Board of Governors of the Federal Reserve System

International Finance Discussion Papers

ISSN 1073-2500 (Print)
ISSN 2767-4509 (Online)

Number 1450

November 2026

Who Built the Service Economy?

Immo Schott; Vladimir Smirnyagin

Please cite this paper as:
Schott, Immo and Vladimir Smirnyagin (2026). "Who Built the Service Economy?," International Finance Discussion Papers 1450. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/IFDP.2026.1450.

NOTE: International Finance Discussion Papers (IFDPs) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the International Finance Discussion Papers Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers. Recent IFDPs are available on the Web at www.federalreserve.gov/pubs/ifdp/. This paper can be downloaded without charge from the Social Science Research Network electronic library at www.ssrn.com.

Who Built the Service Economy?*

Immo Schott† Vladimir Smirnyagin‡

Federal Reserve Board University of Virginia

October 30, 2026

Abstract

We show empirically that the reallocation of employment from manufacturing to services in the U.S. was driven by large incumbent service firms, which expanded by opening additional establishments. Guided by these facts, we develop a two-sector general equilibrium model of firm dynamics in which incumbent firms can scale by operating multiple establishments. Quantitatively, a relative increase in manufacturing productivity and a decline in replication costs—the cost of opening and managing additional establishments—jointly account for more than half of the observed sectoral reallocation. The productivity channel shifts employment toward services but leaves the firm-size distribution unchanged. Lower replication costs, by contrast, generate the disproportionate expansion of large, multi-establishment service firms and, on their own, account for roughly one third of the reallocation generated by the two forces jointly. Our results identify firm dynamics—specifically, incumbent expansion through new establishments—as a quantitatively important force behind structural transformation.

Keywords: Firm dynamics; structural transformation; multi-establishment firms; entry

JEL Codes: L11, L16, O41

*We thank Sina Ates, Gian Luca Clementi, Ryan Decker, Francesco Ferrante, Joseph Kaboski, Igor Livshits, Hannah Rubinton, Felipe Saffie, Eric Sager, Roberto Samaniego, and Todd Schoellman for useful comments. We thank Veronika Penciakova for an excellent discussion.

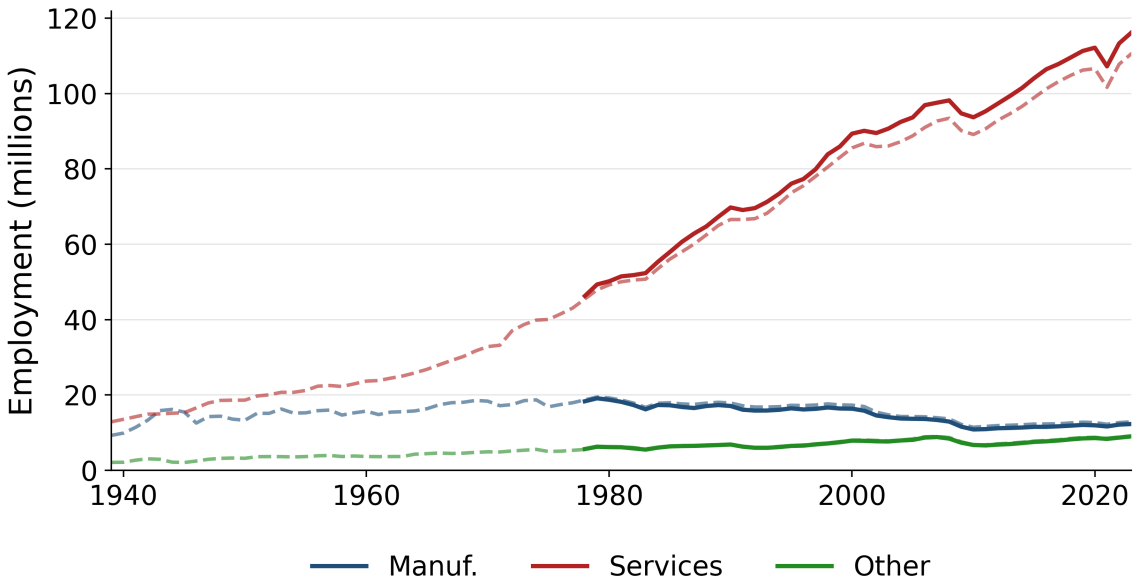
†E-mail: immoschott@gmail.com. Federal Reserve Board. Division of International Finance, 20th and C Streets, NW, Washington, DC 20551, United States. The views expressed here are those of the authors and do not necessarily reflect the views of the Federal Reserve System.

‡E-mail: vladimir.smirnyagin@virginia.edu. Department of Economics, University of Virginia. 248 McCormick Road, Charlottesville, VA 22904, United States.

1 Introduction

Structural transformation is a defining feature of modern growth. Figure 1 shows a steady reallocation of labor from the manufacturing sector toward the service sector, which accelerated around the 1980s. Since 1939, service employment has increased by a factor of nine, while manufacturing employment has increased by only about 40 percent. As a result, the service sector's share of total employment has risen from 50 to over 80 percent, while the manufacturing share has fallen from 40 to 10 percent.[1]

FIGURE 1: SECTORAL EMPLOYMENT



Notes: The figure shows employment in manufacturing, services, and other industries. Solid lines use the U.S. Census Bureau's Business Dynamics Statistics (BDS), which cover establishments with at least one paid employee and are available from 1978. Dashed lines use the Bureau of Labor Statistics' Current Employment Statistics (CES), which extends back to 1939. Appendix Table A.1 provides the mapping of industries to broad sector groups.

How did this large reallocation of labor come about? Did it primarily reflect the rise of new service firms or the expansion of large, established incumbents? And if incumbent firms drove the growth of the service sector, did they expand existing establishments or open new

[1] We decompose the service-sector employment gains by subsector in Appendix Figure A.1. The 'Health Care and Social Assistance' subsector accounts for the largest share of the gains (about 23 percent). Administrative and support services, accommodation and food services, and professional and technical services also expanded substantially, each accounting for more than 10 percent of the total employment growth.

ones? In this paper, we take a firm-dynamics perspective to answer these questions, focusing on how incumbent scaling—particularly the expansion through additional establishments—shaped structural transformation.

The trends shown in Figure 1 are well documented. However, much of the structural transformation literature treats sectors as aggregates—often implicitly as collections of representative firms—and therefore abstracts from the underlying firm dynamics. As a result, it is silent about whether sectoral reallocation reflects changes in the number of firms or the scaling of incumbents, or whether incumbents grow within existing establishments or by opening new ones. The firm dynamics margin is important because a large empirical literature shows that firm turnover, heterogeneous firm growth, and reallocation across firms account for a large share of aggregate employment dynamics (e.g., Davis et al., 1996; Haltiwanger et al., 2013).

Using data from the U.S. Census Bureau's Business Dynamics Statistics (BDS), we find that structural transformation is primarily an incumbent scaling phenomenon. Specifically, we document three findings. First, service-sector employment has become increasingly concentrated in large firms. Second, large service firms expanded primarily by increasing the number of establishments they operate, rather than by increasing employment at existing establishments. Third, service-sector growth is driven by incumbent expansion rather than net firm entry: gross entry and exit flows are sizable, but they largely offset each other. Establishment creation within incumbent firms accounts for a substantial share of service-sector employment growth.

Guided by these facts, we develop a two-sector general-equilibrium model of firm dynamics. Our model extends the canonical framework of Hopenhayn and Rogerson (1993) by allowing incumbent firms to operate multiple establishments. Adding establishments lets a firm grow without running into establishment-level decreasing returns, thereby raising the effective firm-level returns to scale. Expansion requires paying a convex replication cost. It captures the managerial and organizational resources needed to open and coordinate additional locations, such as standardizing procedures, training managers, and monitoring establishments from a distance. The convexity of the cost function reflects span-of-control limits: each additional establishment is harder to manage than the last, which governs how

far incumbents can profitably scale.

Following the structural transformation literature (e.g., Kongsamut et al., 2001; Herrendorf et al., 2014), household preferences are non-homothetic and feature complementarity between manufactured goods and services. These preferences introduce the two standard forces behind structural transformation: rising income shifts expenditure toward services, while faster productivity growth in manufacturing lowers the relative price of manufactured goods and reallocates employment toward services (Ngai and Pissarides, 2007).

We use an analytically tractable version of the model to establish two results. First, under free entry, an increase in sectoral productivity is offset by a proportional decline in the sector's output price, leaving firm policy functions—and hence the within-sector firm-size distribution—unchanged. Relative manufacturing productivity growth therefore cannot reproduce the firm-level facts behind structural transformation. Second, a decline in the convexity of replication costs raises the number of establishments, with larger responses among firms that initially operate more establishments. Provided that large service firms operate more establishments than their manufacturing counterparts in the data, this result predicts a stronger expansion response among large service incumbents even when both sectors experience the same change in establishment-expansion technology.

We calibrate the model to 1979, the first year for which all of our calibration moments are available in the BDS data. In each sector, we target a rich set of moments including the firm-size distribution, the average number of establishments per firm across firm-size groups, and the size of entrants and young firms. In our main experiment, manufacturing productivity rises relative to services and the cost of opening additional establishments falls. We choose the productivity change to match the observed decline in the relative price of manufactured goods between 1979 and 2023 and the change in replication costs to match the increase in establishments per firm among large service firms. Together, the two channels in our main experiment account for 52 percent of the observed sectoral employment reallocation and reproduce the shift toward large, multi-establishment service firms. Lower replication costs alone account for 17 percent of the observed reallocation—roughly one third of the employment reallocation generated by the two channels jointly. Thus, firm dynamics, in particular incumbent expansion through additional establishments, play a quantitatively

important role in structural transformation.

We interpret the decline in replication costs as reflecting advances in information technology (IT) and logistics, as well as the standardization of business processes. Recent models of firm expansion across local markets, branches, and products provide distinct mechanisms through which more productive firms expand disproportionately along the extensive margin (Hsieh and Rossi-Hansberg, 2023; Brynjolfsson et al., 2025; Argente et al., 2025). In each of these models, the number of units a firm operates rises with its productivity. We show that the changes in technology or demand considered in these papers manifest themselves as an increase in the slope of the relationship between firm productivity and the number of units operated. Our model captures the same relationship between establishments and productivity with a single parameter, the convexity of replication costs. We do not take a stand on a particular mechanism of firm expansion but focus on the combined effect of these forces on the sectoral reallocation of employment.

Related literature. Our paper contributes to three strands of the literature. First, it relates to the literature on firm dynamics. Canonical models study how entry, exit, and incumbent growth shape the firm-size distribution (Hopenhayn, 1992; Hopenhayn and Rogerson, 1993; Luttmer, 2007). Subsequent work shows how firm entry and exit amplify and propagate aggregate shocks (Clementi et al., 2015; Clementi and Palazzo, 2016). An important strand of this literature emphasizes heterogeneity in firms' growth potential at entry. Sedláček and Sterk (2017) show that conditions at birth have persistent effects on cohort employment, while Sterk et al. (2021) establish the importance of ex ante heterogeneity for subsequent firm growth. Smirnyagin (2023) highlights firms' returns to scale as a dimension of entrant heterogeneity with persistent aggregate consequences. Recent work links demographic changes to firm entry (Hopenhayn et al., 2022; Karahan et al., 2024; Bornstein, 2025; Peters and Walsh, 2026), while Hopenhayn et al. (2025) study the connection between firm demographics and structural transformation.[2] We contribute to this literature by distinguishing the role of entering cohorts from incumbent expansion in sectoral employment reallocation, and by studying how the cost of operating additional establishments shapes

[2]Relatedly, Dent et al. (2016) decompose structural transformation into differences in the sectoral allocation of startup employment and subsequent cohort life-cycle dynamics.

incumbent growth.

Second, the paper is related to work on structural transformation and multisector growth (see Herrendorf et al., 2014, for a review). That literature explains the reallocation of employment toward services through two channels: relative price effects arising from uneven sectoral productivity growth (Ngai and Pissarides, 2007; Herrendorf et al., 2015; Samaniego and Sun, 2016), and income effects, which stem from non-homothetic demand (Kongsamut et al., 2001; Buera and Kaboski, 2012; Comin et al., 2021). These papers explain the allocation of employment across sectors while typically abstracting from the firm dynamics underlying that reallocation.[3] Our model combines these demand and productivity channels with endogenous establishment expansion. This allows us to assess which forces can jointly account for sectoral employment reallocation and the growing importance of large, multi-establishment service firms.

Third, the paper relates to work on firms' spatial expansion and multi-establishment organization. Hsieh and Rossi-Hansberg (2023) document that productive service firms increasingly scale across local markets. They attribute this expansion to the adoption of new fixed-cost technologies and study its implications for rising concentration.[4] Cao et al. (2022) decompose the secular increase in average U.S. firm size into employment per establishment and establishments per firm. They use a one-sector growth model to attribute the rise in establishments per firm to lower external innovation costs, lower establishment exit, and slower aggregate growth. Our contribution is to connect establishment expansion to structural transformation and to quantify its role in the joint evolution of sectoral employment shares and within-sector employment concentration.

Several technological and organizational changes can lower the cost of expanding through establishment growth. In Aghion et al. (2023), lower overhead costs of spanning product markets allow the most efficient firms to expand into additional markets. Advances in in-

[3] Fort et al. (2018) and Ding et al. (2022) document a rise in nonmanufacturing employment within manufacturing firms. This within-firm reallocation is complementary to the establishment expansion documented in our paper. Relatedly, Hyun et al. (2025) show that sectoral shocks propagate between manufacturing and service establishments within multisector firms, shaping industry-level employment dynamics.

[4] Oberfield et al. (2024) model firms' endogenous choice of the number, size, and location of plants, emphasizing spatial sorting. Kleinman (2025) studies a decline in spatial expansion frictions emphasizing wage inequality and spatial disparities. Xi (2023) also develops a dynamic multi-establishment firm model and studies establishment-level distortions, misallocation, and cross-country productivity.

formation technology can facilitate the replication of efficient routines across establishments and markets (Lashkari et al., 2024; Brynjolfsson et al., 2025), while complementarities between IT and firm scale can make falling IT prices particularly beneficial to large service firms (Eckert et al., 2025). Relatedly, De Ridder (2024) shows that intangible inputs such as software lower marginal costs while raising fixed costs, giving adopting firms a competitive advantage and discouraging entry. Other mechanisms include standardization, which lowers the cost of operating additional establishments (Argente et al., 2025), and reductions in the cost of managing geographically dispersed establishments (Gumpert et al., 2022). A concrete example comes from Holmes (2011), who estimates substantial distribution-cost savings from Wal-Mart's dense network of retail stores, showing how the profitability of an additional location depends on a firm's existing network (for related papers on retail chains see also Foster et al. (2006) and Jarmin et al. (2009) on their growing importance and Basker (2012) on productivity gains from barcode scanners). We show that several of these mechanisms map into our model as changes in the convexity of the replication cost schedule. Our model therefore captures a range of technological and organizational changes without taking a stand on which particular mechanism drives the decline in expansion costs.

The rest of the paper is organized as follows. Section 2 describes the data and documents our main empirical facts. Section 3 develops a two-sector firm dynamics model with multi-establishment firms and derives analytical results on how productivity and replication costs shape firm growth. Section 4 calibrates the model, performs our main quantitative experiment, and considers extensions of the baseline model. Section 5 concludes.

2 Empirical Results

In this section, we document the firm dynamics behind the shift toward the service sector shown in Figure 1. In particular, we highlight three facts. The first is about the importance of large firms: we show that service-sector employment has become increasingly concentrated in large firms. Second, we show that the expansion of these firms is primarily driven by an increase in the number of establishments they operate. Third, incumbent expansion rather than net firm entry accounts for most service-sector employment growth.

2.1 Data

Our empirical analysis draws on the BDS, housed by the U.S. Census Bureau. The Census Bureau constructs the BDS from the Longitudinal Business Database (LBD), the most comprehensive longitudinally linked dataset covering the universe of U.S. establishments with at least one paid employee. The BDS spans the time period 1978–2023 and reports a wide range of establishment- and firm-level statistics across industries, sectors, and geographic regions, including annual measures of job creation and destruction as well as firm startups and shutdowns.

We distinguish between establishments and firms. An establishment is a single physical location at which business is conducted, while a firm consists of one or more establishments under common ownership.[5] Aggregate employment trends in the BDS closely mirror those in the Bureau of Labor Statistics (BLS) data, which extend back to 1939 and appear as the dashed lines in Figure 1.

Throughout the analysis, we classify firms as small (fewer than 20 employees), medium (20–499 employees), or large (500 or more employees).[6] The manufacturing sector spans NAICS industries 31–33. The service sector comprises wholesale and retail trade, transportation, information, FIRE (finance, insurance, and real estate), professional services, and educational services, among other industries. We group the remaining industries—such as agriculture, mining, and construction—under "other".[7]

2.2 Structural Transformation and Large Firms

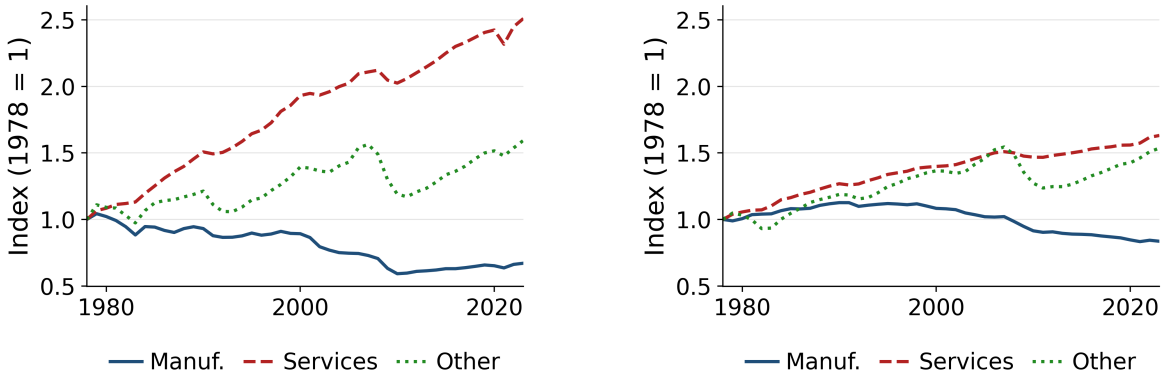
Panel (A) of Figure 2 plots sectoral employment relative to 1978. It complements the longer time series in Figure 1 and shows that relative to 1978 service employment has grown by a factor of 2.5, while manufacturing employment shrank by about one third. Interestingly,

[5]The BDS data record a firm entry when a new firm consists entirely of age-0 establishments, and a firm exit when all of a firm's establishments—regardless of changes in ownership over time—cease operations and exit.

[6]Firm size refers to employment across all establishments of a firm. Sectoral employment counts are assigned by establishment industry. Appendix Figure A.3 bounds the resulting size-bin misclassification and shows that it cannot account for the shift of service employment toward large firms.

[7]Table A.1 in the Appendix provides the mapping between NAICS two-digit industries and these broad sector categories.

FIGURE 2: SECTORAL EMPLOYMENT AND THE NUMBER OF FIRMS



(A) Employment

(B) Number of firms

Notes: All series are normalized to one in 1978. Data source: BDS.

there has been no comparable change in the *number* of firms. Panel (B) shows that the number of service firms increased by only 63 percent, while the number of manufacturing firms follows a hump-shaped path, with an eventual decline of only 16 percent. The fact that firm counts change much less than employment in both sectors implies substantial changes in the firm-size distribution, especially in services.

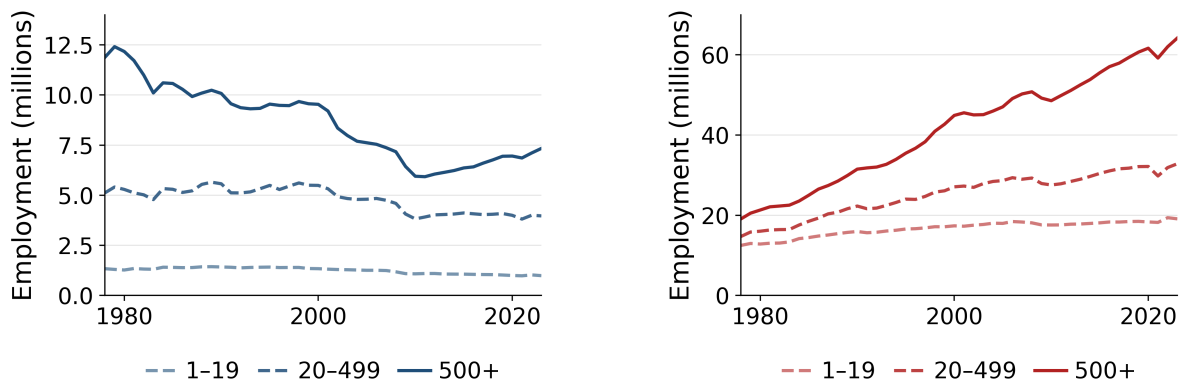
Figure 3 plots each sector's employment by firm-size group. Panel (A) corresponds to the manufacturing sector. Employment in large firms fell by about 4.5 million workers, or roughly 40 percent. Large manufacturing firms thus account for about three quarters of the secular decline in manufacturing employment. Employment in small and medium firms has declined by less. Panel (B) shows a striking concentration of service-sector employment growth in large firms: employment in this group more than tripled, while total employment in the medium and small firm-size groups increased by less. As a result, the largest firms' employment share rose from about 40 to over 55 percent over the sample period.[8]

2.3 Expansion Through Establishments

How did large service firms expand? Firms can increase employment by expanding existing establishments, adding new ones, or both. We now examine these two margins. The top row

[8] Appendix Figure A.2 shows the corresponding employment patterns by firm age. Service-sector growth is concentrated among older firms. The age patterns are complicated by the fact that firms present before the first year of data are left-censored, meaning their exact age is unknown.

FIGURE 3: SECTORAL EMPLOYMENT BY SIZE GROUP



(A) Manufacturing

(B) Services

Notes: Employment is reported in millions, by firm-size group. Data source: BDS.

of Figure 4 shows the average number of establishments per firm across firm-size groups in the manufacturing and service sectors. The key fact is that there has been a large increase in the average number of establishments per large service firm: from about 18 to 33 over the sample period.[9] At the same time, employment per service establishment has remained very stable (panel (D) of Figure 4). The increase in average employment per large service firm thus reflects an increase in establishments per firm, rather than an increase in average employment per establishment.

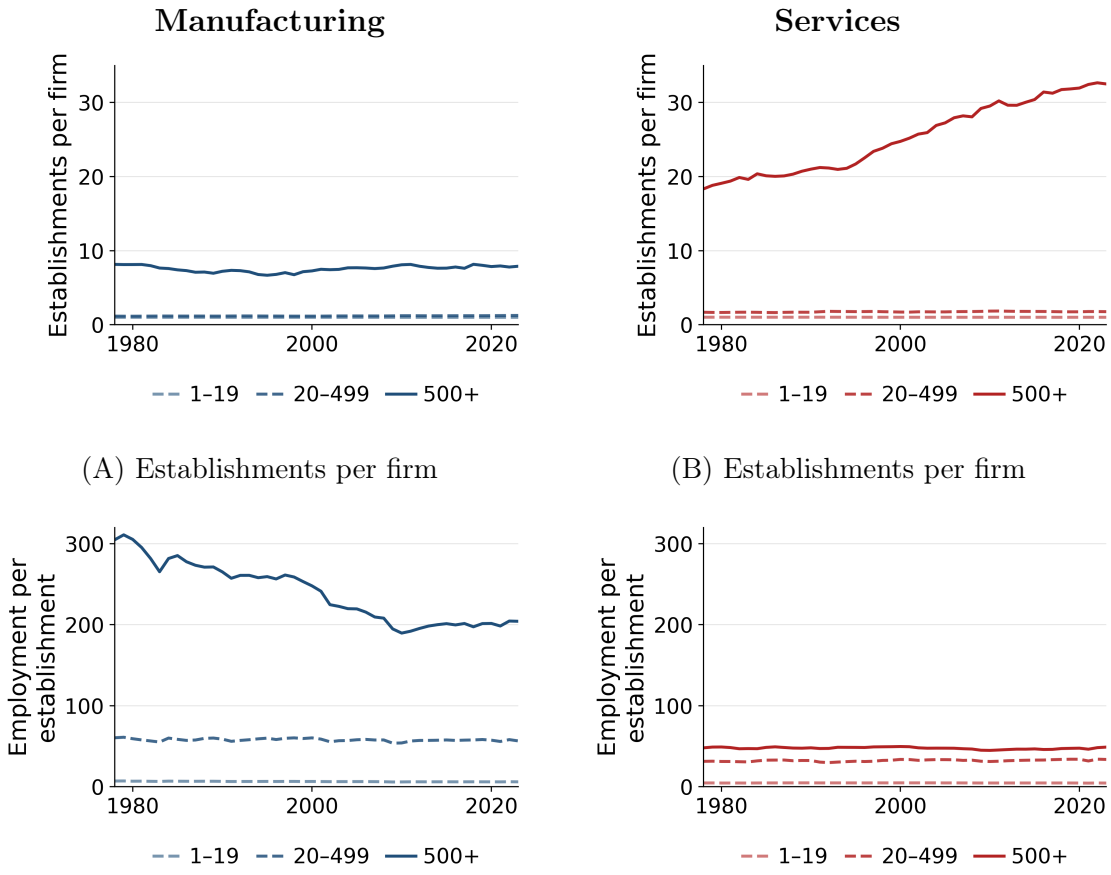
Large manufacturing firms display a different pattern, shown in the left column of Figure 4: the average number of establishments is about eight and has been stable over time, while employment per establishment declined.

The increase in the number of establishments in large service firms could reflect acquisitions of existing locations or new openings. In Appendix Figure A.6 we show that new openings minus permanent shutdowns account for almost all of the increase in establishment counts for the largest firms. This implies that acquisitions are not the main explanation for the increase, even though they might contribute to the growth of individual firms.

For each sector, Appendix Table A.2 summarizes levels of employment, firm-size distributions, and average number of establishments per firm at the beginning, middle, and end of the sample period.

[9]This increase is visible in nearly all two-digit service industries (Appendix Figure A.5).

FIGURE 4: FIRM SIZE AND ESTABLISHMENT STRUCTURE



(A) Establishments per firm

(B) Establishments per firm

(C) Employment per establishment

(D) Employment per establishment

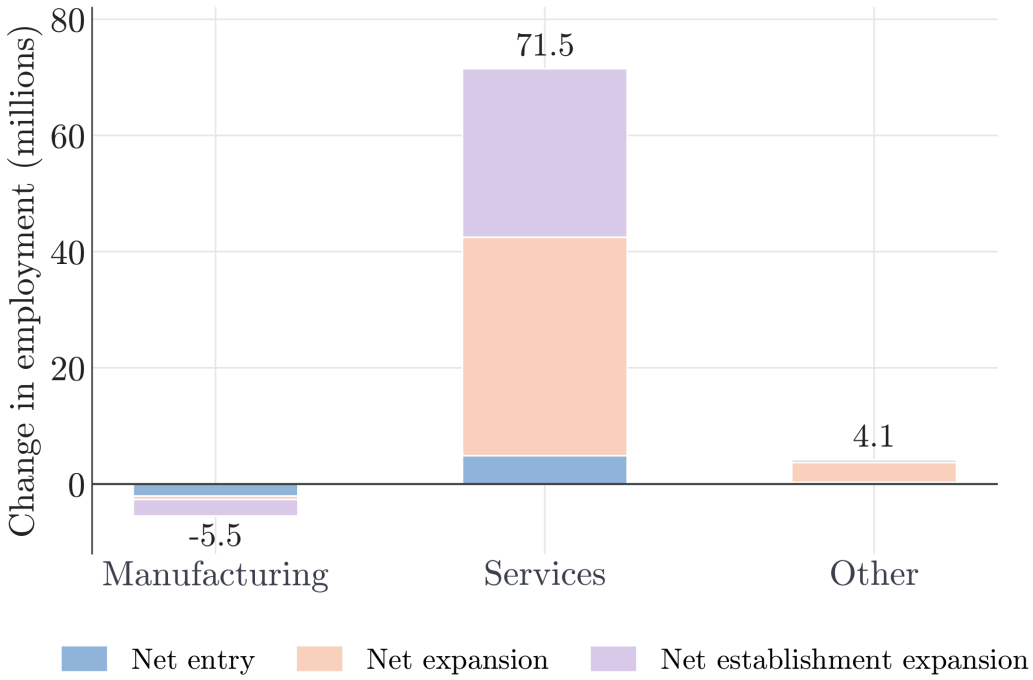
Notes: The top row reports average establishments per firm; the bottom row reports average employment per establishment. Each panel distinguishes small, medium, and large firms. Manufacturing appears on the left and services on the right. The data shows averages within firm-size groups, not in a fixed cohort of firms. Data source: BDS.

2.4 Accounting for Sectoral Employment Growth

We now decompose the sectoral employment changes into three components: net firm entry, employment changes at continuing establishments, and net establishment entry within incumbent firms. We use the following identity:

$$\Delta E_{s,t} = \underbrace{\left(B^F_{s,t} - D^F_{s,t}\right)}_{\text{Net firm entry}} + \underbrace{\left(G^C_{s,t} - L^C_{s,t}\right)}_{\substack{\text{Employment changes at} \\ \text{continuing establishments}}} + \underbrace{\left(B^I_{s,t} - D^I_{s,t}\right)}_{\substack{\text{Net establishment entry} \\ \text{within incumbent firms}}} \tag{1}$$

FIGURE 5: DECOMPOSITION OF EMPLOYMENT CHANGES BY SECTOR, 1978–2023



Notes: Each bar decomposes the change in sector employment into net firm entry, net within-establishment expansion of incumbents, and net establishment expansion within incumbent firms. Numbers next to the bars are total changes. Table A.3 in the Appendix reports the underlying numbers. Data source: BDS.

All variables measure employment in sector s between years $t-1$ and t. $B^F_{s,t}$ denotes employment created by firm births, and $D^F_{s,t}$ denotes employment lost through firm deaths. $G^C_{s,t}$ and $L^C_{s,t}$ denote employment gains and losses at continuing establishments. Finally, $B^I_{s,t}$ denotes employment created by new establishments within incumbent firms, and $D^I_{s,t}$ denotes employment lost through establishment closures within incumbent firms.[10]

Figure 5 shows the decomposition for the period 1978–2023; the underlying numbers are reported in Table A.3 in the Appendix. What stands out is that the large increase in service-sector employment was almost exclusively driven by incumbent firms: the net expansion of continuing establishments plus the net establishment creation within incumbent firms together account for over 90 percent of the increase. Net establishment creation accounts for 41 percent of the increase in service employment. It increases to over 54 percent once we account for employment growth by young establishments within mature firms.[11]

[10] We compute $D^I_{s,t}$ as total employment lost through establishment closures minus $D^F_{s,t}$ in order to exclude closures from firm deaths.

[11] Any post-birth employment growth is by definition attributed to continuing establishments. Adding the

Interestingly, the contribution of net entry to service employment growth was only about 7 percent. This is surprising given the importance of entry for job creation (Haltiwanger et al., 2013) and for the amplification and propagation of business cycles (Clementi et al., 2015; Clementi and Palazzo, 2016).[12] One qualification is that entrants contribute to employment growth well beyond their year of entry, whereas our decomposition assigns their subsequent growth to incumbents. To examine this further, Appendix Figure A.8 follows employment in the 1980, 1990, 2000, and 2010 entry cohorts. In services, cohort employment remains broadly stable as cohorts age. Thus, the expansion of successive entry cohorts does not account for the large increase in service-sector employment.

In manufacturing, net firm entry (38 percent) and establishment closures (51 percent) account for the majority of the employment decline. The cohort analysis shows that manufacturing cohort employment declines with age and that later cohorts enter with less employment.

Taken together, the empirical evidence highlights the expansion of large incumbent service firms through new establishments as an important feature of structural transformation. The next section develops a two-sector model with multi-establishment firms to explain this trend.

3 Model

Time is discrete and indexed by $t = 0, 1, 2, \ldots$. A representative household supplies $\bar{L}$ units of labor inelastically and owns all firms. Production takes place in two sectors, manufacturing (M) and services (S), indexed by $j \in \{M, S\}$. Each sector is comprised of incumbent firms and potential entrants. Incumbent firms are heterogeneous in idiosyncratic productivity z. Firms can operate multiple establishments; all establishments operated by a given firm have the same productivity z and use the same decreasing-returns technology. Expanding the number of establishments requires paying a convex replication cost.

9.9 million jobs created by young establishments within mature service firms to the establishment-creation component raises its contribution from 41 to over 54 percent. Appendix Figure A.9 shows more details.

[12] Gross employment flows associated with firm entry and exit are substantial, but largely offset each other. Appendix Figure A.7 reports the employment contributions of entering and exiting firms in panels (A) and (B), and the corresponding firm entry and exit rates in panels (C) and (D).

3.1 Household

The representative household maximizes the expected discounted sum of flow utilities:

$$U = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \ln(C_t), \tag{2}$$

where $\beta \in (0,1)$ is the discount factor and C_t is a composite of the two sectoral goods. Preferences are non-homothetic: the household must purchase a subsistence quantity $\bar{c}_M > 0$ of the manufacturing good each period. The composite C_t is:

$$C_t = \left[\mu^{\frac{1}{\varepsilon}} \left(C_{Mt} - \bar{c}_M\right)^{\frac{\varepsilon-1}{\varepsilon}} + (1-\mu)^{\frac{1}{\varepsilon}} C_{St}^{\frac{\varepsilon-1}{\varepsilon}}\right]^{\frac{\varepsilon}{\varepsilon-1}}, \tag{3}$$

where $\varepsilon > 0$ is the elasticity of substitution over supernumerary consumption and $\mu \in (0,1)$ is a taste weight. Household income is $I_t = W_t \bar{L} + \Pi_t$, where W_t is the wage and Π_t denotes aggregate firm profits rebated to the household. The budget constraint is:

$$p_{Mt} C_{Mt} + p_{St} C_{St} = I_t. \tag{4}$$

The household first purchases the subsistence quantity $\bar{c}_M$ of the manufacturing good. The remaining supernumerary expenditure is $X_t \equiv I_t - p_{Mt}\bar{c}_M$, which the household allocates across the two sectoral goods according to CES expenditure shares. The associated price index is $P_t = \left[\mu p_{Mt}^{1-\varepsilon} + (1-\mu)\, p_{St}^{1-\varepsilon}\right]^{1/(1-\varepsilon)}$. The resulting expenditure shares are $\phi_{Mt} = \mu \left(p_{Mt}/P_t\right)^{1-\varepsilon}$ and $\phi_{St} = (1-\mu)\left(p_{St}/P_t\right)^{1-\varepsilon}$, with $\phi_{Mt} + \phi_{St} = 1$. The household's demands for the two goods are thus $C_{Mt} = \bar{c}_M + \left(\phi_{Mt}/p_{Mt}\right) X_t$ and $C_{St} = \left(\phi_{St}/p_{St}\right) X_t$.

3.2 Heterogeneous Firms

Our setup extends the framework of Hopenhayn and Rogerson (1993) by allowing firms to own multiple establishments: an incumbent i in sector j with idiosyncratic productivity z_{it} operates $m_{it} \geq 1$ identical establishments. To streamline the exposition, in what follows we consider a firm in a given sector j and drop time and firm subscripts.

Establishments. Each establishment of a given firm hires ℓ workers and produces output according to $q = A_j\, z\, \ell^{\alpha_j}$, where A_j denotes sector-level TFP and $\alpha_j \in (0,1)$ governs (sector-specific) diminishing returns at the establishment level.

Firms. Firm productivity z evolves according to a sector-specific AR(1) process:

$$\log z' = (1-\rho_j)\,\mu_j^z + \rho_j\,\log z + \sigma_j\,\xi', \quad \xi' \sim \mathcal{N}(0,1), \tag{5}$$

where ξ' is i.i.d. across firms and time, μ_j^z is the unconditional mean of $\log z$, and $\sigma_j/\sqrt{1-\rho_j^2}$ is its unconditional standard deviation.

Operating m establishments requires paying the replication cost:

$$G(m) = c_{m,j}\,\frac{m^{1+\eta_j}-1}{1+\eta_j}, \tag{6}$$

where $c_{m,j}$ is a sector-specific scale parameter and $\eta_j > 0$ governs how fast the replication costs rise with the number of establishments. The cost is zero for single-unit firms ($m = 1$).

Given prices (W, P, p_j), each period a firm chooses labor per establishment ℓ and the number of establishments $m \geq 1$ to maximize current-period profit:

$$\max_{m\geq 1,\,\ell\geq 0} \quad m[p_j\,A_j\,z\,\ell^{\alpha_j} - W\,\ell] - P\,G(m) - P f_j. \tag{7}$$

Here f_j is a fixed operating cost. Following Hopenhayn and Rogerson (1993), the fixed and replication costs (as well as the entry costs, introduced below) are denominated in units of the final consumption bundle, whose price is P.

The first-order condition for labor per establishment yields:

$$\ell^*(z) = \left(\frac{p_j\,\alpha_j\,A_j\,z}{W}\right)^{\frac{1}{1-\alpha_j}}, \tag{8}$$

so that, at the optimum, operating profit per establishment is:

$$\pi_v(z) = (1-\alpha_j)\,(p_j\,A_j\,z)^{\frac{1}{1-\alpha_j}}\left(\frac{\alpha_j}{W}\right)^{\frac{\alpha_j}{1-\alpha_j}}. \tag{9}$$

Given ℓ^*, the optimal number of establishments is:

$$m^*(z) = \max\left\{1,\ \left(\frac{\pi_v(z)}{P\,c_{m,j}}\right)^{1/\eta_j}\right\}. \tag{10}$$

This policy function features an extensive-margin threshold: firms with $\pi_v(z) \leq Pc_{m,j}$ operate a single establishment, whereas firms above the threshold operate multiple establishments, with establishment scope responding to per-establishment operating profit with elasticity $1/\eta_j$. Optimal output of a firm with productivity z is therefore $Y_j(z) \equiv m_j^*(z)A_j z[\ell_j^*(z)]^{\alpha_j}$.

Let $V_j(z)$ denote the value of an incumbent firm in sector j. Each period, a firm earns current-period profits, faces the exogenous exit probability δ_j, and, conditional on survival, decides whether to continue operating. The value function satisfies:

$$V_j(z) = \Pi_j(z) + \beta\,(1-\delta_j)\max\{0,\ \mathbb{E}[V_j(z') \mid z]\}, \tag{11}$$

where current-period profit is $\Pi_j(z) = m^*(z)\,\pi_v(z) - P\,G(m^*(z)) - Pf_j$. A firm exits endogenously when the expected continuation value is negative:

$$\text{exit}_j(z) = \mathbf{1}\left\{\mathbb{E}[V_j(z') \mid z] < 0\right\}. \tag{12}$$

Entry. In each sector, a prospective entrant pays a sector-specific entry cost $c_{e,j}$ before learning its idiosyncratic productivity draw. Upon entry, the firm draws productivity from a sector-specific entrant distribution $F_{e,j}$, which is log-normal:

$$\log z \sim \mathcal{N}\left(\mu_{e,j}^z,\ \sigma_{e,j}^2\right). \tag{13}$$

We assume that entrants draw productivity from (13) rather than from the invariant distribution of the idiosyncratic productivity process (5) because entrants are much smaller than incumbents in the data. The parameters $\left(\mu_{e,j}^z, \sigma_{e,j}\right)$ allow the model to reproduce the size

distribution at birth. Free entry requires that the expected value of entry equal its cost:

$$\int V_j(z)\,\mathrm{d}F_{e,j}(z) = Pc_{e,j}. \tag{14}$$

3.3 Stationary Recursive Competitive Equilibrium

The wage is the numeraire, $W = 1$. The stationary recursive competitive equilibrium consists, for each sector $j \in \{M, S\}$, of a value function V_j, decision rules $\ell_j^*(z)$, $m_j^*(z)$, and $\text{exit}_j(z)$, a measure of incumbent firms μ_j and a mass of entrants M_j, an output price p_j, household consumption C_j, and a final-good expenditure share ϕ_j, together with the final-good price index P, aggregate dividends Π, and total final-good demand from firms' cost payments B. These objects satisfy:

1. given prices and income, household consumption $\{C_M, C_S\}$ maximizes the consumption composite in (3) subject to the budget constraint in (4);

2. given prices, in each sector j labor per establishment $\ell_j^*(z)$ is given by (8), the number of establishments $m_j^*(z)$ by (10), the value function $V_j(z)$ by (11), and the exit rule $\text{exit}_j(z)$ by (12);

3. the free entry condition (14) holds in each sector;

4. the final-good price index and expenditure shares are

$$P = \left[\mu p_M^{1-\varepsilon} + (1-\mu)p_S^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}, \qquad \phi_M = \mu\left(\frac{p_M}{P}\right)^{1-\varepsilon}, \qquad \phi_S = (1-\mu)\left(\frac{p_S}{P}\right)^{1-\varepsilon};$$

5. in each sector, the measure of incumbents is stationary:

$$\mathcal{T}_j(\mu_j, M_j) = \mu_j, \qquad j \in \{M, S\},$$

where $\mathcal{T}_j$ is the law of motion implied by entry, the productivity process, and firms' exit decisions, defined explicitly in (B.1).

6. total final-good demand from firms' cost payments is[13]

$$B = \sum_{j \in \{M,S\}} \left[\int \left[f_j + G(m_j^*(z)) \right] \mathrm{d}\mu_j(z) + M_j c_{e,j} \right];$$

7. goods markets clear, where output in each sector is used for household consumption and firms' purchases of the final consumption bundle:

$$\int Y_j(z) \, \mathrm{d}\mu_j(z) = C_j + \phi_j \frac{P}{p_j} B, \qquad j \in \{M, S\};$$

8. the labor market clears:

$$\sum_{j \in \{M,S\}} \int \ell_j^*(z) m_j^*(z) \, \mathrm{d}\mu_j(z) = \bar{L};$$

9. aggregate dividends equal net payouts from both sectors:

$$\Pi = \sum_{j \in \{M,S\}} \left(\int \Pi_j(z) \, \mathrm{d}\mu_j(z) - P \, M_j c_{e,j} \right).$$

3.4 Growth Through Establishments

A central question of this paper is why the employment reallocation from manufacturing to services is concentrated among large, multi-establishment service firms. This section contrasts two forces. The first is faster productivity growth in manufacturing. This traditional driver of structural transformation changes relative prices and reallocates activity across sectors, but does not reshape the within-sector firm-size distribution. The second, a decline in the convexity of service-sector replication costs, by contrast, changes how firms expand and has its largest effects on high-productivity firms. The following propositions formalize this distinction.

To develop intuition, we consider a tractable variant of the baseline model. We make two

[13] We assume that firms pay fixed operating, entry, and replication costs by purchasing a CES bundle of manufacturing goods and services, with the same weights and elasticity of substitution as household consumption above subsistence.

changes which apply only in this section; the quantitative model retains the baseline assumptions stated above. First, we abstract from non-homothetic demand by setting subsistence consumption $\bar{c}_M = 0$.[14] Second, we denominate fixed operating, replication, and entry costs in labor units rather than in units of the final consumption bundle. A firm's static problem becomes

$$\max_{m \geq 1,\, \ell \geq 0} \quad m[p_j\, A_j\, z\, \ell^{\alpha_j} - W\, \ell] - W\, G(m) - W f_j. \tag{15}$$

The optimal labor choice and operating profit per establishment remain as in (8) and (9), while the optimal number of establishments is

$$m^*(z) = \max\left\{1,\ \left(\frac{\pi_v(z)}{W\, c_{m,j}}\right)^{1/\eta_j}\right\}. \tag{16}$$

The incumbent Bellman equation continues to be given by (11), with current-period profit $\Pi_j(z) = m^*(z)\pi_v(z) - WG(m^*(z)) - Wf_j$. The free-entry condition becomes

$$\int V_j(z)\, \mathrm{d}F_{e,j}(z) = W c_{e,j}. \tag{17}$$

Proposition 1 (Invariance of firm policies to sectoral productivity)**.** *For each sector $j \in \{M, S\}$, define $x_j \equiv p_j A_j$. Then:*

(i) the policies $\ell_j^*(z)$ and $m_j^*(z)$, operating profit per establishment $\pi_{v,j}(z)$, current-period profit $\Pi_j(z)$, the value function $V_j(z)$, and the exit rule depend on (p_j, A_j) only through x_j;

(ii) holding all other sectoral primitives fixed, the free-entry condition uniquely determines x_j. Consequently, if $A_j' = \lambda A_j$ for some $\lambda > 0$, then $p_j' = p_j/\lambda$;

(iii) the stationary firm distribution is invariant to A_j.

Proof. Using $x_j = p_j A_j$, Equations (8) and (9) become

$$\ell_j^*(z) = (\alpha_j x_j z)^{1/(1-\alpha_j)}, \qquad \pi_{v,j}(z) = (1-\alpha_j)(x_j z)^{1/(1-\alpha_j)} \alpha_j^{\alpha_j/(1-\alpha_j)}.$$

[14] The firm-level comparative statics derived below continue to hold under the baseline non-homothetic preferences. Setting $\bar{c}_M = 0$ simplifies the analytical characterization of sectoral demand.

Thus, $\ell_j^*(z)$ and $\pi_{v,j}(z)$ depend on (p_j, A_j) only through x_j. Equation (16) then implies the same for $m_j^*(z)$ and for current-period profit $\Pi_j(z) = m_j^*(z)\pi_{v,j}(z) - G(m_j^*(z)) - f_j$. It follows that the Bellman operator, its unique fixed point $V_j(z; x_j)$, and the associated exit rule also depend on (p_j, A_j) only through x_j. Define the value of entry as $\mathcal{E}_j(x_j) \equiv \int V_j(z; x_j)\, \mathrm{d}F_{e,j}(z)$. Since $\mathcal{E}_j(\cdot)$ is strictly increasing, the free-entry condition $\mathcal{E}_j(x_j) = c_{e,j}$ uniquely determines x_j. Thus, if $A_j' = \lambda A_j$, equilibrium requires $p_j' = p_j/\lambda$, so that $p_j' A_j' = p_j A_j$.

All firm policies and the exit rule are therefore unchanged. Since the productivity process is independent of sectoral productivity A_j, the stationary measure per unit mass of entrants, and hence the firm-size distribution, is unchanged. □

Proposition 1 highlights the limitation of the productivity channel. Free entry pins down the composite $p_j A_j$: when sectoral productivity rises, the sector's output price falls proportionally. Firms therefore choose the same labor per establishment and the same number of establishments as before. Productivity growth can still change relative prices and reallocate employment across sectors, but it cannot by itself generate changes in within-sector employment concentration or establishments per firm. This is why the traditional productivity channel cannot account for the distributional facts documented in Section 2.

The next proposition shows that a reduction in η_j, the convexity of replication costs, changes establishment choices and acts like an increase in effective firm-level returns to scale, conditional on prices.

Proposition 2 (Replication costs and effective firm-level returns to scale)**.** *For each sector $j \in \{M, S\}$, conditional on prices, define*

$$\mathcal{I}_j \equiv \{z : \pi_{v,j}(z) > c_{m,j}\}.$$

Then:

(i) for $z \notin \mathcal{I}_j$, $m_j^*(z) = 1$ and is independent of η_j. For $z \in \mathcal{I}_j$,

$$\frac{\partial \log m_j^*(z)}{\partial \eta_j} < 0, \qquad \frac{\partial^2 \log m_j^*(z)}{\partial \eta_j\, \partial \log z} < 0.$$

Thus, a reduction in η_j raises the number of establishments, with a larger proportional effect at higher productivity;

(ii) for $z \in \mathcal{I}_j$, firm employment satisfies

$$L_j(z) \equiv m_j^*(z)\ell_j^*(z) \propto z^{\frac{1+\eta_j}{\eta_j(1-\alpha_j)}} = z^{\frac{1}{1-\alpha_{\text{eff},j}}},$$

where

$$\alpha_{\text{eff},j} \equiv \frac{1+\eta_j\alpha_j}{1+\eta_j} \in (\alpha_j, 1).$$

Moreover, $\alpha_{\text{eff},j}$ is decreasing in η_j, converges to α_j as $\eta_j \to \infty$, and converges to one as $\eta_j \to 0$.

Proof. For $z \in \mathcal{I}_j$, Equation (16) implies

$$\log m_j^*(z) = \frac{1}{\eta_j}\log\left(\frac{\pi_{v,j}(z)}{c_{m,j}}\right), \qquad \frac{\partial \log m_j^*(z)}{\partial \eta_j} = -\frac{1}{\eta_j^2}\log\left(\frac{\pi_{v,j}(z)}{c_{m,j}}\right) < 0.$$

Since $\pi_{v,j}(z) \propto z^{1/(1-\alpha_j)}$,

$$\frac{\partial^2 \log m_j^*(z)}{\partial \eta_j \, \partial \log z} = -\frac{1}{\eta_j^2(1-\alpha_j)} < 0.$$

For $z \notin \mathcal{I}_j$, the corner solution $m_j^*(z) = 1$ establishes the remaining claim in part (i).

For firms in the interior region,

$$\ell_j^*(z) \propto z^{1/(1-\alpha_j)}, \qquad m_j^*(z) \propto z^{1/[\eta_j(1-\alpha_j)]}.$$

Therefore,

$$L_j(z) \propto z^{\frac{1+\eta_j}{\eta_j(1-\alpha_j)}}, \qquad \alpha_{\text{eff},j} = \frac{1+\eta_j\alpha_j}{1+\eta_j}.$$

Finally, $\partial\alpha_{\text{eff},j}/\partial\eta_j = (\alpha_j - 1)/(1+\eta_j)^2 < 0$, and the stated bounds and limits follow directly. □

Proposition 2 describes the direct firm-level mechanism. A lower η_j makes replication costs rise less rapidly with the number of establishments. Firms below the replication threshold remain at a single establishment, while firms above it expand along the establishment

margin. For these firms, the establishment policy implies $\partial \log m_j^*(z)/\partial \log \eta_j = -\log m_j^*(z)$, which says that the same proportional decline in replication-cost convexity induces a larger proportional expansion at firms that initially operate more establishments. Because the response increases with firm productivity, employment shifts toward large firms. Equivalently, a lower η_j raises effective firm-level returns to scale and makes firm employment more responsive to productivity. The increase in effective firm-level returns to scale is larger when establishment-level returns α_j are lower, at a given η_j.

In general equilibrium, a decline in η_j raises the value of entry, so free entry requires p_j to fall. The resulting decline in operating profit per establishment partly offsets the direct increase in the optimal number of establishments, $m_j^*(z)$. At the same time, the change in relative prices reallocates household expenditure across sectors and changes the equilibrium masses of entrants. Section 4.2 quantifies the resulting equilibrium changes in establishment expansion and the sectoral allocation of employment.

Interpretation of replication costs. The parameter η governs the convexity of replication costs and we interpret it as a parsimonious way of capturing technological and organizational changes that facilitated firm expansion. The previous results showed that a decline in η generates a disproportionate expansion of more productive firms. This could be generated by several mechanisms.

One such mechanism is firm-level technology adoption. The adoption cost is paid once, but the benefits extend to every establishment, which makes expansion more attractive to multi-establishment firms. In Hsieh and Rossi-Hansberg (2023), adopting an ICT-type technology lowers production costs across local markets. More productive firms adopt more intensively because they can use the technology in more markets. This makes entry into additional, less profitable markets viable. In Brynjolfsson et al. (2025), IT lets firms transfer the routines of their most efficient branch to other branches. The return to this investment rises with the number of branches, and applying these routines in a new market raises the profit from entering it. Argente et al. (2025) study a related mechanism along the product margin. Standardization shifts expertise toward a component that can be used across products, which reduces the strain that a broader product scope places on a firm's

limited managerial capacity.

Although these models describe different mechanisms, all of them generate a stronger extensive-margin expansion response among more productive and profitable firms. In Appendix B.6 we show for each of these models how the change in the relationship between firm productivity and the number of markets, branches, or products maps into a decline in replication-cost convexity, η, in our model.

The propositions in this section showed how relative productivity growth and lower replication-cost convexity differ in their effects on firm expansion. In the next section, we assess to what extent these forces can explain the shift toward services and the concentration of service employment in large, multi-establishment firms, taking adjustments in prices, demand, and entry into account.

4 Quantitative Results

This section describes our quantitative results. We first solve the initial steady state of the model, which we calibrate to BDS moments from 1979. We then present our main experiment, which combines a relative increase in manufacturing productivity with a proportional decline in the replication-cost parameters η_M and η_S. The experiment reproduces the shift toward large, multi-establishment service firms and accounts for 52 percent of the aggregate employment reallocation. We then provide a decomposition of the results, study counterfactuals, and show robustness to a variety of extensions. Appendix B.1 describes the numerical solution.

4.1 Parameterization

We partition the parameter set into two groups. Those in the first group are assigned externally, while parameters in the second group are calibrated to match empirical moments. Table 1 summarizes all parameter values and, for calibrated parameters, the corresponding empirical targets.

TABLE 1: MODEL PARAMETERS

Parameter	Description	Value	Target	Data	Model
		Parameters assigned outside the model			
β	Discount factor	0.96	Model period represents a year		
ε	Substitution elasticity	0.40	Desmet and Rossi-Hansberg (2014)		
A_M, A_S	Sectoral TFP	1.00	Normalization		
δ_M	Exogenous exit rate	0.08	BDS average firm entry & exit rate		
δ_S	Exogenous exit rate	0.11	BDS average firm entry & exit rate		
p_S/p_M	Relative price of services	0.64	BEA value-added price indexes		
		Household			
$\bar{c}_M$	Manufacturing subsistence	0.32	Goods-share slope wrt log PCE p.c.	−0.23	−0.23
μ	CES preference weight	0.09	Manufacturing employment share	0.28	0.28
		Manufacturing			
μ_M^z	Mean log productivity	1.81	Average firm size	73.21	74.16
σ_M	Productivity volatility	0.44	Employment share, firms with 500+	0.65	0.67
$\mu_{e,M}^z$	Mean log entrant productivity	0.27	Average entrant size	9.04	9.03
$\sigma_{e,M}$	Entrant productivity dispersion	0.88	Entrant firm share with 20+	0.03	0.03
ρ_M	Productivity persistence	0.69	Age 1–5 firm share with 20+	0.15	0.15
α_M	Establishment returns to scale	0.53	Average establishment size	61.72	60.21
$c_{m,M}$	Replication-cost level	52.87	Average establishments per firm	1.19	1.23
η_M	Replication-cost convexity	0.87	Estab./firm: 500+ relative to <20	8.10	8.00
$c_{e,M}$	Entry cost	386.56	Free-entry condition	—	—
		Services			
μ_S^z	Mean log productivity	0.14	Average firm size	16.43	16.43
σ_S	Productivity volatility	0.18	Employment share, firms with 500+	0.42	0.42
$\mu_{e,S}^z$	Mean log entrant productivity	1.95	Average entrant size	4.90	4.90
$\sigma_{e,S}$	Entrant productivity dispersion	0.24	Entrant firm share with 20+	0.01	0.01
ρ_S	Productivity persistence	0.99	Age 1–5 firm share with 20+	0.06	0.06
α_S	Establishment returns to scale	0.47	Average establishment size	13.46	13.46
$c_{m,S}$	Replication-cost level	14.12	Average establishments per firm	1.22	1.22
η_S	Replication-cost convexity	0.85	Estab./firm: 500+ relative to <20	18.44	18.44
$c_{e,S}$	Entry cost	94.86	Free-entry condition	—	—

Notes: The first group of parameters is assigned outside the model. The remaining parameters are calibrated to match empirical targets. Subscripts M and S denote manufacturing and services. Where reported, the Data and Model columns show the empirical target and its model counterpart. BDS firm-level targets are measured in 1979, except for the share of firms aged 1–5 with 20 or more employees, which is measured in 1982.

Fixed parameters. The model period is one year. We set the discount factor $\beta = 0.96$. The elasticity of substitution between manufacturing and services is set to $\varepsilon = 0.40$, consistent with the structural transformation literature (e.g., Desmet and Rossi-Hansberg, 2014). The sectoral TFP levels $A_M = A_S = 1$ are normalized to one in the initial steady state. In the baseline model, we set the fixed operating cost $f_j = 0$, so firms exit only at the exoge-

nous rate δ_j, which we calibrate to the 1979 BDS data.[15] We set the initial relative price of services, $p_S/p_M = 0.64$, equal to the ratio of the BEA value-added price indexes for private service-producing industries and manufacturing.

Calibrated parameters. We proceed in two steps. First, we discipline the two remaining preference parameters. The manufacturing subsistence term $\bar{c}_M$ governs how strongly the household's expenditure share on goods declines as expenditure rises. The CES preference weight μ determines the allocation of spending across manufacturing and services. We pin down the subsistence term using U.S. Personal Consumption Expenditures data.[16] We choose μ to match the manufacturing employment share in 1979.

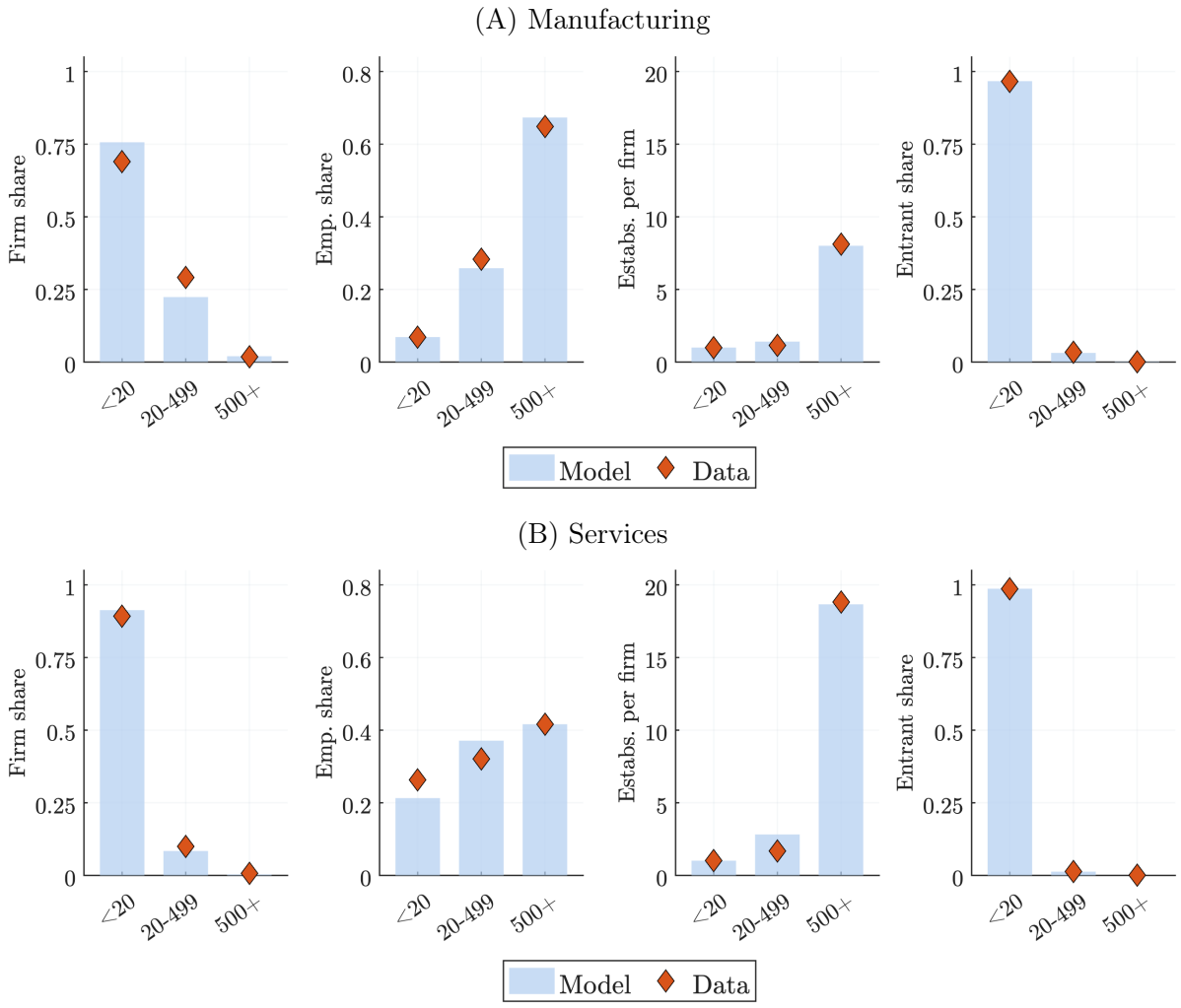
Second, we estimate the firm side parameters for manufacturing and services by minimizing the distance between model moments and the corresponding sector-specific BDS targets. Although each targeted firm-level moment is jointly affected by all parameters, we provide a heuristic identification argument that motivates the choice of empirical targets. The mean of log idiosyncratic productivity, μ_j^z, shifts the level of firm employment and is therefore pinned down by average firm size. The dispersion of the invariant productivity distribution implied by the AR(1) process is given by $\sigma_j/\sqrt{1-\rho_j^2}$ and it governs the thickness of the right tail of the firm-size distribution. It is disciplined by the employment share of firms with 500 or more employees. The entrant-distribution parameters $(\mu_{e,j}^z, \sigma_{e,j})$ pin down the size distribution at birth, targeted by the average entrant size and the share of entrants with 20 or more employees. The persistence ρ_j governs how quickly entering cohorts converge toward the incumbent distribution; it is identified by the share of firms aged one to five with 20 or more employees.[17] The returns-to-scale parameter α_j is identified by the average establishment size. The replication-cost level $c_{m,j}$ sets the productivity threshold at which operating additional establishments becomes profitable and is therefore disciplined by the average number

[15] Section 4.4 shows that allowing for endogenous exit delivers largely the same results.

[16] Specifically, we estimate a reduced-form time-series relationship between the goods expenditure share, real expenditure per capita, and the relative price of goods to services using annual data from 1970 to 2025. The coefficient of interest is the slope of the goods expenditure share with respect to log real expenditure per capita, holding relative prices fixed. We choose the subsistence term so that the model reproduces the estimated slope of -0.227 in the initial steady state. Appendix B.2 describes the data construction and reports alternative samples and specifications. Section 4.4 shows that recalibrating the model using these alternative slope estimates leaves the main quantitative conclusions largely unchanged.

[17] We measure this moment in 1982, the first year in which the BDS fully tracks the age 1–5 group.

FIGURE 6: INITIAL STEADY-STATE FIT (1979)

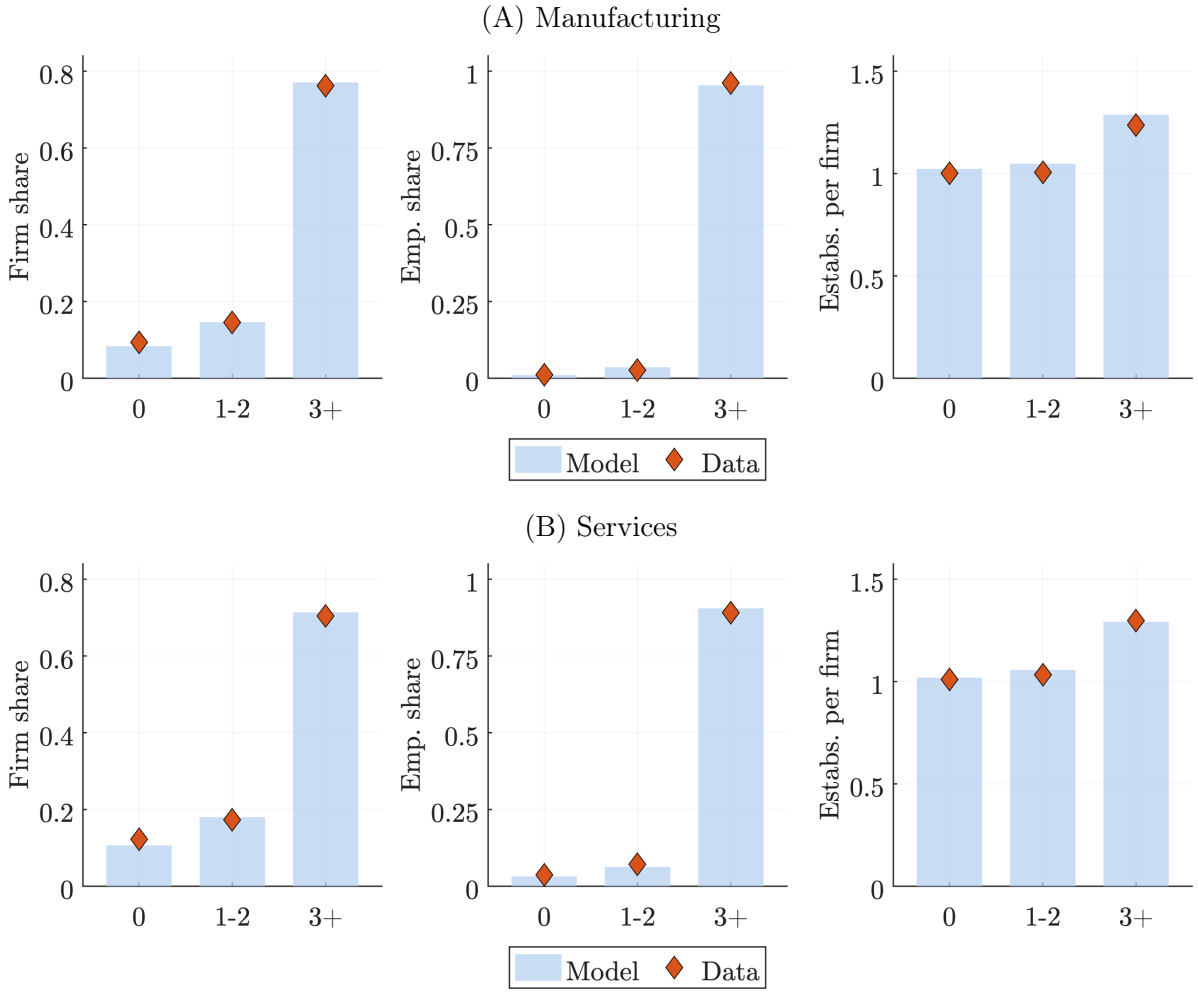


Notes: Within each sector, the four columns show firm share, employment share, establishments per firm, and entrant firm share by size bin, from left to right. Diamonds denote the data and bars denote the model. In each sector, the employment share of firms with 500+ employees and the share of entrants with < 20 employees are targeted. The ratio of establishments per firm in large relative to small firms is also targeted.

of establishments per firm, while the convexity η_j controls how steeply the establishment count rises with productivity and is identified by the average number of establishments per firm among large (500+) relative to small (<20) firms.

Model fit. Figure 6 compares the model's steady state to the 1979 data across three firm-size bins, for manufacturing (top panel) and services (bottom panel). The bars show the model moments, while the red diamonds show the data. Of the 24 data points shown in the figure, only four are direct calibration targets: the employment shares of firms with 500 or

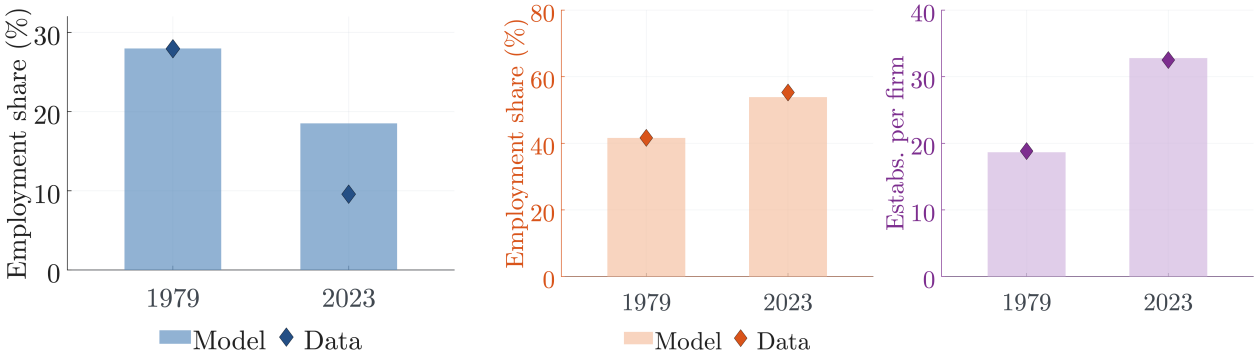
FIGURE 7: INITIAL STEADY-STATE: AGE DISTRIBUTIONS



Notes: The columns show the share of firms, the employment share, and the average number of establishments per firm. Bars denote the model and diamonds denote the data. Because the BDS firm-age panel starts in 1977, exact ages in 1979 are observed only for age 0 and ages 1–2; firms aged three and above are pooled into a single left-censored group both in the model and the data. None of the age moments is directly targeted. A related calibration target is the share of firms aged 1–5 with 20 or more employees, measured in 1982.

more employees and the shares of entrants with fewer than 20 employees. We therefore see it as a success of the model that it closely tracks the main features of the 1979 firm-size profiles in both sectors, including the high share of small service firms, the high employment share of large manufacturing firms, the steep increase in establishments per firm with firm size, and the prevalence of small entrants. The good fit between the model and the data implies that the model provides a credible 1979 starting point for the quantitative experiment that follows.

FIGURE 8: AGGREGATE REALLOCATION AND LARGE SERVICE FIRMS



(A) Manufacturing share

(B) Large service firms

Notes: Panel (A) reports the manufacturing employment share. Panel (B) reports two key service-sector distributional moments: the employment share of firms with 500 or more employees (left chart) and the average number of establishments per firm among those firms (right chart). Bars denote the model and diamonds denote the data.

Relatedly, Figure 7 reports the model's fit to the 1979 firm-age distribution. Because the BDS starts in 1977, in 1979 exact ages are only observed for age 0 and ages 1–2; all older firms are left-censored. We impose the same censoring on the model by pooling firms aged three and above into a single left-censored group. The model reproduces the 1979 age profiles very closely in both sectors. None of the age-bin moments plotted in Figure 7 was targeted: the only age-related calibration moment used is the share of firms aged one to five with 20 or more employees. The good fit therefore provides a strong independent validation of the model's firm-dynamics mechanism.

4.2 Main Experiment: Structural Transformation and the Expansion of Large Service Firms

The empirical evidence in Section 2 shows that structural transformation combined a reallocation of employment toward services with the expansion of large incumbent service firms through additional establishments. Our main experiment asks whether the model can account for both patterns simultaneously. To do so, we combine two forces: an increase in relative manufacturing productivity and a decline in the convexity of establishment-replication costs.

All experiments take the calibrated 1979 steady state as their starting point and compare

it with a counterfactual 2023 steady state. In the main experiment, we choose the increase in manufacturing productivity to match the change in the relative price of services and impose a common proportional decline in η_M and η_S to match the change in establishments per firm among large relative to small service firms.[18] All other primitives remain fixed.

Figure 8 summarizes the main experiment along the two dimensions emphasized in the empirical analysis. Panel (A) reports the manufacturing share of total employment, while panel (B) reports the employment share and average number of establishments per firm among large service firms. The diamonds represent the data, while the bars show the results of the model. The 2023 manufacturing employment share and the employment share of large service firms are not targeted.

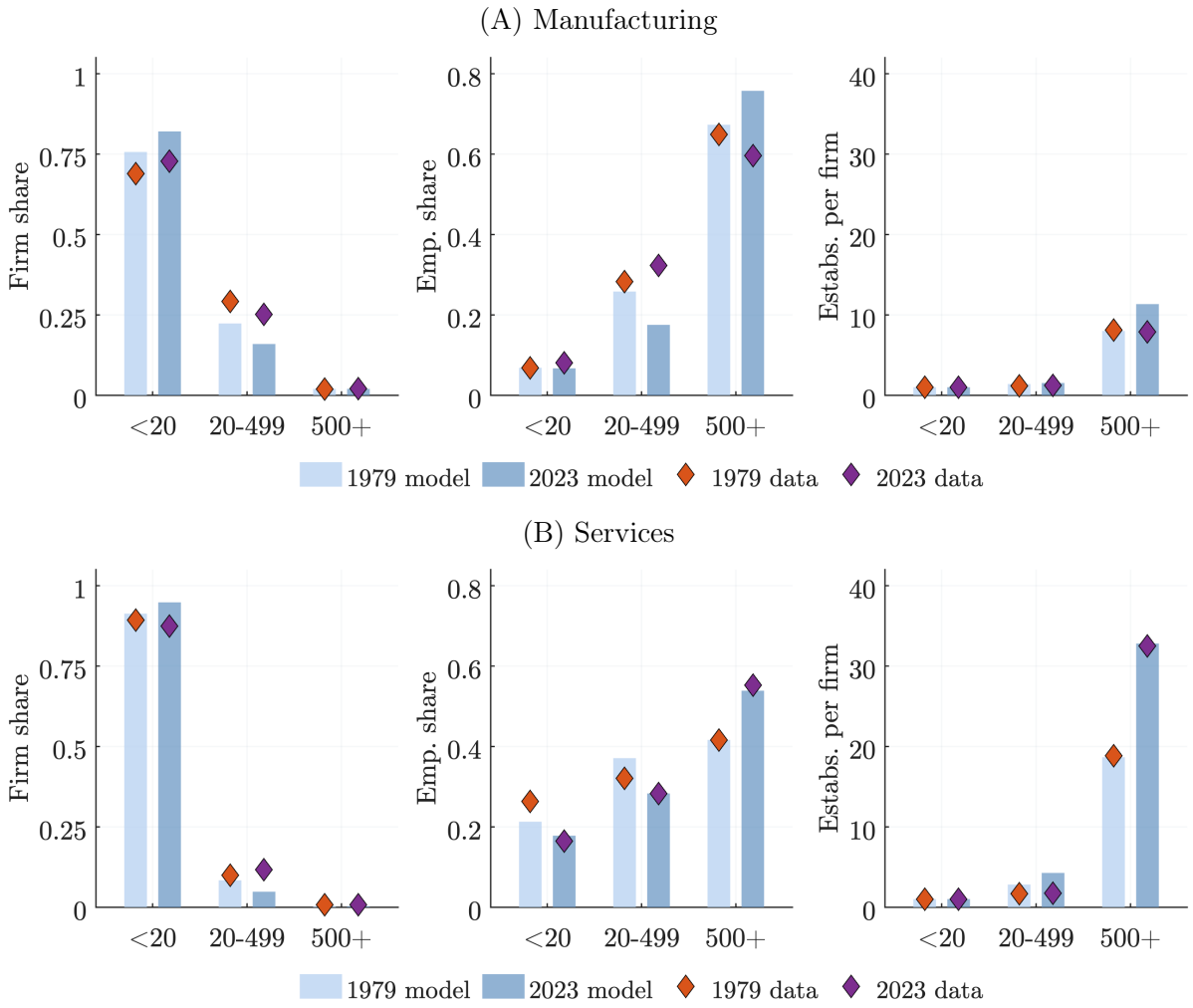
Panel (A) shows that the manufacturing employment share falls from 28.0 percent to 18.5 percent in the model. This implies that the model can account for 51.5 percent of the observed decline. Part of this reallocation reflects higher manufacturing productivity, which shifts expenditure toward services through two demand channels. Cheaper manufacturing goods shift expenditure toward services because the two goods are complements and because the manufacturing subsistence requirement becomes less costly to satisfy. The resulting shift in demand reallocates labor toward services.

Panel (B) of Figure 8 shows that this aggregate employment reallocation is accompanied by the expansion of large, multi-establishment service firms. Consistent with Proposition 2, lower η_j raises effective firm-level returns to scale and disproportionately increases establishment creation by high-productivity firms. Establishments per large service firm rise from 18.7 to 32.8 (right chart), closely matching the 2023 data counterpart. This fit partly reflects the fact that establishments per firm in large relative to small service firms is a calibration target. The employment share of these firms, which is not targeted, rises from 41.6 to 53.9 percent (left chart), close to the observed 55.3 percent.

Figure 9 extends the analysis by reporting the complete firm-size profiles in both sectors. None of the 2023 firm-share or employment-share bins is targeted; the only target is the ratio

[18] We impose the same proportional decline in η in both sectors. This restriction reflects our interpretation that information technology, standardization, and management tools reduced replication costs broadly across sectors; the model must therefore generate the disproportionate expansion of large service firms as an equilibrium outcome.

FIGURE 9: FIRMS, EMPLOYMENT, AND ESTABLISHMENTS BY FIRM SIZE



Notes: Within each sector, the columns show firm share, employment share, and establishments per firm by size bin, from left to right. Bars denote the model and diamonds denote the data. None of the 2023 firm-share or employment-share bins are targeted. We target the ratio of establishments per firm among large relative to small service firms.

of establishments per firm among large relative to small service firms. The model nevertheless captures the reallocation of service employment toward large firms and the pronounced expansion of large service firms along the establishment margin. It also reproduces the much smaller change in establishments per firm in manufacturing, although the model does generate a modest increase where the data show almost no change.

This sectoral asymmetry arises even though we reduce η by the same proportion in both sectors. In the initial steady state, manufacturing firms face higher returns within establishments but a substantially higher level of replication costs. Manufacturing firms can therefore

attain more of their scale within existing establishments, while the higher replication threshold limits expansion through additional establishments. Consistent with Proposition 2, lower establishment-level returns in services strengthen the effect of a decline in η on effective firm-level returns to scale. Moreover, conditional on prices, the same proportional decline in η generates a larger proportional increase in establishments at firms that initially operate more of them. These mechanisms help explain why the quantitative model generates a stronger establishment expansion in services without imposing a sector-specific decline in η.

4.3 Disentangling the Drivers of Structural Transformation

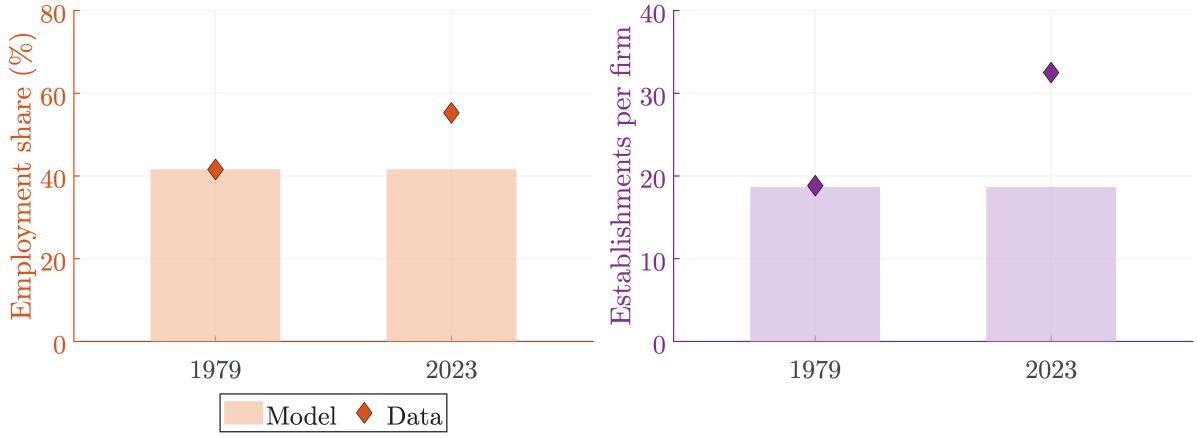
In order to assess the distinct roles of the two forces in the main experiment, we now consider each force in isolation.

Manufacturing productivity only. We begin by isolating the increase in manufacturing productivity. As before, we increase A_M to match the observed change in the relative price of services between 1979 and 2023. Appendix Table B.3 reports the parameter changes and provides the comparison of model and data moments.

With replication costs held fixed, the increase in A_M leads to a decline in the relative price of manufactured goods because of free entry. This price adjustment reallocates employment toward services through the relative-price and subsistence channels described above. The manufacturing employment share falls from 28.0 percent to 19.9 percent, accounting for 44.0 percent of the observed decline.

However, as illustrated in Proposition 1 and shown in Figure 10, the price adjustment that accompanies the change in A_M leaves firm policies and within-sector firm-size distributions essentially unchanged. Productivity growth reallocates employment toward services, but without generating the expansion of large, multi-establishment service firms. Figure 10 plots the employment share of large service firms and the average number of establishments among large service firms (the same two charts as in panel (B) of Figure 8). Both model outcomes are virtually unchanged. Thus, the productivity channel alone misses the firm dynamics margin through which structural transformation takes place, namely the expansion of large service firms through additional establishments.

FIGURE 10: MANUFACTURING PRODUCTIVITY ONLY: LARGE SERVICE FIRMS



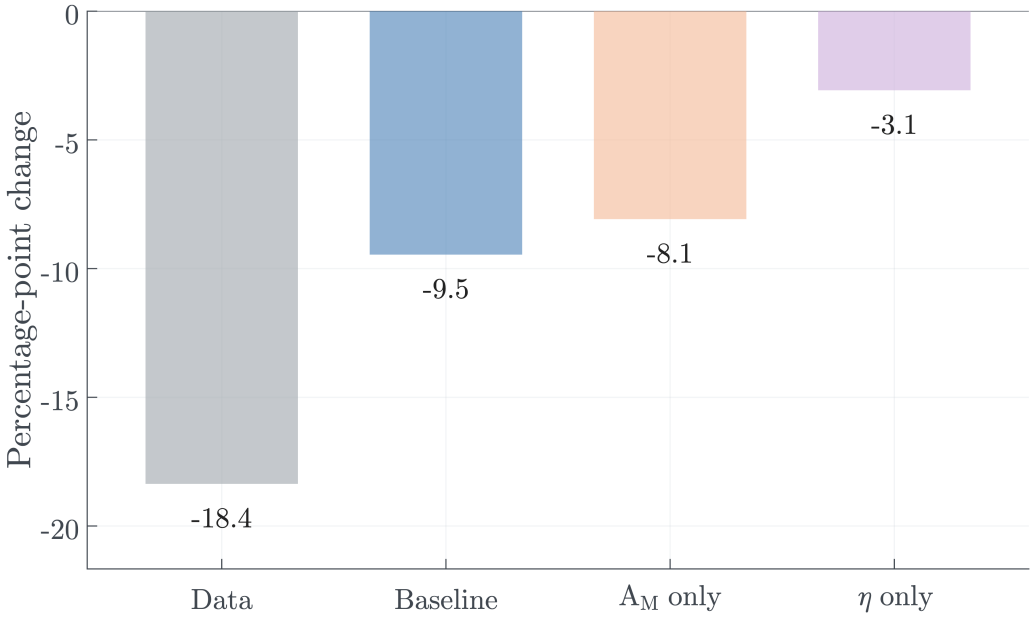
Notes: Bars denote the model and diamonds denote the data.

The reason is that the manufacturing-productivity channel does not generate differential growth across service firms. With η fixed, the equilibrium adjustment leaves establishment choices essentially unchanged, while the changes in sector-level prices scale employment per establishment by the same proportion for a given productivity level. Firm employment therefore shifts in parallel across the productivity distribution. A decline in η, by contrast, changes the slope of the establishment policy and induces high-productivity firms to expand disproportionately. It is this differential response that shifts service employment toward large firms.

Replication costs only. We now isolate this second force by holding manufacturing productivity fixed and allowing only for a proportional change in η_M and η_S. As before, we target the increase in the average number of establishments per firm among large (500+) relative to small (<20) service firms in 2023, which the model matches (*cf.* Table B.3).

A decline in η alone accounts for 16.7 percent of the observed drop in the manufacturing employment share. This is roughly one third of the decline explained by the main experiment. Lower replication costs therefore matter not only for matching the firm-level establishment moments but also for the aggregate reallocation of employment. Figure 11 compares the decline in the manufacturing employment share across the three experiments.

FIGURE 11: MANUFACTURING EMPLOYMENT-SHARE DECLINE ACROSS EXPERIMENTS



Notes: The figure reports the percentage-point decline in the manufacturing employment share between 1979 and 2023 in the data, the main experiment, the manufacturing-productivity-only counterfactual (A_M only), and the replication-cost-only counterfactual (η only). The contributions of A_M and η are not additive because each experiment induces its own equilibrium price adjustment.

4.4 Extensions

We now consider a number of extensions to the baseline model to assess the robustness of the quantitative results. First, we incorporate population and employment growth. Next, we allow for endogenous exit. Third, we relax the assumption of a perfectly elastic supply of entrants by allowing the cost per entrant to increase with the number of entering firms. Finally, we assess the role of Stone-Geary preferences by considering a version of the model with non-homothetic CES preferences.

Labor-force growth. The quantitative experiments considered above held aggregate labor supply fixed. In the data, however, total employment grew substantially over this period. Recent work shows that changes in aggregate employment can be an important source of service-sector growth (Duarte and Restuccia, 2026). In the model, a larger labor force affects sectoral reallocation through the non-homothetic demand channel: higher aggregate income shifts expenditure toward services. Free entry, however, leaves equilibrium prices and firm-level policies unchanged relative to the baseline, so the additional labor is absorbed through

changes in the masses of firms across sectors. As a result, labor-force growth changes sectoral employment shares but leaves within-sector employment shares and establishments per firm unchanged.

We show the full results of the main experiments with a larger labor force in Table B.5. In a first version of this extension, we simply scale aggregate labor supply by 1.88, the growth in BDS total employment between 1979 and 2023. With a larger labor force, the manufacturing employment share falls from 28.0 percent to 14.8 percent. The experiment therefore accounts for 71.8 percent of the observed manufacturing employment decline, compared with 51.5 percent in the main experiment. This calculation holds aggregate subsistence demand fixed. Because population growth also increases subsistence needs, we consider a second alternative that scales labor supply by the growth in employment per person. The population increased by a factor of 1.50 between 1979 and 2023, implying that employment per person increased by a factor of 1.25. Scaling labor supply by this population-adjusted factor, the manufacturing employment share falls to 16.9 percent, accounting for 60.3 percent of the observed decline.

In summary, allowing for labor-force growth improves the fit of the manufacturing employment decline but leaves the employment share and establishments per firm of large service firms unchanged.

Endogenous exit. The quantitative results presented in the previous sections assumed that all firm exit is exogenous. We now show that allowing for endogenous exit leaves our main conclusions unchanged. To do so, we first recalibrate the exogenous exit probability to match the average entry and exit rates among firms with at least 20 employees.[19] Second, we recalibrate the model parameters. Instead of setting the fixed operating cost to zero, we choose it to match the average entry and exit rates in the BDS data. All details, including the re-estimated parameter values, are shown in Appendix B.5. The steady state of the model with endogenous exit fits the data well (*cf.* Table B.6).

We then repeat the main quantitative experiment and the channel decomposition.[20] The results are close to the baseline with exogenous exit. In the main experiment, where manufac-

[19] Those rates are 1.3 percent in manufacturing and 1.4 percent in services. Using entry and exit rates for firms with at least 500 employees yields very similar numbers.

[20] The estimated parameter values and moments used are reported in Table B.7.

turing productivity rises and η falls by a common factor in both sectors, the manufacturing employment share falls to 19.2 percent, or by 8.6 percentage points. The model thus accounts for 47 percent of the observed decline. For comparison, the share fell by 9.5 percentage points in the main experiment with exogenous exit. The model with endogenous exit also generates the shift toward large service firms, driven by an increase in their average number of establishments. The full results, including the channel decomposition, are shown in Table B.8.

Convex entry costs. The baseline model assumes a constant cost per entrant, so that the supply of potential entrants is perfectly elastic. This assumption matters because following an increase in the value of entry, free entry requires sectoral prices to fall in response. This price adjustment is what reallocates employment across sectors. With a less elastic supply of entrants, part of the adjustment can occur through the mass of entrants rather than prices, potentially affecting the strength of the two channels we study. To this end, we follow Karahan et al. (2024) and relax the free-entry assumption by allowing entry costs to increase with the mass of entrants in each sector:

$$c_{e,j}(M_j) = c_{e,j}\left(\frac{M_j}{M_{j,1979}}\right)^{\psi-1}, \qquad \psi \geq 1. \tag{18}$$

Here M_j is the mass of entrants in sector j and $M_{j,1979}$ is its value in the calibrated initial steady state. At $\psi = 1$, the specification reduces to the baseline. For $\psi > 1$, aggregate entry costs are convex and the supply of entrants has elasticity $1/(\psi - 1)$ with respect to the expected value of entry. The entry condition becomes $\int V_j(z)\,\mathrm{d}F_{e,j}(z) = Pc_{e,j}(M_j)$.

Table 2 reports the manufacturing employment shares for $\psi \in \{1.0, 1.5, 2.0\}$. For each value of ψ, we re-estimate the changes in manufacturing productivity and replication-cost convexity using the same targets as in the baseline. All other parameters remain at their calibrated values. The results remain close to the baseline. The main experiments now yield a manufacturing employment share of 18.0 to 18.1 percent, compared with 18.5 percent under constant entry costs, and the contributions of the two channels change little.

Table 2: Manufacturing employment share under convex entry costs

Experiment	$\psi = 1.0$	$\psi = 1.5$	$\psi = 2.0$
Main experiment	18.5	18.0	18.1
A_M only	19.9	20.2	20.3
η only	24.9	24.8	24.7

Notes: Manufacturing employment shares in 2023, in percent. The data counterparts are 27.9 percent in 1979 and 9.6 percent in 2023. ψ governs aggregate entry-cost convexity in Equation (18); $\psi = 1.0$ is the baseline with constant per-entrant costs. All experiments start from the same calibrated 1979 steady state. The main experiment changes A_M and the common multiplier on η_j to match the 2023 relative price of services and the ratio of establishments per firm among large relative to small service firms. The decomposition experiments change one parameter at a time to match its corresponding target.

Role of preferences. Finally, we examine the role of household preferences in three ways. First, we assess how much non-homothetic demand contributes to the model's predicted employment reallocation. Second, we examine whether the quantitative results are robust to replacing Stone-Geary preferences with an alternative non-homothetic demand specification. Third, we vary the expenditure-share slope used in the calibration of the model.

We begin by setting $\bar{c}_M = 0$, thus making household preferences homothetic. In order to obtain the same manufacturing employment share as in the baseline, we recalibrate the taste weight μ. We then apply the same changes to manufacturing productivity and replication costs as in the baseline. With $\bar{c}_M = 0$, at fixed relative prices, expenditure shares now remain constant as income rises: the income-driven reallocation toward services is absent. The results, presented in Table 3, show that manufacturing employment nevertheless falls from 28.0 to 23.1 percent, accounting for 26.4 percent of the observed decline. The results imply that non-homothetic demand amplifies aggregate reallocation, but the model still generates a substantial shift toward services without it.[21]

Next, we replace Stone-Geary preferences with non-homothetic CES preferences, which allow income effects to persist as household expenditure rises (Comin et al., 2021). We calibrate this specification to match the baseline initial expenditure share on goods and its response to rising household expenditures, holding relative prices fixed. Matching these

[21] Setting $\bar{c}_M = 0$ does more than eliminate the income effect because removing subsistence demand also changes price responses and equilibrium feedback. In addition, the recalibrated value of μ enters firms' cost bundle.

targets requires a lower household substitution elasticity, which we set to $\varepsilon = 0.30$. To distinguish the effect of changing the demand specification from that of lowering the substitution elasticity, we also repeat the Stone-Geary experiment at $\varepsilon = 0.30$. Appendix B.2 provides the calibration details.

Table 3 shows that the model specification with non-homothetic CES accounts for 49.1 percent of the observed employment reallocation, compared with 52.6 percent under Stone-Geary at the same household substitution parameter. The experiments where only manufacturing productivity or only the replication costs change explain 40.4 and 17.7 percent, respectively, under non-homothetic CES, compared with 45.1 and 16.7 percent under Stone-Geary preferences at the same household substitution elasticity.

The results are similar because both Stone-Geary and non-homothetic CES preferences reproduce the same initial spending pattern and the same local response of goods expenditure shares to rising household expenditure. In addition, the initial steady state calibration of firm-level parameters stays unchanged, so the mechanisms through which productivity and replication costs affect firm growth remain unchanged.

Finally, we assess the sensitivity to the expenditure-share slope used to discipline the strength of non-homotheticity in Stone-Geary preferences. Using the alternative estimates in Appendix Table B.1, we recalibrate household demand and repeat the main experiment. We find that the two channels jointly explain between 45.9 and 51.1 percent of the observed employment reallocation, compared with 51.5 percent in the baseline, shown in the lower panel of Table 3. These results show that our baseline results are robust to the alternative slope estimates. Appendix Table B.2 also reports the corresponding decomposition experiments.

5 Conclusion

This paper studies structural transformation through the lens of firm dynamics. Our empirical analysis shows that service-sector employment has become increasingly concentrated in large firms, which grew by increasing the number of establishments they operate. Incumbent expansion accounts for most service-sector employment growth and dwarfs the contribution

Table 3: Robustness to household preferences

Specification	Manufacturing employment share, 2023 (%)	Observed decline explained (%)
Panel A: Preference specifications		
Baseline Stone-Geary	18.5	51.5
Homothetic CES ($\bar{c}_M = 0$)	23.1	26.4
Stone-Geary, $\varepsilon = 0.30$	18.3	52.6
Non-homothetic CES, $\varepsilon = 0.30$	19.0	49.1
Panel B: Expenditure-share slope robustness		
Baseline + time trend (-0.223)	18.5	51.1
Longer sample (-0.184)	19.3	47.2
Longer sample + time trend (-0.171)	19.5	45.9

Notes: All specifications in panel A start from a manufacturing employment share of 27.97 percent. Results refer to the joint productivity and replication-cost experiment. The household substitution elasticity is $\varepsilon = 0.40$ unless otherwise indicated. The final two specifications in panel A match the baseline initial goods expenditure share and its response to higher household expenditure, holding relative prices fixed. Panel B varies the expenditure-share slope used to calibrate Stone-Geary preferences. The slopes are shown in parentheses and are taken from Appendix Table B.1. Demand parameters are recalibrated to match the initial manufacturing employment share and the selected slope. The changes in technology and replication costs are re-estimated using the same targets as in the baseline. Appendix Table B.2 shows the decomposition into the two channels.

of net firm entry. Opening new establishments within existing firms therefore plays a central role in the expansion of the service economy.

Motivated by these findings, we develop a two-sector model of firm dynamics in which incumbent firms can grow both within establishments and by opening new ones. The model distinguishes two forces behind structural transformation. Faster manufacturing productivity growth shifts employment toward services but leaves the within-sector firm-size distribution unchanged. Lower replication costs, by contrast, encourage establishment expansion, particularly among large service firms. Quantitatively, the two forces jointly account for about half of the observed decline in the manufacturing employment share and reproduce the shift toward large, multi-establishment service firms. Lower replication costs alone account for roughly one third of the employment reallocation explained by the joint experiment.

One interpretation of the calibrated sectoral differences is that manufactured goods are often tradable, allowing firms to serve additional markets from existing plants. Many services, by contrast, require a local presence, so reaching new customers often involves opening additional establishments. Our model captures the corresponding difference in how firms

scale: large service firms operate more establishments and therefore respond more strongly to improvements in replication technology.

Expansion across locations requires organizational capabilities as well as additional establishments. As Howard Schultz, Starbucks' former chairman and CEO, describes, expansion required systems for site selection, store design, and operations, as well as the transfer of managerial expertise across markets. In our model, improvements in these capabilities take the form of lower replication costs, allowing firms to operate more establishments, with the strongest expansion among large service firms. The capacity of incumbent firms to expand across establishments is therefore a key link between firm growth and structural transformation.

References

Aghion, P., A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li (2023): "A Theory of Falling Growth and Rising Rents," *The Review of Economic Studies*, 90, 2675–2702.

Argente, D., S. Moreira, E. Oberfield, and V. Venkateswaran (2025): "Scalable Expertise: How Standardization Drives Scale and Scope," NBER Working Paper 34160, National Bureau of Economic Research.

Basker, E. (2012): "Raising the Barcode Scanner: Technology and Productivity in the Retail Sector," *American Economic Journal: Applied Economics*, 4, 1–27.

Bornstein, G. (2025): "Entry and Profits in an Aging Economy: The Role of Consumer Inertia," Tech. rep., Wharton.

Brynjolfsson, E., W. Jin, G. Petropoulos, and X. Wang (2025): "Information Technology, Firm Size, and Industrial Concentration," NBER Working Paper 31065, National Bureau of Economic Research, first version March 2023.

Buera, F. J. and J. P. Kaboski (2012): "The Rise of the Service Economy," *American Economic Review*, 102, 2540–69.

Cao, D., H. R. Hyatt, T. Mukoyama, and E. Sager (2022): "Firm Growth through New Establishments," Tech. rep., Working paper, February 2022 version.

Clementi, G. L., A. Khan, B. Palazzo, and J. K. Thomas (2015): "Entry, Exit and the Shape of Aggregate Fluctuations in a General Equilibrium Model with Capital Heterogeneity," Tech. rep., Mimeo.

Clementi, G. L. and B. Palazzo (2016): "Entry, Exit, Firm Dynamics, and Aggregate Fluctuations," *American Economic Journal: Macroeconomics*, 8, 1–41.

Comin, D., D. Lashkari, and M. Mestieri (2021): "Structural Change With Long-Run Income and Price Effects," *Econometrica*, 89, 311–374.

Davis, S. J., J. C. Haltiwanger, and S. Schuh (1996): *Job Creation and Destruction*, vol. 1 of *MIT Press Books*, The MIT Press, 1 ed.

De Ridder, M. (2024): "Market Power and Innovation in the Intangible Economy," *American Economic Review*, 114, 199–251.

Dent, R. C., F. Karahan, B. Pugsley, and A. Şahin (2016): "The Role of Startups in Structural Transformation," *American Economic Review: Papers & Proceedings*, 106, 219–23.

Desmet, K. and E. Rossi-Hansberg (2014): "Spatial Development," *American Economic Review*, 104, 1211–1243.

Ding, X., T. C. Fort, S. J. Redding, and P. K. Schott (2022): "Structural Change Within versus Across Firms: Evidence from the United States," NBER Working Paper 30127, National Bureau of Economic Research.

Duarte, M. and D. Restuccia (2026): "Aggregate Employment and the Rise of Services across Time and Countries," NBER Working Paper 34804, National Bureau of Economic Research.

Eckert, F., S. Ganapati, and C. Walsh (2025): "Urban-Biased Growth: A Macroeconomic Analysis," Working paper, October 2025 revision.

Fort, T. C., J. R. Pierce, and P. K. Schott (2018): "New Perspectives on the Decline of US Manufacturing Employment," *Journal of Economic Perspectives*, 32, 47–72.

Foster, L., J. Haltiwanger, and C. J. Krizan (2006): "Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s," *The Review of Economics and Statistics*, 88, 748–758.

Gumpert, A., H. Steimer, and M. Antoni (2022): "Firm Organization with Multiple Establishments," *The Quarterly Journal of Economics*, 137, 1091–1138.

Haltiwanger, J., R. S. Jarmin, and J. Miranda (2013): "Who Creates Jobs? Small versus Large versus Young," *The Review of Economics and Statistics*, 95, 347–361.

Herrendorf, B., C. Herrington, and Á. Valentinyi (2015): "Sectoral Technology and Structural Transformation," *American Economic Journal: Macroeconomics*, 7, 104–33.

Herrendorf, B., R. Rogerson, and Á. Valentinyi (2014): "Growth and Structural Transformation," in *Handbook of Economic Growth*, ed. by P. Aghion and S. N. Durlauf, Elsevier, vol. 2 of *Handbook of Economic Growth*, 855–941.

Holmes, T. J. (2011): "The Diffusion of Wal-Mart and Economies of Density," *Econometrica*, 79, 253–302.

Hopenhayn, H., J. Neira, and R. Singhania (2022): "From Population Growth to Firm Demographics: Implications for Concentration, Entrepreneurship and the Labor Share," *Econometrica*, 90, 1879–1914.

——— (2025): "Firm Demographics and Sectoral Reallocation," Tech. rep., Working paper, December 2025 draft.

Hopenhayn, H. and R. Rogerson (1993): "Job Turnover and Policy Evaluation: A General Equilibrium Analysis," *Journal of Political Economy*, 101, 915–938.

Hopenhayn, H. A. (1992): "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60, 1127–1150.

Hsieh, C.-T. and E. Rossi-Hansberg (2021): "The Industrial Revolution in Services," Working Paper CES-21-34, U.S. Census Bureau, Center for Economic Studies.

——— (2023): "The Industrial Revolution in Services," *Journal of Political Economy Macroeconomics*, 1, 3–42.

Hyun, J., Z. Park, and V. Smirnyagin (2025): "Shock Propagation within Multisector Firms," *Journal of Money, Credit and Banking*, advance online publication.

Jarmin, R. S., S. D. Klimek, and J. Miranda (2009): *Producer Dynamics: New Evidence from Micro Data*, University of Chicago Press, chap. The Role of Retail Chains: National, Regional and Industry Results, 237–262, NBER Chapters.

Karahan, F., B. Pugsley, and A. Şahin (2024): "Demographic Origins of the Start-up Deficit," *American Economic Review*, 114, 1986–2023.

Kleinman, B. (2025): "Wage Inequality and the Spatial Expansion of Firms," Tech. rep., Stanford University, revise and resubmit, Quarterly Journal of Economics.

Kongsamut, P., S. Rebelo, and D. Xie (2001): "Beyond Balanced Growth," *The Review of Economic Studies*, 68, 869–882.

Lashkari, D., A. Bauer, and J. Boussard (2024): "Information Technology and Returns to Scale," *American Economic Review*, 114, 1769–1815.

Luttmer, E. G. J. (2007): "Selection, Growth, and the Size Distribution of Firms," *The Quarterly Journal of Economics*, 122, 1103–1144.

Ngai, L. R. and C. A. Pissarides (2007): "Structural Change in a Multisector Model of Growth," *American Economic Review*, 97, 429–443.

Oberfield, E., E. Rossi-Hansberg, P.-D. Sarte, and N. Trachter (2024): "Plants in Space," *Journal of Political Economy*, 132, 867–909.

Peters, M. and C. Walsh (2026): "Population Growth and Firm Dynamics," *Journal of Political Economy Macroeconomics*, 4, 256–296.

Samaniego, R. M. and J. Y. Sun (2016): "Productivity Growth and Structural Transformation," *Review of Economic Dynamics*, 21, 266–285.

Sedláček, P. and V. Sterk (2017): "The Growth Potential of Startups over the Business Cycle," *American Economic Review*, 107, 3182–3210.

Smirnyagin, V. (2023): "Returns to Scale, Firm Entry, and the Business Cycle," *Journal of Monetary Economics*, 134, 118–134.

Sterk, V., P. Sedláček, and B. Pugsley (2021): "The Nature of Firm Growth," *American Economic Review*, 111, 547–79.

Xi, X. (2023): "Multi-Establishment Firms, Misallocation, and Productivity," *Journal of Economic Dynamics and Control*, 154, 104705.

Appendix

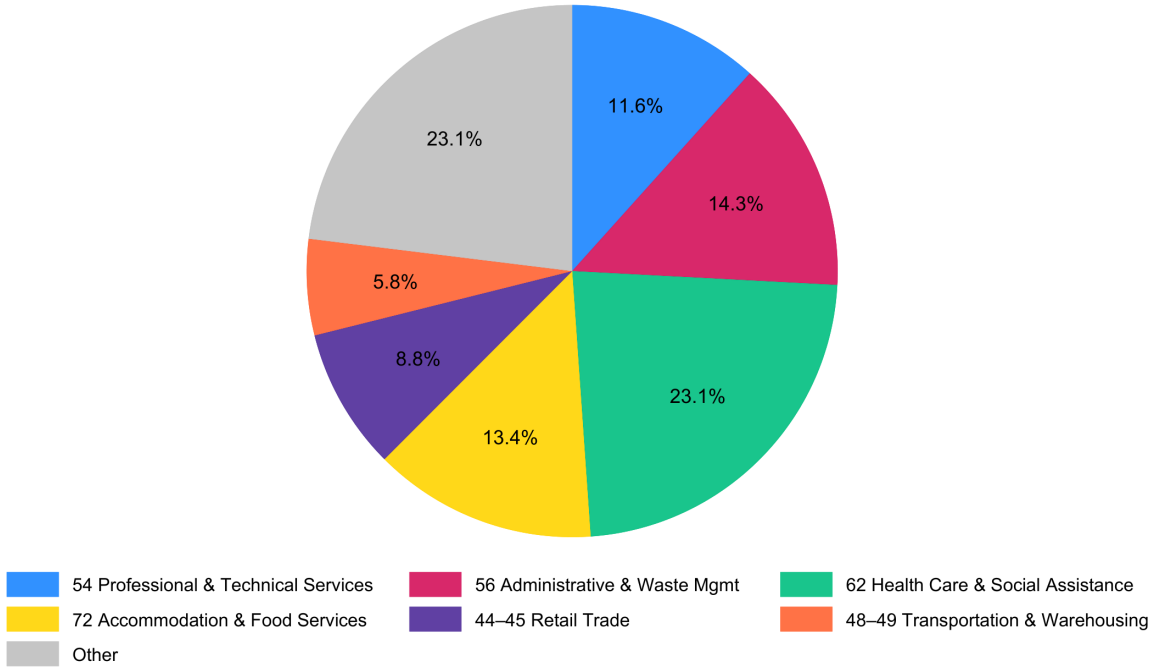
A Empirical Appendix

This appendix collects the sector definitions and additional figures and tables supporting the empirical analysis in Section 2.

A.1 Additional Figures and Tables

Sector coverage and composition. Figure A.1 breaks down service-sector employment gains by two-digit industry. Growth is spread across several industries, with health care and social assistance contributing the largest share.

FIGURE A.1: SERVICE-SECTOR EMPLOYMENT GAINS BY SUBSECTOR



Notes: Slices show each subsector's share of total gains; slices below 5 percent are grouped as Other. Data source: BDS.

Table A.1 shows our baseline definition of service, manufacturing, and other sectors.

Table A.1: Sector classification

Code	**Sector Name**	**Category**
11	Agriculture, Forestry, Fishing and Hunting	Other
21	Mining, Quarrying, and Oil and Gas Extraction	Other
22	Utilities	Other
23	Construction	Other
31–33	Manufacturing	Manufacturing
42	Wholesale Trade	Services
44–45	Retail Trade	Services
48–49	Transportation and Warehousing	Services
51	Information	Services
52	Finance and Insurance	Services
53	Real Estate and Rental and Leasing	Services
54	Professional, Scientific, and Technical Services	Services
55	Management of Companies and Enterprises	Services
56	Administrative and Support and Waste Management Services	Services
61	Educational Services	Services
62	Health Care and Social Assistance	Services
71	Arts, Entertainment, and Recreation	Services
72	Accommodation and Food Services	Services
81	Other Services (except Public Administration)	Services

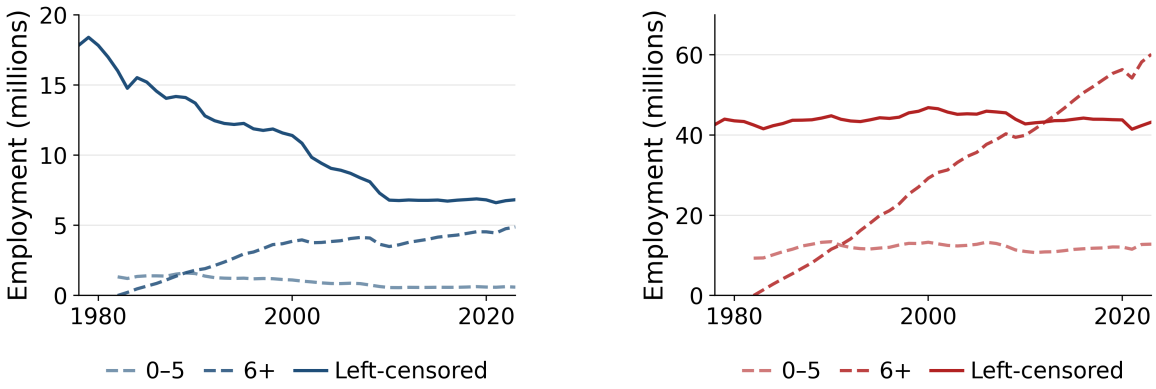
Table A.2 reports summary statistics for both sectors in 1979, 2001, and 2023.

TABLE A.2: SUMMARY STATISTICS BY SECTOR

	Manufacturing			Services		
	1979	2001	2023	1979	2001	2023
Employment (millions)	19.1	15.8	12.3	49.3	90.1	116.2
Small firms (<20 employees), %	6.8	8.3	8.0	26.3	19.2	16.5
Medium firms (20–499 employees), %	28.3	33.6	32.3	32.1	30.3	28.3
Large firms (≥500 employees), %	64.9	58.1	59.7	41.6	50.6	55.3
Number of firms (thousands)	261	285	221	3,002	4,053	4,712
Small firms (<20 employees), %	69.0	71.1	72.8	89.3	87.4	87.5
Medium firms (20–499 employees), %	29.2	27.1	25.2	10.0	11.7	11.7
Large firms (≥500 employees), %	1.9	1.8	2.1	0.7	0.9	0.9
Average number of establishments per firm	1.19	1.17	1.22	1.22	1.32	1.37
Small firms (<20)	1.00	1.00	1.01	1.02	1.01	1.01
Medium firms (20–499)	1.16	1.17	1.26	1.67	1.71	1.77
Large firms (500+)	8.13	7.49	7.90	18.82	25.19	32.51
Average firm size (employees)	73.2	55.5	55.7	16.4	22.2	24.7
Small firms (<20)	7	6	6	5	5	5
Medium firms (20–499)	71	69	71	53	57	60
Large firms (500+)	2,526	1,808	1,614	922	1,245	1,591
Number of establishments (in 1,000)	310	332	268	3,664	5,334	6,464
Employment per establishment	61.7	47.6	45.8	13.5	16.9	18.0
Firm entry rate (%)	9.3	5.9	5.2	12.3	8.9	9.1
Firm exit rate (%)	7.4	7.0	6.9	8.6	8.3	9.4

Notes: Sector definitions follow Table A.1. The firm entry (exit) rate is the number of age-0 firms (number of firm deaths) divided by the total number of firms.

Figure A.2: Sectoral employment by firm age



(A) Manufacturing

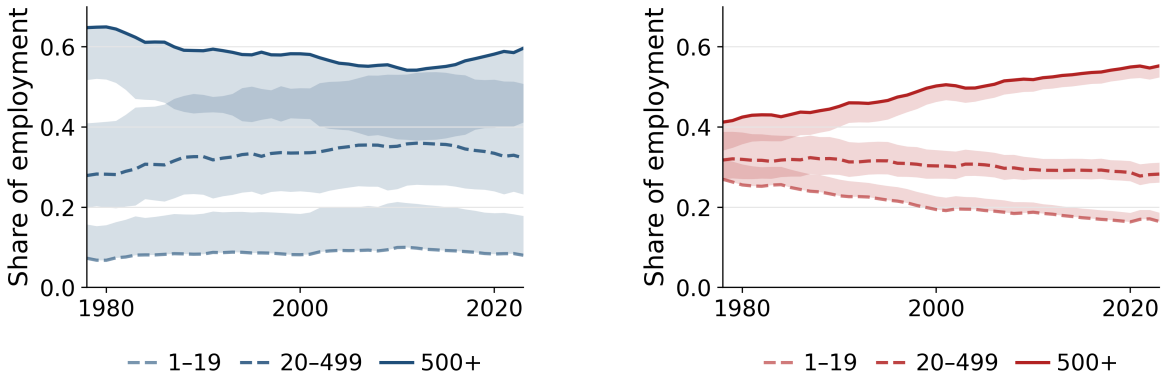
(B) Services

Notes: Data source: BDS.

Multisector firms in the BDS. The BDS measures firm size by a firm's total employment across all sectors. A firm with establishments in several two-digit sectors therefore appears in each of these sectors. This raises the concern that the rise of large service firms, shown in Figure 3, could reflect firms that are large only because of employment outside the services sector. To address this concern, Figure A.3 shows employment shares by firm-size bin in manufacturing and services, together with shaded bands that account for the potential effect of multisector firms.

To construct the shaded bands, we proceed as follows, taking a service firm as an example. A firm's total employment is at least its employment in services, meaning the BDS can place a firm in a size bin that is too large, but never in one that is too small. The measured share of the 500+ bin is thus an upper bound on the true share, and the measured share of the 1–19 bin is a lower bound. Next, consider a firm in the 500+ bin that has fewer than 500 service workers. Each potentially misclassified firm contributes fewer than 500 service workers and must also appear in the 500+ category of at least one non-service sector. Multiplying the summed number of 500+ firms in non-service sectors by 500 therefore gives a conservative ceiling on the service employment attributed to large firms solely because of their employment outside services. We subtract this maximum to obtain the lower bound for the 500+ bin. We apply the same logic to the other size classes; the 20–499 bin has a two-sided band because it can gain firms from the 500+ bin and lose firms to the 1–19 bin.[22]

FIGURE A.3: EMPLOYMENT SHARES BY FIRM SIZE



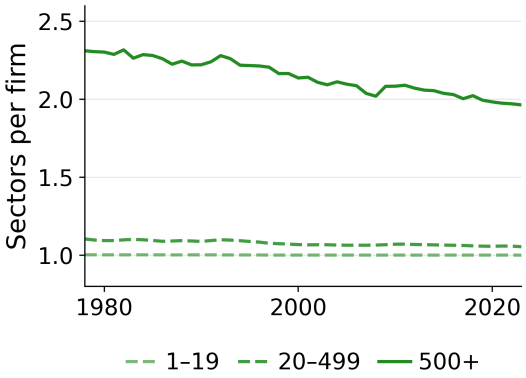
(A) Manufacturing (B) Services

Notes: Lines show the share of sectoral employment in each firm-size bin (based on levels reported in Figure 3). Shaded bands show the range of possible employment shares across firm-size bins. See text for details. Data source: BDS.

Panel (B) of Figure A.3 alleviates the concern that the rise of large service firms was driven by non-service employment. The band for large service firms is narrow, and it shrinks over time. In the scenario that is the most unfavorable to our narrative, the share of service employment in large firms still rises from 41 percent in 1978 to 52 percent in 2023, while the share in small firms falls from 27 percent to 19 percent. Multisector firms therefore cannot

[22]These bands are illustrative rather than formal bounds: the calculation abstracts from changes in size categories arising from changes in establishment ownership.

FIGURE A.4: PRESENCE IN TWO-DIGIT NAICS SECTORS, BY FIRM SIZE BIN



Notes: Number of firms summed over the 19 two-digit NAICS sectors, divided by the economy-wide number of firms, by firm-size bin. A firm with establishments in several sectors is counted once in each sector. Data source: BDS.

account for the shift of service employment toward large firms. The bands in panel (A) of Figure A.3 are wider, because there are relatively few large manufacturing firms.

Additional evidence against the idea that large service firms have primarily grown through employment in other sectors comes from Figure A.4. The BDS counts a firm once in every sector in which it has an establishment. The figure shows the average number of two-digit sectors in which a firm appears, by size bin. Within each size bin, we sum sectoral firm counts and divide by the corresponding economy-wide firm count. The resulting ratio measures how many sectors firms appear in on average. Small and medium firms essentially always operate in a single sector, whereas large firms operated in 2.3 sectors on average in 1978 and in 2.0 sectors in 2023. Presence in several sectors is therefore a feature of large firms. Importantly, if large service firms had grown through employment in other sectors, their presence across sectors would have increased; instead, it fell. Thus, it is unlikely that multisector firms played a central role in the growing concentration of service employment in large firms, or in the expansion of large service firms through additional establishments.

Establishment growth within service industries. Figure A.5 reports establishments per firm within two-digit service industries. The increase among large firms appears across nearly all of these industries, supporting the interpretation that the aggregate pattern reflects expansion within industries.

FIGURE A.5: AVERAGE NUMBER OF ESTABLISHMENTS PER FIRM, BY SERVICE INDUSTRY

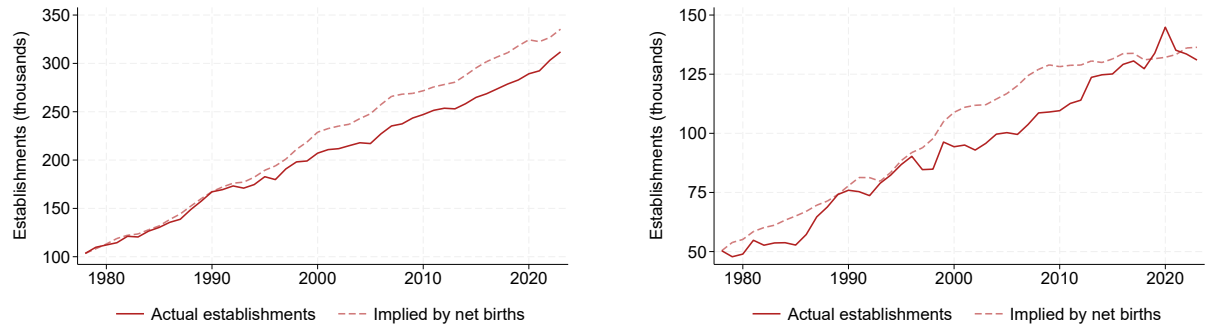


Notes: Data source: BDS.

New openings and ownership changes. Figure A.6 compares actual establishment counts with the counts implied by cumulative openings minus permanent closures in two size groups of large service firms. The close correspondence indicates that net openings account for most of the increase in these groups. The remaining gap includes both ownership changes and movements across size categories.

A caveat is that this comparison does not separately identify the contribution of consolidation to rising establishments per firm. A merger between firms already in the same size group can reduce the firm count without changing the establishment stock, thereby increasing establishments per firm without generating a gap between the two series. The evidence nevertheless shows that net openings closely track establishment-stock growth in the two groups. The comparison covers firms with 500–4,999 employees; firms with 5,000 or more employees are omitted because some underlying flows are suppressed.

FIGURE A.6: ESTABLISHMENT STOCKS AND CUMULATIVE NET OPENINGS AT LARGE SERVICE FIRMS

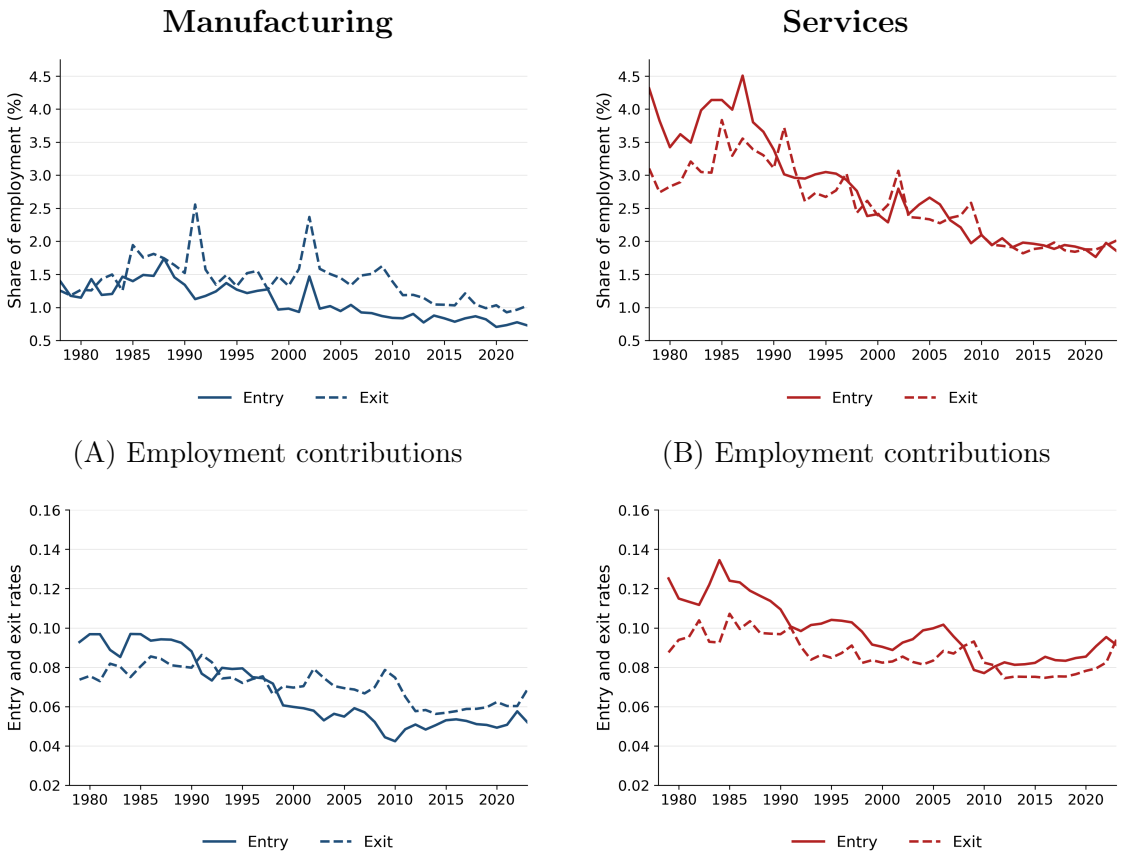


(A) Firms with 500–2,499 employees (B) Firms with 2,500–4,999 employees

Notes: The solid line plots the actual number of establishments at service firms in each size group. The dashed line plots the number of establishments implied by net entry: the 1978 stock of establishments plus cumulative net establishment births within the size group. In the BDS data, establishment births and deaths record newly opened establishments and permanent closures, so the dashed line is robust to ownership changes. A gap between the lines captures acquisitions, as well as firms growing into (or leaving) a given size group. Firms with 5,000 or more employees are not shown because several of the underlying entry and exit counts are suppressed in BDS data due to disclosure risk. Data source: BDS.

Firm turnover and employment flows. Figure A.7 reports the employment contributions of entering and exiting firms in the top row and firm entry and exit rates in the bottom row. Gross employment flows are substantial but largely offset each other, explaining why the net entry contribution is much smaller than either gross flow.

FIGURE A.7: FIRM ENTRY AND EXIT



(A) Employment contributions

(B) Employment contributions

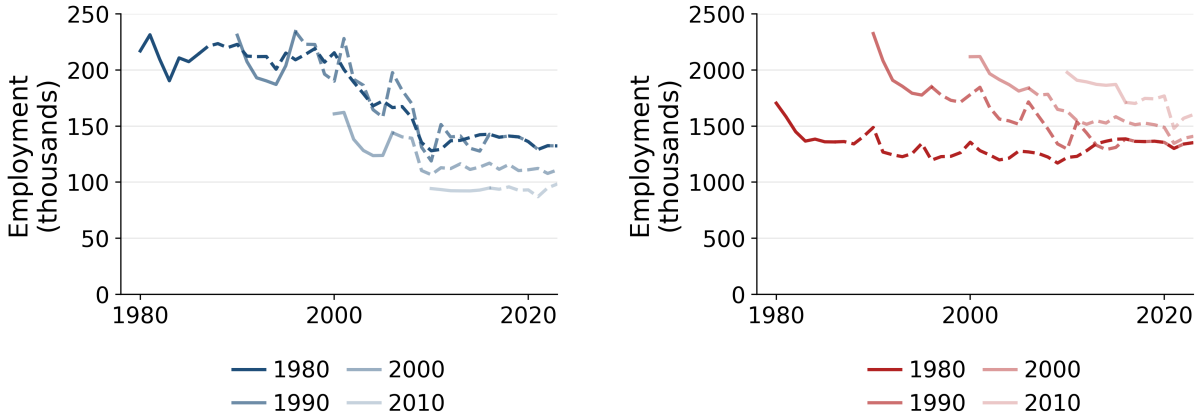
(C) Firm entry and exit rates

(D) Firm entry and exit rates

Notes: The top row reports employment created by entering firms and lost through firm exit, as percentages of sectoral employment using the BDS employment denominator. The bottom row reports the numbers of entering and exiting firms divided by the average of the current and previous year's firm counts. Solid lines show entry; dashed lines show exit. Data source: BDS.

Employment after firm entry. Figure A.8 follows employment in the 1980, 1990, 2000, and 2010 entry cohorts. Service cohort employment remains broadly stable as cohorts age, whereas manufacturing cohort employment declines and later cohorts enter with less employment. Beyond age five, the plotted cohort paths are approximations based on the BDS age bins.

FIGURE A.8: COHORT EMPLOYMENT OVER TIME



(A) Manufacturing (B) Services

Notes: For ages 0–5, the BDS reports employment by exact firm age (solid segments). At older ages, the BDS reports employment only in multi-year age bins, so we approximate cohort employment with the bin average (dashed segments). Data source: BDS.

Employment accounting and young establishments. Table A.3 reports the gross flows and net contributions underlying the main-text decomposition.

Table A.3: Decomposition of employment changes by sector, 1978–2023

Component	Δ Employment (millions)	Share of Total Change (%)
Panel A: Manufacturing		
Net entry	−2.07	37.7
Entry (startups)	7.64	
Exit (firm deaths)	−9.71	
Net expansion	−0.60	11.0
Job creation (continuers)	52.67	
Job destruction (continuers)	−53.27	
Net establishment expansion of incumbent firms	−2.81	51.3
Total change	−5.48	**100.0**
Panel B: Services		
Net entry	4.88	6.8
Entry (startups)	98.41	
Exit (firm deaths)	−93.52	
Net expansion	37.57	52.5
Job creation (continuers)	381.45	
Job destruction (continuers)	−343.88	
Net establishment expansion of incumbent firms	29.05	40.6
Total change	71.50	**100.0**
Panel C: Other		
Net entry	0.31	7.4
Entry (startups)	12.45	
Exit (firm deaths)	−12.15	
Net expansion	3.39	81.6
Job creation (continuers)	46.83	
Job destruction (continuers)	−43.45	
Net establishment expansion of incumbent firms	0.46	11.0
Total change	4.15	**100.0**

Notes: Data source: BDS.

Figure A.9 separates employment changes at continuing establishments into growth at young establishments within mature firms and growth at all other continuing establishments. The first component captures part of the subsequent growth of locations opened by incumbent firms, which equation (1) assigns to the continuing-establishment margin rather than to establishment creation.

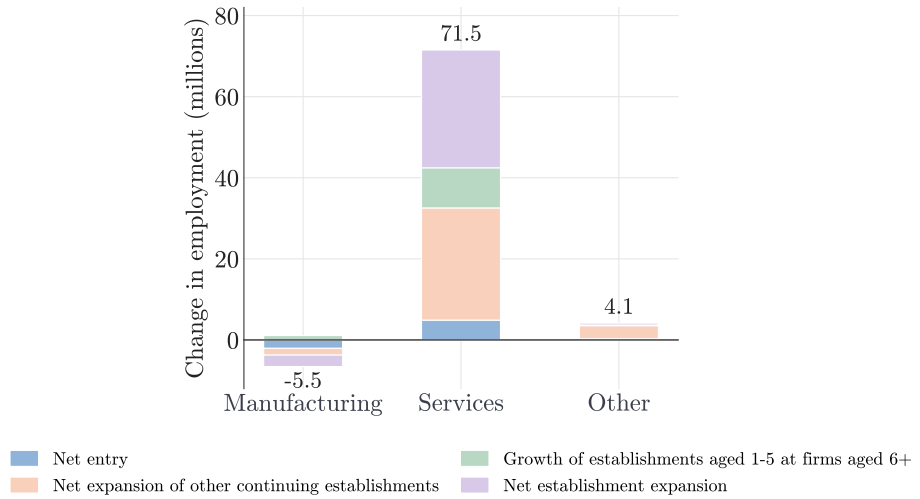
The BDS reports job creation and destruction by establishment age and by firm age separately, but not jointly. We recover the relevant joint category as follows: Let a_E denote establishment age and a_F firm age, and retain $G^C_{s,t}$ and $L^C_{s,t}$ for job creation and destruction at continuing establishments. With parentheses indicating the age restrictions, we compute

$$\begin{aligned}&(G^C_{s,t} - L^C_{s,t})(1 \le a_E \le 5,\ a_F \ge 6)\\ &\quad \approx (G^C_{s,t} - L^C_{s,t})(1 \le a_E \le 5) - (G^C_{s,t} - L^C_{s,t})(1 \le a_F \le 5).\end{aligned} \tag{A.1}$$

The subtraction approximates net job creation at young establishments within mature firms. Firm age is assigned at birth and advances thereafter, so young firms can subsequently acquire older establishments. The approximation is exact when young firms do not own establishments older than five years.

Figure A.9 then sums these annual flows over the sample period and separates them from the remaining continuing-establishment component. Net firm entry and net establishment entry are unchanged. This calculation captures only part of the post-opening growth of establishments created by incumbents: it excludes growth after establishments reach age six and growth while their parent firms are still aged five or younger.

FIGURE A.9: POST-ENTRY GROWTH OF YOUNG ESTABLISHMENTS AT MATURE FIRMS, 1978–2023



Notes: The figure repeats the decomposition in Figure 5, separating net job creation at continuing establishments into two components: establishments aged one to five within firms aged six and older, and all other continuing establishments. Net firm entry and net establishment entry within incumbent firms are unchanged. Numbers next to the bars report total employment changes. The construction of the age-based split is described in Equation (A.1). Data source: BDS.

B Model Appendix

This appendix describes the numerical solution of the model and provides additional details on the demand calibration and quantitative experiments. It then relates the establishment expansion in our model to mechanisms for growth across markets, branches, and products in recent work.

B.1 Computational Details

Outline of the Algorithm. The only state variable is idiosyncratic productivity. We discretize the AR(1) productivity process and then solve the static establishment and labor choices analytically at each grid point. We evaluate the normal density of log entrant productivity on the sector's productivity grid and normalize the resulting probabilities to sum to one. For a given sector and a candidate output price, the algorithm proceeds as follows:

1. Compute optimal labor and operating profit per establishment for each productivity state.
2. Given per-establishment operating profit, solve analytically for the optimal number of establishments $m^*(z)$.
3. Construct current-period profits and solve the incumbent Bellman equation on the discretized Markov chain.
4. Under exogenous exit, the continuation operator is linear and the value function is obtained in closed form. Under endogenous exit, the code iterates on the Bellman equation until convergence.
5. Back out the sector-specific entry cost $c_{e,j}$ from the free-entry condition in the initial steady state.
6. For counterfactual exercises, bisect over the sectoral output price until the free-entry condition holds at the previously backed-out entry cost.

The stationary measure is then obtained from the law of motion implied by the transition matrix, entry, and firms' exit decisions. The linearity of this law makes it possible to obtain the measure associated with a unit mass of entrants and then scale it by the equilibrium entrant mass, as described next.

Stationary Aggregation. Let $q_{j,rs}$ denote the probability that productivity moves from grid point $z_{j,r}$ to grid point $z_{j,s}$. The transition matrix for continuing firms has elements

$$[\mathbf{Q}_j]_{rs} = (1 - \delta_j)\,[1 - \text{exit}_j(z_{j,r})]\,q_{j,rs}.$$

Let $\boldsymbol{\nu}$ be any measure of incumbents over the productivity grid, written as a column vector, and let M be a mass of entrants. The law of motion for the measure of incumbents is

$$\mathcal{T}_j(\boldsymbol{\nu}, M) \equiv \mathbf{Q}_j'\boldsymbol{\nu} + M\boldsymbol{f}_{e,j}, \tag{B.1}$$

where $\boldsymbol{f}_{e,j}$ is the vector of entrant probabilities over the productivity grid corresponding to $F_{e,j}$. The first term is the measure of continuing incumbents and the second is the measure of new entrants. The stationary equilibrium measure therefore satisfies

$$\boldsymbol{\mu}_j = \mathcal{T}_j(\boldsymbol{\mu}_j, M_j).$$

Because (B.1) is linear in the incumbent measure and the mass of entrants, let $\widehat{\boldsymbol{\mu}}_j$ denote the stationary measure associated with a unit mass of entrants. It satisfies

$$\widehat{\boldsymbol{\mu}}_j = (\mathbf{I} - \mathbf{Q}_j')^{-1}\boldsymbol{f}_{e,j}, \qquad \boldsymbol{\mu}_j = M_j\widehat{\boldsymbol{\mu}}_j.$$

Aggregating over $\widehat{\boldsymbol{\mu}}_j$ yields gross output, labor demand, dividends, and model moments per unit mass of entrants in each sector. The two sectoral goods-market-clearing conditions then jointly pin down M_M and M_S; labor-market clearing follows from the remaining equilibrium conditions by Walras' law. When costs are denominated in the final good, fixed, entry, and replication costs generate additional final-good demand from firms' cost payments, allocated across sectors using the same CES expenditure shares as household consumption. When costs are denominated in labor, these costs instead enter labor demand directly.

B.2 Demand Calibration and Additional Quantitative Results

Calibration of Non-Homothetic Demand. This appendix explains how we discipline the manufacturing subsistence term in household demand. The goal is to measure how the expenditure share on goods changes as real expenditure per capita rises, after accounting for movements in relative prices.

We use annual U.S. Personal Consumption Expenditures data from the Bureau of Economic Analysis and population estimates from the Census Bureau. From BEA Table 2.4.5U we collect total personal consumption expenditures, goods expenditures, and services expenditures. From BEA Table 2.4.4U we collect the overall PCE price index, the goods price index, and the services price index. We then construct three objects: the expenditure share on goods, real expenditure per capita, and the relative price of goods to services. The baseline specification regresses the goods expenditure share on log real expenditure per capita and the log relative price of goods to services, using annual data from 1970 to 2025. We treat goods as an empirical proxy for manufacturing.

The estimate of interest is the coefficient on log real expenditure per capita. A more negative coefficient means that, conditional on relative prices, higher expenditure per capita is associated with a lower goods expenditure share and therefore stronger reallocation toward the service sector. In the model, the manufacturing subsistence term is chosen so that the initial steady state reproduces this local expenditure-share slope. We then use the manufacturing employment share to pin down the remaining demand-side weight on supernumerary expenditure.

Table B.1 reports the baseline estimate together with simple robustness checks. Our benchmark calibration uses the 1970–2025 specification, which yields a slope of -0.227. The results consistently imply a negative relationship between the goods expenditure share and real expenditure per capita.

Table B.1: Expenditure-share regressions

Specification	Sample	Income slope	Std. error
Baseline	1970–2025 ($N = 56$)	-0.227	(0.045)
Baseline + time trend	1970–2025 ($N = 56$)	-0.223	(0.054)
Longer sample	1959–2025 ($N = 67$)	-0.184	(0.021)
Longer sample + time trend	1959–2025 ($N = 67$)	-0.171	(0.040)

Notes: The income slope is the coefficient on log real expenditure per capita in a regression of the goods expenditure share on that variable and the log goods-to-services relative price. All specifications include an intercept; the specifications with a time trend additionally include a linear time trend.

Sensitivity to the expenditure-share slope. Table B.2 repeats the full results of the main experiment using alternative expenditure-share slopes. For each slope, we recalibrate the demand weight and subsistence term to match the initial manufacturing employment share and the selected slope. We then re-estimate the productivity and replication-cost changes using the same targets as in the corresponding baseline experiments. Less negative slopes reduce the employment reallocation explained by both channels.

Table B.2: Robustness to the expenditure-share slope

Slope specification	Income slope	Observed decline explained (%)		
		Joint	A_M only	η only
Baseline	−0.227	51.5	44.0	16.7
Baseline + time trend	−0.223	51.1	43.7	16.5
Longer sample	−0.184	47.2	41.2	13.2
Longer sample + time trend	−0.171	45.9	40.3	12.1

Notes: All experiments use Stone-Geary preferences with substitution elasticity $\varepsilon = 0.40$. All slope estimates are reported in Table B.1. The income slope is the estimated coefficient on log real expenditure per capita, conditional on the log goods-to-services relative price and, where included, a linear time trend. The experiments target the relative price and/or the ratio of establishments per large to small service firm.

Calibration of Non-Homothetic CES Preferences. For the alternative household demand specification, we match the baseline initial goods expenditure share, $s_M = 0.346$, and the same expenditure-share slope used to calibrate Stone-Geary preferences, $\partial s_M / \partial \log I = -0.227$, holding relative prices fixed. These targets imply an initial goods income elasticity of

$$1 + \frac{1}{s_M} \frac{\partial s_M}{\partial \log I} = 1 - \frac{0.227}{0.346} \simeq 0.344.$$

In the two-sector non-homothetic CES specification, this elasticity must exceed the substitution parameter. The baseline value $\varepsilon = 0.40$ therefore cannot reproduce both targets. We set the household substitution parameter to $\varepsilon = 0.30$ and choose the taste weight and service non-homotheticity parameter to match them. With the manufacturing non-homotheticity parameter normalized to one, the service parameter is 23.54. Its large value reflects the proximity of the targeted goods income elasticity to the lower bound of 0.30.

For comparison, we also evaluate Stone-Geary household demand at $\varepsilon = 0.30$, matching the same initial expenditure share and expenditure-share slope. In both specifications, firms retain their baseline cost bundle, including its substitution elasticity of 0.40, and the initial firm calibration and allocation remain unchanged. We re-estimate the productivity and replication-cost changes using the baseline targets.

B.3 Parameter Changes Across Experiments

Table B.3 summarizes the parameter changes and target moments in the main experiment and the two single-channel exercises.

Table B.3: Parameter changes and targets across experiments

Panel A: Parameter changes

Experiment	$A_M/A_{M,1979}$	$\eta_j/\eta_{j,1979}$
Main experiment	1.57	0.80
A_M only	1.55	1.00
η only	1.00	0.70

Panel B: Target moments

	Data		Model 2023		
Moment	1979	2023	Main	A_M only	η only
Relative service price, p_S/p_M	0.64	0.99	0.99*	0.99*	0.63
Estabs. per firm, 500+/<20, services	18.44	32.14	32.14*	18.00	32.14*

Notes: All experiments start from the calibrated 1979 steady state. Panel A reports parameter ratios to 1979 values; a value of one indicates no change. The same proportional change in η_j applies to manufacturing and services. Aggregate labor supply is held fixed at its 1979 value. All other primitives remain fixed. Panel B reports the data and model outcomes for the two calibration moments. An asterisk indicates that the moment's 1979–2023 change is targeted in that experiment.

Table B.4: Outcomes across quantitative experiments

	Data		Model 2023		
Outcome	1979	2023	Main	A_M only	η only
Manufacturing employment share (%)	27.9	9.6	18.5	19.9	24.9
Change in manufacturing share (p.p.)	—	−18.4	−9.5	−8.1	−3.1
Employment share of large service firms (%)	41.6	55.3	53.9	41.6	68.9
Establishments per large service firm	18.8	32.5	32.8	18.7	32.6

Notes: All experiments start from the calibrated 1979 steady state. The model columns report the main experiment, the manufacturing-productivity-only experiment, and the replication-cost-only experiment. Aggregate labor supply is held fixed at its 1979 value. Parameter changes and calibration targets are reported in Table B.3. Employment shares in 2023 are not targeted; the establishment target is the ratio of establishments per firm among large (500+) relative to small (<20) service firms.

B.4 Labor-Force Growth

This extension repeats the main experiment allowing for employment growth between 1979 and 2023, while retaining the same productivity and replication-cost changes. The first version increases aggregate labor supply while holding aggregate subsistence demand fixed. The second experiment increases aggregate subsistence demand in proportion to population growth. Aggregate employment grew by a factor of 1.8768 in the BDS, while population grew by a factor of 1.4975, calculated from annual averages of the monthly FRED population series POPTHM. Employment per person therefore grew by a factor of 1.2533. For sectoral employment shares, the second version is equivalent to increasing labor supply by this factor while holding subsistence demand fixed.

Table B.5: Main experiment with labor-force growth

	Data 2023	Main	Main $+\bar{L}$	Main + $\bar{L}/N$
Manufacturing employment share (%)	9.6	18.5	14.8	16.9
Change in manufacturing share (p.p.)	−18.4	−9.5	−13.2	−11.1
Employment share of large service firms (%)	55.3	53.9	53.9	53.9
Establishments per large service firm	32.5	32.8	32.8	32.8

Notes: All experiments start from the calibrated 1979 steady state. Aggregate labor supply is held fixed in the main experiment and grows by a factor of 1.88 in the labor-force-growth extension ("Main + $\bar{L}$"). In all cases, manufacturing productivity rises by a factor of 1.57 and replication-cost convexity falls by a factor of 0.80. The relative service price and establishments per firm among large relative to small service firms are targeted in the main experiment and remain matched in both growth extensions. Large firms have at least 500 employees. The last column ("Main + $\bar{L}/N$") additionally scales aggregate subsistence demand by 1.4975. For employment shares, this is equivalent to increasing labor supply by employment growth per person while holding subsistence demand fixed.

B.5 Endogenous Exit

In this extension, firms face both an exogenous exit shock and the endogenous exit rule in Equation (12). We set the exogenous rate to the 1979–1981 average entry and exit rates among firms with at least 20 employees. These rates are 1.277 percent in manufacturing and 1.361 percent in services. The overall entry-exit average is still a calibration target and now identifies the fixed operating cost. Table B.6 reports the parameters that change relative to Table 1; parameters omitted from the table are unchanged.

Conditional on surviving the exogenous shock, the endogenous exit probabilities are 7.0 and 9.4 percent, respectively, meaning that selection accounts for most steady-state exit in both sectors.

Table B.6: Endogenous-exit calibration

Parameter	Description	Value	Target	Data	Model
		Aggregate and assigned parameters			
δ_M	Exogenous exit rate	0.013	Entry/exit avg., firms with 20+	0.013	0.013
δ_S	Exogenous exit rate	0.014	Entry/exit avg., firms with 20+	0.014	0.014
$\bar{c}_M$	Manufacturing subsistence	0.300	Goods-share slope wrt log PCE p.c.	−0.23	−0.23
μ	CES preference weight	0.118	Manufacturing employment share	0.28	0.28
		Manufacturing			
μ_M^z	Mean log productivity	2.02	Average firm size	73.21	73.26
σ_M	Productivity volatility	0.20	Employment share, firms with 500+	0.65	0.65
$\mu_{e,M}^z$	Mean log entrant productivity	−0.74	Average entrant size	9.04	9.07
$\sigma_{e,M}$	Entrant productivity dispersion	1.20	Entrant firm share with 20+	0.033	0.033
ρ_M	Productivity persistence	0.94	Age 1–5 firm share with 20+	0.153	0.155
α_M	Establishment returns to scale	0.50	Average establishment size	61.72	61.45
f_M	Fixed operating cost	25.49	Average firm entry and exit rate	0.083	0.082
$c_{m,M}$	Replication-cost level	101.47	Average establishments per firm	1.19	1.19
η_M	Replication-cost convexity	0.59	Estab./firm: 500+ relative to <20	8.10	8.08
$c_{e,M}$	Entry cost	68.23	Free-entry condition	—	—
		Services			
μ_S^z	Mean log productivity	1.49	Average firm size	16.43	16.45
σ_S	Productivity volatility	0.38	Employment share, firms with 500+	0.42	0.41
$\mu_{e,S}^z$	Mean log entrant productivity	1.86	Average entrant size	4.90	4.88
$\sigma_{e,S}$	Entrant productivity dispersion	0.33	Entrant firm share with 20+	0.013	0.013
ρ_S	Productivity persistence	0.81	Age 1–5 firm share with 20+	0.063	0.063
α_S	Establishment returns to scale	0.48	Average establishment size	13.46	13.51
f_S	Fixed operating cost	9.15	Average firm entry and exit rate	0.106	0.106
$c_{m,S}$	Replication-cost level	24.22	Average establishments per firm	1.22	1.22
η_S	Replication-cost convexity	0.50	Estab./firm: 500+ relative to <20	18.44	18.47
$c_{e,S}$	Entry cost	29.76	Free-entry condition	—	—

Notes: Under endogenous exit, δ_j is the exogenous exit rate and f_j is calibrated to the overall average sectoral entry and exit rates. Parameters that are unchanged from Table 1 are omitted.

Table B.7 reports the same parameter changes and target moments as Table B.3. The estimated parameter changes are similar to those in the baseline. The joint experiment accounts for 47.1 percent of the observed decline in the manufacturing employment share,

compared with 51.5 percent in the baseline model with exogenous exit. The manufacturing-productivity and replication-cost channels account for 42.0 and 11.5 percent, respectively, compared with 44.0 and 16.7 percent in the baseline. Thus, allowing for endogenous exit preserves the main quantitative conclusions of the model.

TABLE B.7: PARAMETER CHANGES AND TARGETS WITH ENDOGENOUS EXIT

Panel A: Parameter changes

Experiment	$A_M/A_{M,1979}$	$\eta_j/\eta_{j,1979}$
Main experiment	1.54	0.78
A_M only	1.54	1.00
η only	1.00	0.70

Panel B: Target moments

	Data		Model 2023		
Moment	1979	2023	Main	A_M only	η only
Relative service price, p_S/p_M	0.64	0.99	0.99*	0.99*	0.64
Estabs. per firm, 500+/<20, services	18.44	32.14	32.14*	18.40	32.14*

Notes: All experiments start from the recalibrated 1979 endogenous-exit steady state. Panel A reports parameter ratios to 1979 values; a value of one indicates no change. The same proportional change in η_j applies to manufacturing and services. All other primitives remain fixed. Panel B reports the data and model outcomes for the two calibration moments. An asterisk indicates that the moment's 1979–2023 change is targeted in that experiment.

TABLE B.8: OUTCOMES ACROSS ENDOGENOUS-EXIT EXPERIMENTS

	1979		2023			
Outcome	Data	Model	Data	Main	A_M only	η only
Manufacturing employment share (%)	27.9	27.8	9.6	19.2	20.1	25.7
Change in manufacturing share (p.p.)	—	—	−18.4	−8.6	−7.7	−2.1
Employment share of large service firms (%)	41.6	41.3	55.3	53.3	41.3	65.1
Establishments per large service firm	18.8	18.5	32.5	32.3	18.5	32.3

Notes: Large firms have at least 500 employees. All experiments start from the same recalibrated 1979 endogenous-exit steady state. Aggregate labor supply is held fixed at its 1979 value. The main experiment changes both A_M and η_j; the two restricted experiments change one parameter at a time.

B.6 Expansion Mechanisms in Related Models

This appendix compares the predictions of our model against three recent models of firm expansion across markets, branches, and products. In each model, changes in technology or demand lead more productive firms to expand disproportionately along the extensive margin. The purpose of this appendix is to establish how each of these mechanisms maps into a fall in the convexity of replication costs. We do not claim that our model reproduces all the predictions of the detailed models developed in those papers. In order to facilitate the exposition, we consider streamlined versions of these detailed models whenever possible.

In Section 3, we considered an environment in which a firm can operate m establishments and pay the replication cost (we suppress the sector subscript j):

$$G(m) \;=\; c_m \,\frac{m^{1+\eta}-1}{1+\eta}. \tag{B.2}$$

The marginal cost of the m-th establishment is therefore $c_m m^{\eta}$, and a firm at an interior optimum operates

$$\ln m^*(z) \;=\; \frac{1}{\eta}\left[\ln \pi_v(z) - \ln c_m\right] \tag{B.3}$$

establishments, where $\pi_v(z)$ denotes operating profit per establishment.[23] A fall in the level parameter c_m shifts the intercept: the number of establishments rises by the same proportion at every multi-establishment firm. A fall in η rotates the slope: the number of establishments rises disproportionately at high-profit firms, which are also the most productive ones.

One can also derive a direct link between the number of establishments and productivity. Since operating profit per establishment in Equation (9) is a power function of productivity, Equation (B.3) can be rewritten as

$$\ln m^*(z) \;=\; \chi \;+\; \frac{1}{\eta\,(1-\alpha)}\,\ln z, \tag{B.4}$$

where the intercept collects the sectoral price, the sectoral productivity level, the wage, and the level of the replication cost:

$$\chi \;\equiv\; \frac{1}{\eta}\left[\ln(1-\alpha) + \frac{1}{1-\alpha}\ln(pA) + \frac{\alpha}{1-\alpha}\ln\frac{\alpha}{W} - \ln c_m\right]. \tag{B.5}$$

A fall in η therefore has the same effect here as in Equation (B.3). It steepens the relationship between establishments and productivity, so that the number of establishments rises at every multi-establishment firm, and it rises the most at the most productive firms. Establishing this relationship is useful for what follows. Some of the models we discuss below admit a clean characterization of expansion in terms of profit, while others admit one in terms of productivity. Since profit is a monotone function of productivity in our model, either characterization maps into the establishment policy in the same way.

We consider models of firm expansion through local markets in Hsieh and Rossi-Hansberg (2023), through branches in Brynjolfsson, Jin, Petropoulos, and Wang (2025), and through

[23] We have normalized $P = 1$ here for illustrative purposes.

products in Argente, Moreira, Oberfield, and Venkateswaran (2025). References to assumptions, propositions, and corollaries below refer to the corresponding paper.

Expansion through local markets. In Hsieh and Rossi-Hansberg (2023) (HR) a firm produces a non-traded service and can serve a continuum of local markets. Serving market n requires an establishment in that market and a per-market fixed cost f. The firm's productivity is A, which is common to all of its markets. In each market, the firm faces a constant-elasticity demand curve, with price elasticity $\sigma > 1$. From the firm's perspective, markets differ along three dimensions: the local wage w_n; a demand shifter E_n that combines local real expenditure with the local price index; and the firm's idiosyncratic local productivity a_n, which is independently and identically distributed across firms. Conditional on serving market n, the firm's operating profit is:

$$\pi_n \;=\; \bar{\sigma}\,(a_n A)^{\sigma-1}\, w_n^{1-\sigma} E_n^{\sigma}, \tag{B.6}$$

where $\bar{\sigma} \equiv (\sigma-1)^{\sigma-1}/\sigma^{\sigma}$. The firm serves market n if and only if $\pi_n \geq f$. This entry condition can be restated as a productivity cutoff:

$$A \;\geq\; \alpha_n \;\equiv\; \left(\frac{f}{\bar{\sigma}\, a_n^{\sigma-1} w_n^{1-\sigma} E_n^{\sigma}}\right)^{\frac{1}{\sigma-1}}. \tag{B.7}$$

Thus, from the firm's perspective, a market is fully described by how productive the firm must be to serve it profitably. Raising Equation (B.7) to the power $\sigma-1$ and rearranging gives $\bar{\sigma}\, a_n^{\sigma-1} w_n^{1-\sigma} E_n^{\sigma} = f\,\alpha_n^{-(\sigma-1)}$; substituting into Equation (B.6) yields:

$$\pi_n \;=\; f\left(\frac{A}{\alpha_n}\right)^{\sigma-1}. \tag{B.8}$$

Equation (B.8) shows that the cutoff summarizes how attractive a market is. The firm breaks even exactly in the market whose cutoff equals its own productivity, and it earns more in markets with lower cutoffs. It therefore serves every market with $\alpha_n \leq A$, and the measure of markets it serves is $m(A) = \Gamma(A)$, where Γ denotes the distribution of cutoffs the firm faces. HR take this distribution as given and common to all firms in their Propositions 1–3, and so do we.

We use HR's power-density example, $\gamma(\alpha) = \Omega\alpha^{\beta}/f^{\sigma-1}$ with $\beta > \sigma-1$, as stated in Appendix K7 of their 2021 working paper version (Hsieh and Rossi-Hansberg, 2021). Here $\Omega > 0$ captures the abundance of profitable markets and is determined in general equilibrium by wages and expenditures, while a larger β means that more of the available markets are hard to enter. In this special case, a firm with productivity A serves

$$m(A) \;=\; \int_0^A \gamma(\alpha)\,\mathrm{d}\alpha \;=\; \frac{\Omega}{1+\beta}\,\frac{A^{1+\beta}}{f^{\sigma-1}}$$

markets. If we order the firm's markets from best to worst, so that the m-th best market is the one with the m-th lowest cutoff, then $\alpha(m) = \left[(1+\beta)f^{\sigma-1}m/\Omega\right]^{1/(1+\beta)}$. Let $\pi_v \equiv fA^{\sigma-1}$

denote operating profit in a market whose (normalized) cutoff is one. From Equation (B.8), profit in the m-th market is $\pi_v\,\alpha(m)^{-(\sigma-1)}$, and the firm expands until the marginal market breaks even. Solving for m yields:

$$\ln m^*(A) \;=\; \frac{1}{\eta}\left[\ln \pi_v - \ln \tilde{c}_m\right] \;=\; \frac{1}{\eta}\left[\ln \pi_v - \ln f\right] \;-\; \ln\frac{(1+\beta)\,f^{\sigma-1}}{\Omega},$$

where $\eta = \frac{\sigma-1}{1+\beta}$ and $\tilde{c}_m \equiv f\left[(1+\beta)f^{\sigma-1}/\Omega\right]^{\eta}$. This expression is analogous to the establishment policy (B.3) in our paper. The convexity parameter η governs how quickly market quality decays with rank, and the level cost parameter $\tilde{c}_m$ combines the per-market fixed cost f with the abundance of markets Ω.

HR consider an experiment with the arrival of a new technology. A firm can adopt it at any intensity $h \geq 1$: adoption raises the firm-level fixed cost F—paid once, regardless of the number of markets served—from F to $h^{\kappa}F$ with $\kappa > 0$, and, in exchange, raises the firm's productivity from A to hA. HR assume $\kappa > \sigma - 1$.[24] More productive firms adopt weakly more intensively (see Proposition 2 in HR). The reason is that the cost of adoption is paid once, at the firm level, while the benefit accrues in every market the firm serves: a firm with many establishments chooses a high h, and a firm with few establishments adopts little.

Following HR's Propositions 1–3, we hold the distribution of cutoffs fixed, so β, $\tilde{c}_m$ and η are all unchanged. In general equilibrium, Ω responds to wages and expenditures; such a response would shift the intercept $\tilde{c}_m$ but leave β and η unchanged, just as the equilibrium fall in p_j discussed in Section 3.4 enters only the intercept χ in Equation (B.5) and leaves the slope of the establishment policy unchanged. The experiment only affects the profit at a normalized cutoff: it becomes $f(hA)^{\sigma-1}$ in place of $fA^{\sigma-1}$, so that

$$\ln m^*(h, A) \;=\; (1+\beta)\left[\ln h + \ln A\right] \;+\; \text{constant}.$$

Holding h fixed (at the level prior to the arrival of the new technology):

$$\frac{\mathrm{d}\ln m^*}{\mathrm{d}\ln A} \;=\; 1+\beta.$$

This is the counterpart of Equation (B.4) in our model: the number of units is log-linear in productivity, with slope $1+\beta$ in HR and $1/[\eta(1-\alpha)]$ in our model. Once the technology is available, adoption becomes a function of productivity:

$$\frac{\mathrm{d}\ln m^*}{\mathrm{d}\ln A} \;=\; \frac{\sigma-1}{\eta}\left[1 + \frac{\mathrm{d}\ln h(A)}{\mathrm{d}\ln A}\right] \;=\; (1+\beta)\left[1 + \frac{\mathrm{d}\ln h(A)}{\mathrm{d}\ln A}\right].$$

According to Proposition 2, $h(\cdot)$ is a weakly increasing function of productivity. The slope of the number of markets with respect to productivity therefore rises, which is what a fall in η delivers in Equation (B.4). HR state the implication directly: the new technologies raise the number of markets of the most productive firms relative to average firms (Proposition 3).

[24] HR denote the exponent κ by η; we rename it to avoid confusion with the convexity parameter of the replication cost. In the power-density example, firm profit grows with h as $h^{1+\beta}$, so an interior choice of h requires the stronger condition $\kappa > 1+\beta$, which HR use in their Appendix K7. We also impose it.

Thus, for the expansion of firms along the extensive margin, the arrival of the new technology in HR maps into a fall in the convexity of replication costs in our model.

Expansion through branches. Brynjolfsson, Jin, Petropoulos, and Wang (2025) (BJPW) consider an environment in which firms operate branches across a set of cities. A branch in city j draws its marginal cost at random and earns an equilibrium profit π_j^*. Operating in N cities requires an overhead cost $\Psi(N)$ as in Aghion et al. (2023). It is assumed that this function satisfies three properties: a single-city firm pays no overhead, $\Psi(1) = 0$; the increments are increasing, $\Psi(N+1) - \Psi(N) > \Psi(N) - \Psi(N-1)$; and $\Psi(N) \to \infty$ as N approaches the total number of cities. They show that under these conditions the firm expands city by city, adding city j only if the profit of a branch there exceeds the marginal overhead (see Proposition 2):

$$\pi_j^* \;>\; \Psi(N+1) - \Psi(N). \tag{B.9}$$

We consider a special form of the overhead cost, $\Psi(N) = c_m\,(N^{1+\eta}-1)/(1+\eta)$.[25] Treating N as continuous, the marginal overhead is $c_m N^\eta$, and at the margin the expansion condition (B.9) reads $\pi_j^* = c_m N^\eta$; this is our establishment policy (B.3). The two models determine profit differently. In BJPW, π_j^* is the outcome of Cournot competition and depends on the number and costs of local rivals; in this paper, firms are competitive and take prices as given. Yet both models predict that firms expand until the profit of the marginal unit equals the marginal overhead.

Information technology is the key ingredient of the BJPW model: it lets a firm replicate the business processes of its most efficient branch. A firm invests in IT to acquire a transfer ability $\varphi \in [0,1]$. The transfer ability lowers the within-firm marginal cost from its drawn value c to $c - \varphi\,(c - c^{\min})$, where $c^{\min}$ denotes the marginal cost of its most efficient branch. The investment cost is convex in φ. Higher transfer ability lowers the marginal cost of every branch (except for the most efficient one); this increases the profit of each of them. According to (B.9), cities that previously failed to cover the marginal overhead now clear it, so the firm operates more branches (Corollary 2).

Write the marginal condition $\pi_j^* = c_m N^\eta$ above in logs,

$$\ln N^* \;=\; \frac{1}{\eta}\big[\ln \pi^* - \ln c_m\big], \tag{B.10}$$

and let π_0 denote the profit the marginal branch earns before any transfer of ability. In BJPW, cities differ in demand, in the number of rivals, and in the firm's cost draw. We assume that the firm adds cities in order of profitability, so that the marginal branch is its N-th most profitable city and its pre-transfer profit falls with N. We write $\pi_0 = \theta N^{-\kappa}$ with $\kappa \geq 0$, where θ is a firm-level index of branch profitability and κ measures how quickly the profit of the marginal city declines with the number of cities; $\kappa = 0$ if cities are identical from the firm's perspective. Let T be the factor by which transfer raises the profit of the

[25] Our functional form satisfies the single-city and convexity conditions and gives the same marginal overhead condition at an interior choice of N. It does not impose divergence as N approaches the finite number of cities; this difference is immaterial when the optimal N is interior.

marginal branch, so $\pi^* = \pi_0 T$ and $T = 1$ at $\varphi = 0$. Then (B.10) reads

$$\ln N^* \;=\; \frac{1}{\eta+\kappa}\big[\ln\theta - \ln c_m\big] \;+\; \frac{1}{\eta+\kappa}\ln T. \tag{B.11}$$

At $\varphi = 0$ the last term vanishes and the number of branches responds to θ with elasticity $1/(\eta+\kappa)$. In BJPW, the convexity of expansion therefore combines the overhead cost, through η, with the declining profitability of the marginal city, through κ; the latter is the channel that governs expansion in HR.

Transfer ability lowers the marginal cost in the firm's branches, so $\partial \ln T/\partial \ln \varphi > 0$. At the optimum, the marginal investment cost equals the sum of the resulting profit gains across the firm's branches other than the most efficient one (Proposition 3). Because this sum has one term per branch, the return to transfer ability rises with the number of branches. We refer to this as the *branch-count channel*. Two further channels operate through the cost gap $c - c^{\min}$. First, holding rivals and $c^{\min}$ fixed, a branch with a lower own cost earns a higher π_0 but has less to gain from transfer, so T falls with the firm's profitability.[26] We refer to this as the *cost-gap channel*. Second, adding branches adds cost draws and can lower $c^{\min}$, which raises the gains from transfer in all other branches. We refer to this as the *best-branch channel*; it works in the same direction as the branch-count channel. We write $T = T(\varphi, \theta)$.

In our streamlined version, we assume that the firm knows the marginal costs of its branches and chooses φ and N jointly, so that $\varphi = \varphi(N^*, \theta)$. BJPW instead assume that the firm chooses φ before it observes these costs and before it chooses N; we return to this timing below. Under our assumption, a firm operating more branches has a larger return to transfer ability and chooses a higher φ: $\partial \ln \varphi/\partial \ln N^* > 0$. Branch profit therefore rises with the number of branches the firm operates:

$$b \;\equiv\; \frac{\mathrm{d}\ln T}{\mathrm{d}\ln N^*} \;=\; \frac{\partial \ln T}{\partial \ln \varphi}\,\frac{\partial \ln \varphi}{\partial \ln N^*} \;>\; 0.$$

Let a denote the elasticity of T with respect to θ holding N^* fixed; the cost-gap channel makes $a < 0$. Because N^* enters the right-hand side of (B.11) through T, differentiating with respect to $\ln\theta$ yields

$$(\eta+\kappa)\,\frac{\mathrm{d}\ln N^*}{\mathrm{d}\ln\theta} \;=\; 1 \;+\; a \;+\; b\,\frac{\mathrm{d}\ln N^*}{\mathrm{d}\ln\theta}, \qquad \text{so} \qquad \frac{\mathrm{d}\ln N^*}{\mathrm{d}\ln\theta} \;=\; \frac{1+a}{\eta+\kappa-b}.$$

The denominator is positive under the strict second-order condition for the firm's choice of N, $\eta + \kappa > b$: the marginal overhead must rise faster than the marginal profit as the firm adds branches. This elasticity exceeds its zero-investment value $1/(\eta+\kappa)$ if and only if $b + (\eta+\kappa)\,a > 0$, that is, if the branch-count channel dominates the cost-gap channel. The model does not determine the sign of this condition without further structure on demand and cost draws across cities. We assume that the branch-count channel dominates; under this assumption, the investment in transfer ability increases the slope between the number

[26] From BJPW's equations (4) and (5), the profit of the marginal branch is $b_j(q^*)^2$ with $q^* = \big(A_j + \sum_{y\neq i}\tilde{c}_y - n_j\tilde{c}\big)/\big(b_j(n_j+1)\big)$. With $\tilde{c} = c - \varphi(c - c^{\min})$, the multiplier is $T = \big[1 + n_j\varphi\,(c - c^{\min})/(A_j + \sum_{y\neq i}\tilde{c}_y - n_j c)\big]^2$, which is increasing in the branch's own cost c given the rivals' costs and $c^{\min}$.

of branches and profitability.

We now return to BJPW's timing, whereby the firm chooses φ before it draws its costs. Investment becomes a function of the number of branches the firm expects to operate. In BJPW, firms are ex ante identical except for their investment efficiency ξ. To obtain a cross-section in profitability, we let them also differ in θ, which they know when they invest. Proposition 3 expresses the first-order condition for φ in terms of the realized number of branches; with φ chosen in advance, the firm instead sums the expected profit gains over the branches it expects to operate. A more profitable firm expects more branches, so the sum has more terms and φ rises with θ: this is the branch-count channel. The cost-gap channel again works in the opposite direction. Writing $\varphi = \varphi(\theta)$ and $T = T(\varphi(\theta), \theta)$, Equation (B.11) yields

$$\frac{\mathrm{d}\ln N^*}{\mathrm{d}\ln\theta} = \frac{1+\tilde{b}}{\eta+\kappa}, \qquad \tilde{b} \equiv \frac{\mathrm{d}\ln T}{\mathrm{d}\ln\theta} = \frac{\partial \ln T}{\partial \ln \varphi}\frac{\mathrm{d}\ln\varphi}{\mathrm{d}\ln\theta} + \frac{\partial \ln T}{\partial \ln \theta}.$$

The first term in $\tilde{b}$ is the branch-count channel and is positive; the second term is the cost-gap channel and is negative. The elasticity exceeds $1/(\eta+\kappa)$ if and only if $\tilde{b} > 0$, that is, if the branch-count channel dominates, the same condition as before. Therefore, under either timing, investment in transfer ability steepens the relationship between the number of branches and profitability when the branch-count channel dominates the cost-gap channel. Thus, for the expansion of firms along the extensive margin, IT investment in BJPW maps into a fall in the convexity of replication costs in our model.

Expansion through products. Argente, Moreira, Oberfield, and Venkateswaran (2025) (AMOV) consider an environment in which a firm produces a continuum of products of measure N. In what follows, we use their tractable case with $\theta = \varepsilon = 2$ and $F(N) = FN$. The signs we derive below require the elasticity of substitution between the two components of expertise to satisfy $1 < \sigma < 2$.

Each product requires expertise, which has two components: a *scalable* component x, which is firm-wide and shared by all products, and a *local* component y, which must be acquired separately for each product. Products are ex-ante identical, so at the optimum the firm acquires the same local component y for every product. There is a capacity constraint, which can be interpreted as limited managerial attention at the firm level:

$$x + Ny \leq 1. \tag{B.12}$$

Expertise per product $Z(x, y)$ is a CES combination of the two components with elasticity of substitution σ, where $0 < \sigma < 2$ (Assumption 1); this makes $Z(\cdot)$ homogeneous of degree one. The firm sells each product subject to a CES demand, and profit is linear in expertise in the special case we are considering:

$$\Pi = G \cdot A \cdot N \cdot Z(x, y) - F \cdot N, \tag{B.13}$$

where A denotes the firm's productivity, G is a common, sector-wide equilibrium coefficient that scales profits, and F is the fixed operating cost per product. Let $k \equiv x/y$ denote the firm's scalability ratio and $z(k) \equiv Z(k, 1)$; homogeneity of degree one implies $Z(x, y) = y\, z(k)$. The capacity constraint binds at the optimum, so $y = 1/(k+N)$. Expertise per

product is then $z(k)/(k+N)$, and the firm's problem can be rewritten as:

$$\max_{N,k} \; G \cdot A \frac{N\, z(k)}{k+N} \; - \; F \cdot N. \tag{B.14}$$

Equation (B.14) embeds a rising cost of scope. To see this, hold the scalability ratio k fixed and consider adding one more product. Revenue in Equation (B.14) is GA times expertise per product, summed over N products, so the new product earns $GA\, z(k)/(k+N)$. But it also tightens the capacity constraint: $y = 1/(k+N)$ falls, so every product the firm already operates loses some local expertise, and with it some revenue. Standardization alleviates this issue: the scalable component is acquired once, so only the local component must be replicated.

Two first-order conditions characterize the solution. Differentiating Equation (B.14) with respect to k yields:

$$\frac{k\, z'(k)}{z(k)} \; = \; \frac{k}{k+N} \; \equiv \; S. \tag{B.15}$$

The left-hand side is the elasticity of expertise with respect to the scalability ratio. The right-hand side is the share of capacity accounted for by the scalable component (scalable share). Differentiating Equation (B.14) with respect to N yields:

$$G \cdot A \frac{z(k)}{k+N} S \; = \; F. \tag{B.16}$$

In the case of the CES aggregator, the elasticity on the left-hand side of Equation (B.15) is $k^{\frac{\sigma-1}{\sigma}} \big/ \big(1 + k^{\frac{\sigma-1}{\sigma}}\big)$. Substituting this expression into Equation (B.15) gives $k^{\frac{\sigma-1}{\sigma}} = k/N$, so that $\ln k = \sigma \ln N$: a firm that widens its scope also standardizes more. Next, take logs of Equation (B.16) and differentiate with respect to $\ln A$, using $kz'(k)/z(k) = S$ and $\mathrm{d}\ln k = \sigma\, \mathrm{d}\ln N$. Collecting terms yields:

$$\frac{\mathrm{d}\ln N}{\mathrm{d}\ln A} \; = \; \frac{1}{(2-\sigma)\,(1-S)}. \tag{B.17}$$

Equation (B.17) is the counterpart of Equation (B.4) in our model, which yields:

$$\frac{\mathrm{d}\ln m}{\mathrm{d}\ln z} \; = \; \frac{1}{\eta\,(1-\alpha)}. \tag{B.18}$$

In our model, Equation (B.4) is linear in $\ln z$ with a slope of $1/[\eta(1-\alpha)]$. In AMOV, the elasticity in Equation (B.17) equals $1/(2-\sigma)$ for a firm with a negligible scalable share and rises with S. Since S rises with productivity (Proposition 1), $\ln N$ is strictly convex in $\ln A$. AMOV emphasize the convexity that this amplification induces in firm size: their Proposition 6 states that revenue is strictly convex in productivity. The policy has a closed-form expression. Substituting $k = N^{\sigma}$ into Equation (B.16) and using $S/(1-S) = N^{\sigma-1}$,

the condition becomes $GA\,S^{-(2-\sigma)/(\sigma-1)} = F$, so that

$$S \;=\; \left(\frac{GA}{F}\right)^{\frac{\sigma-1}{2-\sigma}} \qquad \text{and} \qquad \ln N \;=\; \frac{1}{2-\sigma}\ln\frac{GA}{F} \;-\; \frac{1}{\sigma-1}\ln(1-S).$$

The first term is log-linear in productivity with slope $1/(2-\sigma)$. The second term is the amplification from standardization: it rises with S and makes $\ln N$ convex in $\ln A$.[27] The two models are therefore compared based on local elasticities. In both cases, the elasticity of the number of units with respect to productivity is the inverse of a curvature term: $(2-\sigma)(1-S)$ in AMOV and $\eta(1-\alpha)$ in our model. In AMOV, the curvature is lower for firms with a higher scalable share, because adding a product drains less expertise from existing products, so it differs across firms. In our model, it is lower when the replication cost is less convex, and it is common to all firms in a sector.

AMOV use the model to study a common rise in demand G. We show that this change manifests itself as an increase in the elasticity in Equation (B.17).

First, demand widens scope. In (B.13), demand and productivity enter only through their product, thus:

$$\frac{\mathrm{d}\ln N}{\mathrm{d}\ln G} \;=\; \frac{\mathrm{d}\ln N}{\mathrm{d}\ln A} \;=\; \frac{1}{(2-\sigma)(1-S)} \;>\; 0.$$

Second, wider scope raises the scalable share. Since $\ln k = \sigma \ln N$ and $S = k/(k+N)$, we have $S/(1-S) = k/N = N^{\sigma-1}$ and

$$\frac{\mathrm{d}\ln S}{\mathrm{d}\ln N} \;=\; (1-S)(\sigma-1) \;>\; 0 \qquad \text{when} \qquad \sigma > 1.$$

Third, a higher scalable share raises the elasticity of scope with respect to demand:

$$\frac{\mathrm{d}}{\mathrm{d}\ln S}\ln\left[\frac{1}{(2-\sigma)(1-S)}\right] \;=\; \frac{S}{1-S} \;>\; 0.$$

Combining the three derivatives, we obtain:

$$\frac{\mathrm{d}}{\mathrm{d}\ln G}\ln\left[\frac{1}{(2-\sigma)(1-S)}\right] \;=\; \frac{S}{1-S}\cdot(1-S)(\sigma-1)\cdot\frac{1}{(2-\sigma)(1-S)} \;=\; \frac{S\,(\sigma-1)}{(2-\sigma)(1-S)} \;>\; 0.$$

The most productive firms hold the highest scalable shares (Proposition 1). Their elasticity of scope with respect to productivity is therefore the highest to begin with (Proposition 2). The last expression shows that it also rises the most when demand rises. A rise in demand thus makes scope more responsive to productivity, especially at the top. A fall in η in Equation (B.4) raises the elasticity of establishments with respect to productivity by the same proportion at every firm. Both changes therefore raise the elasticity of the number of units with respect to productivity, and in both the proportional expansion in scope is largest at the most productive firms. A single sector-wide η captures the common rise in the elasticity but not its heterogeneity across firms. Thus, for the expansion of firms along the

[27] As GA/F approaches one, S approaches one and scope diverges. AMOV restrict parameters so that this corner case is not reached.

extensive margin, a rise in demand in AMOV maps into a fall in the convexity of replication costs in our model.