_H(a)^2 + \ell_L(a)\,x_L(a)^2}{\psi_x}\,da + \frac{z}{\psi_z},
\end{aligned} \tag{A-17}$$

where (ii) requires the high-type innovation flow to sustain the H-held line stock, (iii) capitalizes eligibility into the entrant value through $B_\theta(0)$, and ω, $\{M_k\}$, A, Γ_θ are the explicit functions of (τ, h, L_P) as in Appendix B; the fiscal outlay is

$$\frac{\text{Outlay}}{Y} = \varsigma\,\omega \int_0^{\bar{a}} \frac{\ell_H(a)\,x_H(a)^2 + \ell_L(a)\,x_L(a)^2}{\psi_x}\,da. \tag{A-18}$$

At $\varsigma = 0$, $B_\theta(a) \equiv b_\theta$, $x_\theta(a) \equiv x_\theta$, and (A-17) reproduces the baseline system (A-4). Model A is the same construction with a single type and three outer unknowns (τ, L_P, z).

F.2 Count eligibility ($n \le \bar{n}$): the exact benchmark

A count-contingent subsidy makes the per-line R&D problem depend on the firm's product count, so firm value is no longer a sum of identical per-line values; it remains additive in the quality terms. Guess

$$V_\theta(\hat{q}_1,\ldots,\hat{q}_n;t) = Y(t)\Big[A\sum_{j=1}^{n}\hat{q}_j^{\varepsilon-1} + B_\theta(n)\Big], \qquad B_\theta(0)=0.$$

The guess verifies because the subsidy enters only the expansion surplus, which is independent of the quality portfolio: matching the $\hat{q}_j^{\varepsilon-1}$ coefficients in the firm-level Bellman equation (A-1) reproduces the slope (16) unchanged, while the constants collect the count-dependent terms. Writing $\varsigma_n \equiv \varsigma\,\mathbf{1}\{n \le \bar{n}\}$,

$$\begin{aligned}
\rho\, B_L(n) &= n\Big[\tau\Big(B_L(n{-}1) - B_L(n)\Big) + \Phi_L(n)\Big],\\
(\rho+\nu)\, B_H(n) &= n\Big[\tau\Big(B_H(n{-}1) - B_H(n)\Big) + \Phi_H(n)\Big] + \nu\, B_L(n),
\end{aligned} \tag{A-19}$$

with per-line surplus, intensity, and capture value

$$\Phi_\theta(n) = \frac{\psi_x\, G_\theta(n)^2}{4\omega(1-\varsigma_n)}, \qquad x_\theta(n) = \frac{\psi_x\, G_\theta(n)}{2\omega(1-\varsigma_n)}, \qquad G_\theta(n) = A\,\Gamma_\theta + B_\theta(n{+}1) - B_\theta(n). \tag{A-20}$$

The events behind (A-19): each of the n lines is lost at rate τ, dropping the constant to $B_\theta(n{-}1)$; expansion runs at firm rate $n\,x_\theta(n)$ with marginal gain $G_\theta(n)$—a capture adds the expected quality value $A\Gamma_\theta$ and moves the constant to $B_\theta(n{+}1)$, so at $n=\bar{n}$ the loss of eligibility from growing is priced through $B_\theta(\bar{n}{+}1)$; the type switch flips the whole portfolio's constants at rate ν.

Far field and nesting. If $\varsigma_n \equiv \varsigma^\infty$ is constant in n, (A-19) is solved exactly by the *linear* function $B_\theta(n) = b_\theta\, n$, with slopes solving the per-line quadratics

$$(\rho+\tau)\, b_L = \frac{\psi_x\,(A\Gamma_L + b_L)^2}{4\omega(1-\varsigma^\infty)}, \qquad (\rho+\tau+\nu)\, b_H = \frac{\psi_x\,(A\Gamma_H + b_H)^2}{4\omega(1-\varsigma^\infty)} + \nu\, b_L \tag{A-21}$$

(smaller roots), which at $\varsigma^\infty = 0$ are exactly (A-3): the count system nests the baseline as $B_\theta(n) = B_\theta \cdot n$, the Klette–Kortum additive case. With a finite threshold, $B_\theta(n) = b_\theta n + D_\theta(n)$ with the deviation D_θ localized around $\bar{n}$: for the subcritical low type, the value of *shrinking back into eligibility* from n lines decays like $D_L(n) \sim (\bar{n}/n)^{\rho/(\tau - x_L)}$—the log count drifts down at rate $\tau - x_L$ while the flow is discounted at ρ—and the supercritical high phase runs against the drift, so D_H inherits only the ν-routed low-type component. Numerically, (A-19) is solved by policy iteration on $n = 1, \ldots, N_v$ with the far-field boundary $B_\theta(N_v{+}1) = B_\theta(N_v) + b_\theta$. This value function—a common quality slope plus firm-level constants solving a second-order difference system in n, the per-line intensity driven by the marginal franchise value $B_\theta(n{+}1) - B_\theta(n)$, and the Klette–Kortum model as the linear case—is precisely the structure of Akcigit and Kerr (2018), Proposition 6; there the n-dependence arises from a non-scaling R&D technology, here from the count-contingent subsidy.

Aggregation with count-dependent policies. With $x_\theta(n)$ varying in n, the line shares μ_θ are no longer sufficient statistics: the stationary firm measure $M_\theta(n)$ (the mass of type-θ firms holding n lines) solves the flow-balance system

$$\begin{aligned}
\Big[n\Big(x_H(n) + \tau\Big) + \nu\Big] M_H(n) &= (n{-}1)\, x_H(n{-}1)\, M_H(n{-}1) + (n{+}1)\, \tau\, M_H(n{+}1) \\
&\quad + p_h z\, \mathbf{1}\{n{=}1\}, \\
n\Big(x_L(n) + \tau\Big) M_L(n) &= (n{-}1)\, x_L(n{-}1)\, M_L(n{-}1) + (n{+}1)\, \tau\, M_L(n{+}1) \\
&\quad + \nu\, M_H(n) + (1 - p_h) z\, \mathbf{1}\{n{=}1\},
\end{aligned} \tag{A-22}$$

with exit at $n = 0$; the single-type case is the flow system of Akcigit and Kerr (2018), Section V, and absent the subsidy its solution is the Klette–Kortum logarithmic distribution $M(n) = (z/x)\,(x/\tau)^n/n$. Let $\ell_\theta(n) \equiv nM_\theta(n)$ denote line mass. Two identities keep aggregation exact despite the Pareto count tail of the supercritical high-type phase, whose summand $\ell_H(n) \sim n^{-\nu/(x_H - \tau)}$ decays with an exponent barely above one at the estimates, making direct n-summation of line mass hopeless. First, the *scalar line balances* are exact: a line is captured at rate τ regardless of its holder and becomes H-held with probability τ_H/τ, so in steady state

$$\mu_H\,(\tau + \nu) = \tau_H, \qquad \mu_L\,\tau = \tau_L + \nu\,\mu_H, \qquad \tau_\theta \equiv \sum_n \ell_\theta(n)\, x_\theta(n) + p_\theta\, z, \tag{A-23}$$

with $p_H = p_h$, $p_L = 1 - p_h$; at constant x_θ this is exactly (15). Second, since $x_\theta(n) = x_\theta^\infty$ up to deviations localized near $\bar{n}$, every aggregate splits into a scalar far-field part and a rapidly converging truncated sum,

$$\sum_{n} \ell_\theta(n)\, x_\theta(n) = x_\theta^\infty\, \mu_\theta + \sum_{n \le N_v} \ell_\theta(n) \Big(x_\theta(n) - x_\theta^\infty \Big),$$

and likewise for the R&D labor; the subsidy outlay itself needs only $n \le \bar{n}$,

$$\frac{\text{Outlay}}{Y} = \varsigma\, \omega \sum_{\theta} \sum_{n \le \bar{n}} \ell_\theta(n)\, \frac{x_\theta(n)^2}{\psi_x}. \qquad \text{(A-24)}$$

The outer system is (A-17) with sums over n replacing integrals over a and free entry evaluated at the one-line constants, $\omega/\psi_z = p_h[A\Gamma_H + B_H(1)] + (1-p_h)[A\Gamma_L + B_L(1)]$. The count rule is not a paper exhibit—its pools cannot reach the statutory sales shares, since one-line firms alone are roughly half of all firms—but it is solvable *exactly*, and Section F.3 shows that the size-eligibility solver nests it without approximation error, which makes it the benchmark of the validation battery.

F.3 Size eligibility ($S_f \le \bar{s}\, Y$): the (θ, n, Q) value function

Firm sales are $S_f = \frac{Y}{M_{\varepsilon-1}} Q_f$ with $Q_f \equiv \sum_j \hat{q}_j^{\,\varepsilon-1}$, so "eligible iff $S_f \le \bar{S} = \bar{s}\, Y$" is $Q_f \le \bar{Q} \equiv \bar{s}\, M_{\varepsilon-1}$. The statutory parameter is $\bar{s}$ (a share of output, the scale-free way to state a size statute on a BGP), held fixed across policy equilibria at the value that reproduces the statutory ceiling's eligible sales share in the baseline (Section 6), while $\bar{Q}$ moves with $M_{\varepsilon-1}$ in general equilibrium and the realized pool share is an outcome. Reduce the firm state to (θ, n, Q) and write

$$V_\theta(\hat{q}_1, \ldots, \hat{q}_n) = Y\Big[A\, Q + W_\theta(n, Q)\Big], \qquad W_\theta(0, \cdot) = 0, \qquad \text{(A-25)}$$

with the same slope A. This firm-level value—a common quality slope plus a firm-level constant $W_\theta(n, Q)$—again has the structure of Akcigit and Kerr (2018), Proposition 6, the firm-level dependence now arising from the size-contingent subsidy. The A-block is *exact with no composition assumption*: profits and relative-quality decay are linear in Q, and a creative-destruction hit removes a uniformly random line, so, writing $c_j \equiv \hat{q}_j^{\,\varepsilon-1}$, the A-part of the loss term in (A-1) is

$$\tau \sum_{j=1}^{n} \Big[A\,(Q - c_j) - A\, Q \Big] = -\tau A \sum_{j=1}^{n} c_j = -\tau A\, Q,$$

composition-free. The only place the portfolio composition enters is the W-part of the loss term, $n\tau\Big[\hat{\mathbb{E}}_j W_\theta(n{-}1, Q - c_j) - W_\theta(n,Q)\Big]$, with $\hat{\mathbb{E}}_j$ the uniform draw over the firm's own lines. Replacing the lost line's contribution by its conditional mean, $c = Q/n$ (a mean-field closure), gives

$$\begin{aligned}
\rho\, W_L(n,Q) &= -(\varepsilon{-}1) g\, Q\, \partial_Q W_L + n\tau\Big[W_L(n{-}1, \tfrac{n-1}{n}Q) - W_L(n,Q)\Big] \\
&\quad + n\Phi_L(n,Q), \\
(\rho+\nu) W_H(n,Q) &= -(\varepsilon{-}1) g\, Q\, \partial_Q W_H + n\tau\Big[W_H(n{-}1, \tfrac{n-1}{n}Q) - W_H(n,Q)\Big] \\
&\quad + n\Phi_H(n,Q) + \nu W_L(n,Q),
\end{aligned} \tag{A-26}$$

with

$$\Phi_\theta(n,Q) = \frac{\psi_x\, G_\theta(n,Q)^2}{4\omega\Big(1 - \varsigma\, \mathbf{1}\{Q \le \bar{Q}\}\Big)}, \qquad x_\theta(n,Q) = \frac{\psi_x\, G_\theta(n,Q)}{2\omega\Big(1 - \varsigma\, \mathbf{1}\{Q \le \bar{Q}\}\Big)}, \tag{A-27}$$

$$G_\theta(n,Q) = A\,\Gamma_\theta + \mathbb{E}_{\hat{q}'\sim F}\Big[W_\theta\Big(n{+}1,\ Q + (\hat{q}' + \lambda_\theta)^{\varepsilon-1}\Big)\Big] - W_\theta(n,Q). \tag{A-28}$$

Lemma A-2 (Where the closure is exact)**.** *The mean-field closure in* (A-26) *is exact (i) at $n = 1$ and (ii) wherever $W_\theta(n-1,\cdot)$ is affine in Q. Consequently, (iii) at $\varsigma = 0$, under a uniform subsidy ($\bar{Q} = \infty$), and under the count rule ($\mathbf{1}\{n \le \bar{n}\}$ replacing $\mathbf{1}\{Q \le \bar{Q}\}$ in* (A-27)*), the solution of* (A-26) *is exactly $n\, b_\theta$, $n\, b_\theta(\varsigma)$ from* (A-21)*, and $B_\theta(n)$ from* (A-19)*, respectively—all Q-independent and free of closure error.*

Proof. (i) At $n = 1$ the lost line is the only line and the firm exits: $\hat{\mathbb{E}}_j W_\theta(0, Q - c_j) = 0 = W_\theta(0,\cdot)$, no composition is involved. (ii) If $W_\theta(n{-}1, Q) = \alpha + \beta Q$, then $\hat{\mathbb{E}}_j W_\theta(n{-}1, Q - c_j) = \alpha + \beta(Q - \hat{\mathbb{E}}_j c_j) = W_\theta(n{-}1, \tfrac{n-1}{n}Q)$, since $\hat{\mathbb{E}}_j c_j = Q/n$: the closure replaces an expectation that the affine function passes through exactly. (iii) Under the three stated rules the subsidy indicator is Q-independent, so (A-26) admits Q-independent solutions $W_\theta(n,Q) = B_\theta(n)$: the drift term vanishes, the loss term becomes $n\tau[B_\theta(n{-}1) - B_\theta(n)]$, and (A-26) reduces to (A-19) (with $\varsigma_n \equiv 0$, $\varsigma_n \equiv \varsigma$, or $\varsigma_n = \varsigma \mathbf{1}\{n \le \bar{n}\}$), whose solutions are the stated ones—and unique in the far-field class specified below, by the standard contraction argument at discount $\rho > 0$; by (ii) these Q-independent (hence affine) solutions involve no closure error. $\square$

The closure error is therefore second order—the curvature of W in Q times the within-firm dispersion of c—and confined to the neighborhood of $\bar{Q}$: far below and far above the threshold the value is again affine in Q, so deep-eligible and deep-ineligible firms are both closure-error free. Crucially, the closure touches *only* the policy function; the firm measure is simulated with full portfolios, exactly (Section F.4).

Far fields and numerics. As $Q \to \infty$ at fixed n, $W_\theta(n, Q) \to n\, b_\theta$: an ineligible firm re-enters eligibility through relative-quality erosion at rate $(\varepsilon - 1)g$ (plus line loss), so the deviation decays like

$$W_\theta(n, Q) - n\, b_\theta = O\Big(\big(\bar{Q}/Q\big)^{\rho/((\varepsilon-1)g)}\Big),$$

with exponent ≈ 0.62 at the headline estimates—the size-design analogue of the count design's shrink-back exponent $\rho/(\tau - x_L) \approx 0.85$; in n the far-field increment is the per-line constant, $W_\theta(N_v{+}1, \cdot) = W_\theta(N_v, \cdot) + b_\theta$. Equation (A-26) is solved on a uniform $\ln Q$ grid anchored so that $\ln \bar{Q}$ is a node (the kink sits on the grid; the constant-coefficient drift is discretized upwind) and $n = 1, \ldots, N_v$ with the far-field boundary, by Gauss–Seidel sweeps of alternating direction in n with the policy updated each sweep; the capture expectation in (A-28) uses an equal-mass quadrature compression of a fixed common-random-number sample from F.

F.4 Computing the subsidized balanced growth path

With $x_\theta(n, Q)$ non-constant, no low-dimensional flow system is available—the state includes Q and, for the true measure, the whole portfolio composition. The stationary firm measure is therefore computed by exact event-time simulation of full portfolios (the state-dependent policy evaluated by interpolation of (A-27)), kept separate from the simulation used in estimation; each cross-section record i carries weight z/N, so that the simulated measure integrates to the firm mass. Aggregates use the two exact devices of Section F.2: the scalar line balances (A-23), with the innovation flows split into a far-field part and record-level deviation sums localized near $\bar{Q}$,

$$\tau_\theta = x_\theta^\infty\, \mu_\theta + \hat{\Sigma}_\theta + p_\theta\, z, \qquad \hat{\Sigma}_\theta = \frac{z}{N} \sum_{i:\, \theta_i = \theta} n_i \Big(x_\theta(n_i, Q_i) - x_\theta^\infty\Big), \tag{A-29}$$

and the analogous deviation sums for R&D labor. The subsidy outlay sums over the eligible records only,

$$\frac{\text{Outlay}}{Y} = \varsigma\,\omega\,\frac{z}{N}\sum_{i:\,Q_i\le\bar{Q}} n_i\,\frac{x_{\theta_i}(n_i,Q_i)^2}{\psi_x}, \tag{A-30}$$

and free entry capitalizes eligibility through the one-line value at the entrant's own draw,

$$\frac{\omega}{\psi_z} = \sum_{\theta} p_\theta\Big(A\,\Gamma_\theta + \mathbb{E}_{\hat{q}'\sim F}\Big[W_\theta\Big(1,(\hat{q}'+\lambda_\theta)^{\varepsilon-1}\Big)\Big]\Big). \tag{A-31}$$

The outer system is (A-17) with these objects (three equations in Model A), solved as a short fixed point in rounds: given records simulated under the previous round's aggregates, the outer system is solved with the records *frozen*—the residuals are then smooth in (τ,h,L_P,z)—and the equilibrium is re-simulated under the solved policy; two to four rounds suffice. All seeds are fixed (common random numbers), so the fixed point is deterministic; at $\varsigma=0$ the policy is constant and every deviation sum vanishes identically, so the $\varsigma\to 0^+$ marginals carry relative, not additive, Monte Carlo error.

F.5 Validation and value finiteness

The solver passes four checks. (i) At $\varsigma=0$, the grid solution equals $n\,b_\theta$ to solver tolerance and the outer residuals vanish at the baseline balanced growth path of (A-4). (ii) In the uniform limit $\bar{Q}=\infty$ it reproduces the uniform-subsidy equilibrium of (A-13). (iii) Run with the count rule $\mathbf{1}\{n\le\bar{n}\}$, it reproduces the exact count-eligibility equilibrium of Section F.2—value function, policy, and the full subsidized balanced growth path, including the simulated count marginal against $M_\theta(n)$ from (A-22); by Lemma A-2 this comparison is free of closure error, which is what makes it a genuine test of the (θ,n,Q) machinery. (iv) Grid, quadrature, and cross-section-size refinement at the headline point.

Remark A-2 (Transversality and value finiteness)**.** *The additive valuation solves the Bellman equation; it is the firm's value provided the transversality condition* $\lim_{T\to\infty} e^{-\rho T}\,\mathbb{E}[V(\text{state}_T)]=0$ *holds. Every term of* V *is bounded by a multiple of the product count plus a decaying transient: each newly captured line enters with expected quality value at most* $A\Gamma_H$*, the quality value of the lines already held only erodes there-*

after, and the per-line constants are bounded by B_H. The expected count of a high-type firm grows at rate $x_H - \tau$ only while the high phase lasts—an $\mathrm{Exp}(\nu)$ *time—after which the low-type continuation is subcritical ($x_L < \tau$). Hence $e^{-\rho T}\,\mathbb{E}[n_T] \to 0$ if and only if*

$$x_H - \tau < \rho + \nu.$$

At the headline estimates $x_H - \tau = 0.15 < \nu = 0.18 < \rho + \nu = 0.23$: the expected portfolio of a high-type firm is bounded even without discounting, and firm values, the systems (A-14), (A-19), *and* (A-26), *and the welfare bookkeeping $\mathcal{W} = \ln C_0/\rho + g/\rho^2$ (eq. 21) are all well-defined. The condition is re-checked at every subsidized equilibrium, since the subsidy moves x_H and τ in general equilibrium.*

G The transition after the reform

The economy sits on the baseline balanced growth path when, at $t = 0$, the subsidy is introduced unexpectedly and permanently at the rate ς^* of Table 6; agents have perfect foresight and the economy converges to the subsidized balanced growth path. Holding the *rate* fixed along the path—rather than the outlay—makes the terminal point exactly the subsidized balanced growth path of Table 6, at the cost that the fiscal outlay is below its long-run 1% of output early on, when the eligible pool has not yet expanded.

The fixed point. Normalize firm values by output. Under log utility with $C = Y$—R&D and entry use labor and the subsidy is a lump-sum transfer, so consumption is output at every date—the Euler equation gives $r(t) = \rho + g(t)$ pointwise, so the effective discount rate on normalized values is exactly ρ along the whole path and no interest-rate or wage-growth path has to be iterated on. The relative-quality distribution enters only through its moments: off the balanced growth path the generator computation behind (14) delivers, given the creative-destruction rate $\tau(t)$ and the high-type innovation share $h(t)$, the closed triangular ODE system

$$\dot{M}_k(t) = -k\, g(t)\, M_k(t) + \tau(t) \sum_{i=0}^{k-1} \binom{k}{i} M_i(t)\, m_{k-i}(t), \qquad g(t) = \tau(t)\, m_1(t), \quad \text{(A-32)}$$

with $m_k(t) = h(t)\lambda_H^k + (1 - h(t))\lambda_L^k$; its stationary points are exactly (14). The transition is therefore a fixed point in three scalar paths, $\big(\tau(t), h(t), \omega(t)\big)$ with $\omega = w/Y$. One sweep integrates (A-32) forward; the value functions backward from the new balanced growth path at the horizon $T = 300$; the policies from the values; and the firm measure forward from the baseline distribution, with the entry rate $z(t)$ pinned pointwise by labor-market clearing—linear in z at each date—and $\omega(t)$ by free entry. The sweep is damped adaptively: the response of eligible firms' values to the wage path has a loop gain above one at statutory subsidy rates (the eligible high-type strip's effective discount rate—$\rho + \tau + \nu$ less its subsidized expansion intensity, approximately $\rho + \tau + \nu - x_H/(1 - \varsigma)$ at the baseline intensity x_H—is close to zero), and fixed damping diverges.

The age screen. Additivity is preserved, and the transition inherits the exact structure of (A-14)–(A-16) with time-varying coefficients: the per-line constants $B_\theta(a,t)$ solve the backward first-order system

$$\begin{aligned} \Big(\rho+\tau(t)\Big)B_L(a,t) &= \Phi_L(a,t)+\Big(\partial_a+\partial_t\Big)B_L(a,t), \\ \Big(\rho+\tau(t)+\nu\Big)B_H(a,t) &= \Phi_H(a,t)+\nu B_L(a,t)+\Big(\partial_a+\partial_t\Big)B_H(a,t), \end{aligned} \tag{A-33}$$

integrated along the cohort characteristics $da = dt$ (so the grid diagonals are exact), with boundary value $b_\theta(t)$ at $a = \bar{a}$—the ineligible far field, which solves the a-independent version of (A-33) backward in t—and terminal condition the new balanced growth path at T. The cohort line masses solve the forward transport equations

$$\Big(\partial_t+\partial_a\Big)\ell_H = \Big(x_H(a,t)-\tau(t)-\nu\Big)\ell_H, \qquad \Big(\partial_t+\partial_a\Big)\ell_L = \Big(x_L(a,t)-\tau(t)\Big)\ell_L+\nu\,\ell_H, \tag{A-34}$$

with $\ell_H(0,t) = p_h z(t)$, $\ell_L(0,t) = (1-p_h)z(t)$, and closed-form tails beyond the cutoff.

The size screen. The value $W_\theta(n,Q,t)$ solves the time-dependent form of (A-26), with ρ replaced by $\rho-\partial_t$ on the left side, the aggregates $\tau(t), g(t), \omega(t), A(t)$ dated, and the eligibility threshold $\bar{Q}_t = \bar{s}\,M_{\varepsilon-1}(t)$ moving across the grid in general equilibrium. It is stepped backward by implicit (backward-Euler) steps that reuse the stationary solver's Gauss–Seidel sweeps with $\rho \mapsto \rho + 1/\Delta t$ and source $W(\cdot, t+\Delta t)/\Delta t$, so a stationary value is an exact fixed point of the stepper at any step size and the scheme is L-stable. The firm measure is a weighted cohort panel: firms evolve *independently* against the guessed aggregate paths—the loss hazard $\tau(t)$, capture draws from F at composition $h(t)$, quality decay at $g(t)$—with entering cohorts weighted by $z(t)$. The capture draws are quasi-static: the stationary F at the date-t composition stands in for the slowly moving line-quality distribution, and is exact at both endpoints; since the value side uses the exact moments (A-32), the approximation touches only the assembly of portfolios for eligibility classification. Aggregates never sum over the size tail: they use the exact scalar line balances (A-23), now as ODEs in t, together with deviation sums (A-29) localized near the threshold, where the policy departs from its far field. Each reported path is differenced against a $\varsigma = 0$ null run on the same seeds, which removes the simulation's finite-sample offset.

Validation. Both endpoints are re-solved *on the transition grid*, so they are exact stationary points of the discretized dynamics: the $\varsigma = 0$ null path is flat to machine precision, and the subsidized path lands on the new balanced growth path to eight digits at $T = 300$. The two independent kernels agree where they overlap: for a uniform subsidy—eligibility everywhere, which both designs nest—the size-screen machinery (backward $W_\theta(n, Q, t)$ and the simulated panel) and the exact age kernel (backward ODEs, no simulation) deliver the same growth path to less than 0.001 basis points at every date and the same consumption-equivalent welfare to five decimals.

H The Peters (2020) model

The Peters (2020) model is a state-of-the-art quality-ladder model of firm dynamics with endogenous markups. Firms conduct *expansion* R&D, adding product lines by displacing incumbents, and *internal* R&D, improving their own lines and thereby widening quality gaps and raising markups; new firms enter with a single line and exit when they lose their last one. The final good is Cobb–Douglas over product lines, so expenditure is equalized across lines and a firm's sales are proportional to its product count, $s_f = n_f Y$. The model therefore coincides with the single-type Model A of Section 5 in the Cobb–Douglas limit $\varepsilon \to 1$, with one addition: internal R&D and the endogenous markups under limit pricing it generates. Under Cobb–Douglas, firm sales are proportional to the product count regardless of markups, and internal R&D leaves the product count unchanged, so the firm-size decomposition depends only on the expansion rate x and the creative-destruction rate τ—exactly as in Model A at $\varepsilon = 1$. The internal-R&D and markup margins distinguish Peters (2020) from a bare count model in its *profit* and *markup* dynamics—disciplined here by the markup–age target—but under Cobb–Douglas they leave firm size, and hence the size decomposition, unchanged. The three margins are therefore computed from the count process as in the single-type, Cobb–Douglas case of Appendix C.

Estimation. I estimate the four parameters $(\psi_x, \psi_I, \psi_z, \lambda)$—the expansion, internal, and entry R&D efficiencies and the quality step size—holding $\rho = 0.05$ and $\zeta = 2$ fixed, to match four equally weighted targets: the size–age profile and the firm-age distribution over integer ages 0–24 and the pooled 25+ group (the two objects that pin down the between-age component), aggregate TFP growth, and the markup–age profile. The size–age profile and age distribution are the 2017 Swedish cross-section (Figures 2a and 2b); following Peters (2020), the growth and markup targets are set to $g = 3\%$ and an age-zero-to-seven markup growth of 8.2%, which the internal-R&D margin and the growth equation deliver. The loss is minimized by a global Sobol search followed by local Nelder–Mead refinement, converging to the parameters in Table A-1. As is standard for this class of models, size dispersion conditional on age is *not* among the targets; the model's dispersion by age is therefore an untargeted prediction, and how far it undershoots the data is the point of the exercise.

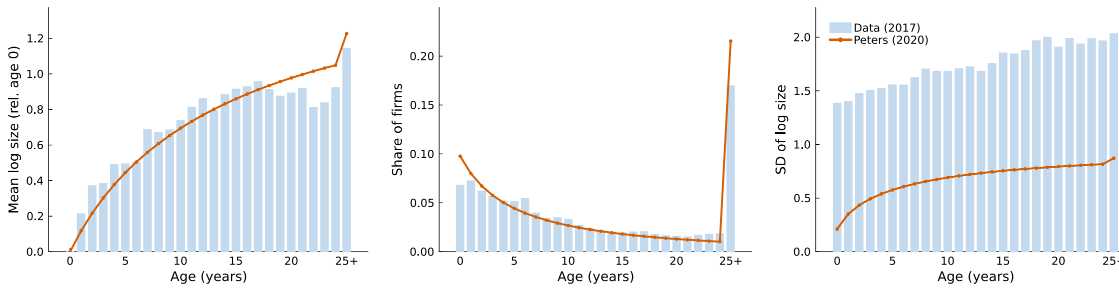
Fit and the between-age share. Table A-1 reports the estimates and Figure A-2 the fit. The model reproduces the mean size–age profile and the age distribution well and matches aggregate growth and the markup–age slope exactly. Applying the decomposition of eq. (1) with the same integer-age grouping as the data, the estimated model attributes 28.5% of the variance of log size to the between-age component, against 3.6% in the data—an overstatement by a factor of about eight—because it generates far too little size dispersion conditional on age (Figure A-2, right panel): its standard deviation of log sales conditional on age is about one-seventh of the data's among entrants, rising to two-fifths among the oldest firms, and its overall standard deviation of log size is 0.77 against 1.75 in the data. This is the Cobb–Douglas ($\varepsilon = 1$) rung of the ladder in Section 5 (a between-age share of 28.8%), confirmed here in an independent, state-of-the-art model.

Table A-1: Estimation of the Peters (2020) model

	Data	Model
Targeted moments		
Firm size–age profile	Figure A-2	
Firm-age distribution	Figure A-2	
TFP growth g in %	3.0	3.0
Markup-age profile, age 0 to 7, in %	8.2	8.2
Estimated parameters		
ψ_x *Expansion R&D efficiency*		1.004
ψ_I *Internal R&D efficiency*		5.522
ψ_z *Entry R&D efficiency*		2.487
λ *Step size of quality improvements*		1.046
Set exogenously		
ρ *Discount rate*		0.05
ζ *R&D cost curvature*		2

Notes: The table reports the targets and the estimated parameters of the Peters (2020) model. The four parameters $(\psi_x, \psi_I, \psi_z, \lambda)$ are estimated to match four equally weighted targets: the firm size–age profile and the firm-age distribution at integer ages 0–24 and the pooled 25+ group in the 2017 Swedish cross-section (Figures 2a and 2b), aggregate TFP growth, and the growth of the firm markup between ages zero and seven. The growth and markup targets follow Peters (2020). The discount rate ρ and the R&D cost curvature ζ are set outside the model. Supplemental Appendix H describes the estimation.

Figure A-2: Estimated Peters (2020) model: fit to the three margins



Notes: The figure compares the estimated Peters (2020) model (line) with the 2017 Swedish cross-section (bars). *Left:* mean log sales by age, relative to age 0 (targeted). *Middle:* the firm-age distribution (targeted). *Right:* the standard deviation of log sales by age, which is *not* targeted and shows the model's structural shortfall in size dispersion conditional on age. Integer ages 0–24 with firms aged 25 and above pooled.

I The cost-heterogeneity variant B^{cost}

The sign of Δg in the policy analysis of Section 6 rests on types entering the innovation *step*. Growth is $g = \tau\bar{\lambda}$, and the sign flip is carried by the composition term $\tau\, d\bar{\lambda}$ of (20): reallocating innovation toward low types lowers the mean step. This channel is special to step heterogeneity. Suppose the types differed instead in the *cost* of innovation at a common step—high types innovate faster per unit of R&D—as in Acemoglu et al. (2018). Then the mean step $\bar{\lambda}$ is constant and the term $\tau\, d\bar{\lambda}$ vanishes. Composition does not vanish with it: reallocating the economy's product lines across types still moves the average rate of innovation τ, exactly as in Model B, where the types already expand at different endogenous rates $x_H > x_L$. What the step form adds is a *second* composition channel, through the mean step, and it is that channel the paper's sign flip runs on. The two forms are otherwise close to observationally equivalent on the moments a firm-dynamics model usually targets—both deliver $x_H > x_L$ and hence size–age–growth correlations, the selection of low types into the small-given-age pool, and a Pareto size tail—so the choice cannot be settled by assertion. This appendix estimates the cost-heterogeneity counterpart of Model B on *exactly* Model B's moments and compares the fit; it then repeats the statutory policy experiments of Section 6 in the estimated variant (the results appear as the Model B^{cost} rows of Table 6). The step form is the closer to the data and the cost form fails most clearly on the level of size dispersion at entry and among young firms.

Specification. Only the innovation technology changes relative to Section 4. Every innovator, of either type, takes the *same* quality step λ; the type instead scales the expansion-R&D cost,

$$c_x^\theta(x) = \frac{x^\zeta}{\psi_{x,\theta}}, \qquad \psi_{x,H} > \psi_{x,L}, \tag{A-35}$$

so high types innovate more cheaply and therefore faster. Everything else is unchanged: entry at rate z with a high type drawn with probability p_h, one-way reversion $H \to L$ at rate ν (now a rise in R&D cost rather than a fall in step size), CES demand, and the labor market. At the cost curvature $\zeta = 2$, this cost-side heterogeneity is exactly the arrival-capacity type of Acemoglu et al. (2018): the exogenous heterogeneity raises the rate at which a firm innovates per unit of R&D, whereas the rate itself, x_θ, is endogenous in both forms. The conjecture $v_\theta(\hat{q}) = A\hat{q}^{\varepsilon-1} + B_\theta$ verifies with the same common slope (16); the capture value loses its type premium, $\Gamma_H = \Gamma_L = \Gamma \equiv \mathbb{E}_F[(\hat{q}' + \lambda)^{\varepsilon-1}]$, so $G_\theta = A\Gamma + B_\theta$ and the type gap in capture values is the franchise gap $B_H - B_L$ alone. At $\zeta = 2$ the policies and constants become

$$x_\theta = \frac{\psi_{x,\theta}\, G_\theta}{2\omega}, \qquad (\rho + \tau)\, B_L = \omega\, \frac{x_L^2}{\psi_{x,L}}, \qquad (\rho + \tau + \nu)\, B_H = \omega\, \frac{x_H^2}{\psi_{x,H}} + \nu B_L, \tag{A-36}$$

and the balanced growth path is the eight-equation system (A-4) under the substitution $\psi_x \to \psi_{x,\theta}$ in the R&D terms, with Model A nested at $\psi_{x,H} = \psi_{x,L}$. The type ordering survives the loss of the capture premium: subtracting the two constants in (A-36) as in Lemma A-1 gives

$$\Big(\rho + \tau + \nu - \tfrac{1}{2}\big(x_H + (\psi_{x,H}/\psi_{x,L})\, x_L\big)\Big)(B_H - B_L) = \frac{(\psi_{x,H} - \psi_{x,L})\, G_L^2}{4\omega} > 0,$$

and the bracket is positive with a wide margin at the estimates (0.33 against 0.46), so $B_H > B_L$, $G_H > G_L$, and $x_H/x_L = (\psi_{x,H}/\psi_{x,L})(G_H/G_L) > \psi_{x,H}/\psi_{x,L}$: the cost advantage is amplified by the franchise gap it creates. Two consequences are structural. First, with a common step,

$$g = \tau\,\lambda \qquad \Longrightarrow \qquad dg = \lambda\, d\tau : \tag{A-37}$$

the step composition channel of (20) is identically zero; composition through the average innovation rate τ remains, as in Model B. Second, with $m_k = \lambda^k$ and $\tau/g =$

$1/\lambda$, the moment recursion (14) collapses to

$$M_k = \frac{1}{k\lambda} \sum_{i=0}^{k-1} \binom{k}{i} M_i \, \lambda^{k-i}, \tag{A-38}$$

a function of λ alone—identical to Model A's at the same λ: type heterogeneity leaves no footprint whatsoever in the quality distribution, because every innovator adds the same step. (Contrast the step model, where $m_k > m_1^k$ for $k > 1$ by Jensen's inequality: at fixed growth, a mean-preserving spread of steps fattens every higher moment of F. That channel is unavailable here by construction.) In particular, an entrant of either type holds one line of quality $\hat{q}' + \lambda$, $\hat{q}' \sim F$, so

$$\mathrm{Var}(\ln S \mid a = 0) = (\varepsilon - 1)^2 \, \mathrm{Var}\Big(\ln(\hat{q}' + \lambda)\Big), \tag{A-39}$$

a function of λ alone (by (A-38), F depends only on λ): the between-type term of (A-5) vanishes, there is no ex-ante dispersion at entry, and every difference between the types works through the endogenous rate $x_H > x_L$, which produces no dispersion until firms have had time to accumulate lines at divergent rates.

Estimation. I estimate the six parameters $\{\psi_{x,L}, \psi_{x,H}, \psi_z, \lambda, \nu, p_h\}$—the same count as Model B—following *exactly* the Model B procedure: the same four targets (the mean size–age profile, the firm-age distribution, the standard deviation of log size by age, and $g = 2\%$), the same weights, the same parameter bounds, and the same Sobol-plus-multistart routine of Appendix E. The cohort simulator is applied verbatim—it already runs with $x_H \neq x_L$, and the cost variant is the special case of a common step—so only the steady-state solver changes. Table A-2 reports the estimates alongside Model B's.

The estimated configuration. The estimates themselves are diagnostic. With the type leaving no mark on size at entry, the estimator compensates on the margins it retains. It stretches the expansion gap (taken up below) close to the edge of the admissible region: the estimate runs to 99% of the supercriticality fence $x_H - \tau \leq 0.95\nu$, imposed identically in both estimations to keep the Pareto tail exponent above one, with $x_H - \tau = 0.166$ against the bound 0.167—where Model B's estimate lies farther inside, at 0.153. Because the routine, targets, weights, and bounds are exactly

Table A-2: The cost variant B^{cost} against the step model

	Model B (step)	Model B^{cost} (cost)	Data
Estimated parameters			
Expansion efficiency, low type $\psi_{x,L}$	0.455	0.265	
Expansion efficiency, high type $\psi_{x,H}$		0.520	
Entry efficiency ψ_z	0.837	0.678	
Step size, low type λ_L	0.025	0.089	
Step size, high type λ_H	0.175		
Switching rate ν	0.176	0.176	
High-type entrant share p_h	0.351	0.880	
Implied balanced growth path			
Expansion rate x_H/x_L	0.409 / 0.197	0.400 / 0.135	
Entry rate z	0.014	0.004	
Creative destruction τ	0.256	0.234	
TFP growth g	0.020	0.021	0.020
Size dispersion (SD of $\ln S$*)*			
Overall	1.45	1.41	1.75
Entrants (a=0)	1.09	0.83	1.39
Oldest (a=25+)	1.61	1.77	2.03
Decomposition of $\text{Var}(\ln S)$			
Between-age share	5.9%	4.2%	3.6%
Within-age share	94.1%	95.8%	96.4%

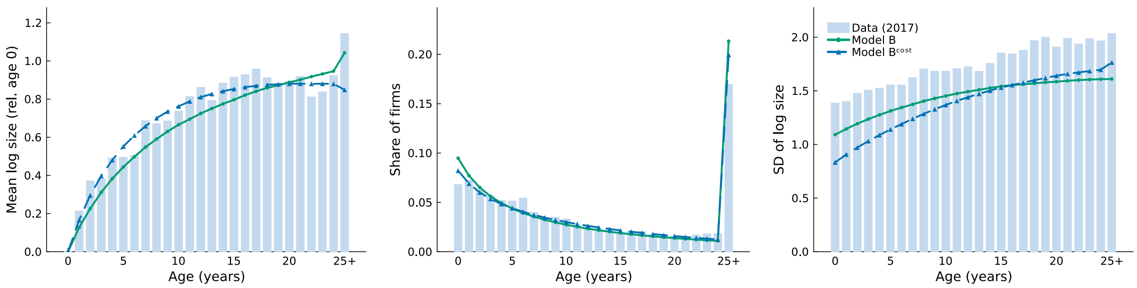
Notes: Both models are estimated to the 2017 Swedish cross-section by the procedure of Appendix E, at $\varepsilon = 5$. A blank means the two types share that parameter: the heterogeneity is in the step ($\lambda_H > \lambda_L$, at common cost) in Model B, and in the cost ($\psi_{x,H} > \psi_{x,L}$, at common step) in B^{cost}. The dispersion summary and the between/within decomposition are as in Table 4.

Model B's, the near-binding fence is not an artifact of a tighter search but part of the diagnosis: the one channel the cost form retains is pushed nearly to the stationarity boundary in place of the at-entry dispersion it cannot produce. And—more telling—it inverts the entrant composition, $p_h = 0.88$ against Model B's 0.35, while the entry rate falls to $z = 0.004$ against 0.014. Nearly nine in ten entrants are now "high types," and the high-type share decays from 81% at age zero to 20% by age ten and 6% by age twenty. This configuration gives up the gazelle reading of the type: a trait carried by almost every entrant and shed with age is not a rare ex-ante difference between firms but a phase that nearly all firms pass through—ex-ante heterogeneity in name, a life-cycle stage in substance. The one parameter on which the two forms agree exactly is the reversion rate, $\nu = 0.176$ in both (identical to four digits): ν is evidently identified by how quickly the type footprint fades with age—the slowing rise of the dispersion schedule, the thinning of fast growth among old large firms—a feature of the data indifferent to which margin the type enters.

Fit to the three targets. Figure A-3 sets the two models against the data on each targeted schedule. On the mean size–age profile (left panel) and the age distribution (middle) the two are comparable, with one telling exception: the cost form runs *too flat* at the top of the size–age profile, reaching only mean log size 0.85 at 25+ against the data's 1.14 and Model B's 1.04. That flatness is why the cost form posts a between-age share, 4.2%, closer to the data's 3.6% than Model B's 5.9% (Table A-2)—an apparent advantage that is a false positive. The between-age component is the variance of the mean size–age profile across ages; a profile that is too flat mechanically carries too little of it. The cost form's lower between-age share is bought with a *worse* fit to the mean profile, not a better fit to either component of the decomposition, and its overall dispersion, 1.41, is if anything slightly farther from the data's 1.75 than Model B's 1.45.

Where the cost form fails: dispersion at entry. The right panel of Figure A-3—size dispersion by age—is where the two forms part. Model B undershoots the dispersion schedule by a roughly constant factor—about 0.8 times the data at every age—because the step gives its types dispersion immediately: high-type entrants take larger quality steps, so size is dispersed by type from birth. The cost form has no such channel. At entry its types are invisible, so its dispersion there, 0.83, barely

Figure A-3: Fit to the three targeted schedules: the step model against the cost variant

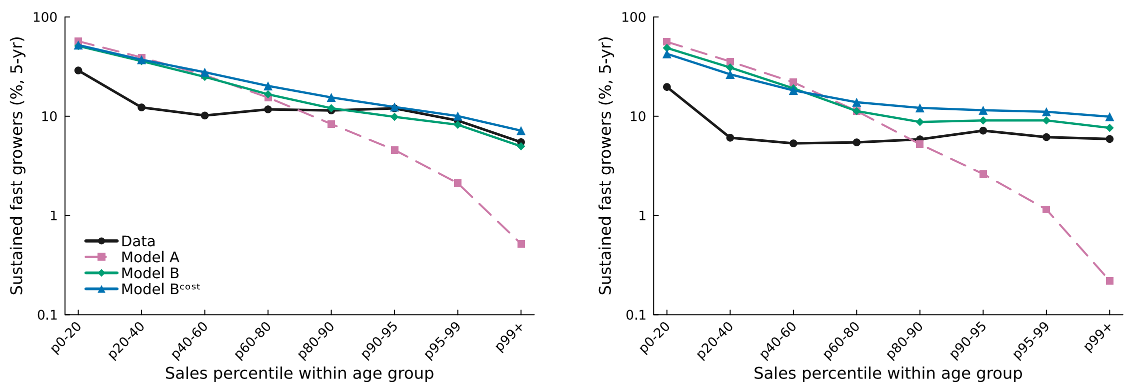


Notes: The three targeted schedules in the 2017 Swedish cross-section (bars) against the estimated step (Model B) and cost (B^{cost}) models at $\varepsilon = 5$. *Left:* mean log sales by age, relative to age 0. *Middle:* the firm-age distribution. *Right:* the standard deviation of log sales by age. Integer ages 0–24 with firms aged 25 and above pooled; the estimation uses every integer age. As in Figure 6, the two models track the mean profile and the age distribution similarly; they part on the dispersion schedule (right panel), where the cost variant starts too low at entry and rises too steeply, crossing the step model among older firms.

exceeds the no-type model's 0.71 and falls far short of Model B's 1.09 and the data's 1.39; the ex-ante share of entry dispersion is identically zero. The level is not an artifact of the estimation but an analytic corollary: with a common step, eq. (19) collapses to its single-type form (A-39), and dispersion at the instant of entry is $(\varepsilon - 1)\,\mathrm{SD}\Big(\ln(\hat{q}' + \lambda)\Big) = 0.79$—a function of λ alone (0.83 in the age-zero bin, the difference being within-bin aging). And λ is not free to fix it: with growth targeted at two percent, $\lambda = g/\tau$, and τ is disciplined by the targeted age distribution, so entry dispersion in the cost form is pinned by the aggregate targets, out of reach of the type parameters $(\psi_{x,H}/\psi_{x,L}, \nu, p_h)$. Its *only* source of dispersion is the divergence of the endogenous rates x_H versus x_L, which puts no dispersion at entry and accumulates only with age. Two things follow, both visible in the panel. The undershoot is worst precisely at entry (a ratio to the data of 0.60, against Model B's 0.79), and the schedule then rises *steeper* than the data: the ratio to the data climbs from 0.60 at entry to 0.87 at 25+, overshooting Model B among the oldest firms. To generate old-age dispersion from expansion alone, the estimator must widen the expansion gap—$x_H/x_L = 3.0$, against Model B's 2.1—and the wider gap is exactly what steepens the age slope. The step form fits better because it splits the work between two channels, an at-entry step and an accumulating expansion gap, and so needs a smaller expansion gap; the cost form must carry everything on the one channel that is silent at entry.

Untargeted growth. The wider expansion gap has a further symptom, in the untargeted growth moments (Figure A-4 and Table A-3), and again Model B is the closer to the data. Both type forms generate sustained fast growth among large firms, through the same transiently supercritical high-type phase—but the cost form *overshoots* it slightly: its five-year fast-grower share among the largest firms of each age runs above both Model B and the data at every age (Figure A-4, the cost line sitting above the step line at the top of the size distribution). One-year survivor growth is the same story: for firms large given their age, the cost form predicts *positive* mean growth at ages 7–19 (+3.2 and +3.4 percent) relative to the economy-wide average, where the data show mild decline and Model B, closer, predicts −2.8 and −1.8. The wider expansion gap the dispersion schedule forces on the cost form spills into these dynamic moments as too many fast growers and too little mean reversion at the top.

Figure A-4: Sustained fast growers by size, with the cost variant (untargeted)



Notes: As in Figure 7—the share of firms sustaining at least 20% annual growth over five years, by within-age sales percentile, for firms aged 1–2 (left) and 3–6 (right)—with the cost variant B^{cost} added. Data 2012–2017; models simulated at their $\varepsilon = 5$ estimates; vertical axes logarithmic. The cost line lies above the step line at the top percentiles: the cost form generates sustained fast growers as readily as the step form, and overshoots the data by more.

The policy experiments. The statutory experiments of Section 6 repeat in the estimated variant with the design held fixed: eligibility below the EU turnover ceilings, mapped by the eligible sales shares (19.6%, 42.2%, 71.0%) on the variant's own balanced-growth cross-section; age pools matched to the same sales footprint; the outlay fixed at 1% of output with the rate ς budget-exhausting. Two mappings shift relative to Model B because the variant's size distribution is more concentrated (its count-tail exponent is 1.06 against Model B's 1.15): the statutory pools are even

Table A-3: Mean and median growth of the largest firms (p95+), with the cost variant (untargeted)

Age bin	Model A	Model B	Model B^{cost}	Data
Mean one-year growth of survivors (percent)				
3–6	−19.0	−10.1	−1.6	−6.3
7–10	−14.1	−2.8	+3.2	−4.6
11–19	−11.1	−1.8	+3.4	−5.4
20+	−8.0	−6.7	−4.7	−4.1
Median one-year growth of survivors (percent)				
3–6	−12.5	−6.8	−0.6	−0.9
7–10	−10.7	−2.4	+3.6	−1.5
11–19	−9.3	−2.4	+4.3	−1.7
20+	−7.0	−7.3	−8.6	−1.9

Notes: As in Table 5, with the cost variant B^{cost} added. Firms in the 2016 registry cross-section (outcomes in 2017) are grouped by age and, within each age bin, by sales percentile; the table reports the mean and median change in log sales among survivors above the 95th size percentile within their age bin (p95+), relative to the economy-wide average. Model cells apply the identical grouping to simulated balanced-growth panels at the $\varepsilon = 5$ estimates. The five-year sustained fast-grower share is shown in Figure A-4.

broader in firm terms—89%, 99%, and essentially 100% of firms across the three ceilings—and the matched age cutoffs are older, 14, 29, and 48 years against Model B's 9, 22, and 41. The subsidized balanced growth paths are solved with the same machinery as Model B's: the size screens with the (θ, n, Q) value function of Appendix F generalized to type-specific costs $\psi_{x,\theta}$—the count-eligibility system (A-19), similarly generalized, validates the grid to the same tolerances as in Model B—and the age screens with the exact per-line system (A-14). All six designs appear as the Model B^{cost} rows of Table 6.

As discussed in the main text, the size screen fails as a screen in the cost form exactly as in the step form. At the statutory ceilings, the eligible pools carry 0.86–0.89 times the ineligible pool's mean innovation rate, with 46–50% of eligible firms at least ten years old; at narrow pools the adverse selection sharpens—the smallest five percent of firms by sales are 14% high types, at 0.81 times the ineligible mean rate, with 64% at least ten years old, against 84% high types among the youngest five percent. Therefore, the low types are disproportionately represented in the eligible pool—exactly as in the step form—and the form of the heterogeneity changes only the price of the size-screen failure.

What the exercise shows. Estimated on Model B's own moments, the cost-heterogeneity variant is farther from the data on every margin the paper examines: it undershoots size dispersion at entry and among young firms, matches the between-age share only by flattening the mean size–age profile it should be fitting, and overshoots the untargeted growth of the largest firms. The forms differ structurally at exactly the point where the cost form is weakest—at entry, where the step form endows firms with dispersion by type while the cost form defers all type differences to later expansion. None of this rejects the cost form outright, and the differences are matters of degree on a few margins rather than a clean discrimination between the two; the balance of them tilts toward step heterogeneity, and with it toward the composition mechanism behind the paper's policy result. The policy implications are accordingly best read in two tiers. The ranking of screens is robust across the forms: in either one the types differ in how fast they innovate, so small-firm eligibility steers the subsidy toward the low types when better-targeted designs are available—the size screen is a poor screen in the cost form too. What depends on the form is the stronger claim, that the misdirected subsidy lowers growth outright rather than merely buying less of it: that claim runs on the step channel, which the data tilt toward without settling. The sign flip is thus estimated rather than assumed—and the case against the size screen does not depend on it.