Icarus and Daedalus: Non-Gaussian Micro Shocks and Aggregate Productivity Risk

Ronit Mukherji
Ashoka University,
Department of Economics
ronit.mukherji@ashoka.edu.in

Alejandro Rojas-Bernal
University of Hawai'i at Manoa,
Department of Economics
rojasber@hawaii.edu

Current draft: September 15, 2026

Abstract

How does productivity evolve at the microeconomic level, and how does that evolution shape aggregate risk? Using restricted-use plant–product data from India's Annual Survey of Industries (2010–11 to 2022–23), we construct annual physical-output Solow-residual growth accounting for markups, inventories, and heterogeneous input costs. Productivity growth is sharply peaked and fat-tailed, with nonlinear mean reversion that varies with producer size. Larger producers recover more strongly after adverse shocks and retain more of their favorable gains—the *Daedalian level effect*—yet their shocks extend farther into both tails conditional on entry. Smaller producers exhibit sharper reversals after large favorable shocks—*Icarian fallout*. We capture these patterns with a three-regime Markov normal mixture conditioned on productivity history and size. Our law of motion yields the distribution of aggregate technology growth. Relative to a size-dependent Gaussian distribution, our regime model has lower aggregate dispersion but substantially thicker standardized tails and raises the probability of an aggregate technology contraction from 0.05% to 0.31%. Removing size dependence while retaining three regimes instead lowers expected growth and raises contraction probability to 2.47%. These separately estimated alternatives show that nonnormality and size dependence have distinct implications for aggregate risk. Granularity is distributional: aggregation weights determine whose shocks matter; productivity dynamics determine the risks they carry.

***JEL* Codes:** C46, D22, D24, D57, E23, E32, L11, L25, O47

Keywords: Firm productivity dynamics; aggregate productivity risk; non-Gaussian productivity shocks; production networks; granular fluctuations; firm-size heterogeneity; physical-output productivity.

1 Introduction

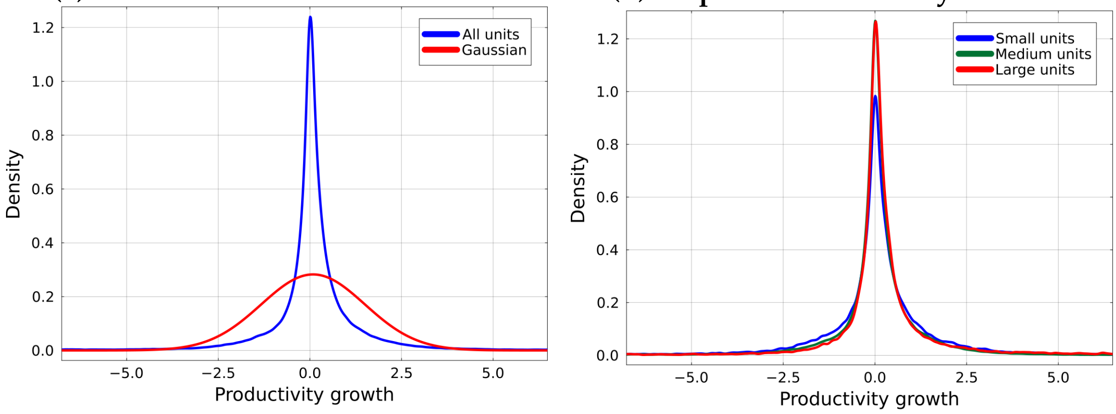
How does productivity evolve at the microeconomic level? How does that evolution translate into macroeconomic risk? These questions are tightly connected: the microeconomic law of motion governs productivity shocks and their evolution, while producer size and network centrality shape their aggregate influence. Together, these forces determine the distribution of aggregate growth. A common assumption is that firm productivity follows a linear Gaussian process—a convenient but restrictive description of microeconomic dynamics.

We show that productivity growth measured from input and output quantities is sharply nonnormal and follows a size-dependent law of motion. The association between productivity risk and producers' centrality reshapes the aggregate growth distribution. Under cost-based Domar aggregation, replacing the Gaussian distribution with the estimated microeconomic law of motion increases the probability of a technological recession from 0.05% to 0.31%, while reducing the probability of growth above 20% from approximately 13% to 4.5%. The Gaussian benchmark is optimistic on both tails by understating contraction risk and overstating the likelihood of large expansions.

We address these questions using restricted-use microdata from India's Annual Survey of Industries (ASI), covering 2010–11 through 2022–23. We use the most disaggregated level available at the ASI—the plant-product pair—and construct annual physical-output Solow-residual productivity growth. Our analysis proceeds in three steps. First, using nonparametric methods, we characterize both the marginal distribution and the conditional dynamics of productivity growth. The distribution is distinctly non-Gaussian, mean reversion is nonlinear, and both its shape and dynamics vary with production-unit size. Second, we estimate a tractable parametric three-regime normal-mixture law of motion in which transition probabilities and conditional distributions depend on lagged productivity growth and size. The model closely reproduces the empirical distribution, conditional moments, tail behavior, and nonlinear mean reversion documented nonparametrically. Finally, we simulate this law of motion with the observed units of production, and using cost-based Domar aggregation estimate the model-implied distribution of aggregate technological growth.

Figure 1: Productivity Growth: Gaussian and Production Unit Size

(a) Pooled and Gaussian densities **(b) Empirical densities by unit size**



Notes: Panel (a) plots the distribution of productivity growth and a fitted Gaussian density. Panel (b) compares empirical densities for small, medium, and large production units.

Figure 1 shows that productivity growth departs sharply from the Gaussian benchmark and varies systematically with production-unit size. The empirical distribution places more probability near its mode and in the far tails, but less in the shoulders. As size increases, the central peak sharpens, yet extreme realizations remain prominent relative to the narrowing core. Greater central concentration thus coexists with higher excess kurtosis.

Size also shapes productivity dynamics. Following comparable shocks, large units exhibit stronger expected recoveries from losses and retain more of their gains—the *Daedalian level effect*. The strength of mean reversion also changes across the growth distribution. As preceding growth moves from strongly negative toward moderately negative realizations, improvements increasingly reduce the rebound expected next period—the *Icarian liftoff*. This effect is sharper among large units, but subsequently gives way to particularly weak reversal within the *Daedalian corridor*. Following large favorable shocks, renewed reversal—the *Icarian fallout*—is more pronounced among small units, whereas large units preserve more of each additional gain. These dynamic advantages do not eliminate extreme outcomes. Large units enter the left tail less frequently than small units but the right tail more often. Their shocks also extend farther beyond the tail threshold in both directions—the *Daedalian range pattern*. Larger units thus combine stronger buffering with tail incidence tilted toward favorable shocks, while remaining exposed to farther-reaching realizations on both sides.

Our main microeconomic contribution is a tractable, size-dependent law of motion for measured physical-output productivity growth. It combines three conditionally normal latent regimes among which production units move over time. *Neutralization* forms a narrow, low-dispersion core: expected growth remains close to zero and varies little with preceding growth. *Stabilization* is similarly prevalent but more dispersed, occupying the shoulders between the core and the far tails. It combines recovery after adverse shocks with comparatively mild reversal of favorable gains. *Correction* is less frequent and much more dispersed, accounting for nearly all outcomes in the farthest tails. It also exhibits the strongest two-sided mean reversion: large adverse shocks predict the largest subsequent rebounds, while large favorable shocks predict the sharpest contractions. The resulting law of motion captures the concentrated core and heavy tails of productivity growth, together with the size-dependent conditional moments, buffering, and nonlinear mean reversion documented above.

Our main macroeconomic contribution is to show how nonnormality and size dependence in microeconomic productivity dynamics jointly shape the distribution of aggregate technological growth. Cost-based Domar aggregation gives greater influence to units whose productivity process differs systematically from that of the average unit. We first validate this micro-to-macro mapping. Domar-weighted aggregates of measured productivity growth and its unpredictable component comove with nine independent indicators of Indian manufacturing activity. Their correlations with NSO real manufacturing GVA are 0.76 and 0.87, respectively. These correlations provide descriptive evidence of the aggregate relevance of our microeconomic shocks.

Table 1 compares the full size-dependent process with two separately estimated alternatives that remove either regime heterogeneity or size dependence. Replacing the three regimes with a single Gaussian regime, while retaining size dependence, raises aggregate mean growth and dispersion but produces much thinner tails. Relative to the full model, the Gaussian implied distribution of aggregate risk places more probability in the intermediate growth intervals, particularly strong expansions, and less in contractions. It therefore understates contraction risk and overstates the likelihood of large expansions: the full model assigns nearly six times as much probability to a technological recession and only about one-third as much probability to growth above 20%. Removing size dependence while retaining three regimes instead lowers average aggregate growth, raises recession risk nearly eightfold, and reduces the probability of growth above 20% by more than half. Thus, ignoring nonnormality produces an overly optimistic assessment of these aggregate events, whereas ignoring size dependence misses the more favorable productivity dynamics of highly exposed units. Both features reshape aggregate moments and the allocation of probability between contractions and expansions.

Table 1: Tails in the Aggregate Technology Growth Distribution

Aggregate Tail	Gaussian	Size-independent	Size-dependent
Below 0%	0.0553%	2.4670%	0.3150%
0% to 5%	1.7560%	14.3900%	1.1900%
20% to 25%	11.3900%	1.1130%	2.7930%
Above 25%	1.6580%	0.7760%	1.7450%

Contribution to the Literature

We contribute to the literature on productivity measurement by constructing annual physical-output Solow-residual growth at the plant–product level. Our focus on physical output builds on the distinction between physical and revenue productivity in Foster, Haltiwanger, & Syverson (2008). Related plant–product studies use physical output quantities to recover TFPQ or technical efficiency as residuals from parametrically estimated multiproduct production functions (Orr, 2022; Dhyne, Petrin, Smeets, & Warzynski, 2024). Closer to our growth-accounting approach, McNerney, Savoie, Caravelli, Carvalho, & Farmer (2021) and Fadinger, Ghiglino, & Teteryatnikova (2022) construct cost-share-based gross-output productivity levels for country–sector pairs. Our measure combines plant–product physical-output granularity with cost-share growth accounting.

We derive the residual from producer first-order conditions that permit heterogeneous markups, inventory accumulation, and unit–input–year variation in input costs. We recover local output elasticities from effective cost shares, with capital valued at its user cost. Elasticities may vary by units, input, and year, and the construction requires no parametric function. Our preferred measure nets out the contributions of growth in labor, capital, domestically produced manufacturing intermediates, direct imports, and other nonmanufacturing primary inputs. Our plant–product physical-output productivity growth measure combines a comprehensive input structure with local elasticities recovered under heterogeneous markups, inventories, and input costs.

Our paper also contributes to the literature using India's ASI to study productivity and production organization. Existing work examines static misallocation (Hsieh & Klenow, 2009), plant life-cycle growth (Hsieh & Klenow, 2014), and how contract enforcement shapes input sourcing and production organization (Boehm & Oberfield, 2020). Related studies examine production fragmentation and specialization (Boehm & Oberfield, 2023), and how joint production and within-firm productivity heterogeneity affect the identification of firm- and product-level markups (Cairncross, Morrow, Orr, & Rachapalli, 2025). We characterize the distribution and size-dependent dynamics of plant–product productivity growth and trace their implications for aggregate risk.

Our distributional analysis informs productivity specifications in quantitative models. Gaussian innovations in log productivity feature in analyses of employment protection (Hopenhayn & Rogerson, 1993), plant investment and stimulus policy (Khan & Thomas, 2008; Winberry, 2021), credit-shock propagation (Khan & Thomas, 2013), and financial frictions, misallocation, and technology adoption (Midrigan & Xu, 2014). Related specifications are used to study the amplification of aggregate shocks through entry and exit (Clementi & Palazzo, 2016) and monetary transmission through heterogeneous investment decisions (Gnewuch & Zhang, 2025). In continuous time, Bilal, Engbom, Mongey, & Violante (2022) model incumbent log productivity as Brownian motion to study firm growth and worker reallocation. These applications motivate examining the full productivity transition law: even with persistence and innovation variance fixed, the shape of the shock distribution can change aggregate event probabilities.

Recent work develops empirically richer alternatives to this Gaussian distribution. Melcangi & Sarpietro (2024) estimate a nonlinear persistent–transitory process for Compustat revenue-based TFP to study aggregate fluctuations. Salgado, Guvenen, & Bloom (2023) calibrate a two-normal mixture of productivity innovations to sales-growth moments, with the innovation distribution varying across common aggregate risk states. Jaimovich, Terry, & Vincent (2023) nonparametrically estimate firm-revenue transitions to study firm values, selection, and responses to firm subsidies, while Boehm, South, Oberfield, & Waseem (2024) generate fat-tailed sales growth through supplier matching and switching. We estimate a three-regime Markov normal mixture for physical-output productivity growth. Producers transition among persistent regimes, yielding a parametric representation of non-Gaussian productivity dynamics.

Finally, we connect the granularity and production network literatures. Hulten aggregation maps producer-level technology changes into aggregate productivity at first order; in distorted economies, cost-based Domar weights capture this aggregation (Hulten, 1978; Baqaee & Farhi, 2020). Heterogeneity in producer size and network position can limit diversification of idiosyncratic risk, while changes in the distribution and dynamics of large producers can alter aggregate volatility (Gabaix, 2011; Acemoglu, Carvalho, Ozdaglar, & Tahbaz-Salehi, 2012; V. Carvalho & Gabaix, 2013; V. M. Carvalho & Grassi, 2019). Non-Gaussian shocks and heterogeneous or state-contingent network exposure can generate macroeconomic tail risk (Acemoglu, Ozdaglar, & Tahbaz-Salehi, 2017; Dew-Becker, 2023). Aggregation weights tell us whose shocks matter most. We estimate how these producers' productivity risks shape the full distribution of aggregate growth, including contraction and expansion probabilities.

Our microeconomic law of motion provides an empirically disciplined tool for aggregation: Cost–Domar weights determine whose shocks matter most; the productivity law determines the risks they carry. Aggregate outcomes reflect the association between size-conditioned productivity risk and these exposures. We compare equal and Cost–Domar weighting under three separately estimated laws: a size-dependent Gaussian, a size-independent three-regime mixture, and the full size-dependent mixture. The resulting stationary distributions quantify contraction, expansion, and tail probabilities for the technological component of aggregate productivity growth.

Outline. Section 2 presents the data and productivity measurement framework. Section 3 documents the distribution and size-dependent dynamics of productivity growth. Section 4 develops and estimates the regime model. Section 5 examines its implications for aggregate technology risk. Section 6 concludes.

2 Data and Model

We use restricted-use microdata from India's Annual Survey of Industries (ASI) for 2010–11 through 2022–23. Item-level physical quantities and values of outputs and inputs let us map plant–commodity production units into a commodity-level production network and construct a physical-output Solow residual accounting for multiproduct input allocation, inventories, intermediates, domestic and imported factors, markups, labor and capital.

2.1 Model

A production unit is a plant-product pair: an ASI plant may be multiproduct, and we treat each product line as an independent unit. Many production units may supply the same good, and their outputs are combined by a competitive good aggregator. Units use labor, capital, and intermediate goods to produce. An input-output link is observed whenever a commodity appears as an input for one unit and as an output supplied by other units.

Commodity Aggregation. For each commodity g, at period t, we distinguish the aggregation from the production. Let $\mathcal{N}_{g,t}$ be the set of plant-product pairs that produce g. A competitive aggregator combines these units into the good that is used for consumption, intermediate inputs, and investment. We write this aggregator in normalized CES form around the observed allocation:

$$Y_{g,t} = \overline{Y}_{g,t}\Big[\sum_{i\in\mathcal{N}_{g,t}} \sigma_{i,t}\big(\widetilde{y}_{i,t}/\overline{\widetilde{y}}_{i,t}\big)^{\frac{\eta_{g,t}-1}{\eta_{g,t}}}\Big]^{\frac{\eta_{g,t}}{\eta_{g,t}-1}},$$

where $\widetilde{y}_{i,t}$ is the quantity sold by unit i, $p_{i,t}$ is its price, $Y_{g,t}$ is the aggregate quantity for g, and $\eta_{g,t}$ is the aggregator's elasticity of substitution. $\sigma_{i,t}$ is the observed sales share of unit i among producers of good g: $\sigma_{i,t} = \frac{p_{i,t}\widetilde{y}_{i,t}}{\sum_{m\in\mathcal{N}_{g,t}} p_{m,t}\widetilde{y}_{m,t}}$.

In the data, $p_{i,t}$ is measured as sales divided by quantity sold. The aggregator turns plant-product output into commodity-level supply while allowing producers of the same good to differ in prices, quantities, and input use. Transactions run from units to aggregators and back to units as intermediates.

For a competitive aggregator, the price $p_{g,t}$ faced by its buyers is its unit cost. Given $\{p_{i,t}\}_{i\in\mathcal{N}_{g,t}}$ and $\{\sigma_{i,t}\}_{i\in\mathcal{N}_{g,t}}$, the price index for the aggregator is

$$p_{g,t} = \Big[\sum_{i\in\mathcal{N}_{g,t}} \sigma_{i,t}^{\eta_{g,t}}\big(\overline{\widetilde{y}}_{i,t}/\overline{Y}_{g,t}\big)^{1-\eta_{g,t}} p_{i,t}^{1-\eta_{g,t}}\Big]^{\frac{1}{1-\eta_{g,t}}}.$$

Thus $p_{i,t}$ is the producer-specific price observed in the ASI, while $p_{g,t}$ is the commodity-level price implied by the aggregator.

Production Units. Let $y_{i,t}$ denote the quantity manufactured, and $\widetilde{y}_{i,t}$ quantity sold to the aggregator. Hence, inventory accumulation is $\Delta s_{i,t} = y_{i,t} - \widetilde{y}_{i,t}$. Technologies are *nonparametric*. Output is $y_{i,t} = A_{i,t}Q_{i,t}\left(L_{i,t}, X_{i,t}\right)$, where $A_{i,t}$ is productivity, $L_{i,t}$ is a primary-factor composite, and $X_{i,t}$ is an intermediate-input composite. The primary-factor composite combines labor and capital, $L_{i,t} = Q^{\ell}_{i,t}\left(\{\ell_{if,t}\}_{f\in\mathcal{F}}, \{k_{ic,t}\}_{c\in\mathcal{K}}\right)$, where $\ell_{if,t}$ is primary factor of type f and $k_{ic,t}$ is capital of type c. The intermediate-input composite combines commodity inputs, $X_{i,t} = Q^{x}_{i,t}\left(\{x_{ig,t}\}_{g\in\mathcal{G}}\right)$, where $x_{ig,t}$ is the quantity of commodity g used by unit i. The functions $Q_{i,t}$, $Q^{\ell}_{i,t}$, and $Q^{x}_{i,t}$ are otherwise unrestricted, except for constant returns to scale and positive marginal products for inputs that are used.

The unit takes the input prices and aggregator's demand system as given, and chooses its price, output, input use, and investment. Hence profits are

$$\pi_{i,t} = p_{i,t}\left(\widetilde{y}_{i,t} + \Delta s_{i,t}\right) - \sum_{f\in\mathcal{F}} \tau^{\ell}_{if,t} w_{f,t}\ell_{if,t} - \sum_{g\in\mathcal{G}} \tau^{x}_{ig,t} p_{g,t} x_{ig,t} - \sum_{m\in\mathcal{G}} p_{m,t} \sum_{c\in\mathcal{K}} I_{icm,t}.$$

The first term is the value of manufactured output, measured as sales revenue plus the value of net inventory accumulation. The next two terms are labor and intermediate-input costs. The wedges $\tau^{\ell}_{if,t}$ and $\tau^{x}_{ig,t}$ allow effective input costs to differ across units and input types. The last term is investment expenditure: $I_{icm,t}$ is the quantity of commodity m used to produce new type c capital with $I_{ic,t} = Q^{k}_{ic,t}\left(\{I_{icm,t}\}_{m\in\mathcal{G}}\right)$. Capital evolves as $k_{ic,t+1} = (1-\delta_{ic,t})k_{ic,t} + I_{ic,t}$, where $\delta_{ic,t}$ is the depreciation rate of capital type c. This makes capital formation part of the same commodity network: goods can be consumed, or used to produce current output or to build future productive capacity.

For any choice variable q, holding inventory accumulation fixed, $\frac{\partial \Delta s_{i,t}}{\partial q} = 0$, so that $\frac{\partial \widetilde{y}_{i,t}}{\partial q} = \frac{\partial y_{i,t}}{\partial q}$. The CES inverse-demand system then implies

$$\frac{\partial\left(p_{i,t}y_{i,t}\right)}{\partial q} = \mu_{i,t}p_{i,t}\frac{\partial y_{i,t}}{\partial q}, \qquad \text{where} \qquad \mu_{i,t} = 1 - \frac{1}{\eta_{g,t}}\frac{y_{i,t}}{\widetilde{y}_{i,t}}, \quad i \in \mathcal{N}_{g,t}.$$

The term $\mu_{i,t}$ combines the markup effect generated by the elasticity of substitution with the distinction between manufactured output and sales created by inventory accumulation. When $\Delta s_{i,t} = 0$, we get the usual inverse markup $1 - 1/\eta_{g,t}$. This wedge scales the first-order conditions below and is central for mapping observed cost shares into production elasticities.

Let $e(a,b)$ denote the elasticity of a with respect to b. For primary factor type f and intermediate good g, the first-order conditions imply

$$e(y_{i,t}, \ell_{if,t}) = \frac{\tau^{\ell}_{if,t}w_{f,t}\ell_{if,t}}{\mu_{i,t}p_{i,t}y_{i,t}}, \qquad e(y_{i,t}, x_{ig,t}) = \frac{\tau^{x}_{ig,t}p_{g,t}x_{ig,t}}{\mu_{i,t}p_{i,t}y_{i,t}}.$$

Thus, after adjusting revenue by the wedge $\mu_{i,t}$, observed expenditures on labor and intermediates identify their output elasticities.

Capital is the only input disciplined by an inter-temporal condition, with each unit choosing investment to maximize $\sum\limits_{s=0}^{\infty}\frac{\pi_{i,t+s}}{(1+r)^{s}}$, subject to the production technology, the capital law of motion, and the real return r. Hence, the user cost of capital type c is $R^{k}_{ic,t} = (1+r)p^{k}_{ic,t} - (1-\delta_{ic,t})p^{k}_{ic,t+1}$, capturing the opportunity cost of tying $p^{k}_{ic,t}$ in capital, $(1+r)p^{k}_{ic,t}$, net of the resale value of leftover capital, $(1-\delta_{ic,t})p^{k}_{ic,t+1}$. Therefore, $e(y_{i,t}, k_{ic,t}) = \frac{R^{k}_{ic,t}k_{ic,t}}{\mu_{i,t}p_{i,t}y_{i,t}}$.

Euler's theorem and the first-order conditions gives

$$C_{i,t} \equiv \sum_{f\in\mathcal{F}}\tau^{\ell}_{if,t}w_{f,t}\ell_{if,t} + \sum_{g\in\mathcal{G}}\tau^{x}_{ig,t}p_{g,t}x_{ig,t} + \sum_{c\in\mathcal{K}}R^{k}_{ic,t}k_{ic,t} = \mu_{i,t}p_{i,t}y_{i,t}.$$

Thus, cost equals the value of wedge-scaled manufactured output, $\mu_{i,t}p_{i,t}y_{i,t}$.

Define the intermediate-input share and the primary-factor share as

$$\omega^X_{i,t} = \frac{\sum_{g\in\mathcal{G}} \tau^x_{ig,t} p_{g,t} x_{ig,t}}{\mu_{i,t} p_{i,t} y_{i,t}}, \qquad \omega^L_{i,t} = 1 - \omega^X_{i,t}.$$

Within intermediates and primary factors, define

$$\omega^x_{ig,t} = \frac{\tau^x_{ig,t} p_{g,t} x_{ig,t}}{\omega^X_{i,t}\mu_{i,t} p_{i,t} y_{i,t}}, \qquad \alpha^\ell_{if,t} = \frac{\tau^\ell_{if,t} w_{f,t} \ell_{if,t}}{\omega^L_{i,t}\mu_{i,t} p_{i,t} y_{i,t}}, \qquad \alpha^k_{ic,t} = \frac{R^k_{ic,t} k_{ic,t}}{\omega^L_{i,t}\mu_{i,t} p_{i,t} y_{i,t}}.$$

Definition 1. Let $\Delta_t z_i \equiv \log z_{i,t+1} - \log z_{i,t}$. Productivity growth for i is

$$g_{i,t+1} = -\omega^L_{i,t}\left(\sum_{f\in\mathcal{F}} \alpha^\ell_{if,t}\Delta_t \frac{\ell_{if}}{y_i} + \sum_{c\in\mathcal{K}} \alpha^k_{ic,t}\Delta_t \frac{k_{ic}}{y_i}\right) - \omega^X_{i,t}\sum_{g\in\mathcal{G}} \omega^x_{ig,t}\Delta_t \frac{x_{ig}}{y_i}.$$

This is a unit-specific Solow residual: productivity rises when a unit uses less inputs for a given quantity of output. We next describe its measurement.

2.2 Mapping the Data to the Model

Output, primary factor and intermediate quantities. Primary-factor quantities, $\ell_{if,t}$, combine labor and non-produced input items. For labor, we use days worked for male and female permanent, contract, and managerial workers. We then add item-level quantities for inputs used that are not manufactured. We include directly imported inputs as primary factors. When an input is also produced by some plant, we treat it as an intermediate, $x_{ig,t}$. We use purchase value to construct primary factor, $\tau^\ell_{if,t} w_{f,t} \ell_{if,t}$, and intermediate costs, $\tau^x_{ig,t} p_{g,t} x_{ig,t}$. Commodities are modeled at the most disaggregated available level, the seven-digit NPCMS code, which avoids unit-of-measurement conversions. In 2022-23, the 83,500 establishment-product units manufactured 4,731 goods. Of these goods, 4,035 are intermediates, generating 388,583 unit-by-good links. The factor block contains four types of labor, 912 domestic input categories not produced within the sample, and 3,171 imported input categories.

Multiproduct plants. If plant e produces goods $g \in \mathcal{G}_{e,t}$, product g's allocation weight is $\varpi_{eg,t} = \frac{S_{eg,t}}{\sum_{h\in\mathcal{G}_{e,t}} S_{eh,t}}$, where $S_{eg,t}$ is the sales value of product g. We assign the same share $\varpi_{eg,t}$ of plant-level costs to the production unit (e, g, t). This procedure preserves plant totals while allowing the model to treat each plant-product pair as a distinct producer. In 2022–23, 35% of plants are multiproduct, accounting for 67.1% of sales and 62.3% of production units. Among these plants, the median sales share of the largest product is 76.9%, and the largest product accounts for at least 75% (90%) of their sales in 52.4% (31.0%) of them.

Capital costs. We measure capital costs through its dynamic condition for seven types: land, buildings, plant and machinery, transport equipment, computer equipment including software, pollution-control and environmental-improvement equipment, and other fixed assets. For capital type c, we assume a stable price between adjacent years ($p^k_{ic,t} = p^k_{ic,t+1}$), so that $R^k_{ic,t}k_{ic,t} = (r + \delta_{ic,t})p^k_{ic,t}k_{ic,t}$. We use the closing gross value as the measure of $p^k_{ic,t}k_{ic,t}$, $r = 6\%$, and measure $\delta_{ic,t}$ as the depreciation rate during the current year.

Inventory adjustment. Using reported quantities produced and sold, we infer net inventory accumulation as $\Delta s_{i,t} = y_{i,t} - \widetilde{y}_{i,t}$. Whenever the absolute implied inventory change exceeds 30% of production, $|\Delta s_{i,t}|/y_{i,t} > 0.30$, we set $\widetilde{y}_{i,t} = y_{i,t}$ and hence $\Delta s_{i,t} = 0$. This rule applies to both inventory accumulation and depletion. We retain reported physical output and manufactured-product sales and recompute unit prices using the adjusted quantities sold. The adjustment limits the influence of extreme discrepancies between reported quantities manufactured and sold that may reflect measurement or reporting errors. The cutoffs of -0.30 and 0.30 approximate the median annual 2nd and 98th percentiles, respectively, of the signed inventory-to-production ratio $\Delta s_{i,t}/y_{i,t}$.

Markups. Let $C_{i,t}$ denote total cost, and $V_{i,t} = p_{i,t}y_{i,t}$ be the value of manufactured output. The inverse markup is $\mu^*_{i,t} = \frac{C_{i,t}}{V_{i,t}}$. To prevent high costs from driving the elasticity estimates, we impose $\mu_{i,t} = \min\{\mu^*_{i,t}, \bar{\mu}\}$, with $\bar{\mu} = 10$. In 2022–23, production units with cost-to-output ratios above 10 and above 1 account for 0.0032% and 11.85% of manufactured-product sales, respectively.

Within-good sales shares. Let $S_{i,t} = p_{i,t}\widetilde{y}_{i,t}$ denote revenue from manufactured sales. The price $p_{i,t}$ is measured by dividing revenue by quantity sold $\widetilde{y}_{i,t}$. For each good g, aggregate sales are $S_{g,t} = \sum_{i\in\mathcal{N}_{g,t}} S_{i,t}$, and $\sigma_{i,t} = \frac{S_{i,t}}{S_{g,t}}$ for $i \in \mathcal{N}_{g,t}$.

Elasticities of substitution. $\mu_{i,t}$ is related to the aggregator's $\eta_{g,t}$ through $\mu_{i,t} = 1 - \frac{1}{\eta_{g,t}}\frac{y_{i,t}}{\widetilde{y}_{i,t}}$ for $i \in \mathcal{N}_{g,t}$. Summing over $\mathcal{N}_{g,t}$, gives $\eta_{g,t} = \frac{\sum_{i\in\mathcal{N}_{g,t}} V_{i,t}/S_{g,t}}{1-\sum_{i\in\mathcal{N}_{g,t}}\sigma_{i,t}\mu_{i,t}}$.

Good-level quantities, prices, and wages. The commodity-level price is then

$p_{g,t} = \frac{S_{g,t}}{Y_{g,t}}$ with $Y_{g,t} = Y^*_{g,t}/\bar{\tau}^x_{g,t}$, where $Y^*_{g,t}$ comes from the CES aggregator. We leave the multiplicative scale of the CES aggregator unrestricted and choose it through a quantity-weighted mean-one normalization of intermediate-input wedges. We exploit this degree of freedom by choosing a normalization $\bar{\tau}^x_{g,t}$ so that quantity-weighted average intermediate-input wedge equals one, thereby pinning down $Y_{g,t}$ and $p_{g,t}$, without changing nominal sales $S_{g,t}$. Starting from the price $p^0_{g,t} = S_{g,t}/Y^*_{g,t}$, define $\bar{\tau}^x_{g,t} = \frac{\sum_i p^0_{g,t} x_{ig,t} \tau^{x,0}_{ig,t}}{\sum_i p^0_{g,t} x_{ig,t}} = \frac{\sum_i C^x_{ig,t}}{\sum_i p^0_{g,t} x_{ig,t}}$, where $C^x_{ig,t}$ is expenditure by unit i on input g. Equivalently $p_{g,t} = \bar{\tau}^x_{g,t} p^0_{g,t}$. The resulting wedges $\tau^x_{ig,t} = \frac{C^x_{ig,t}}{p_{g,t} x_{ig,t}}$ have quantity-weighted mean one within each commodity. We apply an analogous normalization to primary factors. For each factor type f, let $C^\ell_{if,t}$ denote the cost and $\ell_{if,t}$ the quantity used by production unit i. We define the factor price as $w_{f,t} = \frac{\sum_i C^\ell_{if,t}}{\sum_i \ell_{if,t}}$. The factor wedge is then $\tau^\ell_{if,t} = \frac{C^\ell_{if,t}}{w_{f,t} \ell_{if,t}}$. By construction, $\frac{\sum_i w_{f,t} \ell_{if,t} \tau^\ell_{if,t}}{\sum_i w_{f,t} \ell_{if,t}} = 1$. Thus, $w_{f,t}$ captures the common price of factor f, while $\tau^\ell_{if,t}$ captures unit-specific deviations from that common price.

Unit-value adjustment. Extreme input unit values can have disproportionate effects on measured productivity. The ratios $\tau^x_{ig,t}$ and $\tau^\ell_{if,t}$ capture unit-specific deviations from the reference prices. We clip both wedges to $[0.1, 10]$, defining $\tau^{x,*}_{ig,t} = \max\{0.1, \min\{10, \tau^x_{ig,t}\}\}$, and $\tau^{\ell,*}_{if,t} = \max\{0.1, \min\{10, \tau^\ell_{if,t}\}\}$, allowing unit values to differ from the common price by up to one order of magnitude. Holding nominal expenditures and the common reference prices fixed, we adjust quantities according to $x^*_{ig,t} = x_{ig,t} \frac{\tau^x_{ig,t}}{\tau^{x,*}_{ig,t}}$, and $\ell^*_{if,t} = \ell_{if,t} \frac{\tau^\ell_{if,t}}{\tau^{\ell,*}_{if,t}}$. In 2022–23, the adjustment affects only 0.003% of intermediate input expenditure below the lower bound and 0.047% above the upper bound; for primary factors, the corresponding shares are 0.012% and 0.240%.

Capital Measurement. The data reports fixed assets in value terms, not in physical units. To use capital in the productivity calculation, we convert these reported values into asset-specific quantities. This conversion requires a price index for each type of capital. We use the 1997 U.S. capital-flow table, which measures commodity composition in capital formation; comparable Indian data on the commodity composition of capital formation are not available. We aggregate the U.S. capital-flow categories into the seven asset types. This gives, for each asset type c, a set of commodity weights describing the goods used to produce that type of capital. We then map the U.S. commodity categories into the seven-digit NPCMS goods. When a U.S. commodity maps to multiple NPCMS goods, we allocate its weight across the matched NPCMS goods using their sales shares. Let $\alpha^k_{cg,t}$ denote the weight of commodity g in the price index for capital asset type c, and use $p^k_{c,t}(\theta) = \left(\sum_{g\in\mathcal{G}} \alpha^k_{cg,t} p^{1-\theta}_{g,t}\right)^{\frac{1}{1-\theta}}$ to measure capital prices, with the Cobb-Douglas limit when $\theta = 1$. We implement four values, $\theta \in \{0, 0.5, 1, 2\}$, to assess the sensitivity of measured capital quantities to substitution. The physical capital stock of production unit i, asset type c, and elasticity θ is $k_{ic,t}(\theta) = \frac{V^k_{ic,t}}{p^k_{c,t}(\theta)}$, where $V^k_{ic,t}$ is the closing value of capital for asset type c. Thus, capital quantities are measured by deflating reported asset values with an asset-specific investment price index. The capital-flow table supplies the commodity-composition weights used to construct asset-specific price indices. We apply the same price index to all production units within each asset type and year. Consequently, within an asset type and year, relative measured capital quantities across units reflect relative reported asset values.

Robustness in the Measure of Productivity. Our preferred productivity measure is the elasticity-weighted growth in all measured inputs: labor, domestic non-produced primary factors, imported primary factors, intermediate inputs, and capital. Because some components require additional measurement assumptions, especially capital, we also construct a sequence of seven alternative productivity measures that use different input sets. The main analysis reports the all-input productivity computed with $\theta = 1$. This specification is the closest empirical counterpart to the model's production technology and uses the full item-level structure of the data. The alternative measures verify that the results are not driven by any single input set or by the substitution parameter used to construct capital quantities. The single-product exercise separately shows that the principal patterns recur where within-plant input allocation across products is unnecessary. The Online Appendix D reports the main facts under ten alternative measures, for each value of θ in capital-inclusive measures: seven vary the inputs entering measured productivity while holding $\theta = 1$, three retain the all-input measure while setting $\theta = 0$, $\theta = 0.5$ and $\theta = 2$, and one exclusively for single-product plants. The main facts remain stable across these alternatives.

2.3 Sales-Based Validation of Measured Productivity

We next examine whether productivity growth is associated with economically meaningful changes in production-unit activity. For each horizon h, Panel (a) of Figure 2 estimates

$$\log S_{ip,t+h} - \log S_{ip,t-1} = \alpha^{h}_{pt} + \beta_h \, g_{ipt} + u_{ipth}, \tag{1}$$

where $S_{ip,t}$ denotes production-unit sales, g_{ipt} is productivity growth realized in year t, and α^{h}_{pt} denotes seven-digit-product-by-realization-year fixed effects. It plots the response associated with a 0.10-log-point positive realization, expressed as $100 \times 0.10\,\widehat{\beta}_h$, over horizons $h = -4, \ldots, 4$.

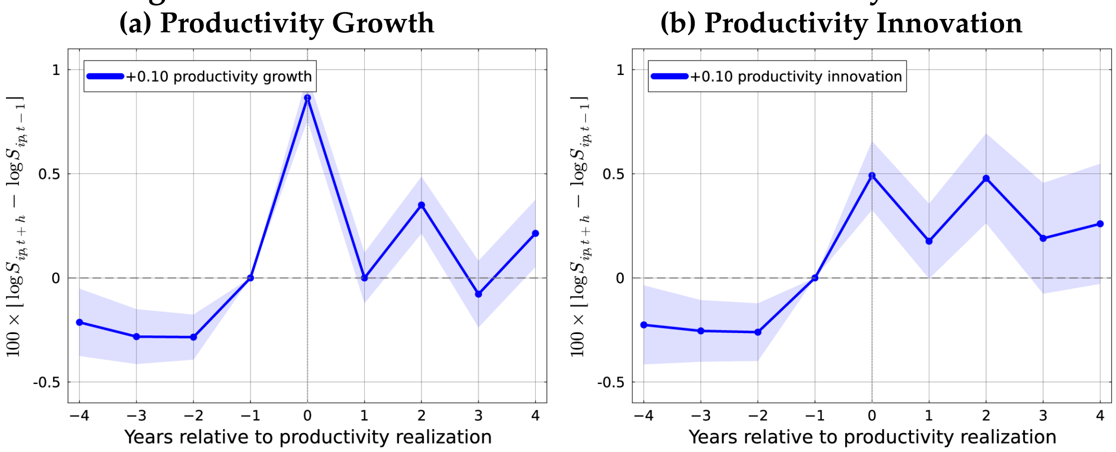
Panel (b) repeats equation (1), replacing productivity growth with its innovation, defined as observed growth minus its out-of-fold prediction from a linear AR(2) model. We construct these predictions using five-fold cross-fitting at the plant level, so that each forecast is estimated exclusively on observations from other plants. The innovation is the out-of-fold prediction error,

$$\varepsilon_{ipt} = g_{ipt} - \widehat{m}_{-k(i)}\left(g_{ip,t-1}, g_{ip,t-2}\right), \tag{2}$$

where $\widehat{m}_{-k(i)}$ denotes the fitted AR(2) prediction function estimated without observations from plants i's fold. The two lagged predictors are winsorized at the 1st and 99th percentiles of the training sample. Each panel retains shock realizations within its own 1st–99th percentile interval.

Panel (a) shows that positive productivity realizations are associated with significant sales increases on impact and at $t + 2$ and $t + 4$. After removing predictable productivity dynamics, Panel (b) shows a smaller contemporaneous increase, but subsequent estimates remain positive through $t + 4$, displaying greater consistency in sign across horizons. Innovation responses are significant on impact and at $t + 2$; the $t + 1$ estimate narrowly misses the 5% threshold. Both panels also reveal favorable preceding sales dynamics, consistent with

Figure 2: Production-Unit Sales Around Productivity Shocks



Notes: Panel (a) draws on 398,459 eligible g_{ipt}. Panel (b) uses 168,653 observations with five-fold, firm-held-out AR(2) innovations. These counts follow sales matching and panel-specific pooled first–ninety-ninth percentile shock trimming. Regressions absorb seven-digit-product-by-realization-year fixed effects. Shaded regions show firm-clustered, pointwise 95% confidence intervals, conditional on the estimated innovations in Panel (b). Plotted values equal $100 \times 0.10\widehat{\beta}_h$, representing 100 times the log-sales response to a positive 0.10-log-point realization. Horizon-specific sample sizes depend on sales availability and the removal of product-year singletons.

productivity improvements occurring among units already experiencing gains in activity. These pretrends connect measured productivity growth to broader improvements in production-unit performance.

3 The Microeconomics of Productivity Shocks

This section characterizes the distribution and conditional dynamics of productivity growth. We first document nonnormality, then examine heterogeneity by production-unit size and nonlinear mean reversion. We subsequently separate tail incidence from conditional tail reach and assess. The final subsection collects these results into the facts that discipline the model developed next.

3.1 Nonnormality, Higher Moments, and Time Variation

We use a sample consisting of adjacent-year transitions for production units. Table 2 reports distributional statistics for each year and for the pooled sample. The pooled panel of Figure 3 compares the empirical density with a Gaussian distribution having the same pooled mean and variance.

The distribution of productivity growth is far from Gaussian. In the pooled sample, mean growth is 0.087 log points, or about 9.1%, and its standard deviation is 1.412. The distribution combines a pronounced concentration of

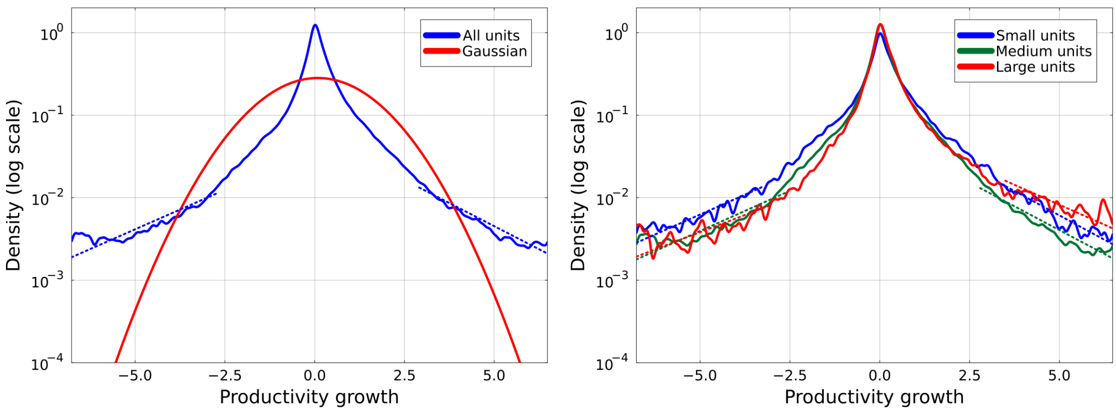
Table 2: Moments and Nonnormality of Productivity Growth

Year	Moments					Tail statistics			Exponential-tail tests			
	N	Mean	SD	Skew.	Kurt.	$p_{\lvert z\rvert>2}$	$p_{\lvert z\rvert>3}$	QR	$\widehat{\zeta}_L$	p_L^{AD}	$\widehat{\zeta}_R$	p_R^{AD}
Gaussian	–	0.087	1.412	0.000	0.000	4.55%	0.27%	3.45	1.897	0.5	1.897	0.5
Pooled	435,206	0.087	1.412	-0.250	14.747	5.12%	2.70%	16.83	0.440	0.5	0.516	0.5
11–12	15,065	-0.487	2.867	-0.400	3.097	8.32%	1.47%	8.12	0.628	0.5	0.493	14.0
12–13	15,797	0.152	2.078	-0.064	6.835	7.17%	3.83%	16.58	0.432	0.5	0.454	0.5
13–14	26,060	0.074	1.859	-0.427	8.030	6.60%	3.63%	17.46	0.425	0.5	0.489	0.5
14–15	32,335	-0.099	1.797	-0.707	8.504	6.40%	3.25%	16.31	0.470	0.5	0.463	0.5
15–16	33,802	0.329	1.477	0.235	12.608	5.27%	2.79%	13.39	0.457	0.5	0.492	7.5
16–17	36,049	0.062	1.490	0.550	13.190	5.38%	2.62%	13.69	0.467	0.5	0.509	0.5
17–18	40,926	0.279	1.577	0.405	9.248	5.69%	2.52%	10.86	0.541	9.0	0.563	1.5
18–19	42,652	0.176	1.007	0.531	21.201	4.73%	1.93%	12.28	0.700	0.5	0.745	0.5
19–20	43,206	-0.067	1.015	-0.276	17.229	5.04%	2.10%	12.86	0.742	0.5	0.751	0.5
20–21	46,681	0.087	0.981	0.424	24.716	4.67%	1.98%	13.23	0.730	0.5	0.708	0.5
21–22	49,903	0.142	0.904	0.337	21.615	5.07%	2.01%	12.20	0.827	0.5	0.863	0.5
22–23	52,730	0.065	0.886	0.353	27.234	4.96%	1.97%	14.70	0.830	0.5	0.815	0.5

Notes: For each row, z uses the row-specific mean and standard deviation, $p_{|z|>k} = \Pr(|z| > k)$, and $QR = (p_{99} - p_1)/(p_{75} - p_{25})$. SD denotes standard deviation, Skew. denotes skewness, and Kurt. is excess kurtosis. The Gaussian row uses the pooled mean and standard deviation and standard-normal population values for $p_{|z|>k}$ and QR. For each row, the tail threshold is two row-specific standard deviations from the mean. If $Y_R = g - \bar{g} - 2\sigma_g$ conditional on $g > \bar{g} + 2\sigma_g$, then $\widehat{\zeta}_R = 1/\bar{Y}_R$; $\widehat{\zeta}_L$ is defined analogously from left-tail exceedances. The Anderson–Darling columns report p_L^{AD} and $p_R^{\mathrm{AD}} \times 10^3$, so displayed entries are p-values in thousandths. The tests assess exponential exceedances using 2,000 replications with rate re-estimated in each sample.

Figure 3: Log Density of Productivity Growth and Tail Fits

(a) Pooled and Gaussian densities **(b) Size Densities and Tail Fits**



Notes: The right panel compares small, medium, and large production units, defined by beginning-year log sales standardized within each year: below -1.33, between -1.33 and 1.33, and above 1.33, respectively. Empirical densities are kernel-density estimated with the pooled-data bandwidth. Dotted lines show the exponential-tail densities implied by the individual exceedances. For each group, the tail thresholds are $\bar{g} \pm 2\sigma_g$, the tail exponent is $\widehat{\zeta} = 1/\bar{Y}$, and the fitted density is normalized using the empirical probability of entering the corresponding tail. The displayed range is bounded by the pooled 0.5th and 99.5th percentiles.

observations near its mode with a small but economically meaningful number of extreme values. Pooled skewness is -0.250, while excess kurtosis is 14.747, compared with zero under a Gaussian. Excess kurtosis is positive and large in every annual transition, ranging from 3.097 to 27.234. Nonnormality is common to all years. The pattern of departure from normality is particularly informative. In the pooled sample, $\Pr(|z| > 3) = 2.7\%$, compared with 0.27% under normality. Realizations more than three standard deviations from the mean are almost ten times as frequent as under the Gaussian. By contrast, $\Pr(|z| > 2) = 5.12\%$ is only modestly above the Gaussian probability of 4.55%. Consequently, the mass between two and three standard deviations is smaller than under normality: 2.42%, against 4.28%. Consequently, the Gaussian assigns too much probability to the shoulders and too little near the mode and in the tails of the distribution.

Quantile comparisons deliver the same conclusion. In the pooled sample, $QR \equiv \frac{p_{99}-p_1}{p_{75}-p_{25}} = 16.83$, compared with 3.45 under normality. The distance between the 1st and 99th percentiles is thus unusually large relative to the width of the central half of the distribution. Formal tests reject normality in every year.[1]

Figure 3 makes these differences in probability allocation visible. A Gaussian log density is quadratic, whereas the empirical density is more peaked near its mode and declines more slowly in the outer regions. Over substantial portions of the tails, the empirical log density appears linear. However, the shape of nonnormality changes over time. The standard deviation falls from 2.867 in 2011–2012 to 0.886 in 2022–2023, while $\Pr(|z| > 2)$ declines from 8.32% to 4.96%. Standardized far-tail measures are nonmonotonic. The probability of exceeding three standard deviations rises from 1.47% in 2011–2012 to 3.83% in 2012–2013 and then settles near 2% in the final years. Similarly, QR rises from 8.12 to a peak of 17.46 in 2013–2014, fluctuates thereafter, and ends at 14.70. Later distributions are therefore markedly narrower, while standardized far-tail prominence moderates from its early-sample peaks rather than declining monotonically

[1] For each year, both the Jarque–Bera and Anderson–Darling tests reject the Gaussian with $p < 0.001$. Given the size of the sample, statistical rejection by itself is not surprising. What matters is that the departures are large, persistent across years, and visible in the density.

over the full sample. Asymmetry also changes direction. Skewness is negative through 2014–2015 and predominantly positive thereafter, with a negative value in 2019–2020, ending at 0.353 in 2022–2023. Excess kurtosis remains positive and large throughout, rising from 3.097 in the first transition to 27.234 in the last, although its annual path is nonmonotonic. The moderation in standardized tail incidence and the higher endpoint kurtosis are not contradictory: kurtosis is particularly sensitive to the magnitude of the most extreme realizations.

3.2 Size Heterogeneity: Icarus and Daedalus Dynamics

The pooled distribution hides substantial heterogeneity across production units of different sizes. We measure size using sales, which are predetermined relative to productivity growth. Let $d_{i,t}$ denote standardized log sales for unit i at year t, where the mean and standard deviation are calculated separately for each year. This standardization removes changes in the level and dispersion of nominal sales across years and makes the size classification comparable over time. We classify $d_{i,t} < -1.33$ and $d_{i,t} > 1.33$ as the thresholds for small, medium, and large production units. The objective is to compare units that are economically far apart in scale while retaining sufficiently large samples in both tails of the size distribution.[2] Table 3 reports the pooled statistics for the three size groups, while Online Appendix Table 1 reports them separately for each year.

The first finding is that size is associated with both the location and the scale of the productivity-growth distribution. Average productivity growth increases with size in the pooled sample, from 0.056 among small units to 0.077 among medium units and 0.235 among large units. The annual evidence, however, does not reveal an invariant size ranking. Small units have higher mean growth than medium units in nine of twelve transitions and than large units in seven; large units have higher mean growth than medium units in eight. The growth advantage of large units is especially pronounced in 2016–2017 and 2017–2018, during the early years following the launch of *Make in India*. Overall dispersion

[2] The annual standard deviation of beginning-year log sales, $\sigma_{\log S,t}$, ranges from approximately 2.51 to 3.04. The ratio between the two thresholds is $S_t^L / S_t^S = \exp(2.66\,\sigma_{\log S,t})$, ranging from approximately 784 to 3,285 across years. Thus, the lower boundary of the large group is roughly three orders of magnitude above the upper boundary of the small group.

Table 3: Tail Shape by Production-Unit Size

	Distributional statistics									Tail tests			
Group	N	Sh.	Mean	SD	Skew.	Kurt.	$p_{\lvert z\rvert>2}$	$p_{\lvert z\rvert>3}$	QR	$\widehat{\zeta}_L$	p_L^{AD}	$\widehat{\zeta}_R$	p_R^{AD}
All	435,206	1.00	0.09	1.41	-0.25	14.75	5.12%	2.70%	16.83	0.44	0.5	0.52	0.5
Small	42,508	0.10	0.06	1.66	-0.52	10.48	5.75%	2.71%	13.81	0.44	0.5	0.54	18.0
Medium	361,408	0.83	0.08	1.36	-0.32	15.25	5.00%	2.62%	16.57	0.45	0.5	0.53	0.5
Large	31,290	0.07	0.24	1.63	0.54	14.35	5.73%	3.26%	20.62	0.38	0.5	0.44	0.5

Notes: The All row uses the same pooled sample as Table 2. Sh. denotes share. Mean is average productivity growth, $\bar{g}$. Within each row, z standardizes productivity growth, $p_{|z|>k} = \Pr(|z| > k)$, SD denotes standard deviation, Kurt. is excess kurtosis, and $QR = (p_{99} - p_1)/(p_{75} - p_{25})$. Tail exponents use individual exceedances beyond two row-specific standard deviations: $\widehat{\zeta}_R = 1/\bar{Y}_R$, where $Y_R = g - \bar{g} - 2s_g$ conditional on $g > \bar{g} + 2s_g$; $\widehat{\zeta}_L$ is defined analogously. The Anderson–Darling columns report p_L^{AD} and p_R^{AD} multiplied by 10^3, so displayed entries are p-values in thousandths. The tests assess exponential exceedances using 2,000 replications.

is nonmonotonic in the pooled sample: standard deviations are similar among small and large units. Nevertheless, small units have higher dispersion than medium units in every annual transition and than large units in nine of twelve. Dispersion also contracts markedly over time within every size group. Between the first and final transitions, the standard deviation falls among all size groups.

These differences in location and overall dispersion do not translate mechanically into a corresponding ranking of standardized tail incidence. In the pooled sample, the probability of a realization beyond two group-specific standard deviations is nearly identical among small and large units, at 5.75% and 5.73%, respectively. However, the allocation of this probability differs. Large units place less mass in the shoulders, $2 < |z| \leq 3$, and more beyond three standard deviations: the latter probability is 3.26%, compared with 2.71% among small units. The annual results do not reproduce this pooled ordering uniformly. Small units have higher two-standard-deviation incidence than large units in seven of twelve transitions and higher three-standard-deviation incidence in eight. The pooled large-unit far-tail premium therefore does not represent a persistent annual ranking in the frequency of crossing a threshold.

The higher moments reveal a more persistent distinction between small units and the other two size groups in the prominence of extreme realizations. Pooled excess kurtosis increases from 10.475 among small units to 15.249 among medium units, before declining to 14.349 among large units. The QR increases monotonically, from 13.81 to 16.57 and 20.62, respectively. The annual evidence broadly preserves the small-unit comparison. Excess kurtosis is higher among medium than small units in eleven of twelve transitions and among large than small units in ten. Similarly, the QR is higher among medium than small units in every transition and among large than small units in ten.

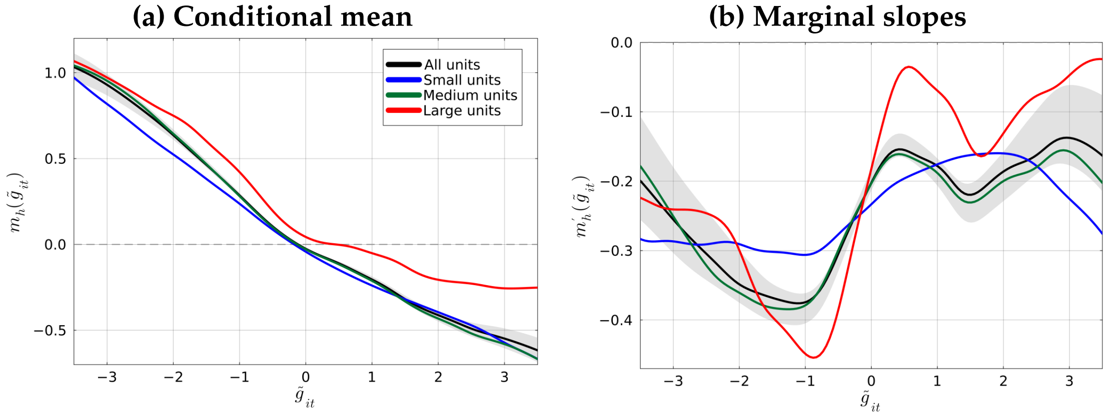
Asymmetry displays a size ranking with substantial annual variation. Skewness increases with size, from -0.520 among small units to -0.316 among medium units and 0.541 among large units. The annual results show, however, that these signs are not fixed characteristics of the size groups. Small-unit growth is positively skewed in six of twelve transitions, while large-unit growth is positively skewed in seven. Large units are left-skewed in each of the first four transitions, but right-skewed in seven of the subsequent eight, with 2020–2021 the exception. Their positive pooled skewness thus accompanies a pronounced change in asymmetry over time rather than a time-invariant size effect.

Taken together, these results imply that size heterogeneity has more than one dimension. Small and large units have comparable overall dispersion in the pooled sample, but differ in asymmetry and in the allocation of probability between shoulders and far tails. Large units place less probability in the shoulders and more beyond three standard deviations, while their higher QR indicates greater separation between the outer percentiles relative to the width of the central half. Figure 3 makes this distinction visible: the large-unit density combines a sharp central peak with a prominent right tail.

Size heterogeneity also appears in the intertemporal dependence of productivity growth. We first remove year-specific means, $\widetilde{g}_{i,t} = g_{i,t} - \overline{g}_t$, where $\overline{g}_t$ is average productivity growth in year t among observations contributing to the consecutive-year sample. For size group h, define the conditional-mean function as $m_h(x) = \mathbb{E}[\widetilde{g}_{i,t+1} \mid \widetilde{g}_{i,t} = x,\, h_{it} = h]$. Figure 4 plots local polynomial estimates of $m_h(x)$ and its derivative, $m_h'(x)$, for the pooled sample and separately by size.

These estimates describe conditional dynamics without imposing linearity or the state structure introduced later in the paper. The pooled estimates reveal pro-

Figure 4: Nonparametric Mean Reversion in Productivity Shocks



Notes: The figure pools consecutive-year observations. Current and subsequent productivity growth are separately demeaned within each annual paired sample before splitting by size. The estimation grid spans the pooled 1st–99th percentiles of demeaned current growth, approximately $[-4.175, 4.057]$; the displayed horizontal range is $[-3.5, 3.5]$. At each grid point x, separately for the pooled sample and each size group, we estimate a Gaussian-kernel local cubic regression of $\tilde{g}_{i,t+1}$ on powers of $\tilde{g}_{i,t} - x$. Each series uses a group-specific bandwidth schedule obtained from a plug-in AMSE criterion for the local-slope estimator and smoothed across x. The left panel reports the estimated conditional mean, $m_h(x)$, and the right its estimated local slope, $m_h'(x)$. Gray ribbons report 95% pointwise robust bias-corrected intervals for the pooled estimates. Bias is estimated using a local-quintic fit, and standard errors are clustered by plant and incorporate the covariance between the cubic estimate and its bias correction.

nounced mean reversion: large negative deviations predict positive deviations in the following transition, whereas large positive deviations predict subsequent declines. The estimated slope is negative throughout the displayed range, and its pointwise confidence interval remains below zero. The strength of this reversal is nonlinear. The slope reaches approximately -0.38 around $\tilde{g}_{i,t} = -1.1$, then becomes less negative and varies nonmonotonically over positive realizations, reaching approximately -0.14 near $\tilde{g}_{i,t} = 3$. A single linear autoregressive coefficient therefore does not capture the variation in these conditional dynamics.

We interpret the size differences in the conditional-mean schedules as the level dimension of the *Icarus–Daedalus pattern*. The large-unit schedule lies above those of both small and medium units. Following negative productivity realizations, this higher schedule implies a stronger expected recovery. Following positive realizations, it implies greater resilience of the preceding gains, with milder subsequent contractions. For example, at $\tilde{g}_{i,t} = -1$, expected subsequent growth is approximately 0.42 for large units and 0.24 for small units; at $\tilde{g}_{i,t} = 1$, the corresponding estimates are -0.05 and -0.24. Large units thus exhibit two-sided buffering: stronger recovery following adverse realizations and greater retention of gains following favorable realizations.

This is the *Daedalian level effect*. In the mythological analogy, large units occupy the position of Daedalus, the master builder and inventor: their conditional dynamics exhibit resilience through stronger recovery after setbacks and greater preservation of gains after expansions. The corresponding *Icarian level effect* is the weaker buffering exhibited by smaller units relative to large units. At a common preceding realization, their lower conditional-mean schedules imply less recovery following adverse outcomes and more pronounced reversals following favorable outcomes. Although the small- and medium-unit schedules cross, the large-unit advantage holds relative to both groups. The defining level pattern is therefore the two-sided resilience associated with large size, rather than a uniform ranking across all three size groups.[3]

The second manifestation of the *Icarus–Daedalus pattern* is a slope effect. Whereas the level effect compares size-specific conditional means, the derivative $m_h'(x)$ measures how expected subsequent productivity growth changes as the preceding realization improves. A more negative slope implies a larger marginal offset in the following transition: expected cumulative growth, $x + m_h(x)$, has marginal response $1 + m_h'(x)$. The slope is negative throughout the displayed range, and its pointwise robust bias-corrected interval remains below zero.

The slope dimension extends the analogy from Daedalus's craftsmanship to the flight enabled by his wings: more favorable productivity realizations represent greater altitude, while production scale represents the mass carried in flight. The negative-growth region exhibits an *Icarian liftoff*. Moving inward from strongly negative realizations, the slope profiles become more negative before reaching their principal troughs around $\widetilde{g}_{i,t} = -1$. This steepening is particularly pronounced among large units, whose slope reaches approximately -0.45 near -0.9, compared with troughs of about -0.31 for small units and

[3]Related heterogeneity in conditional dynamics appears in the "*butterfly pattern*" documented by Guvenen et al. (2021) for earnings. Workers with low recent earnings exhibit stronger reversal of negative changes and greater persistence of positive changes than workers with high recent earnings. In our setting, two-sided buffering is instead associated with large units.

-0.38 for medium units. In the analogy, production units lift off from the sea with wings still damp from the preceding adverse realization. Around the liftoff trough, carrying a greater mass makes the ascent more demanding: as large units gain altitude along the realization distribution, each additional improvement sacrifices more of the expected rebound following the original loss. Smaller units, carrying less weight, experience a milder marginal withdrawal of this buffering. The demanding large-unit liftoff therefore concerns the erosion of their buffering advantage, not its disappearance: their conditional-mean schedule remains above those of smaller units, even where improvements in the preceding realization narrow that advantage.

After the liftoff trough, the schedules enter the *Daedalian corridor*. Slopes become less negative, indicating a region of controlled flight in which additional growth is associated with weaker marginal reversal in the following year. The pooled and medium-unit slopes reach initial local maxima of approximately -0.15 and -0.16 near $\widetilde{g}_{i,t} = 0.4$. Large units exhibit a particularly sharp moderation after their deeper liftoff trough, reaching approximately -0.04 near 0.6. Small units approach the corridor more gradually: their slope continues to become less negative until approximately 1.9, where it reaches -0.16. Daedalian moderation therefore appears at every size, but is most pronounced among large units. In the flight analogy, their demanding liftoff gives way to the controlled ascent associated with Daedalus's craftsmanship: within this corridor, additional gains in altitude entail comparatively little sacrifice of buffering at the margin. Small units experience a milder transition, with marginal reversal diminishing more gradually and reaching its weakest point farther into the positive range.

Beyond the Daedalian corridor, the size-specific profiles diverge, with small units exhibiting a more pronounced *Icarian fallout*. Their slope falls from approximately -0.16 near $\widetilde{g}_{i,t} = 1.9$ to -0.28 at 3.5, indicating a sustained intensification of the marginal withdrawal of buffering. Thus, as favorable realizations become larger, a progressively greater share of each additional gain is offset in the following transition. Large units display the opposite upper-tail pattern.

After a temporary steepening around 1.7, their slope approaches zero, reaching approximately -0.02 at the right boundary. Additional favorable realizations therefore encounter progressively less subsequent reversal, preserving more of the incremental productivity gain. Medium units follow a nonmonotonic path and also steepen near the boundary, although less sharply than small units. In the flight analogy, large units embody Daedalus's craftsmanship: after their demanding liftoff, they can fly closer to the sun without experiencing the sustained erosion of buffering observed among small units. Small units face a milder liftoff, but their flight becomes harder to sustain at high positive realizations, where additional altitude entails a growing sacrifice of buffering. The Icarian fallout is therefore more pronounced among small units, whereas large units retain more of each additional productivity gain.

Taken together, the evidence supports a multidimensional interpretation of size heterogeneity. Large units combine higher average growth with stronger two-sided buffering: their expected subsequent shock is more favorable following both adverse and favorable shocks. These advantages coexist with greater far-tail incidence, a higher QR, and a more demanding liftoff. Around the slope trough, improvements in productivity entail a sharper marginal withdrawal of buffering among large units. Beyond this trough, however, their demanding ascent gives way to pronounced Daedalian moderation and comparatively weak marginal reversal at high positive shocks. Small units experience a milder liftoff but exhibit weaker buffering in levels and a more pronounced upper-tail Icarian fallout. Larger size is therefore associated with stronger resilience, despite more prominent extreme shocks and a sharper sacrifice of buffering during liftoff.

3.3 Tail Incidence and Conditional Reach

The preceding evidence establishes that productivity growth is sharply non-Gaussian and that its distribution varies substantially with unit size. To characterize the tails more precisely, we separate two margins that are often combined under the description *fat-tailed*. The first is *tail incidence*: the probability that a realization crosses a specified outer threshold. The second is *conditional tail reach*:

the distribution of the excess beyond that threshold, conditional on crossing it. For example, for a right-tail cutoff u and distance $y > 0$,

$$\Pr(g > u + y) = \Pr(g > u)\Pr(g - u > y \mid g > u).$$

The first factor measures tail entry; the second describes how much farther realizations extend beyond the cutoff. A group may cross a threshold relatively infrequently yet generate large excesses conditional on entry, or cross it frequently while exhibiting comparatively limited conditional reach.

We use exponential exceedances as a parametric benchmark for conditional tail reach. For size group h and productivity-growth cutoff u, define the right-tail exceedance as $Y_{R,i}(u) = g_i - u$ for $g_i > u$, and the left-tail exceedance as $Y_{L,i}(u) = u - g_i$ for $g_i < u$. Under the exponential distribution, $\Pr(Y_{R,i}(u) > y \mid g_i > u, h_i = h) = \exp(-\zeta_{R,h}(u)y)$, $y \geq 0$, with an analogous expression for the left tail. Let $N_{R,h}(u)$ and $N_{L,h}(u)$ denote the numbers of observations in group h above and below u, respectively. We estimate the decay rates as

$$\widehat{\zeta}_{R,h}(u) = \left[\frac{1}{N_{R,h}(u)} \sum_{i \in h: g_i > u} (g_i - u)\right]^{-1}, \quad \widehat{\zeta}_{L,h}(u) = \left[\frac{1}{N_{L,h}(u)} \sum_{i \in h: g_i < u} (u - g_i)\right]^{-1}.$$

These estimators are the reciprocals of the corresponding sample mean exceedances. A smaller decay rate implies slower conditional decay and greater expected reach beyond the cutoff. An exponential right tail corresponds to a Pareto upper tail in the gross productivity-growth ratio $R_i = e^{g_i}$. An exponential left tail similarly corresponds to a Pareto upper tail in R_i^{-1}, or equivalently to power-law behavior near zero in R_i. The estimators above therefore take the inverse-Hill form for R_i and R_i^{-1}, respectively (Hill, 1975). Exponential decay in both tails implies double-Pareto-type tail behavior in productivity growth.

Tables 2 and 3 evaluate the exponential-exceedance null at thresholds located two row-specific standard deviations from the mean. The straight tail-fit segments superimposed on the log densities in Figure 3 display the exponential densities implied by these same estimates. Under the exponential distribution, the log density is linear beyond the threshold with slope $-\widehat{\zeta}_{R,h}$ on the right and $\widehat{\zeta}_{L,h}$ on the left. A larger decay rate implies a steeper outward decline and shorter conditional tail reach. In the pooled sample, the estimated rates are $\widehat{\zeta}_L = 0.440$ and $\widehat{\zeta}_R = 0.516$, indicating greater conditional reach in the left tail. Across years, the left-tail estimates range from 0.425 to 0.830, and the right-tail estimates from 0.454 to 0.863. Later years generally exhibit higher rates, consistent with a distributional contraction. These fitted rates summarize conditional reach, but a single exponential rate does not fully capture the shape of the empirical tails. Parametric-bootstrap Anderson–Darling tests reject exponential exceedances at the 5% level on both sides in every annual sample and each pooled size group. Rejection also holds at the 1% level except for the right tails of the 2011–2012 sample and the pooled small-unit group.

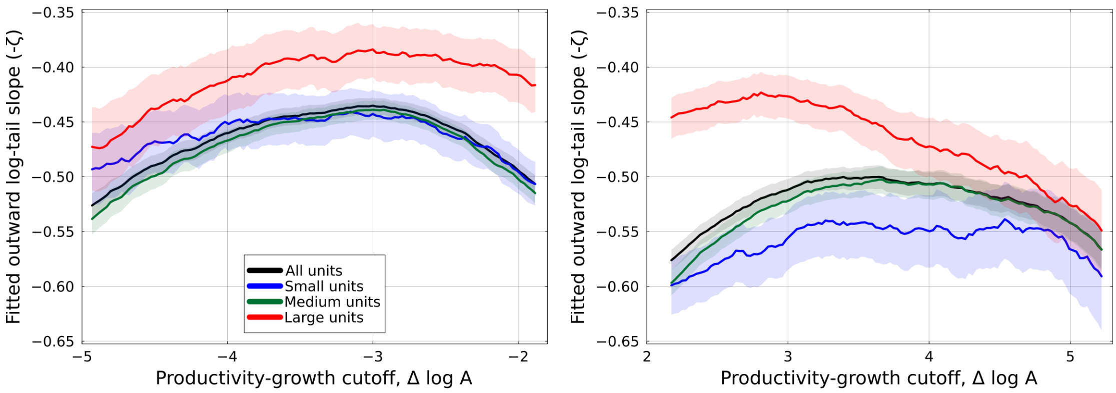
The size comparison makes the incidence–reach distinction concrete. The pooled probability of crossing two group-specific standard deviations is nearly identical for small and large units, at 5.75% and 5.73%, respectively. Beyond three standard deviations, however, incidence is higher among large units: 3.26%, compared with 2.71% among small units. Large units also exhibit greater conditional reach on both sides. The left-tail decay rate falls from 0.444 among small units to 0.384 among large units, while the right-tail rate falls from 0.543 to 0.445. Since $1/\widehat{\zeta}$ equals the sample mean exceedance, large-unit realizations extend farther beyond their respective thresholds, conditional on entering the tail. Incidence conceals an important asymmetry: large units enter the left tail less frequently than small units, at 2.13% versus 3.04%, but enter the right tail more frequently, at 3.61% versus 2.71%. Greater conditional reach therefore coexists with lower incidence on the left and higher incidence on the right.

Figure 5 compares conditional reach across size groups using different cutoffs. Throughout both tails, the large-unit point estimate lies closer to zero than those of small and medium units. The profiles are nevertheless distinctly nonhorizontal. On the left, decay is slowest around $u = -3$ and accelerates toward the end. On the right, the large-unit slope reaches approximately -0.42 near $u = 2.8$ before declining to -0.55 near 5.2. The small-unit profile remains comparatively flat over an intermediate range before steepening toward the outer boundary. A single constant rate cannot summarize either tail across thresholds.

We call this the *Daedalian range pattern*: larger units exhibit a broader conditional flight envelope once a tail boundary has been crossed. This range dimension complements their buffering. Large units reach farther into both tails, even while displaying stronger buffering. Greater conditional resilience thus coexists with greater conditional tail reach.

Figure 5: Threshold-Specific Tail Decay in Productivity Growth

(a) Left-tail decay by firm size **(b) Right-tail decay by firm size**



Notes: Let $N_{R,h}(u)$ and $N_{L,h}(u)$ be observation count above and below u, respectively. The fitted right-tail decay rate is $\widehat{\zeta}_{R,h}(u) = N_{R,h}(u)[\sum_{i \in h: g_i > u}(g_i - u)]^{-1}$, and left-tail rate is $\widehat{\zeta}_{L,h}(u) = N_{L,h}(u)[\sum_{i \in h: g_i < u}(u - g_i)]^{-1}$. The curves plot the corresponding outward log-tail slopes, $-\widehat{\zeta}_{R,h}(u)$ and $-\widehat{\zeta}_{L,h}(u)$. Shaded regions are pointwise 95 percentile intervals from 1,000 plant-cluster bootstrap replications. Same cluster resamples are used for every curve.

3.4 Empirical Benchmarks

The preceding analysis establishes five empirical facts for a statistical model of productivity growth.

I. Core concentration and far-tail mass. Productivity growth is not well represented by a Gaussian distribution. The real distribution combines greater concentration near its mode, less probability between two and three standard deviations, and greater probability beyond three standard deviations. The model must reproduce this combination of a concentrated core, relatively thin shoulders, and prominent far tails.

II. Multidimensional size heterogeneity. Mean growth and the quantile ratio increase with size, while skewness shifts from negative among small and medium units to positive among large units. Dispersion is nonmonotonic, and large units have the highest far-tail incidence.

III. Nonlinear, size-dependent mean reversion. Large adverse shocks predict rebounds, while large favorable shocks predict subsequent contractions. Local slopes remain negative, but their magnitude varies substantially. Large units exhibit stronger two-sided buffering in conditional means, alongside a deeper *Icarian liftoff* near the negative-growth trough. This demanding ascent gives way to pronounced moderation within the *Daedalian corridor* and comparatively weak marginal reversal at high positive realizations. Small units experience a milder liftoff but weaker buffering in levels and a more pronounced, sustained upper-tail *Icarian fallout*. Neither a common nor a size-specific constant autoregressive coefficient captures these conditional dynamics.

IV. Distinct tail-incidence and conditional-reach margins. Similar two-sided incidence beyond two group-specific standard deviations masks directional asymmetry: large units enter the left tail less frequently than small units, but the right tail more frequently. Conditional reach, however, is greater among large units on both sides—the *Daedalian range pattern*. The nonconstant decay profiles are inconsistent with a fixed-exponent Pareto-type representation of the tails. The model must therefore distinguish tail-entry probabilities from conditional reach and accommodate both asymmetric incidence and decay.

V. Time-varying scale and asymmetry. Dispersion contracts substantially between the first and final year, while standardized far-tail incidence follows a nonmonotonic path and remains well above the Gaussian distribution. Excess kurtosis is positive and large throughout, and skewness changes sign across years. Changes in scale therefore coexist with changes in distributional shape.

4 A Regime Model of Productivity Growth

The preceding section documents a concentrated core and prominent tails, time-varying dispersion and asymmetry, and nonlinear conditional dynamics that differ systematically with production-unit size. We now develop a dynamic normal-mixture model that brings these distributional and dynamic features into a unified framework.

4.1 A Unified Regime Normal-Mixture Framework

Let $R_{it} \in \mathcal{R}_K \equiv \{1, \ldots, K\}$ denote the latent regime governing $g_{i,t}$. Let $\mathcal{H}_{it}$ contain the production unit's observed productivity-growth and size histories through t, $(g_{i,1}, \ldots, g_{i,t})$ and $(d_{i,1}, \ldots, d_{i,t})$. Define $\pi_{is,t|t} \equiv \Pr(R_{it} = s \mid \mathcal{H}_{it})$ as the probability of regime s conditional on $\mathcal{H}_{it}$, and $\pi_{ir,t+1|t} \equiv \Pr(R_{i,t+1} = r \mid \mathcal{H}_{it})$ as the predicted probability of regime r before observing $g_{i,t+1}$.

For regime r, let $\mu_{r,t+1}(g_{i,t}, d_{i,t})$ and $v_{r,t+1}(g_{i,t}) > 0$ denote the conditional mean and standard deviation of $g_{i,t+1}$. Following McLachlan & Peel (2000), we use finite normal mixtures to approximate non-Gaussian conditional distributions and specify the productivity-growth process as follows:

$$\textbf{Regime transition: } \Pr(R_{i,t+1} = r \mid R_{it} = s, \mathcal{H}_{it}) = p_{sr}(g_{i,t}, d_{i,t}), \; s, r \in \mathcal{R}_K, \tag{3}$$

$$\textbf{Regime distribution: } g_{i,t+1} \,|\, R_{i,t+1} = r, \mathcal{H}_{it} \sim \mathcal{N}\left(\mu_{r,t+1}(g_{i,t}, d_{i,t}), v^2_{r,t+1}(g_{i,t})\right), \tag{4}$$

$$\textbf{Regime probabilities: } \pi_{ir,t+1|t} = \sum_{s=1}^{K} \pi_{is,t|t}\, p_{sr}(g_{i,t}, d_{i,t}), \tag{5}$$

$$\textbf{Predictive density: } f_{i,t+1|t}(y) = \sum_{r=1}^{K} \frac{\pi_{ir,t+1|t}}{v_{r,t+1}(g_{i,t})} \phi\left(\frac{y - \mu_{r,t+1}(g_{i,t}, d_{i,t})}{v_{r,t+1}(g_{i,t})}\right). \tag{6}$$

Equations (3)–(6) separate the productivity process into two primitives and two definitions. Equation (3) specifies the transition law. Conditional on current regime s, the probability that the next realization is generated by regime r is $p_{sr}(g_{i,t}, d_{i,t})$, with $\sum_{r=1}^{K} p_{sr}(g_{i,t}, d_{i,t}) = 1$. Dependence on $(g_{i,t}, d_{i,t})$ allows lagged productivity growth and unit size to change subsequent regime probabilities. Conditional on these variables and the current regime, earlier observations in $\mathcal{H}_{it}$ have no additional effect on the transition law. Equation (4) specifies the distribution generated by each destination regime. Its conditional mean $\mu_{r,t+1}(g_{i,t}, d_{i,t})$ may depend on both lagged productivity growth and size, while $v_{r,t+1}(g_{i,t})$ denotes its conditional standard deviation.

Because the regime is unobserved, $\pi_{is,t|t}$ gives the probability that production unit i occupies regime s, conditional on its observed history through t. Equation (5) propagates these filtered probabilities through the transition law to obtain $\pi_{ir,t+1|t}$, the probability of destination regime r before observing $g_{i,t+1}$.

Equation (6) then averages the regime-conditional densities using these predicted probabilities. Here, y denotes a possible realization of $g_{i,t+1}$, and $\phi(\cdot)$ is the standard normal density. The mixture weights vary across production units and their observed histories.

For each regime $r \in \mathcal{R}_K$, we decompose the mean and standard deviation as

$$\mu_{r,t+1}(g,d) = m_r(g,d) + \delta_{r,t+1}, \qquad v_{r,t+1}(g) = \sigma_r \tau_{r,t+1} \sqrt{\mathcal{V}_r(z)},$$

where $z = (g - \bar{g})/s_g$, and $\bar{g}$ and s_g are the pooled mean and standard deviation of lagged productivity growth in the sample. The additive term $\delta_{r,t+1}$ and multiplicative factor $\tau_{r,t+1} > 0$ capture regime-year shifts in location and volatility, respectively. The parameter $\sigma_r > 0$ provides the regime-specific baseline scale, while $\mathcal{V}_r(z) > 0$ captures variation in conditional variance with lagged growth.

The autoregressive location functions and multinomial-logit transition indices are cubic in lagged productivity growth and linear in current size. Both use the eight-term tensor-product basis $\Phi(g,d) = (1,d)' \otimes (1,g,g^2,g^3)'$. The regime-specific conditional mean function is

$$m_r(g,d) = \boldsymbol{\beta}_r' \Phi(g,d) = \sum_{j=0}^{3} \left(\beta_{jr}^0 + \beta_{jr}^d \, d \right) g^j.$$

Size affects both the intercept and the interactions with each polynomial term g.

For a production unit currently in regime s, the destination-regime indices are $\eta_{sr}(z,d) = \boldsymbol{\kappa}_{sr}' \Phi(z,d)$ for $r = 1, \ldots, K-1$, with $\eta_{sK}(z,d) = 0$. The corresponding multinomial-logit transition probabilities are

$$p_{sr}(g,d) = \frac{\exp\{\eta_{sr}(z,d)\}}{\sum_{\ell=1}^{K} \exp\{\eta_{s\ell}(z,d)\}}.$$

To allow volatility to vary flexibly with lagged productivity growth while remaining positive, we specify a cubic Bernstein variance multiplier. Let z_L and z_U denote the pooled first and ninety-ninth percentiles of standardized lagged growth in the estimation sample, and define $u(z) = \min\left\{1, \max\left\{0, \frac{z - z_L}{z_U - z_L}\right\}\right\}$. This transformation maps the central support to $[0,1]$ and clips realizations outside that interval. For regime r, the variance multiplier is

$$\mathcal{V}_r(z) = a_{0r}(1-u(z))^3 + 3a_{1r}u(z)(1-u(z))^2 + 3a_{2r}u(z)^2(1-u(z)) + a_{3r}u(z)^3.$$

Because the Bernstein basis weights are nonnegative and sum to one, $\mathcal{V}_r(z)$ is a convex combination of the controls.[4]

We estimate the model by maximizing the observed-data conditional likelihood, evaluated through the forward recursion. Optimization uses a bound-constrained quasi-Newton algorithm with multiple starting points. Online Appendix B details the estimation procedure and convergence criteria.[5]

4.2 Regime Dynamics and Mean Reversion

We first study conditional dynamics along the lagged-productivity dimension. We then examine how the same law of motion varies with unit size.

1. Selecting the three-state model. Figure 6 and Online Appendix Table 2 compare one-, two-, and three-state specifications estimated on the same sample. The models use the same functional-form families for growth–size dependence, conditional volatility, and annual effects. The one-state conditionally Gaussian model fails to reproduce the combination of a sharply concentrated center and prominent tails. A second state substantially improves the fit, but leaves systematic discrepancies in the shoulders and tails. A third state delivers a further substantial improvement.

In the pooled sample, the CDF Kolmogorov–Smirnov (KS) statistic, which measures the supremum distance between the empirical and model-implied outcome CDFs, falls from 16.17% under one state to 3.22% under two states and 1.61% under three states. The probability integral transform (PIT) KS statistic measures departures of observation-specific predictive probabilities from uniformity and falls from 15.34% to 2.60% and 0.98%, respectively. Finally, the PIT AD statistic, which measures deviations from PIT uniformity while placing greater weight on observations near the tails, falls from 542.48 to 15.89 and then to 1.84.

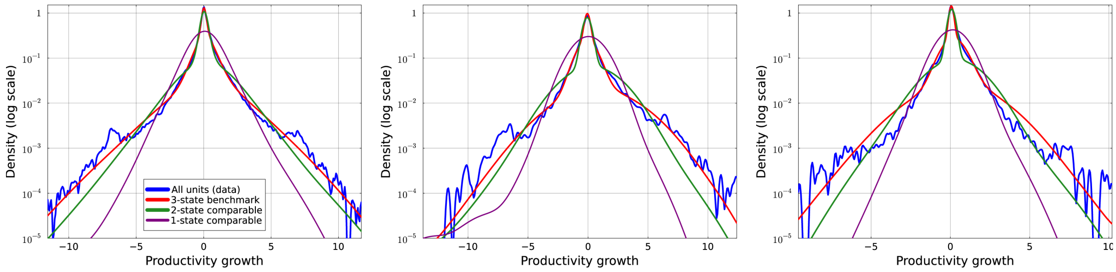
The three-state model achieves the lowest values of all three diagnostics in

[4]We fix $a_{3r} = 1$ to anchor the variance multiplier at the upper boundary of the lagged-growth support. Separately, we restrict $a_{\ell r} \in [1/3, 3]$ for $\ell = 0, 1, 2$ to limit variation in conditional variance to $1/3 \leq \mathcal{V}_r(z) \leq 3$. These bounds constrain conditional volatility. Clipping makes the multiplier constant outside the central support: $\mathcal{V}_r(z) = a_{0r}$ for $z \leq z_L$ and $\mathcal{V}_r(z) = 1$ for $z \geq z_U$.

[5]Forward recursion makes the conditional likelihood computable, integrating out the latent regimes. Maximum likelihood evaluates each observation using its model-implied predictive density. Simulated method of moments would instead require selecting and weighting moments and approximating their model counterparts by simulation. Likelihood-based estimation therefore avoids these additional choices and simulation noise in the estimation objective.

Figure 6: Conditional productivity-growth densities, $f(g_{i,t} \mid \mathcal{H}_{i,t-1})$

(a) Pooled sample **(b) Best PIT 16–17 | 15–16** **(c) Worst PIT 18–19 | 17–18**



Notes: Densities use a log scale over each sample's 0.01st–99.99th percentiles. Blue denotes the empirical kernel density. Red, green, and purple show one-step-ahead predictive densities $\bar{f}_{K,t}(g) = N_t^{-1} \sum_{i=1}^{N_t} f_K(g \mid \mathcal{H}_{i,t-1})$ for $K = 3, 2, 1$, respectively, where N_t counts AR-eligible pairs. Models are estimated on the full panel. Pooled model densities weight annual densities by sample shares. The $K = 3$ PIT KS statistic selects 2018–2019 | 2017–2018 as the worst-calibrated transition and 2016–2017 | 2015–2016 as the best.

every annual transition. It also provides the closest tail match in the pooled sample and in most annual comparisons. These results support the three regime model on both marginal distributional fit and predictive calibration. Figure 6 illustrates the improvement: the three-state predictive distribution better captures the central peak and the decay of both tails. Even in 2018–2019, conditional on information through 2017–2018—the transition with the highest three-state PIT KS statistic—the three-regime model outperforms simpler specifications.

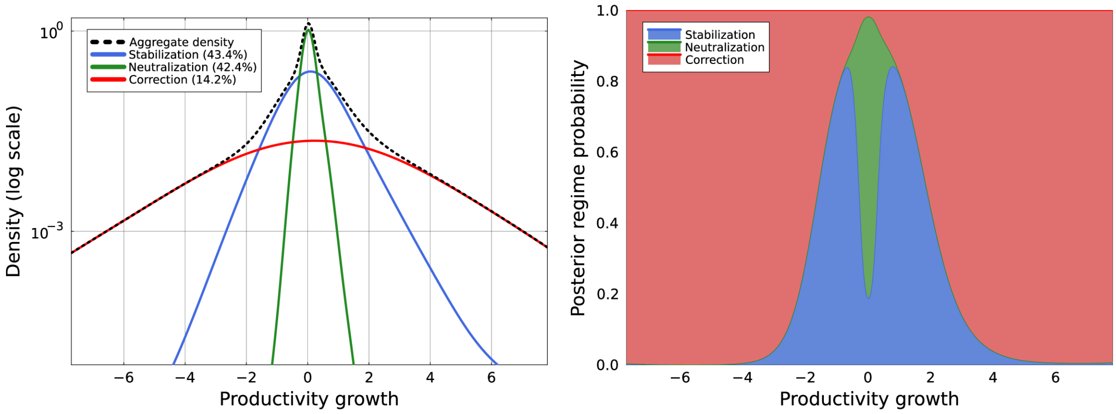
2. Regime anatomy and the Icarus–Daedalus response. Figure 7 and Table 4 describe each regime's contribution to the predictive distribution, while Figure 8 characterizes its dynamic response. Together, these properties motivate the labels *Neutralization*, *Stabilization*, and *Correction*. These labels describe statistical roles rather than structural shocks or permanent production-unit types. Units can move among all three regimes.

Neutralization forms the concentrated core. It accounts for 42.4% of predictive probability, has a standard deviation of 0.186, and contributes negligible probability to the tails. Its conditional-mean schedule remains close to zero and nearly flat. Conditional on this destination regime, preceding productivity growth has little influence on expected growth in the following year.

Stabilization is almost equally prevalent, accounting for 43.4% of predictive probability, but has substantially greater dispersion, with a standard deviation of 0.805. It occupies the regions between the narrow core and the far tails,

Figure 7: Regime Densities and Posterior Probabilities

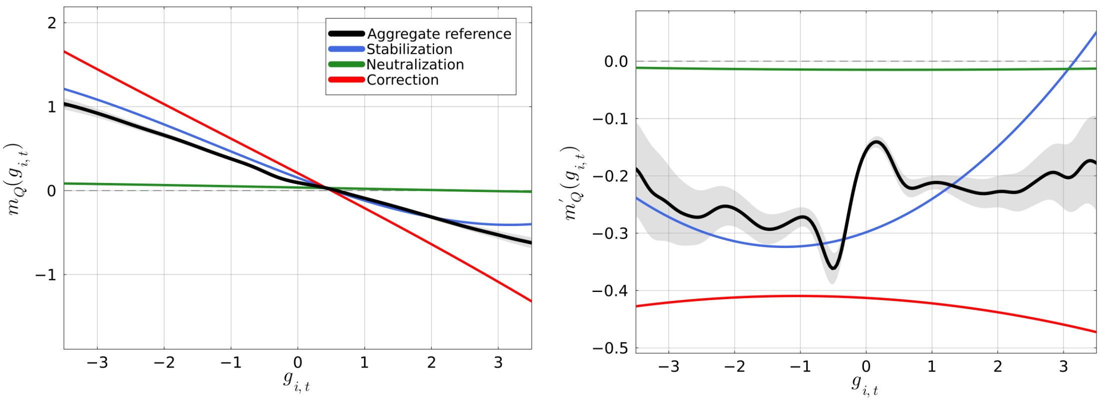
(a) Probability-weighted densities **(b) State probabilities**



Notes: The left panel shows regime k's contribution to the pooled one-step-ahead predictive density, $h_k(g) = N^{-1} \sum_{(i,t)\in\mathcal{A}} \pi_{ik,t|t-1} f_k(g \mid \mathcal{H}_{i,t-1})$, where $\mathcal{A}$ contains the N AR-eligible observations. Legend percentages are the average pre-outcome probabilities $\bar{\pi}_k = N^{-1} \sum_{(i,t)\in\mathcal{A}} \pi_{ik,t|t-1}$, with $\int_{\mathbb{R}} h_k(g)\, dg = \bar{\pi}_k$. The dotted curve is the aggregate density $h(g) = \sum_k h_k(g)$. The right panel shows pooled posterior regime probabilities $q_k(g) = h_k(g)/h(g)$ conditional on growth g, integrating over the empirical distribution of lagged growth, beginning-year size, year, and filtering histories. The displayed range spans the pooled sample's 0.1st–99.9th percentiles.

Figure 8: Conditional Mean Reversion in the Three-State Model

(a) Conditional mean functions **(b) Marginal mean-reversion slopes**



Notes: Colored curves show $m_r(x) = \beta_r' \Phi(x, \bar{d}_r)$ and $m_r'(x)$, net of state–year location effects, where $\bar{d}_r$ is regime r's pooled predictive-probability-weighted mean size. Gaussian-kernel weights $\theta_{it}(x) = K_h(g_{it} - x) / \sum_{j,s} K_h(g_{js} - x)$ give $\bar{\pi}_r(x) = \sum_{i,t} \theta_{it}(x) \pi_{ir,t+1|t}$. Black curves show the aggregate reference $m_Q(x) = \sum_r \bar{\pi}_r(x) m_r(x)$ and $m_Q'(x) = \sum_r [\bar{\pi}_r(x) m_r'(x) + \bar{\pi}_r'(x) m_r(x)]$, combining within-regime responses and changing regime composition. The bandwidth is $h = 5(1.06) \min\{s_g, \mathrm{QR}_g/1.34\} N^{-1/5}$, using all N AR-eligible pairs and the pooled lagged-growth standard deviation and interquartile range. The displayed range is $[-3.5, 3.5]$. Gray regions apply only to the black curves and give the outer envelope of two 95% simultaneous linearized production-unit-cluster multiplier bands, each calibrated jointly across the mean and slope grids using 10,000 draws from either the mean-one exponential or Rademacher family.

contributing 24.7% and 22.8% of the probability in the lower and upper 5% tails, respectively. Its dynamic response is asymmetric. Adverse realizations predict a subsequent rebound, while the marginal reversal of favorable realizations

Table 4: Pooled and Annual Predictive Moments of the Three Regimes

Panel A. Marginal fit and conditional calibration

	Distributional statistics					Tail contributions (%)			
Group	%	Mean	SD	Skew.	Kurt.	Left 1%	Right 1%	Left 5%	Right 5%
Aggregate	100	0.103	1.164	0.366	18.480	100.0	100.0	100.0	100.0
Stabilization	43.4	0.133	0.805	1.287	44.546	0.3	1.6	24.7	22.8
Neutralization	42.4	0.029	0.186	0.264	1.428	0.000	0.000	0.000	0.000
Correction	14.2	0.230	2.727	−0.048	0.916	99.7	98.4	75.3	77.2

Panel B: Annual regime shares, means, and standard deviations

	Aggregate		Stabilization			Neutralization			Correction		
Outcome	Mean	SD	%	Mean	SD	%	Mean	SD	%	Mean	SD
12–13	0.159	1.810	44.2	0.333	1.310	39.8	0.048	0.200	16.0	−0.048	3.933
13–14	0.075	1.529	43.0	0.199	0.975	42.0	0.059	0.193	15.0	−0.237	3.555
14–15	−0.057	1.573	43.4	0.088	0.959	41.6	0.021	0.202	15.0	−0.690	3.640
15–16	0.342	1.372	43.1	0.431	0.891	41.8	0.184	0.274	15.0	0.528	3.149
16–17	0.047	1.353	44.9	−0.079	0.880	40.4	−0.060	0.220	14.7	0.723	3.065
17–18	0.248	1.507	44.8	0.293	1.128	40.8	0.053	0.231	14.5	0.661	3.363
18–19	0.167	1.064	44.4	0.212	0.757	40.9	0.062	0.141	14.6	0.323	2.424
19–20	−0.067	0.936	43.0	−0.089	0.668	43.5	−0.049	0.158	13.6	−0.059	2.228
20–21	0.076	0.936	42.9	0.095	0.624	43.4	0.026	0.137	13.6	0.174	2.261
21–22	0.134	0.832	42.6	0.184	0.574	43.9	0.054	0.144	13.5	0.232	1.994
22–23	0.058	0.839	42.4	0.068	0.526	44.0	0.002	0.114	13.6	0.209	2.057

Notes: Predictive regime probabilities are formed before each outcome by filtering within contiguous AR-eligible spells. Panel A reports moments of the pooled predictive distribution and its normalized regime components. For regime k, the left and right q% columns report $100\,\mathrm{Pr}(R = k \mid g \leq Q_{q/100})$ and $100\,\mathrm{Pr}(R = k \mid g \geq Q_{1-q/100})$. Contributions below 0.001% are displayed as zero. The pooled predictive quantiles are $Q_{0.01} = -3.671$, $Q_{0.99} = 4.150$, $Q_{0.05} = -1.338$, and $Q_{0.95} = 1.757$. Panel B reports annual predictive means, standard deviations, and regime shares; 12–13 uses pre-outcome information through 11–12. Means, SDs, and cutoffs are in log-growth units. Skew. denotes skewness and Kurt. denotes excess kurtosis.

weakens as growth increases. Stabilization thus combines recovery after negative realizations with comparatively greater preservation of favorable shocks.

Correction accounts for only 14.2% of predictive probability but has a standard deviation of 2.727. It generates 99.7% of the lower 1% tail and 98.4% of the upper 1% tail, together with approximately three-quarters of both 5% tails. Its dominance concerns outcomes on both tails. Its conditional-mean schedule also exhibits the strongest two-sided reversal: adverse growth predicts the largest rebound, whereas favorable growth predicts the largest subsequent contraction. Correction combines broad tail risk with forceful mean reversion. Stabilization, by contrast, preserves more of preceding favorable gains while providing a smaller recovery after adverse realizations.

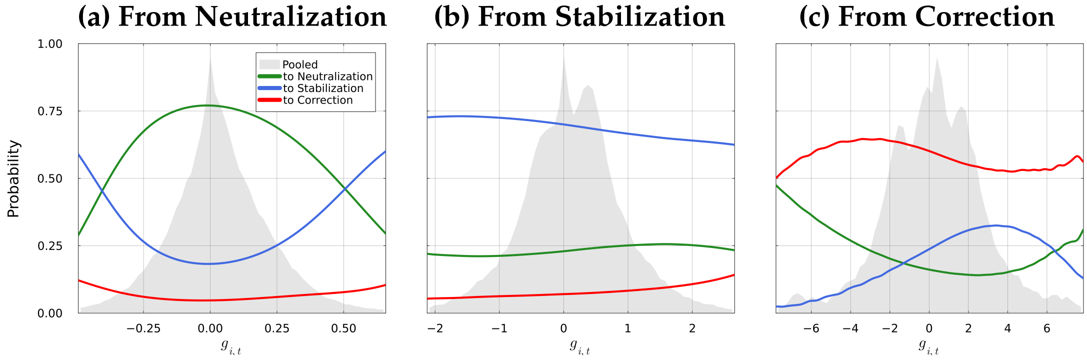
The black curves in Figure 8 combine these responses into an aggregate conditional mean schedule. Each regime's mean function is evaluated at its fixed, probability-weighted mean size, $\bar{d}_r$. Letting $\bar{\pi}_r(g)$ denote locally averaged pre-outcome regime probabilities, the mean and its derivative are $m_Q(g) = \sum_{r=1}^{3} \bar{\pi}_r(g) m_r(g, \bar{d}_r)$ and $m'_Q(g) = \sum_{r=1}^{3}(\bar{\pi}_r(g) m'_r(g, \bar{d}_r) + \bar{\pi}'_r(g) m_r(g, \bar{d}_r))$. The first derivative component captures within-regime adjustment; the second captures variation in regime composition along the growth distribution.

These schedules recover the qualitative sequence of the *Icarus–Daedalus pattern*. The *Icarian liftoff* culminates in a slope of approximately -0.36 near $g = -0.51$: improvements in g withdraw part of the subsequent rebound. At this trough, within-regime adjustment contributes approximately -0.22 and changing composition contributes -0.14. The slope then moderates to approximately -0.14 near $x = 0.14$, defining the *Daedalian corridor*. Around zero, Neutralization receives approximately 54% of predictive probability, giving substantial weight to its nearly flat response. Farther into positive growth, increasing exposure to Correction accompanies renewed marginal reversal, with the aggregate slope reaching approximately -0.23 near $g = 1.77$. This *Icarian fallout* is not a uniformly intensifying upper-tail contraction. Stabilization's response becomes progressively flatter, and the aggregate slope subsequently moderates to approximately -0.18 near the right boundary despite Correction's increasing weight. The aggregate pattern therefore reflects the interaction between regime composition and distinct within-regime responses.

The annual evidence preserves this division of roles. Neutralization remains the least dispersed regime and Correction the most dispersed in every year. Regime shares remain comparatively stable across years. State-year location and scale effects absorb an important part of the time variation while maintaining a concentrated core, an intermediate stabilization component, and a less frequent correction regime that dominates both far tails.

3. Transitions between regimes. Figure 9 reports destination probabilities separately for each origin regime. The gray distributions show productivity

Figure 9: Transition Probabilities in the 3-State Model



Notes: For origin regime r, curves report $\bar{p}_{rs}(g)$, averaging $p_{rs}(g, d_{i,t})$ over observed size using Gaussian-kernel weights proportional to $K_h(g_{i,t} - g)\pi_{ir,t|t}$. Here, $\pi_{ir,t|t}$ is the filtered origin-regime probability. Green, blue, and red denote transitions to Neutralization, Stabilization, and Correction, respectively. The gray area ("Pooled") shows the origin-probability-weighted distribution of g_{it}, rescaled to peak at 0.95. Each panel spans its origin-weighted 0.5th–99.5th percentiles. The common bandwidth is $h = 5(1.06)\min\{s_g, \mathrm{QR}_g/1.34\}N_T^{-1/5}$, where s_g and QR_g are the pooled lagged-growth standard deviation and interquartile range, and N_T counts contiguous transition-eligible pairs.

growth weighted by the filtered probabilities. Each panel spans the corresponding 0.5th–99.5th percentiles, so the displayed ranges differ across regimes.

Panel (a) identifies Neutralization as a persistent central regime. Within its narrow range, from -0.49 to 0.66, the probability of remaining in Neutralization peaks near zero at 0.77, compared with 0.18 for Stabilization and 0.05 for Correction. Departures from the center reduce persistence and raise the probability of Stabilization. At both boundaries, Stabilization becomes the most likely destination, receiving approximately 0.59–0.60 of transition probability, while Neutralization falls to about 0.29. Direct transitions to Correction remain comparatively infrequent, although their probability rises toward both boundaries. Departures from the neutral core lead primarily toward intermediate dispersion rather than directly into the tail regime.

Panel (b) shows that Stabilization is persistent throughout its support, from -2.12 to 2.66. Its retention probability ranges from 0.62 to 0.73 and is approximately 0.70 near zero. Remaining in Stabilization is always the most likely outcome. Neutralization is the principal alternative, with probabilities around 0.21–0.26, while movement into Correction becomes more likely as growth increases, rising from approximately 0.05 at the lower bound to 0.14 at the upper bound. Stabilization is a persistent adjustment regime, with exits favoring the neutral core but increasing exposure to Correction following favorable shocks.

Panel (c) shows substantial persistence in Correction across a much wider range, from -7.80 to 7.86. Remaining in Correction is the most likely destination throughout, with probabilities between 0.50 and 0.65 and approximately 0.60 near zero. Exit routes depend on the preceding realization. Neutralization is the main alternative following strongly negative growth. Stabilization becomes the more likely exit around the center and over much of the positive-growth region, while Neutralization regains prominence near the far-right boundary. High dispersion can persist across multiple periods, with subsequent adjustment occurring within Correction or through transitions to less dispersed regimes.

Together, the panels distinguish central persistence in Neutralization, sustained adjustment with favorable buffering in Stabilization, and persistent exposure to high dispersion and strong mean reversion in Correction. The conditional-mean functions determine expected subsequent growth within each destination regime, while the transition law assigns probabilities to regimes with sharply different dispersions and tail contributions. Adjustment therefore combines destination-specific mean responses with movements among regimes, allowing the model to distinguish an isolated extreme realization from a persistent episode of elevated tail risk.

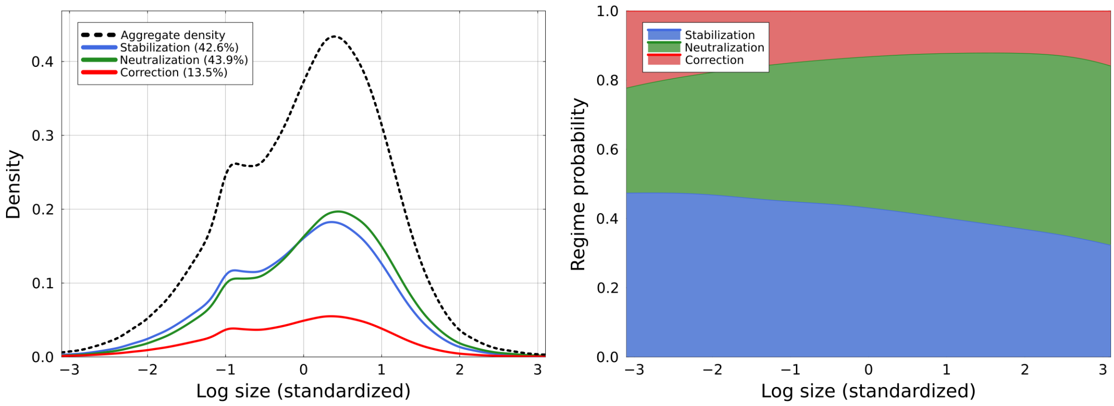
4.3 Firm Size, Regime Composition, and Productivity Risk

We now examine the same estimated three-state model along the size dimension. The specification retains cubic growth terms and linear size interactions; no separate model is estimated for small, medium, or large units.

1. Size and regime composition. Figure 10 summarizes how regime composition varies with size. Panel (a) decomposes the empirical size density into contributions $f_d(d)\pi_r(d)$, where $\pi_r(d)$ is the stationary regime probability implied by the locally averaged transition matrix at size d. Averaging over the displayed size distribution gives shares of 43.9% for Neutralization, 42.6% for Stabilization, and 13.5% for Correction. Panel (b) shows a reallocation toward Neutralization across the central size range. Between the small- and large-unit thresholds, $d = -1.33$ and $d = 1.33$, its probability rises from approximately 38.7% to 48.8%. Stabilization declines from 45.5% to 39.1%, while Correction falls from 15.7% to 12.1%. Larger size is therefore associated with greater weight on the narrow, low-dispersion core regime.

Figure 10: Firm Size and Model-Implied Regime Composition

(a) Regime-weighted size densities **(b) Stationary regime probabilities**

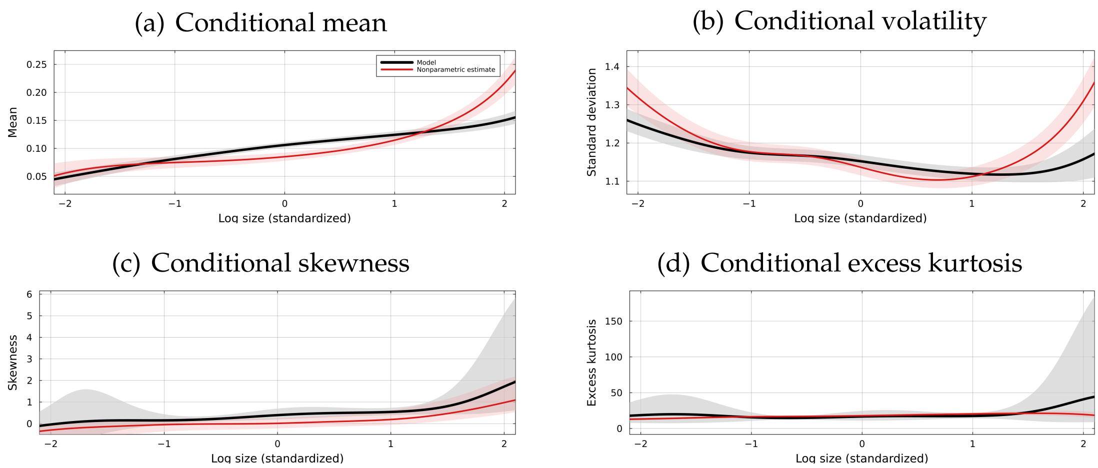


Notes: At each d, the fitted transition matrix is averaged over the kernel-local empirical distribution of lagged productivity, using filtered origin-state probabilities. Let $\bar{P}(d)$ denote this averaged matrix; the stationary probabilities satisfy $\pi(d)'\bar{P}(d) = \pi(d)'$. Panel (a) plots the empirical size density $f_d(d)$ and the regime contributions $f_d(d)\pi_r(d)$, which sum pointwise to the aggregate density. Shares are their integrals over the support, normalized to sum to 100%. Panel (b) plots $\pi_r(d)$, which sums to one at every d. Both panels cover $d \in [-3.1, 3.1]$. *Notes*: At each size d, $\bar{P}(d)$ averages the fitted transition probabilities $p_{rs}(g_{i,t}, d)$ row by row over contiguous transition-eligible pairs, using Gaussian size-kernel weights multiplied by filtered origin-regime probabilities. The resulting local stationary probabilities satisfy $\pi(d)'\bar{P}(d) = \pi(d)'$. Panel (a) plots the empirical size kernel density $f_d(d)$ and the contributions $f_d(d)\pi_r(d)$, which sum pointwise to $f_d(d)$. Legend percentages integrate these contributions over the displayed support and normalize them to sum to 100%. Panel (b) plots $\pi_r(d)$, with $\sum_r \pi_r(d) = 1$ at every d. Both panels cover standardized log size $d \in [-3.1, 3.1]$.

This reallocation is not monotonic throughout the size distribution. Correction reaches its minimum near $d = 1.6$ and subsequently rises, reaching 15.8% at the upper boundary. Neutralization continues increasing to approximately 52.0% near $d = 2.8$, then edges slightly lower, while Stabilization declines monotonically. Thus, the shift toward a concentrated core coexists with renewed exposure to Correction among the largest units.

2. Moments across the size distribution. Figure 11 compares model-implied and empirical moments across size. For each observation, the model constructs the one-step-ahead predictive mixture using pre-outcome regime probabilities. We kernel-average its first four raw moments over $d_{i,t}$ and then obtain the mean, standard deviation, skewness, and excess kurtosis, shown in black. Applying the same kernel weights to realized productivity growth yields the red curves.

Figure 11: Moments over the Firm-Size Distribution



Notes: For each AR-eligible pair, let $M^P_{it,k} = \mathbb{E}[g^k_{i,t+1} \mid \mathcal{H}_{it}]$, $k = 1, \ldots, 4$, denote the three-state model's raw predictive moments. We smooth these over current size using $\widehat{M}^P_k(d) = (\sum_{i,t} K_h(d_{i,t} - d))^{-1} \sum_{i,t} K_h(d_{i,t} - d) M^P_{it,k}$. Black curves transform the model's moments into the mean, standard deviation, skewness, and excess kurtosis. Red curves apply the same procedure with $g^k_{i,t+1}$ replacing $M^P_{it,k}$. The Gaussian bandwidth is $h = 5(1.06)\min\{s_d, \mathrm{QR}_d/1.34\}N^{-1/5}$, where s_d and QR_d describe standardized log size across all $N = 267{,}285$ eligible observations. Gray and pale-red regions are 95% production-unit-cluster multiplier bands based on 10,000 mean-one exponential draws. Bands are simultaneous over size within each panel and calibrated separately for the model and nonparametric estimates. Model bands combine constrained linearized updates of all 146 parameters with empirical reweighting; nonparametric bands recompute the kernel-weighted moments under the same cluster multipliers.

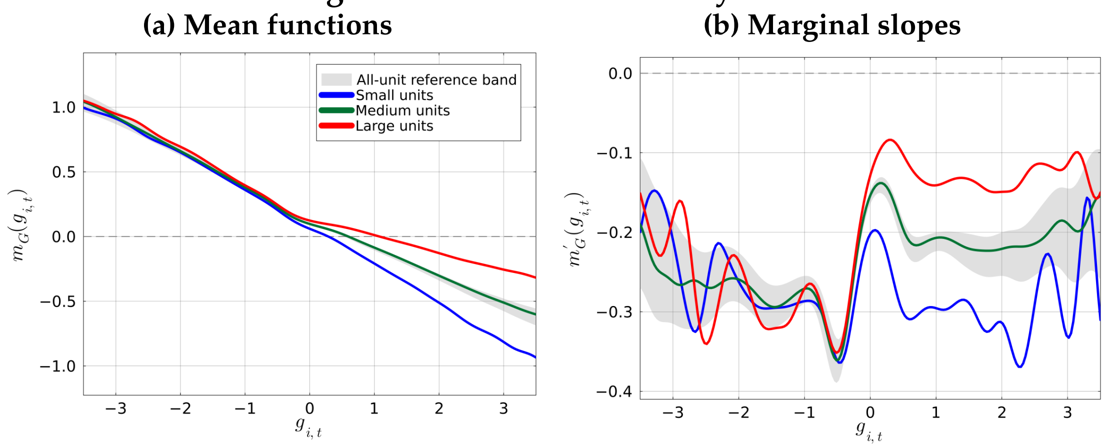
Both estimates associate larger size with higher expected productivity growth. Across the displayed support, the model-implied mean rises from 0.045 to 0.155, compared with 0.051 to 0.239 in the nonparametric estimate. Conditional volatility follows a U-shaped profile. The model-implied standard deviation falls from 1.26 to a minimum of 1.12 before rising to 1.17; the empirical estimate falls from 1.34 to 1.10 and then rises to 1.36. The model thus reproduces the positive size gradient in growth and the broad U-shaped pattern in dispersion.

Skewness increases broadly with size, changing from negative to positive in both estimates, while excess kurtosis remains large throughout. The model and nonparametric confidence bands overlap across the displayed support. The model therefore captures two central features of the higher moments: the shift toward right-skewness with size and the persistent prominence of extreme realizations. Differences in their point estimates should be viewed alongside the uncertainty surrounding these moments, particularly among the largest units, where the wider bands make their precise magnitudes less tightly determined.

These predictive-moment profiles reflect more than the stationary regime shares in the preceding figure. Size enters both transition probabilities and conditional means, changing predictive regime weights, regime location, and regime dispersion. The distributions of lagged productivity, years, and filtering histories also vary across size, altering the composition of conditional means and variances. Within-regime volatility has no direct size term: its contribution varies through lagged productivity and regime-year effects.

3. Size-dependent conditional mean reversion. Figure 12 provides the model-based counterpart to the *Icarus–Daedalus pattern*. In Panel (a), the large-unit schedule lies above those of small and medium units throughout the displayed support. Large units exhibit a stronger expected rebound following adverse realizations and milder subsequent contractions following favorable ones. Their reference mean crosses zero around $g_{i,t} = 1.08$, compared with 0.60 for medium units and 0.29 for small units. The model thus reproduces the *Daedalian level effect*: larger units display stronger two-sided buffering.

Figure 12: Mean Reversion by Size



Notes: Panel (a) reports size-group conditional means. Regime-specific functions are evaluated at each group's regime-specific, predictive-probability-weighted mean sizes. Mixture weights are Gaussian-kernel averages of the pre-outcome probabilities $\pi_{ir,t+1|t}$ within each group. Panel (b) differentiates these means, including both within-regime responses and changes in regime composition. State–year location effects are excluded. The common bandwidth is $h = 5(1.06)\min\{s_g, \mathrm{QR}_g/1.34\}N^{-1/5}$, using the pooled lagged-growth standard deviation and interquartile range across all $N = 267{,}285$ AR-eligible observations. The displayed range is $[-3.5, 3.5]$. Gray regions reproduce the all-firm reference envelope from Figure 8, combining the exponential and Rademacher 95% simultaneous linearized production-unit-cluster bands.

Panel (b) shows that this level ordering coexists with nonlinear marginal adjustment. All three profiles exhibit a pronounced trough near $g_{i,t} = -0.5$, with slopes of approximately -0.36 for small and medium units and -0.35 for large units. This *Icarian liftoff* therefore has a similar depth across size groups in the model. Near the center, marginal reversal moderates sharply. The slopes reach local maxima of approximately -0.20, -0.14, and -0.08 for small, medium, and large units, respectively. The *Daedalian corridor* is consequently most pronounced among large units. Beyond the corridor, renewed marginal correction is strongest among small units. At $g_{i,t} = 2$, their slope is approximately -0.31, compared with -0.22 for medium units and -0.15 for large units. Large units retain weaker marginal reversal than small units throughout the positive-growth region. The model therefore associates the *Icarian fallout* primarily with small units, while larger units combine more favorable conditional means with greater preservation of incremental gains following positive shocks.

4. Size and productivity-risk exposure. Taken together, Figures 10–12 associate larger size with several favorable features of productivity dynamics. First, larger units place more stationary probability in the narrow Neutralization regime. Second, they exhibit higher expected growth and more favorable tail exposure: right-tail incidence is higher, while left-tail incidence is lower. Third, larger units display stronger two-sided buffering: adverse realizations are followed by stronger expected recoveries, while favorable realizations encounter milder subsequent contractions. Their weaker upper-tail marginal reversal also allows them to preserve more of each incremental productivity gain.

Larger units combine higher expected growth, more favorable tail exposure, and greater conditional resilience. These advantages do not eliminate extreme outcomes: distributions remain strongly leptokurtic, and both volatility and Correction exposure turn upward among the largest units. Nor does unit-level resilience necessarily imply limited aggregate risk. An extreme realization at a large unit can have substantial aggregate consequences even when such episodes are infrequent. The next section evaluates how these size-dependent productivity dynamics interact with aggregation weights.

5 The Macroeconomics of Productivity Risk

The preceding sections characterize the nonlinear, size-dependent dynamics of unit-level productivity growth. We now examine their aggregate implications. First, we assess whether the microeconomic measures track manufacturing fluctuations. Second, we study the distribution of aggregate outcomes implied by the estimated productivity process. The first exercise provides external validation; the second shows that the estimated microeconomic law of motion has aggregate distributional effects.

5.1 From Microeconomic Productivity to Aggregate Outcomes

Our aggregation follows the distorted-economy growth-accounting framework of Baqaee & Farhi (2020), which separates changes in aggregate TFP into technological and allocative-efficiency components. We focus on the technological component, $d \log \mathrm{TFP}^{\mathrm{tech}}_{t+1} = \sum_i \widetilde{\lambda}_{i,t}\, g_{i,t+1}$, where $\widetilde{\lambda}_{i,t}$ is the cost-based Domar weight of production unit i. Applying the same aggregation operator to the cross-fitted AR(2) innovations yields the corresponding aggregate innovation index. We leave the allocative-efficiency component outside this exercise because measuring it would require additional structural and parametric assumptions.[6]

Within each year, we remove the bottom and top 7.5% of observations separately for productivity growth and innovations. For either series x, let $\mathcal{R}^x_t$ denote the retained observations and $\mathcal{M}_t$ the sample before trimming. We construct $G^x_t = \frac{100}{\rho^x_t} \sum_{i \in \mathcal{R}^x_t} \widetilde{\lambda}_{i,t-1} g_{i,t}$ with $\rho^x_t = \frac{\sum_{i \in \mathcal{R}^x_t} \tilde{\lambda}_{i,t}}{\sum_{i \in \mathcal{M}_t} \tilde{\lambda}_{i,t}}$; dividing by ρ^x_t restores the scale of the matched sample's total Domar weight after trimming.

Table 5 compares these aggregates with nine manufacturing measures from KLEMS, ASI, NSO, and the manufacturing IIP over 2013–2014 through 2022–2023. Correlations are positive for every index. The strongest association is with NSO real manufacturing GVA: 0.756 for productivity growth and 0.870 for in-

[6] The cost-based Domar weights are equilibrium objects computed annually as $\widetilde{\lambda}_t = \widetilde{\Psi}'_{x,t} \beta_t \chi^c_t$, where χ^c_t is the final expenditure share in GDP, β_t is the vector of final-expenditure shares, and $\widetilde{\Psi}_{x,t}$ is a cost-based Leontief inverse that accounts for inventories, capital investment and market-clearing-consistent constraints. At the plant–commodity level, the corresponding weight is $\widetilde{\lambda}_{ipt} = \widetilde{\lambda}_{gt} \sigma_{ipt}$, where σ_{ipt} is the unit's share of sales of commodity g.

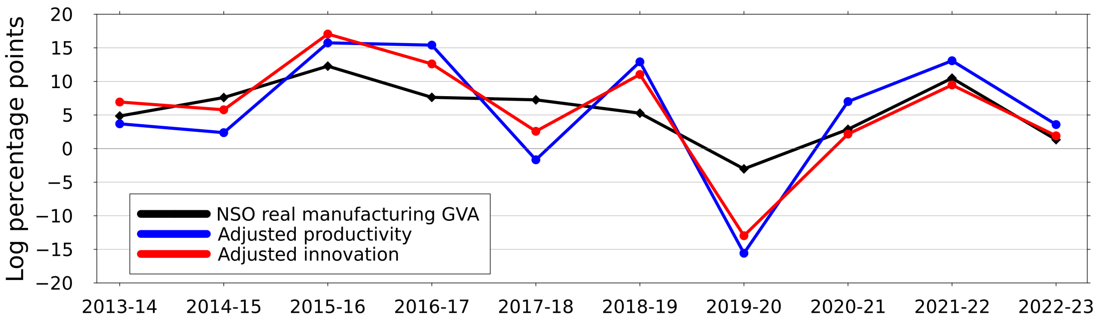
novations. The corresponding correlations are 0.708 and 0.839 with KLEMS gross-output TFP, and 0.738 and 0.854 with KLEMS value-added TFP. Figure 13 plots both aggregates alongside NSO real manufacturing GVA growth. For visual comparison, each aggregate is shifted to match the ten-year mean: $G_t^{P,\text{adjusted}} = G_t^P + 1.044$ and $G_t^{\varepsilon,\text{adjusted}} = G_t^\varepsilon + 7.137$. These constant shifts leave correlations and the timing of fluctuations unchanged. Both aggregates track important cyclical movements in manufacturing activity, with particularly strong comovement for innovations. Given the ten annual observations, we interpret this evidence as descriptive external validation.

Table 5: Correlation With Aggregate Data

	Productivity Shock	Innovation Shock
Aggregate Measure	Correlation	Correlation
KLEMS gross-output TFP	0.708	0.839
ASI real output per person	0.574	0.517
KLEMS real gross output	0.446	0.463
ASI real output	0.449	0.447
NSO real manufacturing GVA	0.756	0.870
KLEMS VA-TFP	0.738	0.854
KLEMS real GVA	0.697	0.827
ASI real output per worker-day	0.508	0.380
Manufacturing IIP	0.291	0.403

Notes: Entries are correlations between Cost–Domar-weighted aggregates of productivity growth $g_{i,t}$ and innovations ε_{it} and aggregate manufacturing indices across ten annual observations, 2013–14 through 2022–23. Each shock is trimmed separately by year, removing the bottom and top 7.5%. Weighted sums are divided by the retained share of matched Cost–Domar weight.

Figure 13: Domar Productivity and Innovation Aggregation and NSO real manufacturing GVA



Notes: The productivity (blue) and innovation (red) aggregates are shifted upward by 1.044 and 7.137 log percentage points, respectively, to match the sample mean of NSO real manufacturing GVA growth (black). The sample covers 2013–14 through 2022–23.

5.2 Granularity and the Distribution of Aggregate Technology

Gabaix (2011) shows how firm-size concentration limits the diversification of idiosyncratic shocks, while Acemoglu et al. (2012) show how asymmetries in production-network linkages allow these shocks to generate aggregate fluctuations. We complement these results by examining how our nonnormal and size-dependent productivity law of motion shapes the full distribution of aggregate technology growth.

Figure 14 crosses two aggregation rules with three microeconomic specifications. The *size-dependent* three-regime model is our benchmark. The *Gaussian* alternative retains size dependence and the lagged-growth and volatility specifications, but restricts the process to one conditionally normal regime. The *size-independent* alternative retains three regimes but removes size from the conditional means and transition probabilities. Both alternatives are separately estimated at the microeconomic level; neither is constructed by matching aggregate moments. These comparisons assess the aggregate implications of restricting regime heterogeneity or size dependence.

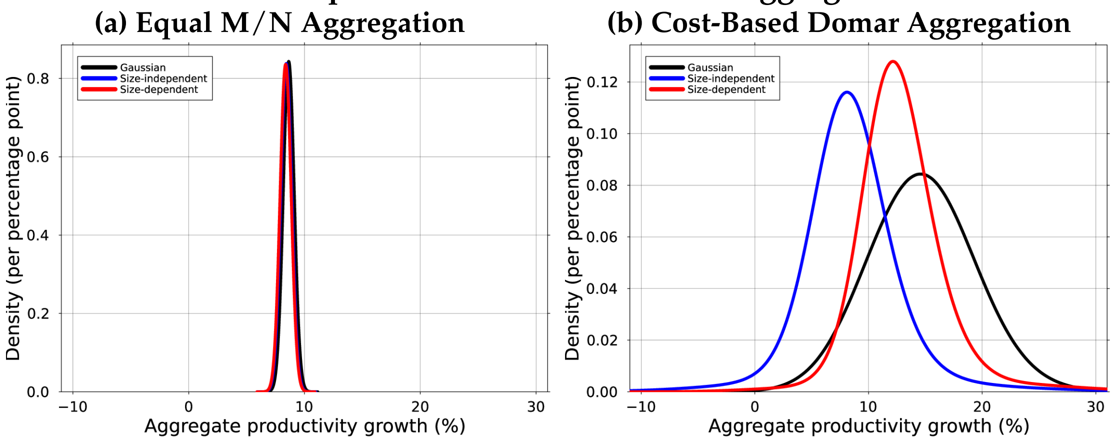
For each specification, we draw independently across units from numerically approximated stationary microeconomic distributions to simulate one billion economies containing the 52,730 production units in the 2022 economy. Unit sizes and the 2022–23 year effects remain fixed. Equal aggregation assigns each unit weight M/N, where $M = \sum_i \widetilde{\lambda}_i$ is the cost-based input-output multiplier. Cost–Domar aggregation instead retains the heterogeneous weights $\widetilde{\lambda}_i$.[7] The two rules therefore hold total exposure constant. For aggregation rule a, economy b, and microeconomic law ℓ, aggregate growth is

$$G^a_{b\ell} = \sum_{i=1}^{N} \omega^a_i g_{ib\ell}, \qquad X^a_{b\ell} = 100 \left[\exp\left(G^a_{b\ell}\right) - 1\right], \qquad \omega^a_i \in \left\{ \frac{M}{N}, \widetilde{\lambda}_i \right\}.$$

Panel (a) shows that M/N aggregation largely diversifies the microeconomic distributional differences. Mean growth is 8.38% under the size-dependent

[7] Although the estimated law conditions on sales rather than on cost-based Domar weights, the two measures are strongly correlated. Across 267,285 AR-eligible plant–commodity observations, their correlation is 0.956 in logs and 0.949 in levels. Thus, sales conditioning captures the main size-exposure sorting relevant for cost-based Hulten aggregation.

Figure 14: One Billion Simulated Economies: Aggregate Productivity Growth under Equal and Cost-Based Domar Aggregation



Notes: Panel (a) uses equal M/N weights, where $M = \sum_i \widetilde{\lambda}_i$; panel (b) uses cost-based Domar weights $\widetilde{\lambda}_i$, preserving total exposure. Black, blue, and red represent separately estimated models: Gaussian ($K = 1$, size-dependent), Size-independent ($K = 3$, no size terms), and Size-dependent ($K = 3$, own-size conditioning). All use cubic lagged-growth terms. Particle simulations approximate stationary microeconomic distributions after 1,000 burn-in transitions for the Gaussian and Size-independent models and 2,000 for the Size-dependent model. Each simulated economy is one joint realization of productivity growth for the 52,730 units in the 2022 sample, drawn independently across units from these distributions. We generate one billion such economies per model, holding sizes, weights, and 2022–23 year effects fixed. The horizontal axis reports $100[\exp(G_{b\ell}) - 1]$. No centering, additive calibration, or trimming is applied.

Table 6: Aggregate Statistics under Cost-Based Domar Aggregation

Statistic	Gaussian	Size-independent	Size-dependent
Mean (%)	14.69	8.53	12.95
Standard deviation	4.73	4.64	4.25
Skewness	0.084	0.676	1.382
Excess kurtosis	0.021	5.724	6.917
$\Pr(\lvert z\rvert > 2)$	4.54%	4.91%	4.47%
$\Pr(\lvert z\rvert > 3)$	0.29%	1.90%	1.89%

Notes: Summarizes the cost-Domar-distributions in Figure 14(b). Statistics refer to aggregate productivity growth, $X_\ell = 100[\exp(G_\ell) - 1]$. Within each distribution, $z_\ell = (X_\ell - \mu_\ell)/s_\ell$ uses its own mean and standard deviation. Skewness is $\mathbb{E}[z_\ell^3]$ and excess kurtosis is $\mathbb{E}[z_\ell^4] - 3$.

benchmark, compared with 8.64% under the Gaussian and 8.43% under the size-independent specification. Standard deviations lie between 0.47 and 0.48 percentage points, and all three aggregate distributions have little skewness or excess kurtosis. Size dependence retains a modest location effect, but the non-Gaussian features have little aggregate effect.

Under cost–Domar weighting, these differences become substantial. Table 6 shows that the size-dependent distribution combines mean growth of 12.95%

with a standard deviation of 4.25 percentage points, skewness of 1.382, and excess kurtosis of 6.917. Relative to these moments, the Gaussian raises mean growth to 14.69% and the standard deviation to 4.73, while producing an almost symmetric distribution with negligible excess kurtosis. Yet greater dispersion does not imply greater far-tail risk: the probability of a realization beyond three distribution-specific standard deviations is only 0.29% under the Gaussian distribution, compared with 1.89% under our benchmark. Incidence beyond two standard deviations is much closer, at 4.54% and 4.47%, respectively.

The event probabilities introduced in Table 1 make the aggregate economic implications explicit. Defining a productivity recession as negative aggregate productivity growth, the size-dependent distribution assigns this event 0.315% probability, compared with 0.0553% under the Gaussian specification—a 5.69-fold difference. Conversely, the Gaussian assigns 11.39% probability to growth between 20% and 25%, compared with 2.79% under the benchmark. Normality understates contraction risk while substantially overstating the frequency of expansions. This optimism does not extend uniformly into the far-right tail: the probability of growth above 25% is similar across the two distributions, and slightly higher under the size-dependent law, at 1.74% versus 1.66%. Non-normality changes the allocation of probability across economically distinct outcomes, not simply the aggregate variance.

Removing size dependence has a different effect. Mean growth falls from 12.95% to 8.53%, while the standard deviation rises from 4.25 to 4.64 percentage points. Skewness falls from 1.382 to 0.676, although excess kurtosis remains substantial at 5.724 and incidence beyond three standard deviations remains close to 1.9%. The event probabilities nevertheless change sharply. Productivity-recession probability rises from 0.315% to 2.47%, and the probability of growth between 0% and 5% rises from 1.19% to 14.39%. At the upper end, the probability of growth above 25% falls from 1.74% to 0.78%. Retaining the size-conditioned process therefore combines higher expected growth and lower contraction risk with greater positive asymmetry, while leaving substantial standardized far-tail risk. This comparison is consistent with the preceding evidence of more favorable large-unit dynamics: suppressing size dependence replaces that sorting with a common productivity law.

The contrast between the two aggregation rules clarifies how granularity interacts with the productivity law of motion. Equal weighting disperses the influence of individual units, whereas cost–Domar weighting gives greater aggregate importance to large units. The close association between exposure and size makes large-unit productivity dynamics particularly consequential for aggregate moments and event probabilities. Relative to the full size-dependent specification, ignoring regime heterogeneity understates contraction risk and overstates the probability of expansions. Ignoring size dependence instead understates average growth and overstates contraction risk. Granularity thus transmits the interaction between heterogeneous exposure and size-conditioned productivity dynamics into the full distribution of aggregate technology growth.

6 Conclusion

Aggregate productivity risk depends on whose shocks matter and how those shocks evolve. Using plant–commodity physical-quantity data, we recover Solow productivity growth residuals that account for the complete input structure, accommodating for heterogeneous markups, inventories, and input costs. Productivity growth combines a concentrated core with heavy tails, while dispersion and nonlinear mean reversion vary systematically with producer size. Our three-regime Markov normal mixture captures these patterns through transitions among persistent Neutralization, Stabilization, and Correction regimes. The resulting parametric law of motion describes how productivity risk changes with a producer's history and size, providing a tractable representation of the distribution and dynamics of physical-output productivity growth.

These microeconomic differences reshape the distribution of aggregate technology growth. Our size-dependent model has lower aggregate dispersion than the Gaussian, yet more than six times the probability of outcomes beyond three standard deviations. It also assigns a 0.31% probability to an aggregate technology contraction, compared with 0.05% under the Gaussian. Removing size while retaining three regimes has a different effect: expected aggregate growth falls, positive skewness declines, and contraction probability rises to 2.47%. These comparisons show how regime heterogeneity and size-dependent dynamics shape aggregate productivity. Measuring whose shocks matter is only half the task; the other half is measuring the risks those producers carry.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., & Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, *80*(5), 1977–2016. doi: 10.3982/ECTA9623

Acemoglu, D., Ozdaglar, A., & Tahbaz-Salehi, A. (2017). Microeconomic origins of macroeconomic tail risks. *American Economic Review*, *107*(1), 54–108. doi: 10.1257/aer.20151086

Baqaee, D. R., & Farhi, E. (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics*, *135*(1), 105–163. Retrieved from https://doi.org/10.1093/qje/qjz030 doi: 10.1093/qje/qjz030

Bilal, A., Engbom, N., Mongey, S., & Violante, G. L. (2022). Firm and worker dynamics in a frictional labor market. *Econometrica*, *90*(4), 1425–1462. doi: 10.3982/ECTA17955

Boehm, J., & Oberfield, E. (2020). Misallocation in the market for inputs: Enforcement and the organization of production. *The Quarterly Journal of Economics*, *135*(4), 2007–2058. doi: 10.1093/qje/qjaa020

Boehm, J., & Oberfield, E. (2023). *Growth and the fragmentation of production* (Tech. Rep.).

Boehm, J., South, R., Oberfield, E., & Waseem, M. (2024). The network origins of firm dynamics: Contracting frictions and dynamism with long-term relationships.

Cairncross, J., Morrow, P., Orr, S., & Rachapalli, S. (2025). *Identifying firm vs. product markups using production data: Micro estimates and aggregate implications.* (Working paper, September 2025)

Carvalho, V., & Gabaix, X. (2013). The great diversification and its undoing. *American Economic Review*, *103*(5), 1697–1727.

Carvalho, V. M., & Grassi, B. (2019). Large firm dynamics and the business cycle. *American Economic Review*, *109*(4), 1375–1425.

Clementi, G. L., & Palazzo, B. (2016). Entry, exit, firm dynamics, and aggregate fluctuations. *American Economic Journal: Macroeconomics*, *8*(3), 1–41.

Dew-Becker, I. (2023). Tail risk in production networks. *Econometrica*, *91*(6), 2089–2123.

Dhyne, E., Petrin, A., Smeets, V., & Warzynski, F. (2024). *Multiproduct firms, import competition and productivity* (Tech. Rep.). CESifo Working Paper.

Fadinger, H., Ghiglino, C., & Teteryatnikova, M. (2022). Income differences, productivity, and input-output networks. *American Economic Journal: Macroeconomics*, *14*(2), 367–415.

Foster, L., Haltiwanger, J., & Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, *98*(1), 394–425. doi: 10.1257/aer.98.1.394

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica*, *79*(3), 733–772.

Gnewuch, M., & Zhang, D. (2025). Monetary policy, firm heterogeneity, and the distribution of investment rates. *Journal of Monetary Economics*, *149*, 103721. doi: 10.1016/j.jmoneco.2024.103721

Guvenen, F., Karahan, F., Ozkan, S., & Song, J. (2021). What do data on millions of us workers reveal about lifecycle earnings dynamics? *Econometrica*, *89*(5), 2303–2339.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, *3*(5), 1163–1174.

Hopenhayn, H., & Rogerson, R. (1993). Job turnover and policy evaluation: A general equilibrium analysis. *Journal of Political Economy*, *101*(5), 915–938. doi: 10.1086/261909

Hsieh, C.-T., & Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics*, *124*(4), 1403–1448. doi: 10.1162/qjec.2009.124.4.1403

Hsieh, C.-T., & Klenow, P. J. (2014). The life cycle of plants in India and Mexico. *The Quarterly Journal of Economics*, *129*(3), 1035–1084. doi: 10.1093/qje/qju014

Hulten, C. R. (1978). Growth accounting with intermediate inputs. *The Review of Economic Studies*, *45*(3), 511–518.

Jaimovich, N., Terry, S. J., & Vincent, N. (2023). *The empirical distribution of firm dynamics and its macro implications* (Tech. Rep.). National Bureau of Economic Research.

Khan, A., & Thomas, J. K. (2008). Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica*, *76*(2), 395–436. doi: 10.1111/j.1468-0262.2008.00837.x

Khan, A., & Thomas, J. K. (2013). Credit shocks and aggregate fluctuations in an economy with production heterogeneity. *Journal of Political Economy*, *121*(6), 1055–1107. doi: 10.1086/674142

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. John Wiley & Sons.

McNerney, J., Savoie, C., Caravelli, F., Carvalho, V. M., & Farmer, J. D. (2021). How production networks amplify economic growth. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(1), e2106031118.

Melcangi, D., & Sarpietro, S. (2024). Nonlinear firm dynamics. *FRB of New York Staff Report*(1088).

Midrigan, V., & Xu, D. Y. (2014). Finance and misallocation: Evidence from plant-level data. *American Economic Review*, *104*(2), 422–458. doi: 10.1257/aer.104.2.422

Orr, S. (2022). Within-firm productivity dispersion: Estimates and implications. *Journal of Political Economy*, *130*(11), 2771–2828. doi: 10.1086/720465

Salgado, S., Guvenen, F., & Bloom, N. (2023). Skewed business cycles.

Winberry, T. (2021). Lumpy investment, business cycles, and stimulus policy. *American Economic Review*, *111*(1), 364–396. doi: 10.1257/aer.20161723

Online Appendix to
Icarus and Daedalus: Idiosyncratic Regimes and Fat Tails in Firm Productivity

Ronit Mukherji
Ashoka University,
Department of Economics
ronit.mukherji@ashoka.edu.in

Alejandro Rojas-Bernal
University of Hawai'i at Manoa,
Department of Economics
rojasber@hawaii.edu

Current draft: September 15, 2026

Contents

A Tables

Table 1: Tail Shape by Production-Unit Size and Transition

		Distributional statistics									Tail tests			
Year	Group	N	Sh.	Mean	SD	Skew.	Kurt.	$p_{\|z\|>2}$	$p_{\|z\|>3}$	QR	$\widehat{\zeta}_L$	p_L^{AD}	$\widehat{\zeta}_R$	p_R^{AD}
11–12	Small	1,503	0.10	-0.27	2.94	-0.25	1.69	6.59%	1.33%	5.42	0.56	224.0	0.58	513.5
11–12	Medium	12,517	0.83	-0.48	2.82	-0.39	3.33	8.36%	1.55%	8.55	0.62	0.5	0.47	7.0
11–12	Large	1,045	0.07	-0.91	3.24	-0.59	2.39	7.37%	1.05%	7.69	0.61	727.0	0.56	240.5
12–13	Small	1,556	0.10	0.37	2.33	0.37	4.01	8.10%	2.57%	11.02	0.55	0.5	0.52	23.5
12–13	Medium	13,127	0.83	0.14	2.04	-0.09	7.03	7.03%	3.91%	17.03	0.43	0.5	0.45	0.5
12–13	Large	1,114	0.07	0.01	2.18	-0.69	8.94	7.36%	4.13%	21.44	0.38	22.0	0.45	10.0
13–14	Small	2,513	0.10	0.20	2.15	-0.02	4.83	7.36%	3.06%	11.34	0.49	0.5	0.55	532.5
13–14	Medium	21,511	0.83	0.06	1.82	-0.48	8.33	6.47%	3.62%	17.79	0.42	0.5	0.48	0.5
13–14	Large	2,036	0.08	0.03	1.83	-0.80	10.50	6.78%	3.29%	21.02	0.44	13.0	0.48	504.0
14–15	Small	3,145	0.10	-0.01	2.04	-0.48	6.29	6.90%	3.28%	13.69	0.47	50.5	0.48	3.0
14–15	Medium	26,724	0.83	-0.11	1.77	-0.75	8.19	6.40%	3.30%	16.36	0.48	0.5	0.46	0.5
14–15	Large	2,466	0.08	-0.07	1.73	-0.75	16.24	5.43%	2.80%	20.08	0.40	295.0	0.43	361.5
15–16	Small	3,374	0.10	0.36	1.70	0.36	7.64	6.16%	2.79%	10.78	0.52	15.0	0.52	864.0
15–16	Medium	28,009	0.83	0.32	1.44	0.14	13.25	5.17%	2.73%	13.22	0.46	0.5	0.50	6.5
15–16	Large	2,419	0.07	0.43	1.61	0.71	14.03	5.13%	3.35%	17.33	0.35	0.5	0.40	51.0
16–17	Small	3,151	0.09	-0.23	1.90	-1.74	13.45	5.52%	2.60%	11.92	0.35	346.0	0.63	85.5
16–17	Medium	30,328	0.84	0.01	1.30	0.52	12.82	5.23%	2.42%	12.36	0.55	0.5	0.59	6.0
16–17	Large	2,570	0.07	1.08	2.39	1.12	3.54	7.12%	1.63%	6.16	0.41	70.5	0.65	154.5
17–18	Small	3,705	0.09	-0.20	1.95	-1.49	7.95	5.51%	2.43%	8.43	0.40	251.5	0.66	324.0
17–18	Medium	34,366	0.84	0.27	1.45	0.52	9.01	5.61%	2.29%	10.19	0.62	0.5	0.62	165.0
17–18	Large	2,855	0.07	0.99	2.16	1.56	4.03	7.29%	1.93%	7.73	0.64	287.5	0.61	295.5
18–19	Small	4,323	0.10	0.23	1.23	0.61	12.73	5.02%	2.22%	11.31	0.55	15.0	0.65	345.0
18–19	Medium	35,203	0.83	0.17	0.99	0.45	21.03	4.68%	1.89%	12.36	0.71	0.5	0.76	0.5
18–19	Large	3,126	0.07	0.18	0.89	1.22	50.68	4.25%	1.41%	11.46	0.84	1.0	0.81	0.5
19–20	Small	3,955	0.09	-0.30	1.41	-1.56	11.16	5.39%	2.15%	9.94	0.49	781.0	0.92	537.0
19–20	Medium	36,161	0.84	-0.05	0.95	-0.02	17.65	5.08%	1.97%	12.51	0.82	0.5	0.82	0.5
19–20	Large	3,090	0.07	0.03	1.14	1.78	13.32	5.11%	2.46%	12.83	0.93	139.0	0.55	573.0
20–21	Small	4,706	0.10	0.17	1.22	1.29	17.05	4.72%	2.21%	11.40	0.67	34.5	0.55	124.5
20–21	Medium	38,618	0.83	0.08	0.96	0.21	25.84	4.66%	1.97%	13.27	0.73	0.5	0.74	0.5
20–21	Large	3,357	0.07	0.08	0.81	-0.29	20.93	5.69%	2.17%	13.20	1.00	1.5	1.09	442.5
21–22	Small	5,113	0.10	0.15	1.15	0.02	13.29	5.10%	2.13%	11.81	0.55	135.5	0.76	179.5
21–22	Medium	41,284	0.83	0.14	0.88	0.33	22.76	5.11%	1.93%	11.96	0.87	0.5	0.90	0.5
21–22	Large	3,506	0.07	0.17	0.75	2.04	27.41	5.22%	2.00%	12.24	1.56	29.5	0.82	377.0
22–23	Small	5,464	0.10	0.12	1.06	0.14	14.94	5.22%	2.21%	13.17	0.66	2.0	0.80	326.5
22–23	Medium	43,560	0.83	0.06	0.87	0.32	28.71	4.87%	1.97%	14.85	0.84	0.5	0.80	0.5
22–23	Large	3,706	0.07	0.06	0.76	1.53	46.19	4.70%	1.62%	14.21	0.90	12.5	0.98	0.5

Notes: Sh. denotes each size group's share of the corresponding year. Mean is average productivity growth, $\bar{g}$. For each cell, z uses the cell-specific mean and standard deviation, the columns report$p_{|z|>k} = \Pr(|z| > k)$, $QR = (p_{99} - p_1)/(p_{75} - p_{25})$, SD denotes standard deviation, Skew. denotes skewness, and Kurt. is excess kurtosis. Tail exponents use individual exceedances beyond two cell-specific standard deviations. The Anderson–Darling columns report p_L^{AD} and p_R^{AD} multiplied by 10^3, so displayed entries are p-values in thousandths. The tests assess exponential exceedances using 2,000 replications.

Table 2: Distributional Match Across Models with 1, 2, and 3 Regimes

Panel A. Marginal fit and conditional calibration

		Three regimes			Two regimes			One regime		
Sample	N	CDF_{KS}	PIT_{KS}	PIT_{AD}	CDF_{KS}	PIT_{KS}	PIT_{AD}	CDF_{KS}	PIT_{KS}	PIT_{AD}
Pooled	267,285	**1.61%**	**0.98%**	**1.84**	3.22%	2.60%	15.89	16.17%	15.34%	542.48
12–13	6,739	**2.40%**	**1.47%**	**1.23**	4.94%	2.73%	21.84	20.14%	14.76%	512.37
13–14	8,865	**1.54%**	**1.70%**	**1.83**	3.85%	3.97%	19.05	19.34%	17.87%	634.35
14–15	15,176	**1.91%**	**1.77%**	**3.78**	3.59%	3.54%	22.53	19.45%	18.36%	677.36
15–16	19,376	**1.62%**	**1.39%**	**3.33**	3.54%	3.27%	27.13	16.22%	14.22%	491.23
16–17	22,894	**1.20%**	**1.31%**	**1.35**	3.32%	3.32%	20.64	17.48%	17.06%	602.19
17–18	25,840	**2.36%**	**1.64%**	**2.55**	3.84%	3.50%	12.48	14.17%	13.70%	435.59
18–19	29,190	**3.33%**	**2.21%**	**6.54**	4.64%	3.07%	14.36	17.28%	13.72%	426.16
19–20	30,295	**1.44%**	**1.54%**	**2.10**	3.21%	3.04%	13.00	15.71%	15.01%	531.93
20–21	32,422	**2.00%**	**1.81%**	**4.86**	3.44%	3.20%	23.02	17.55%	16.70%	625.37
21–22	36,253	**1.88%**	**1.52%**	**3.30**	3.69%	3.35%	21.82	16.90%	15.90%	546.26
22–23	40,235	**1.75%**	**1.43%**	**3.17**	3.70%	3.04%	22.11	18.51%	17.23%	687.77

Panel B. Tail quantile agreement and distance

		Three regimes				Two regimes				One regime			
Sample	N	IA_L	TW_L	IA_R	TW_R	IA_L	TW_L	IA_R	TW_R	IA_L	TW_L	IA_R	TW_R
Pooled	267,285	**0.96**	**0.99**	**0.97**	**0.68**	0.85	2.05	0.89	1.66	0.58	4.25	0.59	4.00
12–13	6,739	**0.95**	**0.71**	0.80	1.05	0.71	2.21	**0.96**	**0.63**	0.39	4.62	0.54	2.53
13–14	8,865	**0.93**	**0.79**	**0.95**	**0.79**	0.68	2.35	0.83	1.58	0.35	5.15	0.43	4.01
14–15	15,176	**1.00**	**0.21**	**0.92**	**0.95**	0.81	1.70	0.87	1.24	0.40	4.69	0.46	3.41
15–16	19,376	**0.88**	**1.06**	**0.96**	**0.65**	0.65	2.12	0.86	1.32	0.36	3.96	0.46	3.57
16–17	22,894	**0.85**	**1.56**	**0.97**	**0.62**	0.77	1.90	0.69	2.32	0.58	3.60	0.41	4.69
17–18	25,840	**0.99**	**0.28**	**0.98**	**0.23**	0.90	0.98	0.82	1.17	0.59	2.44	0.47	3.01
18–19	29,190	**0.92**	1.64	**0.94**	**1.42**	0.84	**1.58**	0.83	1.52	0.58	3.10	0.56	3.87
19–20	30,295	**0.94**	**0.94**	**0.93**	**0.91**	0.80	1.77	0.72	1.98	0.57	4.11	0.51	4.14
20–21	32,422	**0.88**	**1.61**	**0.87**	**1.47**	0.69	2.25	0.70	2.32	0.51	4.69	0.50	5.08
21–22	36,253	**0.90**	**1.20**	**0.90**	**1.22**	0.74	1.84	0.76	1.61	0.52	3.99	0.53	4.04
22–23	40,235	**0.85**	**1.98**	**0.90**	**1.57**	0.70	2.27	0.72	2.41	0.51	4.79	0.50	5.67

Notes: Annual rows label the outcome transition; for example, 12–13 conditions on 11–12. N counts AR-eligible lag–outcome pairs. All specifications share the linear-size by cubic-growth mean basis, bounded Bernstein variance, and regime-year effects; the multi-regime models also share the transition basis. Let $F_{K,it}$ denote the pre-outcome predictive CDF under $K \in \{1, 2, 3\}$ regimes. CDF KS compares the empirical CDF with the sample-average predictive CDF. PIT KS and $\text{PIT}^{\text{AD}} = 10^4 A^2/N$ assess uniformity of $u_{it} = F_{K,it}(g_{i,t})$, where A^2 is the Anderson–Darling statistic. Both KS statistics are percentages; lower values indicate better fit for all three diagnostics. Panel B compares empirical and predictive quantiles within each outer 1% tail. IA is the Willmott index of agreement (target 1); TW is the mean absolute quantile discrepancy divided by QR (target 0). Bold identifies the best specification for each statistic within a row.

Table 3: Full-Panel Matching and Sample Accounting ($\theta = 1$)

Transition	Matched pairs: Exact 7-digit	Matched pairs: Unique 5-digit fallback	Matched pairs: Total	Productivity calculation: Valid $\widehat{A}$	Productivity calculation: Excluded Cost share $> 5\%$
11–12	12,954	2,956	15,910	15,065	845
12–13	14,389	2,371	16,760	15,797	963
13–14	23,433	4,815	28,248	26,060	2,188
14–15	31,517	3,711	35,228	32,335	2,893
15–16	33,960	2,553	36,513	33,802	2,711
16–17	36,569	2,502	39,071	36,049	3,022
17–18	38,601	2,331	40,932	40,926	6
18–19	40,644	2,010	42,654	42,652	2
19–20	41,306	1,906	43,212	43,206	6
20–21	44,867	1,818	46,685	46,681	4
21–22	48,341	1,562	49,903	49,903	0
22–23	51,346	1,388	52,734	52,730	4
Pooled	417,927	29,923	447,850	435,206	12,644

Notes: Counts refer to firm–commodity transition pairs for the full five-block productivity measure ($\theta = 1$). Matching first uses complete seven-digit commodity codes within each firm. When no exact match exists, a five-digit prefix match is accepted only if it identifies a single next-year commodity within the same firm. This allows potential continuity within a broader commodity group when detailed codes differ, without choosing among multiple candidate products. Missing or invalid time-t input quantities prevent calculation of their log changes. We impute zero quantity growth for these inputs and retain the pair only if their combined share of time-t total cost is at most 5%, limiting reliance on imputed input movements. The final column counts pairs excluded for exceeding this threshold. Pooled counts sum across transitions, so production units may appear more than once.

B Likelihood, Filtering, and Estimation

We estimate the model by direct maximum likelihood. Because regimes are unobserved, each observation contributes to several possible regime histories. The forward recursion evaluates their combined likelihood without enumerating every history. A backward recursion then provides the posterior probabilities needed to calculate the likelihood's derivatives. These derivatives guide the joint optimization of all model parameters.

B.1 Observation Densities and Transition Probabilities

Each usable observation contains next-period productivity growth, $g_{i,t+1}$, and the observed predictors $(g_{i,t}, d_{i,t})$. Here, $d_{i,t}$ is log sales at the beginning of the year, standardized within the annual sample. The size-dependent specifications use $\Phi(x,d) = (1,d)' \otimes (1,x,x^2,x^3)'$. For destination regime r, the conditional mean and standard deviation are

$$\begin{aligned} \mu_{r,t+1}(g_{i,t},d_{i,t}) &= \boldsymbol{\beta}_r'\Phi(g_{i,t},d_{i,t}) + \delta_{r,t+1}, \\ v_{r,t+1}(g_{i,t}) &= \sigma_r\tau_{r,t+1}\mathcal{V}_r(z_{i,t})^{1/2}, \qquad z_{i,t} = \frac{g_{i,t}-\bar{g}_{\text{lag}}}{s_{\text{lag}}}. \end{aligned} \tag{1}$$

Lagged growth is standardized using its pooled mean and standard deviation across all eligible observations and years. These constants are common to every year and regime and remain fixed throughout estimation. Size enters the mean linearly, including interactions with all three growth powers, but does not enter volatility directly.

The emission is the regime-specific density for $g_{i,t+1}$conditional at $(g_{i,t},d_{i,t})$:

$$e_{ir,t+1} = \frac{1}{v_{r,t+1}(g_{i,t})}\phi\left(\frac{g_{i,t+1}-\mu_{r,t+1}(g_{i,t},d_{i,t})}{v_{r,t+1}(g_{i,t})}\right). \tag{2}$$

Here, ϕ is the standard normal density. A larger emission likelihood means that regime r assigns greater density to a realization.

For origin regime s, transition probabilities are

$$p_{sr}(g_{it},,d_{i,t}) = \frac{\exp\{\boldsymbol{\kappa}_{sr}'\Phi(z_{i,t},d_{i,t})\}}{\sum_{\ell=1}^{K}\exp\{\boldsymbol{\kappa}_{s\ell}'\Phi(z_{i,t},d_{i,t})\}}, \qquad \boldsymbol{\kappa}_{sK} = \mathbf{0}. \tag{3}$$

Each row sums to one. Parameters are common across production units, but probabilities vary with their observed growth and size.

We estimate size-dependent models with $K = 1,2,3$, and a separate size-independent model with $K = 3$. The latter replaces $\Phi(x,d)$ with $(1,x,x^2,x^3)'$ in both means and transition indices. All specifications use the same estimation sample. For $K = 1$, the regime probability is identically one.

B.2 Variance Restrictions and Annual Normalizations

The variance function allows dispersion to change with preceding growth while remaining positive. Let z_L and z_U be the pooled 1st and 99th percentiles of standardized lagged growth, and define $u(z) = \min\left\{1, \max\left\{0, \frac{z-z_L}{z_U-z_L}\right\}\right\}$.

Using the cubic Bernstein weights $B_{\ell,3}(u) = \frac{3!}{\ell!(3-\ell)!}u^\ell(1-u)^{3-\ell}$, we set

$$\mathcal{V}_r(z) = \sum_{\ell=0}^{3} a_{\ell r}B_{\ell,3}(u(z)), \qquad a_{3r} = 1, \qquad a_{\ell r} \in [1/3, 3], \quad \ell = 0,1,2. \tag{4}$$

The weights are nonnegative and sum to one, so the multiplier is a weighted average of the control values and remains between 1/3 and 3. Fixing the final control to one separates the variance profile from the baseline scale σ_r; bounding the remaining controls limits how much lagged growth can change that profile. Clipping makes the multiplier constant outside the percentile range: $\mathcal{V}_r(z) = a_{0r}$ below z_L and $\mathcal{V}_r(z) = 1$ above z_U. This clipping applies only to volatility.

The annual effects also require normalizations. Index the full set of years by $c = 1, \ldots, J$, independently of the length of any unit's history. We impose $\sum_{c=1}^{J} \delta_{rc} = 0$ and $\frac{1}{J} \sum_{c=1}^{J} \tau_{rc} = 1$. These restrictions assign the time-invariant location to the mean intercept and the time-invariant scale to σ_r. The restrictions hold automatically during optimization. We estimate the first $J - 1$ location effects and set $\delta_{rJ} = -\sum_{c<J} \delta_{rc}$. Annual scale multipliers are parameterized as $\tau_{rc} = J \frac{\exp(\chi_{rc})}{\sum_{h=1}^{J} \exp(\chi_{rh})}$ with $\chi_{rJ} = 0$. The annual scale multipliers are therefore positive and average one, so they measure year-specific volatility relative to the baseline scale σ_r. To keep the baseline scale positive during optimization, we estimate its logarithm, $s_r = \log \sigma_r$. Likewise, we estimate $\nu_{\ell r} = \log a_{\ell r}$ for $\ell = 0, 1, 2$ with $\nu_{\ell r} \in [-\log 3, \log 3]$.

B.3 Observed-Data Likelihood and Contiguous Spells

Let $\mathcal{O}_i$ contain the usable outcome dates for production unit i. An outcome is usable only when its lagged growth and required size covariate are also available. The common estimation sample contains 267,285 observations from 85,803 production units over eleven annual outcome transitions.

We divide each unit's usable dates into contiguous spells. A gap in eligibility ends one spell and starts another; the algorithm does not propagate an unobserved regime transition across that gap. Every spell begins from the same estimated probability vector $\boldsymbol{\pi}^0$. This vector is estimated jointly with the remaining parameters.

Consider a spell b of production unit i, beginning at outcome date a_b and ending at T_b. For a latent regime path $(r_{a_b}, \ldots, r_{T_b})$, the joint likelihood is

$$L_b(\vartheta; r_{a_b}, \ldots, r_{T_b}) = \pi^0_{r_{a_b}} e_{ir_{a_b}, a_b} \prod_{t=a_b}^{T_b - 1} p_{r_t r_{t+1}}(g_{i,t}, d_{i,t}) e_{ir_{t+1}, t+1}. \tag{5}$$

The first emission conditions on the available $(g_{i,a_b-1}, d_{i,a_b-1})$, but no transition probability precedes it. The lag preceding the spell is treated as given.

Because the path is latent, the spell likelihood sums over all possible paths:

$$L_b(\vartheta) = \sum_{r_a=1}^{K} \cdots \sum_{r_b=1}^{K} L_b(\vartheta; r_a, \ldots, r_b), \qquad \ell(\vartheta) = \sum_b \log L_b(\vartheta). \tag{6}$$

The parameter vector ϑ contains the initial probabilities, transition coefficients, conditional-mean coefficients, variance controls, and regime–year effects. Directly summing the K^{e-a+1} paths is unnecessary: the forward recursion evaluates the same likelihood much more efficiently.

B.4 Forward Filtering and Likelihood Evaluation

The forward recursion alternates between prediction and updating. Before observing an outcome, the model assigns a probability to each destination regime. After observing it, Bayes' rule increases the probabilities of regimes that better explain the shock.

At the first outcome a of each spell, $\pi_{ir,a|a-1} = \pi_r^0$. At subsequent dates, the filtered probabilities are propagated with the transition law:

$$\pi_{ir,t+1|t} = \sum_{s=1}^{K} a_{is,t}\, p_{sr}(g_{i,t}, d_{i,t}), \qquad a_{is,t} \equiv \pi_{is,t|t}. \tag{7}$$

For every outcome, including a spell's first observation, the updating step is

$$\widetilde{a}_{ir,t} = \pi_{ir,t|t-1} e_{ir,t}, \qquad c_{it} = \sum_{r=1}^{K} \widetilde{a}_{ir,t}, \qquad a_{ir,t} = \frac{\widetilde{a}_{ir,t}}{c_{it}}. \tag{8}$$

The normalizing constant c_{it} is the predictive density evaluated at the observed outcome. Consequently, $\ell(\vartheta) = \sum_i \sum_{t \in \mathcal{O}_i} \log c_{it}$.

Repeated multiplication of small densities can exceed numerical precision. Consequently, the implementation instead stores logarithms and evaluates sums with the stable identity $\log\left(\sum_r e^{x_r}\right) = m + \log\left(\sum_r e^{x_r - m}\right)$ with $m = \max_r x_r$. This log-scale calculation produces the same filtered probabilities and likelihood while avoiding numerical underflow. Its computational cost grows linearly with the number of observations, rather than exponentially with the length of a spell.

B.5 Backward Recursion and Smoothed Probabilities

Filtering uses outcomes only through the current date. Estimation also benefits from later observations: a subsequent realization can help identify which regime generated an earlier one. The backward recursion incorporates this additional information within each contiguous spell.

Let $b_{ir,t}$ summarize the likelihood of the observations following t, conditional on regime r at t, using the same scaling as the forward recursion. At the spell's final date e, set $b_{ir,e} = 1$. Working backward,

$$b_{is,t} = \frac{\sum_{r=1}^{K} p_{sr}(g_{i,t}, d_{i,t}) e_{ir,t+1} b_{ir,t+1}}{c_{i,t+1}}. \tag{9}$$

Combining forward and backward information gives the smoothed probability:

$$\gamma_{ir,t} = \frac{a_{ir,t} b_{ir,t}}{\sum_{\ell=1}^{K} a_{i\ell,t} b_{i\ell,t}}. \tag{10}$$

Unlike a filtered probability, $\gamma_{ir,t}$ uses all outcomes in the same spell.

For adjacent usable outcomes, the smoothed probability of a transition s to r is

$$\xi_{isr,t+1} = \frac{a_{is,t} p_{sr}(g_{i,t}, d_{i,t}) e_{ir,t+1} b_{ir,t+1}}{c_{i,t+1}}. \tag{11}$$

These probabilities satisfy

$$\sum_r \xi_{isr,t+1} = \gamma_{is,t}, \qquad \sum_s \xi_{isr,t+1} = \gamma_{ir,t+1}, \qquad \sum_{s,r} \xi_{isr,t+1} = 1.$$

Thus, γ allocates each observation across possible regimes, while ξ allocates each adjacent pair across possible regime transitions. Smoothing stops at spell boundaries. Neither smoothed quantity is substituted for the pre-outcome probabilities used to construct predictive distributions.

B.6 Analytic Likelihood Scores

The score is the vector of derivatives of the observed-data log likelihood with respect to the parameters. It measures how the likelihood changes when each parameter changes. We calculate these derivatives analytically and supply them to the joint optimizer. Let $\mathcal{O} = \{(i,t) : t \in \mathcal{O}_i\}$ denote the usable outcome observations, $\mathcal{F} \subseteq \mathcal{O}$ the spell starts, and $\mathcal{C}$ the adjacent usable outcome pairs, indexed by their origin dates. The score is

$$\begin{aligned}\nabla_\vartheta \ell(\vartheta) = & \sum_{(i,t)\in\mathcal{F}} \sum_r \gamma_{ir,t} \nabla_\vartheta \log \pi_r^0 + \sum_{(i,t)\in\mathcal{O}} \sum_r \gamma_{ir,t} \nabla_\vartheta \log e_{ir,t} \\ & + \sum_{(i,t)\in\mathcal{C}} \sum_{s,r} \xi_{isr,t+1} \nabla_\vartheta \log p_{sr}(g_{i,t}, d_{i,t}).\end{aligned} \tag{12}$$

The three terms correspond to initial regime probabilities, observation densities, and regime transitions. If the regimes were observed, each observation would contribute only to its realized regime and each adjacent pair only to its realized transition. Because regimes are unobserved, the smoothed probabilities allocate these contributions across the possible regimes and transitions.

Equation (12) follows by differentiating the likelihood summed over latent regime paths. It expresses the result as a posterior-weighted average of complete-data scores. No additional derivatives of γ or ξ are required in this first-derivative identity. All posterior probabilities are evaluated at the current parameter vector, using the observations within each contiguous spell. They are recomputed whenever the optimizer evaluates a new parameter vector. The following subsections derive the individual score blocks; these blocks are assembled into one vector.

B.7 Initial-Probability and Conditional-Mean Scores

The common spell-start probabilities are parameterized by logits:

$$\pi_r^0 = \frac{\exp(\alpha_r)}{\sum_{\ell=1}^{K} \exp(\alpha_\ell)}, \qquad \alpha_K = 0. \tag{13}$$

This parameterization keeps the probabilities positive and ensures that they sum to one. The reference logit α_K is fixed because adding the same constant to every logit would leave the probabilities unchanged.

For each free logit, the score is

$$\frac{\partial \ell}{\partial \alpha_r} = \sum_{(i,t)\in\mathcal{F}} \left(\gamma_{ir,t} - \pi_r^0\right), \qquad r < K. \tag{14}$$

Each spell start compares its posterior probability of regime r with the common initial probability. A positive total score supports increasing that regime's initial logit. When these scores are zero, the initial probabilities equal the average smoothed regime probabilities at spell starts: $\pi_r^0 = \frac{1}{|\mathcal{F}|}\sum_{(i,t)\in\mathcal{F}} \gamma_{ir,t}$.

For the emission scores, abbreviate the observation-specific mean and standard deviation as $\mu_{ir,t} = \mu_{r,t}(g_{i,t-1}, d_{i,t-1})$ and $v_{ir,t} = v_{r,t}(g_{i,t-1})$, and define $\epsilon_{ir,t} = g_{i,t} - \mu_{ir,t}$. The normal log density is $\log e_{ir,t} = -\frac{1}{2}\log(2\pi) - \log v_{ir,t} - \frac{\epsilon_{ir,t}^2}{2v_{ir,t}^2}$. Its derivatives with respect to the mean and log standard deviation are

$$\frac{\partial \log e_{ir,t}}{\partial \mu_{ir,t}} = \frac{\epsilon_{ir,t}}{v_{ir,t}^2}, \qquad \frac{\partial \log e_{ir,t}}{\partial \log v_{ir,t}} = \frac{\epsilon_{ir,t}^2}{v_{ir,t}^2} - 1. \tag{15}$$

It is useful to collect their posterior-weighted contributions:

$$M_{ir,t} = \gamma_{ir,t} \frac{\epsilon_{ir,t}}{v_{ir,t}^2}, \qquad W_{ir,t} = \gamma_{ir,t} \left(\frac{\epsilon_{ir,t}^2}{v_{ir,t}^2} - 1 \right). \tag{16}$$

The location contribution $M_{ir,t}$ favors moving the mean toward the observation. It gives greater weight to observations more likely to belong to the regime and to those with smaller conditional variance. The scale contribution $W_{ir,t}$ favors greater dispersion when the squared standardized residual exceeds one, and lower dispersion when it is below one.

Because the conditional mean is linear in $\boldsymbol{\beta}_r$, its coefficient score is

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}_r} = \sum_{(i,t)\in\mathcal{O}} M_{ir,t} \Phi(g_{i,t-1}, d_{i,t-1}). \tag{17}$$

This has the posterior- and precision-weighted residual structure of weighted least squares. Here, however, it supplies derivatives to the joint optimizer rather than defining a separate regression update.

B.8 Variance and Regime–Year Scores

The conditional standard deviation is $v_{ir,t} = \sigma_r \tau_{r,t} \mathcal{V}_r(z_{i,t-1})^{1/2}$. Its three components describe the baseline regime scale, annual scale variation, and dependence on preceding growth. Their scores follow from $W_{ir,t}$ in Equation (16).

Baseline scales. Since $s_r = \log \sigma_r$ enters $\log v_{ir,t}$ with coefficient one,

$$\frac{\partial \ell}{\partial s_r} = \sum_{(i,t)\in\mathcal{O}} W_{ir,t}. \tag{18}$$

The scale responds to the pooled discrepancy weighted by regime membership.

Bernstein controls. Recall that $\mathcal{V}_r(z) = \sum_{\ell=0}^{3} a_{\ell r} B_{\ell,3}(u(z))$, $a_{\ell r} = \exp(\nu_{\ell r})$, for $\ell = 0, 1, 2$ and $a_{3r} = 1$. For each estimated log control,

$$\frac{\partial \log v_{ir,t}}{\partial \nu_{\ell r}} = \frac{1}{2} \frac{a_{\ell r} B_{\ell,3}(u(z_{i,t-1}))}{\mathcal{V}_r(z_{i,t-1})}.$$

Consequently,

$$\frac{\partial \ell}{\partial \nu_{\ell r}} = \frac{1}{2} \sum_{(i,t)\in\mathcal{O}} W_{ir,t} \frac{a_{\ell r} B_{\ell,3}(u(z_{i,t-1}))}{\mathcal{V}_r(z_{i,t-1})}, \qquad \ell = 0, 1, 2. \tag{19}$$

Each control receives the scale discrepancy weighted by its contribution to the local variance multiplier. The factor $1/2$ appears because $\mathcal{V}_r$ multiplies variance, whereas the underlying emission derivative is with respect to log standard deviation. The fixed control a_{3r} has no estimated coefficient. At an active control bound, convergence is assessed using the projected-score criterion rather than requiring the raw score to vanish.

Annual location effects. Let $\mathcal{O}_c$ contain observations in calendar outcome transition c, for $c = 1, \ldots, J$, and define the annual totals

$$M^Y_{rc} = \sum_{(i,t)\in\mathcal{O}_c} M_{ir,t}, \qquad W^Y_{rc} = \sum_{(i,t)\in\mathcal{O}_c} W_{ir,t}. \tag{20}$$

The location normalization sets $\delta_{rJ} = -\sum_{c<J}\delta_{rc}$. Increasing a free δ_{rc} therefore raises the mean in year c and lowers the final year's mean by the same amount. Its score accounts for both changes:

$$\frac{\partial \ell}{\partial \delta_{rc}} = M^Y_{rc} - M^Y_{rJ}, \qquad c < J. \tag{21}$$

Annual scale effects. The positive multipliers are parameterized as

$$\tau_{rc} = J\frac{\exp(\chi_{rc})}{\sum_{h=1}^{J}\exp(\chi_{rh})}, \qquad \chi_{rJ} = 0.$$

Changing one annual logit affects every year's multiplier through:

$$\frac{\partial \log \tau_{rh}}{\partial \chi_{rc}} = \mathbf{1}\{h = c\} - \frac{\tau_{rc}}{J}.$$

The resulting score is

$$\frac{\partial \ell}{\partial \chi_{rc}} = W^Y_{rc} - \frac{\tau_{rc}}{J}\sum_{h=1}^{J} W^Y_{rh}, \qquad c < J. \tag{22}$$

The first term captures the scale discrepancy in year c; the second accounts for the accompanying adjustment across all years required to preserve the mean-one normalization. These restrictions are incorporated directly into the derivatives, rather than imposed after separate annual updates.

B.9 Transition-Probability Scores

The reference-destination normalization leaves $K-1$ free coefficient vectors for each origin regime. For destination $r < K$, differentiation gives

$$\frac{\partial \ell}{\partial \boldsymbol{\kappa}_{sr}} = \sum_{(i,t)\in\mathcal{C}} \left[\xi_{isr,t+1} - \gamma_{is,t}p_{sr}(g_{i,t}, d_{i,t})\right]\Phi(z_{it}, d_{i,t}). \tag{23}$$

The first term inside the brackets is the posterior probability that the observed pair followed transition $s \to r$. The second is the transition law's probability of that destination, weighted by the posterior probability of occupying origin regime s. The origin weight is the smoothed probability, since $\sum_r \xi_{isr,t+1} = \gamma_{is,t}$.

The initial-probability, emission, annual-effect, and transition scores are assembled into the full vector $S(\vartheta) = \nabla_\vartheta \ell(\vartheta)$. Contributions are also retained by production unit, summing across all of its spells, for the numerical scaling and convergence checks described below.

All parameter blocks are optimized jointly, and changing any block can change the smoothed probabilities entering the others. The implementation therefore recomputes filtering, smoothing, and the complete score at every candidate parameter vector; it does not perform separate initial-probability, regression, variance, or transition-gate maximizations.

B.10 Data-Based Initialization

Mixture likelihoods can have several local maxima, so a single starting point may miss a better-fitting solution. We use four reproducible, data-based starts for each specification. Each model is initialized separately, without importing estimates from another specification or an earlier run.

We first regress growth on the mean basis and annual location effects. Let $\epsilon_n^{(0)}$ denote the resulting residual for observation n. For $K > 1$, these residuals initialize a mixture of zero-mean normal components with different standard deviations. Initial component probabilities are $1/K$; initial scales are proportional to the residual standard deviation, with ratios $(0.45, 2.10)$ for $K = 2$ and $(0.90, 0.32, 2.80)$ for $K = 3$.

Writing their probabilities and scales as $\widetilde{\pi}_r$ and $\widetilde{\sigma}_r$, we calculate observation-specific weights $w_{nr} = \frac{\widetilde{\pi}_r \widetilde{\sigma}_r^{-1} \phi(\epsilon_n^{(0)}/\widetilde{\sigma}_r)}{\sum_{\ell=1}^K \widetilde{\pi}_\ell \widetilde{\sigma}_\ell^{-1} \phi(\epsilon_n^{(0)}/\widetilde{\sigma}_\ell)}$. A residual receives greater weight in components that assign it greater density. We then update

$$\widetilde{\pi}_r \leftarrow \frac{\sum_n w_{nr} + 1/2}{N + K/2}, \qquad \widetilde{\sigma}_r^2 \leftarrow \frac{\sum_n w_{nr} (\epsilon_n^{(0)})^2}{\sum_n w_{nr}},$$

subject to a small positive scale floor. Eight rounds of these weight and scale calculations produce the initialization weights.

Using these weights, we estimate initial regime-specific mean coefficients and annual location effects by weighted regression. Let R_{rc} be the weighted root-mean-square residual for regime r and transition c, after subtracting that fitted mean, including its annual effect. Initial scales are

$$\sigma_r^{(0)} = \frac{1}{J}\sum_{c=1}^{J} R_{rc}, \qquad \tau_{rc}^{(0)} = \frac{R_{rc}}{\sigma_r^{(0)}}.$$

Thus, annual standard-deviation multipliers average one.

Initial regime probabilities use the sum of provisional weights at all spell starts, including restarts after gaps. Transition counts use products of the provisional origin and destination weights across adjacent eligible outcomes. We add 1/2 to each initial and transition count, normalize the counts into probabilities, and convert them into initial logits and transition intercepts. Before perturbation, all other transition coefficients are zero.

The first start uses the residual-scale ratios given above. For the second, third, and fourth starts, we increase those ratios by 8%, 16%, and 24%, respectively, before calculating the provisional mixture weights. These changes affect only the initial guesses, not the model's specification or restrictions.

After constructing the parameter vector for each start, we leave the first vector unchanged. For the other three, we add independent normal draws with mean zero and standard deviation 0.025 to the free parameters. The draws use fixed random seeds, making the starting values reproducible. Probabilities are perturbed through their logits and positive scales through their logarithms, so they remain admissible when converted back. The Bernstein controls are excluded from these perturbations: all start at one, giving a constant initial variance multiplier. They are subsequently estimated along with the other parameters. These preliminary calculations construct starting values only. Their adjustments do not enter the subsequently maximized likelihood. Once joint estimation begins, regime probabilities are obtained from the full forward–backward recursion rather than these provisional weights.

B.11 Joint Likelihood Maximization

Starting from each initial parameter vector, we maximize the observed-data log likelihood jointly over all free parameters. The analytic score $S(\vartheta) = \nabla_\vartheta \ell(\vartheta)$ supplies the derivatives used to improve the parameter estimates.

Because parameters have different numerical scales, we first rescale them using production-unit scores. Let $s_{ij}(\vartheta)$ be unit i's score contribution for parameter j, summed over its spells. Define

$$D_j(\vartheta) = \left[\max\left\{\sum_i \big(s_{ij}(\vartheta) - \bar{s}_j(\vartheta)\big)^2, \epsilon_{\mathrm{m}}\right\}\right]^{1/2}, \qquad \bar{s}_j = \frac{1}{I}\sum_i s_{ij}, \tag{24}$$

where I is the number of production units and ϵ_{m} is machine precision.

At the beginning of each optimization pass, let ϑ^0 be the current estimate and set $c_j = \max\{D_j(\vartheta^0), 10^{-4}\}$ and $C = \operatorname{diag}(c_1, \ldots, c_p)$. The solver works with $\psi = C(\vartheta - \vartheta^0)$, holding C and ϑ^0 fixed within the pass. It minimizes the equivalent negative log likelihood

$$\begin{aligned} F(\psi) &= -\ell(\vartheta^0 + C^{-1}\psi), \\ \nabla_\psi F(\psi) &= -C^{-1} S(\vartheta^0 + C^{-1}\psi). \end{aligned} \tag{25}$$

Thus, each score component is divided by its numerical scale and changes sign because the solver minimizes rather than maximizes.

L-BFGS combines this gradient with changes in previous gradients to approximate the objective's curvature. The Bernstein-control bounds are enforced through a logarithmic barrier $B(\psi)$. Its weight $\eta > 0$ is progressively reduced. Writing H_k for the limited-memory approximation to inverse curvature, an inner iteration takes the form

$$\begin{aligned} h^{(k)} &= -H_k \nabla_\psi \big[F(\psi^{(k)}) + \eta B(\psi^{(k)})\big], \\ \psi^{(k+1)} &= \psi^{(k)} + \alpha_k h^{(k)}, \\ \vartheta^{(k+1)} &= \vartheta^{(k)} + \alpha_k C^{-1} h^{(k)}. \end{aligned} \tag{26}$$

The direction $h^{(k)}$ specifies how parameters move together; backtracking chooses the step length α_k, shortening the step until it remains feasible and sufficiently decreases the solver's objective. Nonfinite evaluations are rejected. The barrier is a numerical device for enforcing bounds.

Each start permits up to six passes, with inner and outer iteration limits of 1,000 per pass. Between passes, score scales are recomputed and the optimizer's numerical history is restarted. Acceptance is checked against the original log likelihood and the projected-score criterion defined below. A candidate must preserve the best likelihood reached, up to floating-point precision, and either improve the likelihood or reduce the convergence residual. Failed passes leave the last accepted estimate unchanged.

B.12 Convergence, Model Selection, and Reporting

We assess convergence directly from the likelihood score, using the current score scales in Equation (24). Let $S_j = \sum_i s_{ij}$ be the total score for parameter ϑ_j. For an unrestricted parameter, convergence requires this derivative to be close to zero. At a parameter bound, however, a derivative pointing outside the admissible range does not indicate an available improvement. We account for this distinction with the projected residual

$$R_j = \begin{cases} \max\{S_j, 0\}/D_j, & \vartheta_j \text{ is at its lower bound,} \\ \max\{-S_j, 0\}/D_j, & \vartheta_j \text{ is at its upper bound,} \\ |S_j|/D_j, & \text{otherwise.} \end{cases} \qquad \max_j R_j \leq 10^{-4}. \quad (27)$$

Only the Bernstein controls have finite bounds; the remaining parameters are unrestricted. Bounds are treated as active within 10^{-8}. This is a numerical first-order, or Karush–Kuhn–Tucker, check: it asks whether any admissible local parameter change would materially increase the likelihood. Normalizing by D_j makes the criterion comparable across parameters. Among the different starts, we retain the highest-likelihood estimate satisfying Equation (27). If none satisfies it, no final estimate is accepted.

With eleven calendar transitions, the size-dependent $K = 1, 2, 3$ models contain 32, 81, and 146 parameters, respectively; the size-independent $K = 3$ model contains 110. After fitting, the three-state labels are assigned consistently: Neutralization has the smallest absolute mean slope at $g = d = 0$, Correction has the more negative slope among the remaining regimes, and Stabilization is the third regime. Multiple starts reduce sensitivity to initialization. We retain the likelihood, convergence residual, and optimization history for every start.

C Derivation of the Disaggregated Solow Residual

For completeness, we derive Definition 1. Write $\widehat{z} = d \log z$ and suppress the unit and time subscripts where doing so causes no ambiguity. The first-order conditions in Section 2.1 identify the local output elasticities as

$$e(y_i,\ell_{if})=\frac{\tau^{\ell}_{if}w_f\ell_{if}}{\mu_i p_i y_i},\quad e(y_i,k_{ic})=\frac{R^k_{ic}k_{ic}}{\mu_i p_i y_i},\quad e(y_i,x_{ig})=\frac{\tau^x_{ig}p_g x_{ig}}{\mu_i p_i y_i}. \tag{28}$$

By the chain rule for the nested technology, these elasticities can be written as

$$e(y_i,\ell_{if})=\omega^L_i\alpha^{\ell}_{if},\qquad e(y_i,k_{ic})=\omega^L_i\alpha^k_{ic},\qquad e(y_i,x_{ig})=\omega^X_i\omega^x_{ig}.$$

Log-differentiating $y_i=A_iQ_i(L_i,X_i)$ and two input composites gives

$$\widehat{y}_i-\widehat{A}_i=\omega^L_i\left(\sum_{f\in\mathcal{F}}\alpha^{\ell}_{if}\widehat{\ell}_{if}+\sum_{c\in\mathcal{K}}\alpha^k_{ic}\widehat{k}_{ic}\right)+\omega^X_i\sum_{g\in\mathcal{G}}\omega^x_{ig}\widehat{x}_{ig}. \tag{29}$$

Constant returns of Q_i, Q^{ℓ}_i, and Q^x_i, together with Euler's theorem, imply

$$\omega^L_i+\omega^X_i=1,\qquad \sum_{f\in\mathcal{F}}\alpha^{\ell}_{if}+\sum_{c\in\mathcal{K}}\alpha^k_{ic}=1,\qquad \sum_{g\in\mathcal{G}}\omega^x_{ig}=1.$$

Thus the coefficients on the disaggregated inputs in (29) sum to one. Adding and subtracting output growth inside each input term yields

$$\widehat{A}_i=-\omega^L_i\left[\sum_{f\in\mathcal{F}}\alpha^{\ell}_{if}(\widehat{\ell}_{if}-\widehat{y}_i)+\sum_{c\in\mathcal{K}}\alpha^k_{ic}(\widehat{k}_{ic}-\widehat{y}_i)\right]-\omega^X_i\sum_{g\in\mathcal{G}}\omega^x_{ig}(\widehat{x}_{ig}-\widehat{y}_i). \tag{30}$$

Evaluating the elasticities at t and replacing log differentials by annual log changes gives the first-order discrete-time counterpart used in Definition 1:

$$g_{i,t+1}=-\omega^L_{i,t}\left[\sum_{f\in\mathcal{F}}\alpha^{\ell}_{if,t}\Delta_t\left(\frac{\ell_{if}}{y_i}\right)+\sum_{c\in\mathcal{K}}\alpha^k_{ic,t}\Delta_t\left(\frac{k_{ic}}{y_i}\right)\right]-\omega^X_{i,t}\sum_{g\in\mathcal{G}}\omega^x_{ig,t}\Delta_t\left(\frac{x_{ig}}{y_i}\right).$$

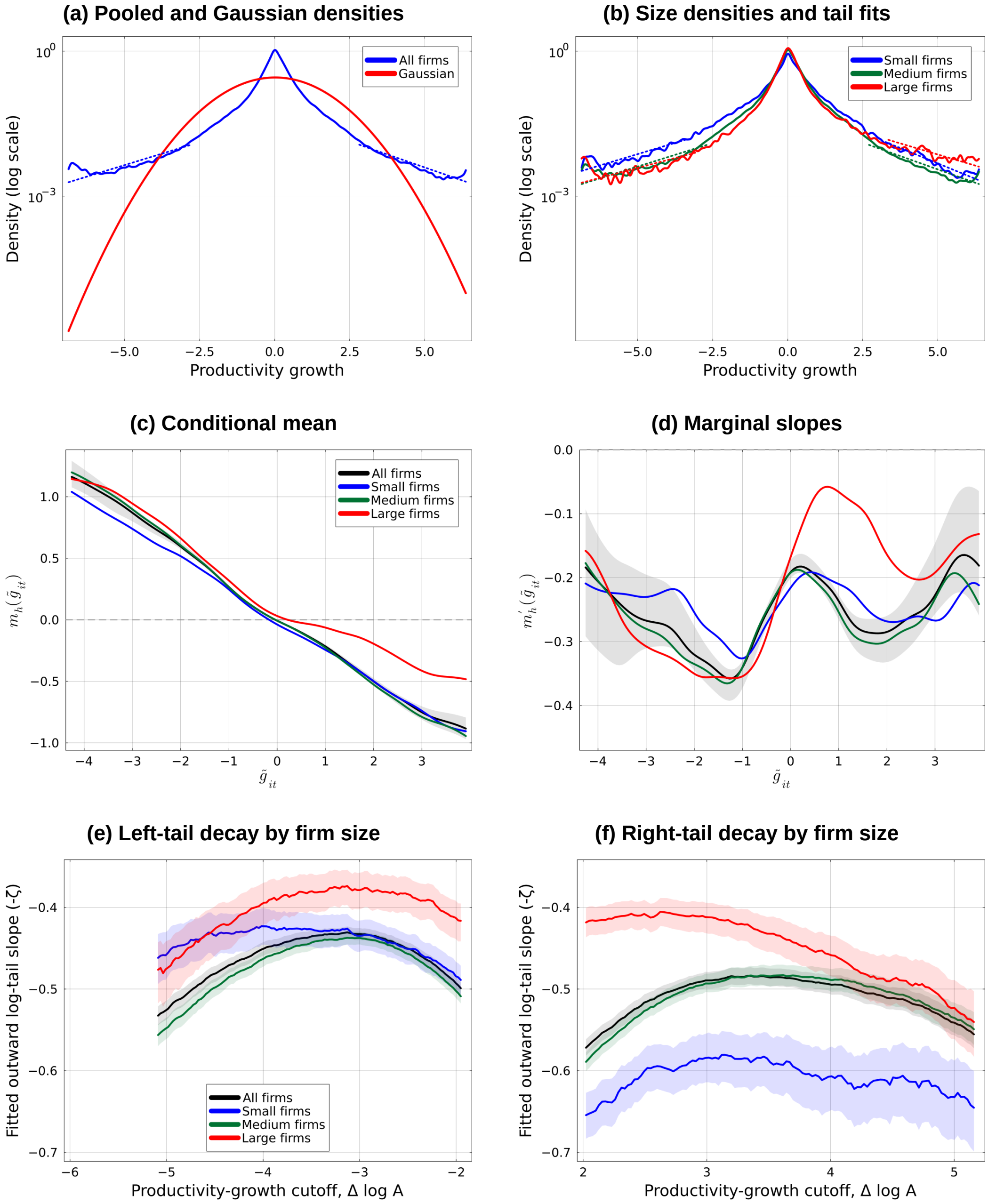
D Robustness-Alternative Productivity Measures

Figures 1–11 repeat Figures 3–5 under eleven robustness exercises. The first seven vary the inputs entering measured productivity under the benchmark $\theta=1$ specification. Figures 8–10 retain the all-input measure and full available sample while setting $\theta=0$, $\theta=0.5$, and $\theta=2$, respectively. Finally, Figure 11 retains the all-input measure and $\theta=1$ but restricts the sample to firms producing only one good in every observed year during 2011–2023, excluding firms ever observed producing multiple goods. In each figure, panels (a) and (b) compare the pooled productivity-growth density with a matched Gaussian and show the size-specific densities and tail fits. Panels (c) and (d) report the year-residualized nonparametric conditional mean and its marginal slope. Panels (e) and (f) report fitted outward log-tail slopes over alternative left- and right-tail cutoffs. The same standard-deviation-based cutoff rule is used throughout, although numerical cutoffs vary with each specification's growth distribution.

The principal findings are broadly stable across all eleven figures. Productivity growth remains sharply concentrated around zero and substantially more fat-tailed than the matched Gaussian. Small units generally place relatively more mass in the shoulders, although relative peak heights and density rankings in the outer tails vary across specifications. The pooled conditional means remain downward sloping and cross near the origin, while the pooled marginal slopes remain negative and nonlinear. Thus, in the pooled relationship, large negative realizations predict subsequent improvement and large positive realizations subsequent decline, with variation in reversal strength. At common right-tail cutoffs, estimated large-unit slopes remain closer to zero than small-unit slopes, indicating greater conditional tail reach among large units once a tail is entered. This ordering generally also holds in the left tail, but with cutoff-specific exceptions.

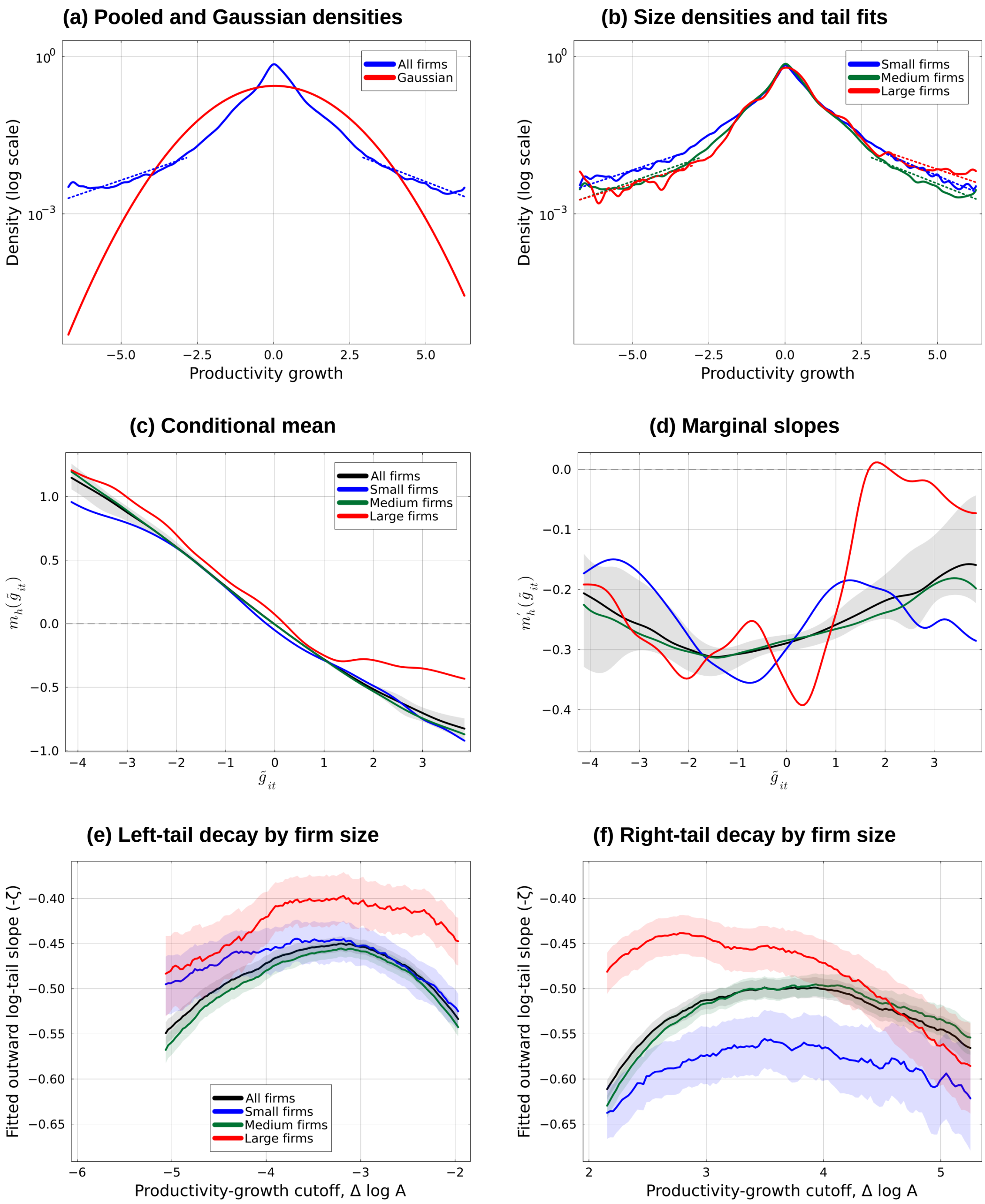
In the labor-only and labor-and-capital panels (Figures 1 and 2), the marginal-slope curves remain predominantly negative and nonlinear but display sharper size-specific curvature over parts of the support. In Figures 5, 6, and 7, the size-specific conditional means differ in local ordering and curvature, particularly over positive growth realizations. Changing θ (Figures 8, 9, and 10) shifts peak heights, turning points, and tail-slope levels while preserving the negative pooled marginal-slope profile; some size-specific slopes are positive over limited intervals. Restricting the sample to single-product firms (Figure 11), for which within-plant input allocation across products is unnecessary, preserves the sharply peaked density, pooled mean-reversion profile, and large-versus-small conditional tail-reach ordering on both sides. The size-specific conditional means nevertheless differ in local ordering and curvature, and large-unit marginal slopes are positive over parts of the support. Thus, the single-product evidence reinforces the core distributional, pooled-dynamic, and tail findings without implying exact pointwise invariance of every size-specific curve.

Figure 1: Productivity-Growth Evidence in the Labor-Only Panel ($\theta = 1$)



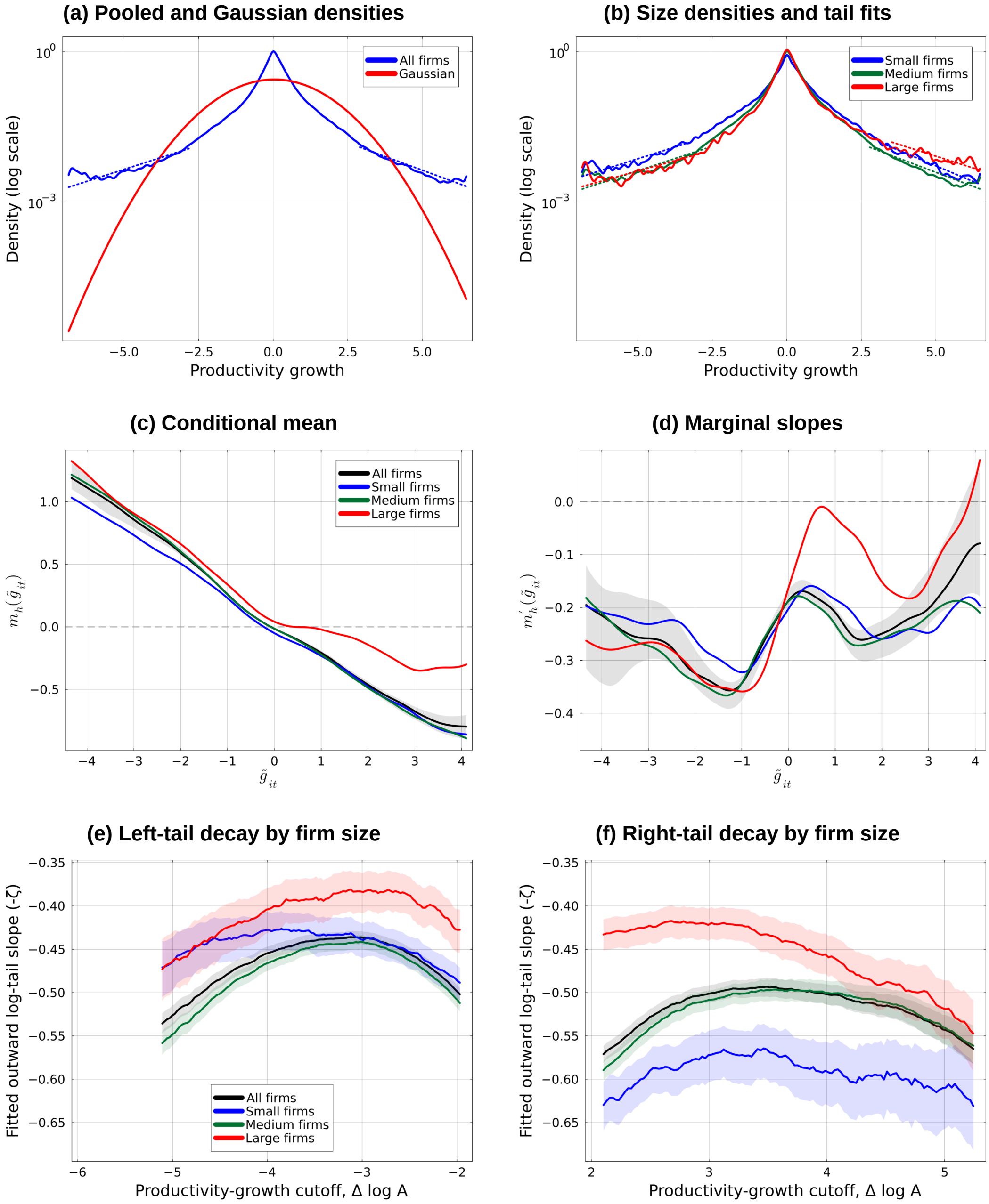
Notes: Panel (a) compares the pooled kernel density with a Gaussian matching its mean and variance; panel (b) shows size-specific densities. All kernel densities share the pooled-data bandwidth; dotted lines fit exponential tails beyond each group's mean ± 2 standard deviations. Establishment–product size is beginning-year log sales standardized within each annual sample: small below -1.33, medium between -1.33 and 1.33, and large above 1.33. Panels (c) and (d) use consecutive growth observations, remove transition-year means from both variables, and report Gaussian-kernel local-cubic conditional means and marginal slopes. The grey band is the pooled 95% pointwise robust bias-corrected interval, with plant-clustered standard errors; size-specific curves are descriptive. Panels (e) and (f) show outward log-tail slopes at SD-based cutoffs shared across size groups, with pointwise 95% intervals from 1,000 plant-cluster bootstrap replications. This figure uses labor-only productivity under the $\theta = 1$.

Figure 2: Productivity-Growth Evidence in the Labor-and-Capital Panel ($\theta = 1$)



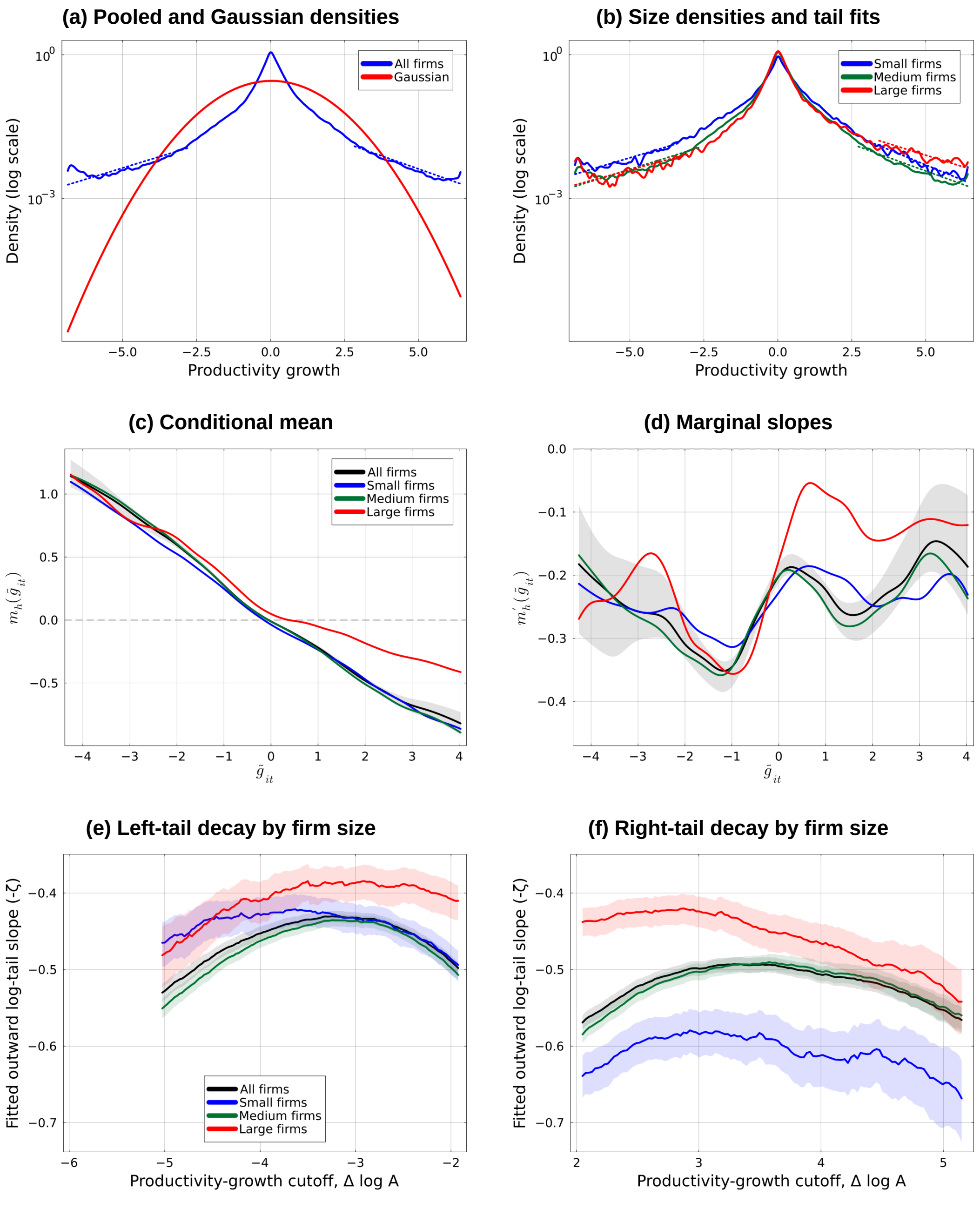
Notes: Panel (a) compares the pooled kernel density with a Gaussian matching its mean and variance; panel (b) shows size-specific densities. All kernel densities share the pooled-data bandwidth; dotted lines fit exponential tails beyond each group's mean ± 2 standard deviations. Establishment–product size is beginning-year log sales standardized within each annual sample: small below -1.33, medium between -1.33 and 1.33, and large above 1.33. Panels (c) and (d) use consecutive growth observations, remove transition-year means from both variables, and report Gaussian-kernel local-cubic conditional means and marginal slopes. The grey band is the pooled 95% pointwise robust bias-corrected interval, with plant-clustered standard errors; size-specific curves are descriptive. Panels (e) and (f) show outward log-tail slopes at SD-based cutoffs shared across size groups, with pointwise 95% intervals from 1,000 plant-cluster bootstrap replications. This figure uses labor-and-capital productivity with $\theta = 1$.

Figure 3: Productivity-Growth Evidence in the Labor-and-Imported-Input Panel ($\theta = 1$)



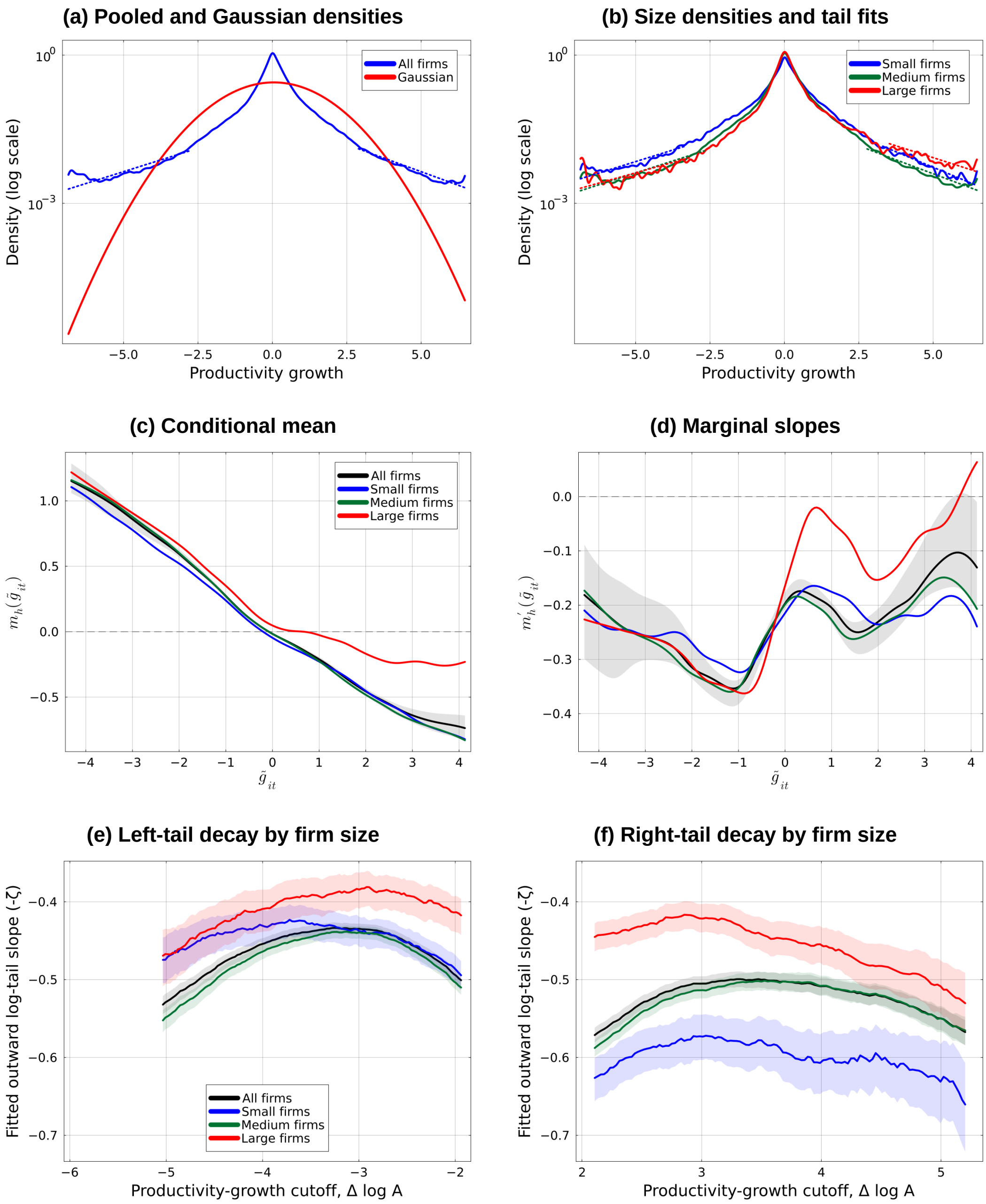
Notes: Panel (a) compares the pooled kernel density with a Gaussian matching its mean and variance; panel (b) shows size-specific densities. All kernel densities share the pooled-data bandwidth; dotted lines fit exponential tails beyond each group's mean ± 2 standard deviations. Establishment–product size is beginning-year log sales standardized within each annual sample: small below -1.33, medium between -1.33 and 1.33, and large above 1.33. Panels (c) and (d) use consecutive growth observations, remove transition-year means from both variables, and report Gaussian-kernel local-cubic conditional means and marginal slopes. The grey band is the pooled 95% pointwise robust bias-corrected interval, with plant-clustered standard errors; size-specific curves are descriptive. Panels (e) and (f) show outward log-tail slopes at SD-based cutoffs shared across size groups, with pointwise 95% intervals from 1,000 plant-cluster bootstrap replications. This figure uses labor and imported inputs under the $\theta = 1$.

Figure 4: Productivity-Growth Evidence in the Labor-and-Primary-Factor Panel ($\theta = 1$)



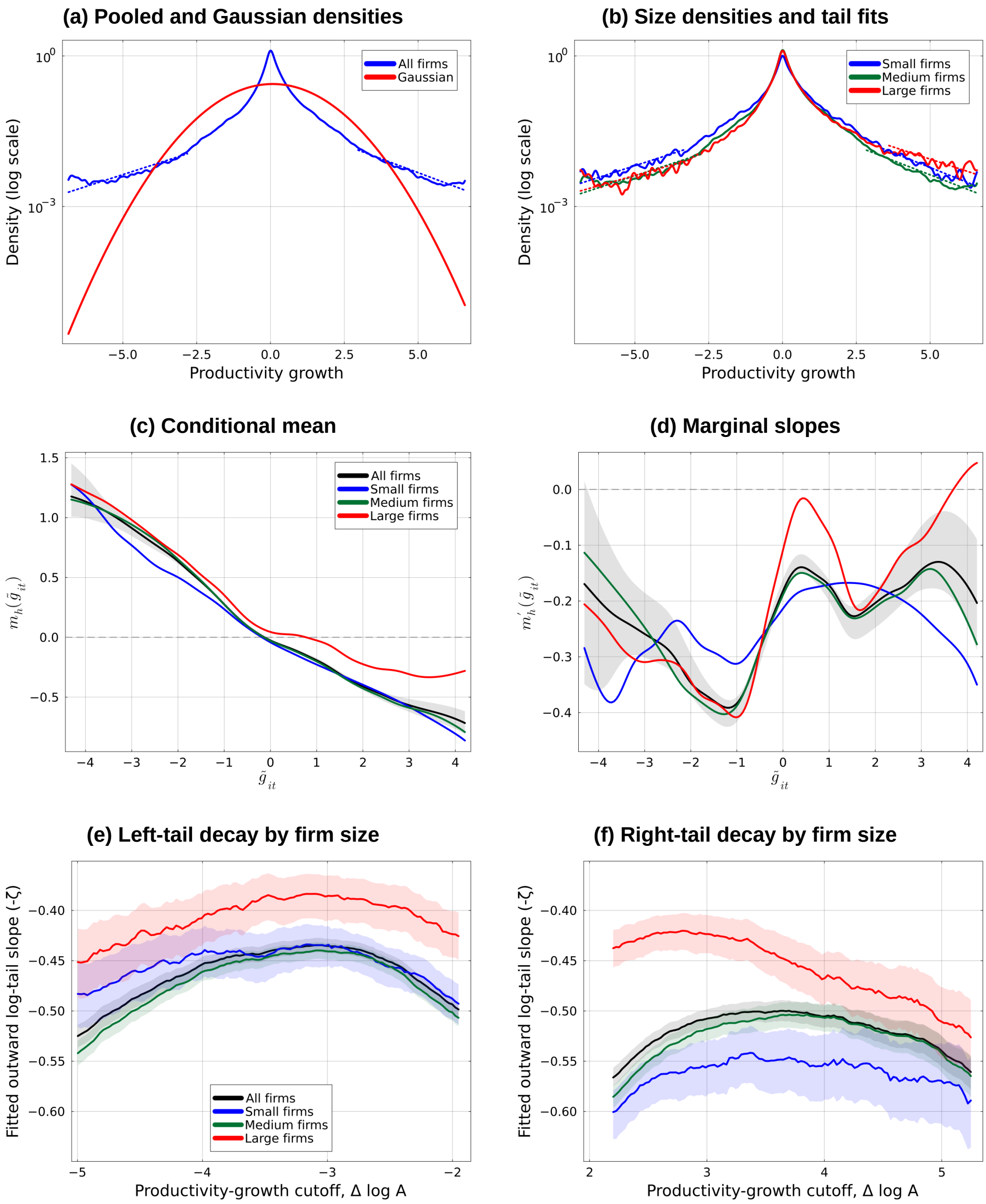
Notes: Panel (a) compares the pooled kernel density with a Gaussian matching its mean and variance; panel (b) shows size-specific densities. All kernel densities share the pooled-data bandwidth; dotted lines fit exponential tails beyond each group's mean ± 2 standard deviations. Establishment–product size is beginning-year log sales standardized within each annual sample: small below -1.33, medium between -1.33 and 1.33, and large above 1.33. Panels (c) and (d) use consecutive growth observations, remove transition-year means from both variables, and report Gaussian-kernel local-cubic conditional means and marginal slopes. The grey band is the pooled 95% pointwise robust bias-corrected interval, with plant-clustered standard errors; size-specific curves are descriptive. Panels (e) and (f) show outward log-tail slopes at SD-based cutoffs shared across size groups, with pointwise 95% intervals from 1,000 plant-cluster bootstrap replications. This figure uses labor and nonproduced domestic primary factors under the $\theta = 1$ benchmark.

Figure 5: Productivity-Growth Evidence in the Labor, Primary-Factor, and Imported-Input Panel ($\theta = 1$)



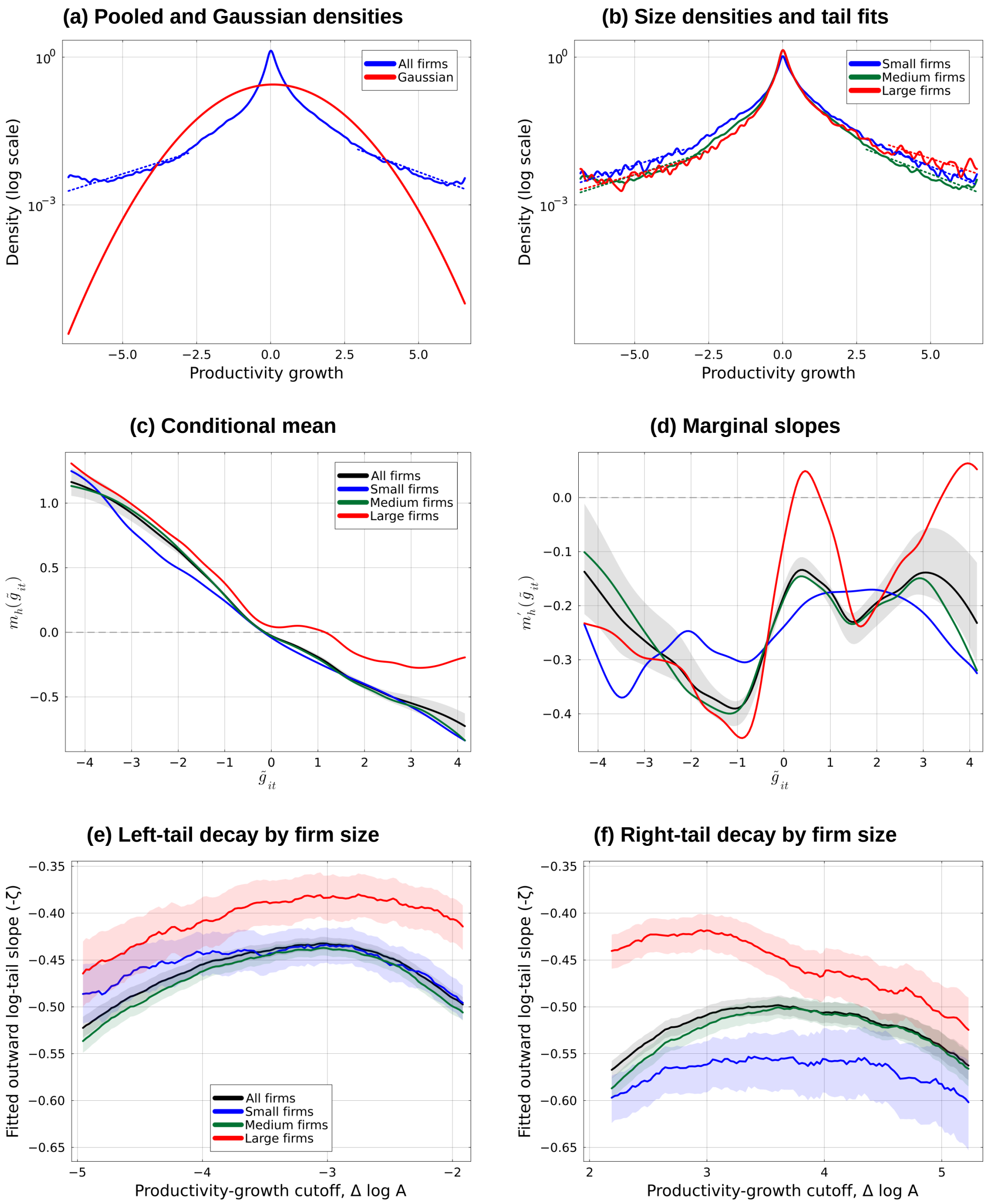
textitNotes: Panel (a) compares the pooled kernel density with a Gaussian matching its mean and variance; panel (b) shows size-specific densities. All kernel densities share the pooled-data bandwidth; dotted lines fit exponential tails beyond each group's mean ± 2 standard deviations. Establishment–product size is beginning-year log sales standardized within each annual sample: small below -1.33, medium between -1.33 and 1.33, and large above 1.33. Panels (c) and (d) use consecutive growth observations, remove transition-year means from both variables, and report Gaussian-kernel local-cubic conditional means and marginal slopes. The grey band is the pooled 95% pointwise robust bias-corrected interval, with plant-clustered standard errors; size-specific curves are descriptive. Panels (e) and (f) show outward log-tail slopes at SD-based cutoffs shared across size groups, with pointwise 95% intervals from 1,000 plant-cluster bootstrap replications. This figure uses labor, nonproduced domestic primary factors, and imports under the $\theta = 1$ benchmark.

Figure 6: Productivity-Growth Evidence in the Labor-and-Intermediate-Input Panel ($\theta = 1$)



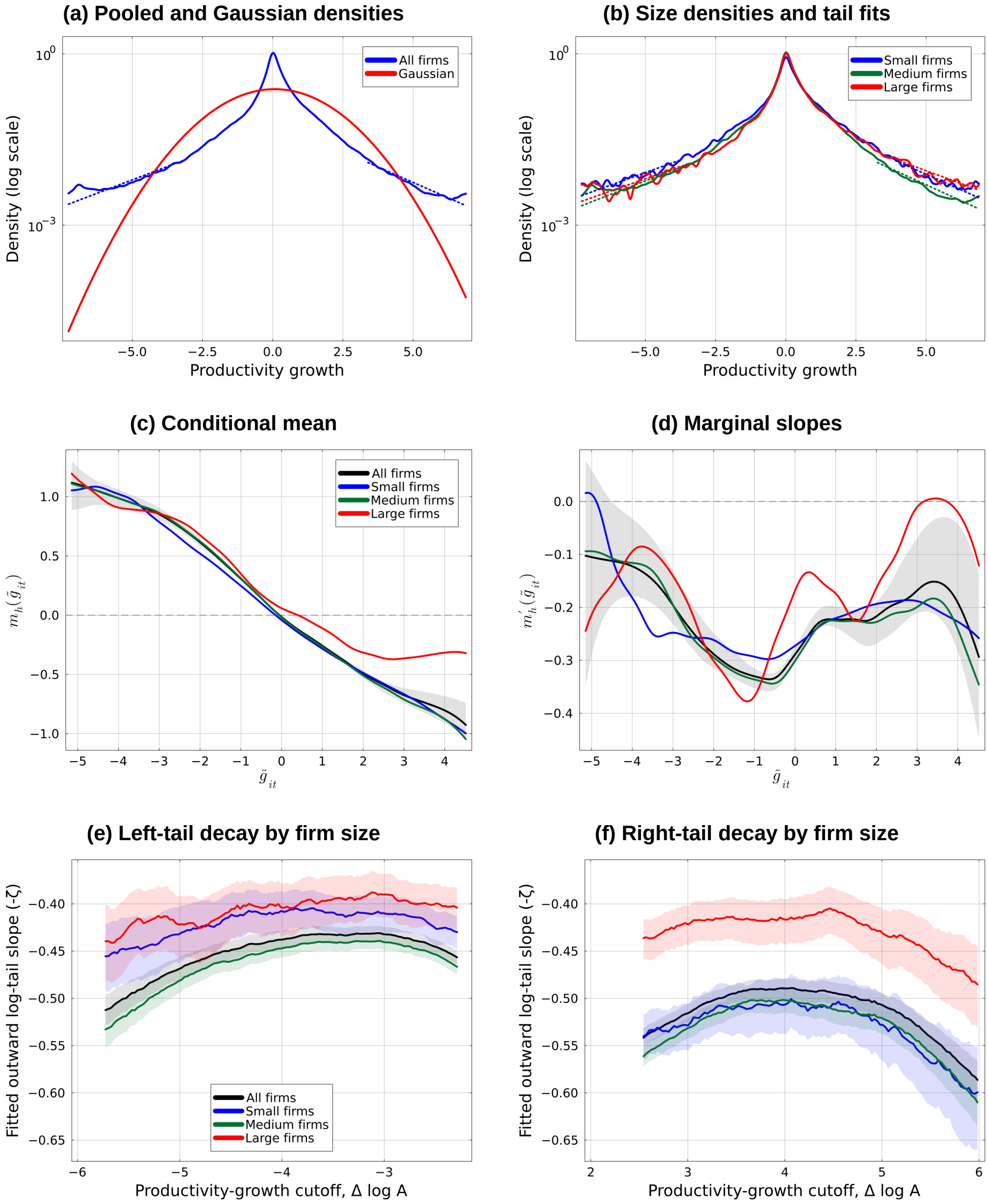
Notes: Panel (a) compares the pooled kernel density with a Gaussian matching its mean and variance; panel (b) shows size-specific densities. All kernel densities share the pooled-data bandwidth; dotted lines fit exponential tails beyond each group's mean ± 2 standard deviations. Establishment–product size is beginning-year log sales standardized within each annual sample: small below -1.33, medium between -1.33 and 1.33, and large above 1.33. Panels (c) and (d) use consecutive growth observations, remove transition-year means from both variables, and report Gaussian-kernel local-cubic conditional means and marginal slopes. The grey band is the pooled 95% pointwise robust bias-corrected interval, with plant-clustered standard errors; size-specific curves are descriptive. Panels (e) and (f) show outward log-tail slopes at SD-based cutoffs shared across size groups, with pointwise 95% intervals from 1,000 plant-cluster bootstrap replications. This figure uses labor and domestic intermediate inputs under the $\theta = 1$ benchmark

Figure 7: Productivity-Growth Evidence in the Labor, Primary-Factor, Intermediate-Input, and Imported-Input Panel ($\theta = 1$)



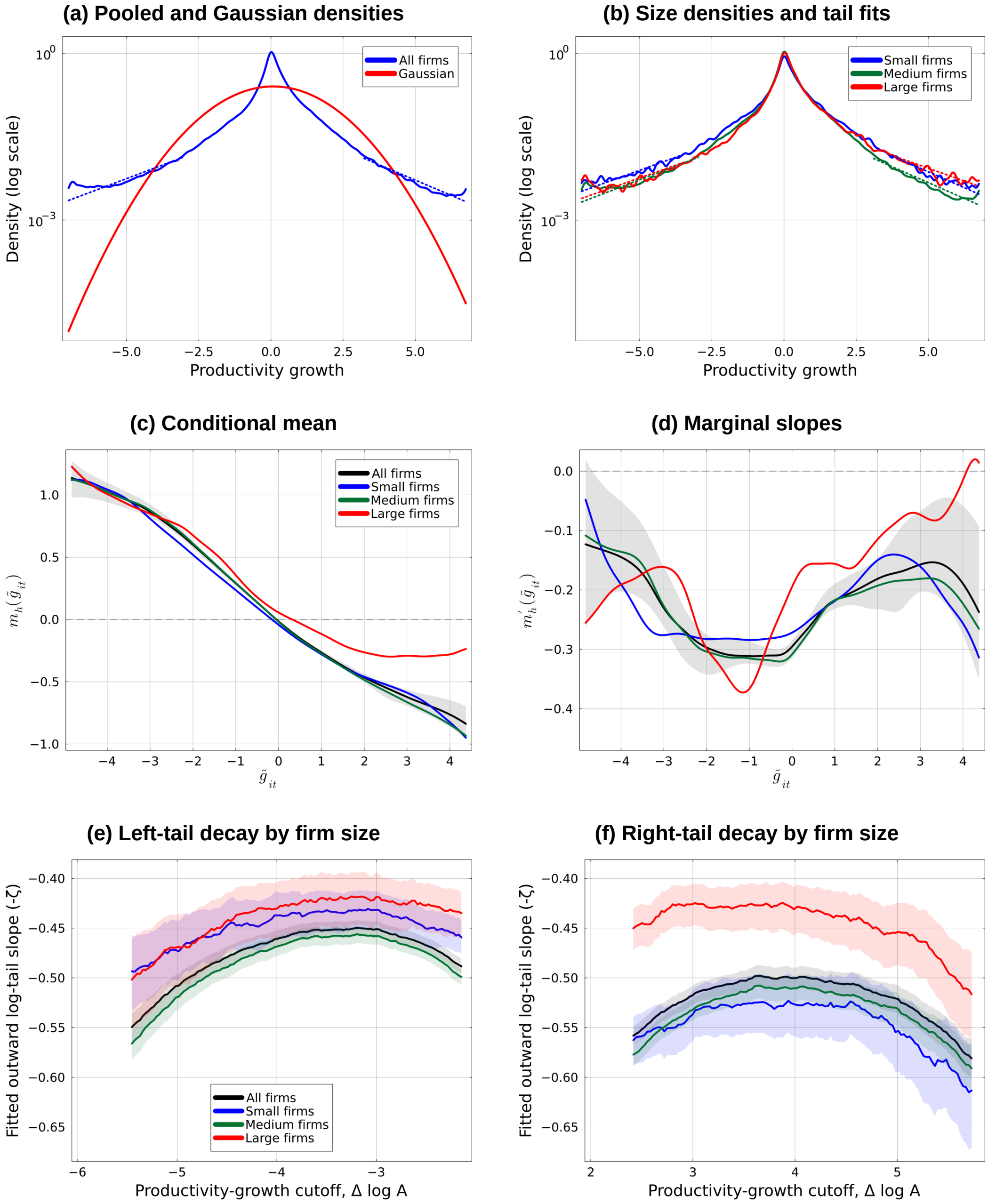
Notes: Panel (a) compares the pooled kernel density with a Gaussian matching its mean and variance; panel (b) shows size-specific densities. All kernel densities share the pooled-data bandwidth; dotted lines fit exponential tails beyond each group's mean ± 2 standard deviations. Establishment–product size is beginning-year log sales standardized within each annual sample: small below -1.33, medium between -1.33 and 1.33, and large above 1.33. Panels (c) and (d) use consecutive growth observations, remove transition-year means from both variables, and report Gaussian-kernel local-cubic conditional means and marginal slopes. The grey band is the pooled 95% pointwise robust bias-corrected interval, with plant-clustered standard errors; size-specific curves are descriptive. Panels (e) and (f) show outward log-tail slopes at SD-based cutoffs shared across size groups, with pointwise 95% intervals from 1,000 plant-cluster bootstrap replications. This figure uses labor, nonproduced domestic primary factors, domestic intermediate inputs, and imports, excluding capital, under the $\theta = 1$ benchmark.

Figure 8: Productivity-Growth Evidence in the Full Panel ($\theta = 0$)



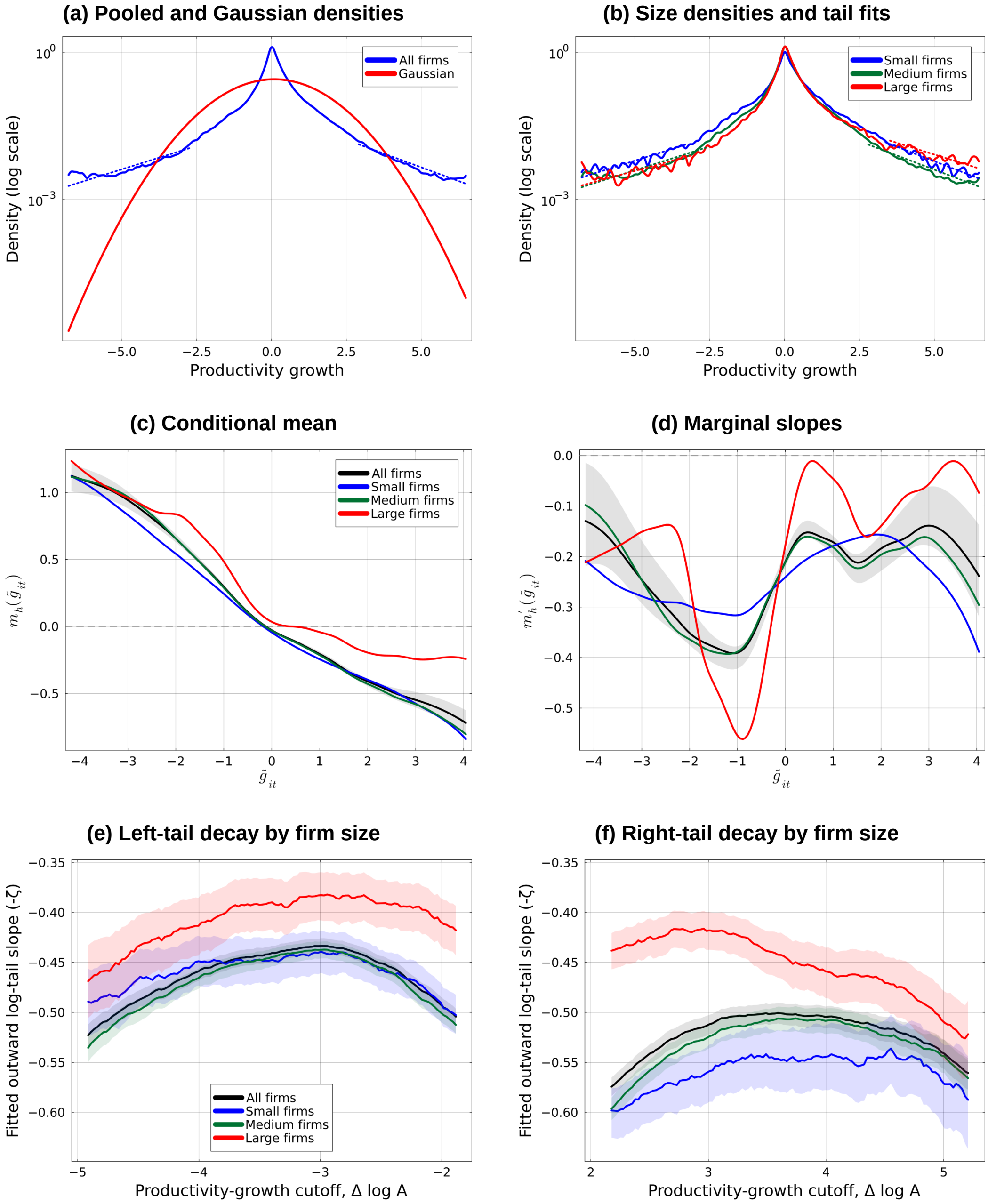
Notes: Panel (a) compares the pooled kernel density with a Gaussian matching its mean and variance; panel (b) shows size-specific densities. All kernel densities share the pooled-data bandwidth; dotted lines fit exponential tails beyond each group's mean ± 2 standard deviations. Establishment–product size is beginning-year log sales standardized within each annual sample: small below -1.33, medium between -1.33 and 1.33, and large above 1.33. Panels (c) and (d) use consecutive growth observations, remove transition-year means from both variables, and report Gaussian-kernel local-cubic conditional means and marginal slopes. The grey band is the pooled 95% pointwise robust bias-corrected interval, with plant-clustered standard errors; size-specific curves are descriptive. Panels (e) and (f) show outward log-tail slopes at SD-based cutoffs shared across size groups, with pointwise 95% intervals from 1,000 plant-cluster bootstrap replications. This figure uses all-input productivity with $\theta = 0$.

Figure 9: Productivity-Growth Evidence in the Full Panel ($\theta = 0.5$)



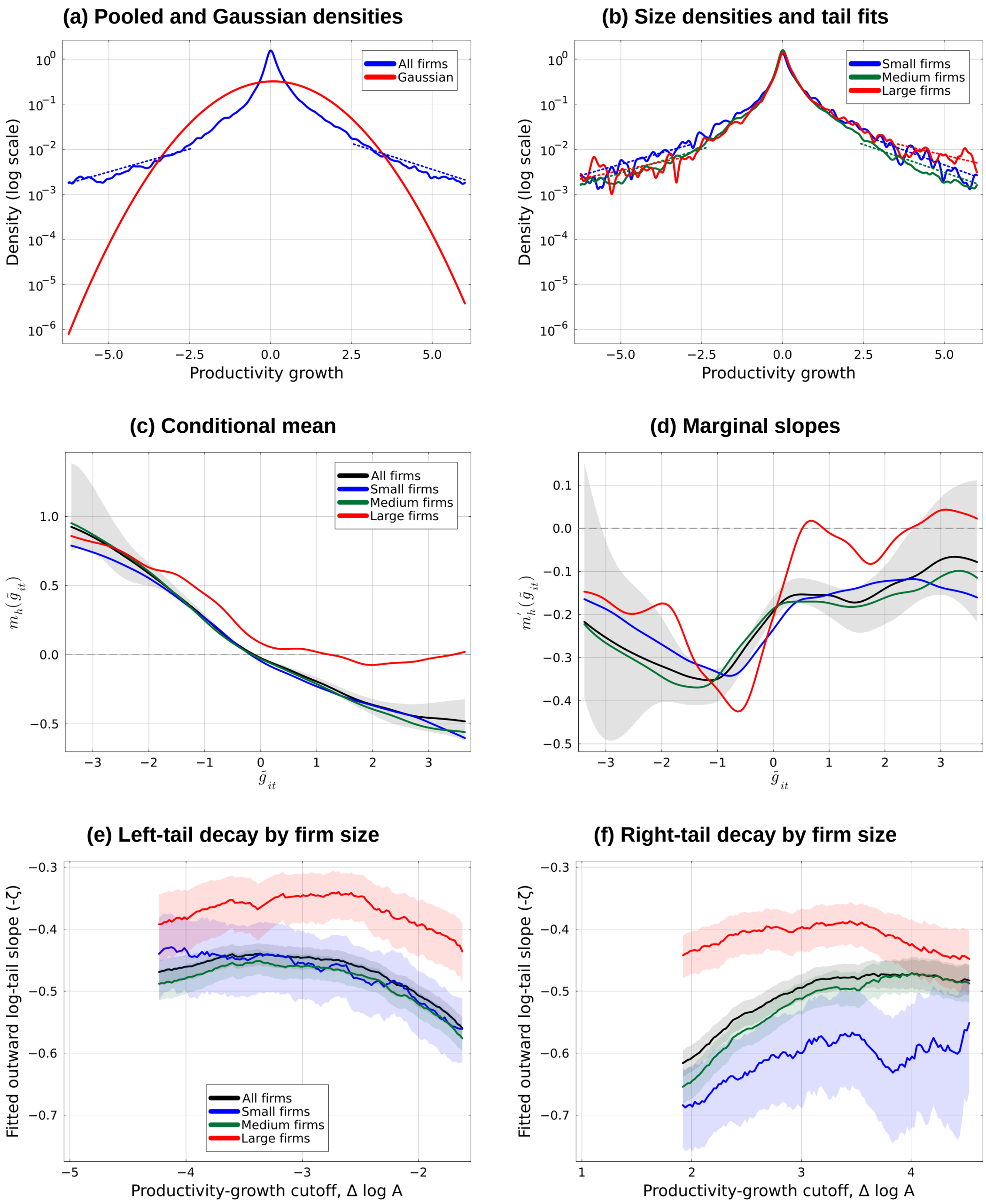
Notes: Panel (a) compares the pooled kernel density with a Gaussian matching its mean and variance; panel (b) shows size-specific densities. All kernel densities share the pooled-data bandwidth; dotted lines fit exponential tails beyond each group's mean ± 2 standard deviations. Establishment–product size is beginning-year log sales standardized within each annual sample: small below -1.33, medium between -1.33 and 1.33, and large above 1.33. Panels (c) and (d) use consecutive growth observations, remove transition-year means from both variables, and report Gaussian-kernel local-cubic conditional means and marginal slopes. The grey band is the pooled 95% pointwise robust bias-corrected interval, with plant-clustered standard errors; size-specific curves are descriptive. Panels (e) and (f) show outward log-tail slopes at SD-based cutoffs shared across size groups, with pointwise 95% intervals from 1,000 plant-cluster bootstrap replications. This figure uses all-input productivity with $\theta = 0.5$.

Figure 10: Productivity-Growth Evidence in the Full Panel ($\theta = 2$)



Notes: Panel (a) compares the pooled kernel density with a Gaussian matching its mean and variance; panel (b) shows size-specific densities. All kernel densities share the pooled-data bandwidth; dotted lines fit exponential tails beyond each group's mean ±2 standard deviations. Establishment–product size is beginning-year log sales standardized within each annual sample: small below −1.33, medium between −1.33 and 1.33, and large above 1.33. Panels (c) and (d) use consecutive growth observations, remove transition-year means from both variables, and report Gaussian-kernel local-cubic conditional means and marginal slopes. The grey band is the pooled 95% pointwise robust bias-corrected interval, with plant-clustered standard errors; size-specific curves are descriptive. Panels (e) and (f) show outward log-tail slopes at SD-based cutoffs shared across size groups, with pointwise 95% intervals from 1,000 plant-cluster bootstrap replications. This figure uses all-input productivity with $\theta = 2$.

Figure 11: Productivity-Growth Evidence for Single-Product Firms in the Full Panel ($\theta = 1$)



Notes: Panel (a) compares the pooled kernel density with a Gaussian matching its mean and variance; panel (b) shows size-specific densities. All kernel densities share the pooled-data bandwidth; dotted lines fit exponential tails beyond each group's mean ± 2 standard deviations. Establishment–product size is beginning-year log sales standardized within each annual sample: small below -1.33, medium between -1.33 and 1.33, and large above 1.33. Panels (c) and (d) use consecutive growth observations, remove transition-year means from both variables, and report Gaussian-kernel local-cubic conditional means and marginal slopes. The grey band is the pooled 95% pointwise robust bias-corrected interval, with plant-clustered standard errors; size-specific curves are descriptive. Panels (e) and (f) show outward log-tail slopes at SD-based cutoffs shared across size groups, with pointwise 95% intervals from 1,000 plant-cluster bootstrap replications. This figure uses all-input productivity with $\theta = 1$, excluding firms ever observed with multiple positive-sales products during 2011–2023. Retained firms may switch products between years and need not appear in every year.