Firm Growth through New Establishments*

Dan Cao
Georgetown University

Henry R. Hyatt
U.S. Census Bureau

Toshihiko Mukoyama†
Georgetown University

Erick Sager
Federal Reserve Board

February 2022

Abstract

This paper analyzes firm-level employment along the extensive margin (the number of establishments in a firm) and the intensive margin (the number of workers per establishment in a firm). Utilizing administrative datasets, we document that the firm-size distribution, and both extensive and intensive margins, exhibit fat tails; and the growth in average firm size between 1990-2014 was primarily driven by an expansion along the extensive margin, particularly in very large firms. We develop a tractable general equilibrium growth model with external innovation that leads to extensive-margin firm growth, and internal innovation that leads to intensive-margin growth. The model generates fat-tailed distributions of firm size, establishment size, and the number of establishments per firm. We estimate the model to uncover the fundamental forces that caused the distributional changes from 1995-2014 and highlight the importance of declining external innovation costs, establishment exit rates, and aggregate productivity growth rate.

Keywords: firm growth, firm-size distribution, establishment, innovation
JEL Classifications: E24, J21, L11, O31

*First version: February 2017. We are grateful to Ufuk Akcigit, Sina Ates, Andy Atkeson, Robert Axtell, Nick Bloom, Stephane Bonhomme, John Earle, Jeremy Greenwood, Hubert Janicki, Sam Kortum, Makoto Nirei, David Ratner, Margit Reischer, Immo Schott, John Shea, and Mike Siemer for helpful comments and suggestions. We also thank seminar participants at Alberta, ASU, Bank of Portugal, FRB Atlanta, George Mason, Hitotsubashi, IMF, Loughborough, NC State, Pontificia Universidad Católica de Chile, Queensland, UConn, UCSD, UIUC, University of Maryland, University of Tokyo, Uppsala, UT Austin, Warwick, and conference participants at ASSA, Barcelona GSE, CICM, GCER, KER Conference, Midwest Macro at U. Wisconsin, SAET, SCE, and SED in St Louis. We thank Jess Helfand and Mike LoBue for help with the QCEW. All results using restricted access microdata from the Bureau of Labor Statistics and Census Bureau have been reviewed to ensure no confidential data are disclosed. Results using Census Bureau microdata have been approved by the Disclosure Review Board with authorization number DRB-B0063-CED-20190628. The views expressed within are those of the authors and do not necessarily reflect those of the Federal Reserve System, Census Bureau, Bureau of Labor Statistics or the U.S. government. All errors are our own.

†Corresponding author: tm1309@georgetown.edu

1 Introduction

Understanding the process of firm growth is essential in the analysis of macroeconomic performance. Firms that innovate and expand are the driving force of output and productivity growth. Recent studies of macroeconomic productivity emphasize the role of innovation and reallocation at the firm level, from both the theoretical and empirical standpoint.

In this paper, we focus on a particular aspect of firm growth: growth through adding new establishments. A firm can increase its size along two margins; it can add more workers to its existing establishments or it can build new establishments. We call the former the *intensive margin* of firm growth and the latter the *extensive margin* of firm growth. This distinction is important because these margins typically imply different reasons for expansion. A new manufacturing plant is often built to produce a new product. In the service sector, building a new store or a new restaurant implies venturing into a new geographical market. Creating a new establishment is also different in that it typically requires a significant amount of investment in equipment and structures. In the following, we characterize the firm size distribution through these extensive and intensive margins, both empirically and theoretically.

Empirically, we find that the two margins (the extensive margin and the intensive margin) of the firm-size distribution, in addition to the entire firm size, exhibit Pareto tails in the U.S. economy. Along the time series, we find that the average firm size has grown in recent years, as is consistent with the findings of recent studies.[1] These studies suggest that this increase in size has had important implications on other changes in macroeconomic variables. We find that this expansion is driven by the extensive margin growth: firms are growing by adding new establishments.

To investigate what changes in the economic environment have contributed to this phenomenon, we build a macroeconomic model of endogenous firm growth. Our model extends previous work by Klette and Kortum (2004) and Luttmer (2011). In these papers, each individual firm grows by adding production units ("product lines"

[1] Autor et al. (2020) document a series of empirical patterns on the emergence of superstar firms, while De Loecker et al. (2020) document increasing market power (measured as markups) of large publicly-listed firms, and Gutiérrez and Philippon (2017) note the recent increase in concentration in the US corporate sector.

in Klette and Kortum (2004) and "blueprints" in Luttmer (2011)). In our model, we call this margin of firm growth *external innovation*. These production units can be naturally interpreted as establishments. The major departure of our model, compared to these two papers, is the recognition that each establishment can grow in size. We introduce this technological improvement at the establishment level, which we call *internal innovation*, and explicitly compare our model outcomes with the data on establishments. We find that the model can replicate the three Pareto tails (the firm size distribution, the extensive margin, and the intensive margin), and we are able to characterize the thickness of the tails both analytically and quantitatively.

Utilizing our model, we investigate the cause of the U.S. firm-size distribution over the recent years. We estimate model parameters to match U.S. firm-size distributions in 1995 and 2014. The difference in estimated parameter values uncovers the fundamental forces that have caused the changes in the firm size distribution and its components over this time period. We find that the largest contributors to the increase in the number of establishments per firm are the decline in innovation cost for adding new products (establishments), the decline in the establishment exit rates, and the slowdown of the aggregate productivity growth.

In the recent theoretical literature, our work shares the most similarities with Akcigit and Kerr (2018). The two papers differ in how the model is mapped to the data (we focus on establishment-level employment whereas Akcigit and Kerr (2018) look at patent data). More importantly, Akcigit and Kerr's (2018) model does not allow for Pareto tails, which is an important feature in our data. On this point, our paper is also related to the theoretical literature studying the firm-size distribution and its fat right tail. Luttmer (2011) and Acemoglu and Cao (2015) build endogenous growth models with fat-tailed distributions arising from only extensive- or intensive-margin growth. Our model features heterogeneity in two dimensions: the extensive margin and the intensive margin. Moreover, we analytically characterize the Pareto tail of firm size distribution in the presence of the (endogenous) two-dimensional heterogeneity, despite the challenge that these two dimensions are potentially *correlated*.[2] This

[2]The characterization is technically challenging and we are able to obtain analytical results on the Pareto-tail index of the distribution by employing recent advances in the literature on regular variations and their applications (see, for example, Bingham et al. (1987), Mimica (2016), and Gabaix

novel characterizations can open up various applications beyond the firm-dynamics literature.[3]

Our paper is also related to the emerging (and concurrent) literature on differentiating firms and establishments (or plants) in which firms are collections of establishments, along with Aghion et al. (2019) and Hsieh and Rossi-Hansberg (2019), as well as Fattal Jaef (2018), Atalay et al. (2019), Gumpert et al. (2019), Kehrig and Vincent (2019), Oberfield et al. (2020), and Moreira et al. (2021). While commonly emphasizing the distinction between firms and establishments, these papers focus on different substantive issues from ours (such as labor share, firm organization, and misallocation) and thus do not put emphasis on the quantitative features of the extensive-margin distribution as in our paper. For example, Kehrig and Vincent (2019) and Gumpert et al. (2019) only allow for up to two establishments per firm. Our paper is unique in that we empirically document the fat (Pareto) tail of the establishment number distribution and our model allows us to match this distribution almost exactly including its Pareto tail.

The paper is organized as follows. Section 2 describes the empirical patterns of firm growth in our dataset. Section 3 sets up the model. Section 4 provides a characterization of the model's stationary firm-size distributions, including extensive- and intensive-margin distributions. Section 5 estimates model parameters and uses the model to perform a quantitative decomposition of 1995-2014 U.S. firm growth. Section 6 concludes.

2 Empirical facts

In this section, we first describe our data and empirical framework to decompose firm size into intensive and extensive margins. We then use the decomposition to docu-

et al. (2016)). Perhaps due to this challenge, firm size distributions have not been characterized analytically in the literature using models with both dimensions of heterogeneity such as Akcigit and Kerr (2018). One exception is Peters (2019), however his model's firm size distribution in sales is exactly the same as the distribution of establishment numbers due to log utility.

[3] Pareto tails are also important in the empirical and theoretical analyses of income and wealth inequality as surveyed in Atkinson et al. (2011) and more recently, Gabaix et al. (2016), Cao and Luo (2017), and Jones and Kim (2018).

ment cross-section and time series facts. Additional empirical results are provided in Appendix B.

2.1 Data

This paper utilizes two restricted access datasets, the Quarterly Census of Employment and Wages (QCEW) and the Longitudinal Employer-Household Dynamics (LEHD), that use common source data that contains a near census of establishments in the U.S. The source data are collected for the QCEW by U.S. states in partnership with the Bureau of Labor Statistics (BLS) for the official administration of state unemployment insurance programs. States also provide establishment-level QCEW microdata to the U.S. Census Bureau's LEHD program as part of the Local Employment Dynamics federal-state partnership.[4] Appendix A details further information of the datasets.

We use the employer identification number (EIN) as the definition of the firm. Song et al. (2018) discuss issues surrounding the use of EIN at length in their study of inequality. One potential concern with using EINs as a firm identifier is the possibility that firms could switch EINs for either accounting reasons or merger activity. Where available, we augment the EIN with auxiliary information provided by the BLS and Census Bureau that control for inconsistencies over time. Changing identifiers does not affect our measurement of distributions, however, as cross sectional statistics do not require repeated use of any EIN over time.

2.2 Conceptual framework for the firm-size decomposition

We define *firm size* as the total number of workers employed by a firm. In what follows, we decompose average firm size into the *extensive margin* and the *intensive margin*. The extensive margin is the total number of establishments owned by a firm, and the intensive margin is the average size of establishments owned by a firm (e.g., the number of workers per establishment in a firm).

[4]Data from the BLS contains information on 38 states, and data from the U.S. Census Bureau contains information on 28 states. For each figure and table, we explicitly indicate the data source that is used.

For the sake of exposition, suppose F_t firms exist at time t, indexed by j. Let N_{jt} be the total number of workers employed by firm j and let E_{jt} be the total number of establishments owned and operated by firm j at time t. We define average establishment size within firm j as N_{jt}/E_{jt} so that we decompose size at the firm-level as

$$N_{jt} = \left(\frac{N_{jt}}{E_{jt}}\right) \times E_{jt}.$$

Accordingly, our measures of average firm size, the average intensive margin and the average extensive margin are

$$\left(\frac{1}{F_t}\sum_{j=1}^{F_t} N_{jt}, \frac{1}{F_t}\sum_{j=1}^{F_t} \frac{N_{jt}}{E_{jt}}, \frac{1}{F_t}\sum_{j=1}^{F_t} E_{jt}\right),$$

respectively.

Publicly-available datasets such as Business Dynamics Statistics of the U.S. Census Bureau contain the size distributions of establishments and firms separately. Compared to studies that utilize these publicly-available datasets, our study has several advantages. First, the use of microdata enables us to characterize the entire size distribution of firms, particularly at the right tail. Second, one cannot decompose firm growth into intensive and extensive margins without the information contained in the microdata that we utilize.

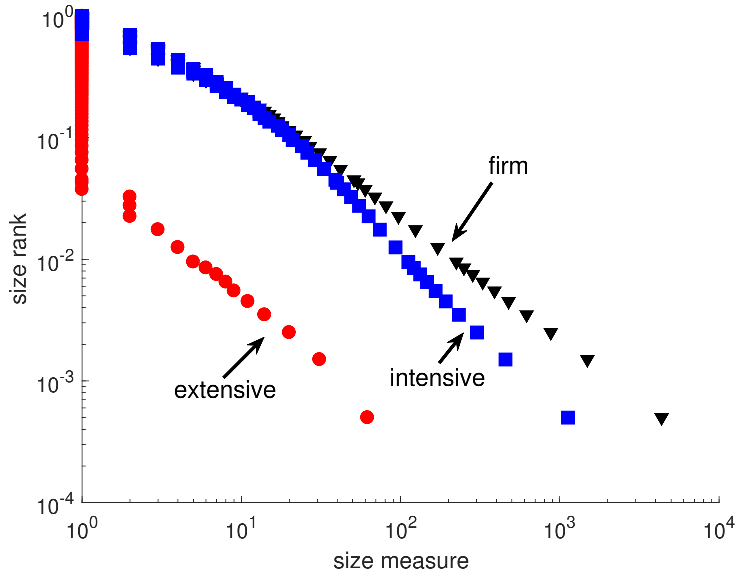
2.3 Cross-sectional properties

We first describe the cross-sectional distributions of firm size, the intensive margin, and the extensive margin. For this analysis, we use LEHD microdata and focus on 2005.

Figure 1 plots the complementary cumulative distribution function in log-log scale, a type of figure commonly used in the literature (Axtell, 2001; Gabaix, 2009) to demonstrate whether the data are consistent with Pareto's Law.[5] All three series

[5]There are, by construction, very few firms at the upper tail of the firm rank distributions we consider. To limit the disclosure risk associated with the release of these data, we rely on polynomial approximation of the underlying distributions to construct Figure 1, see Online Appendix C for details. For the set of polynomial estimates, see Online Appendix Table A.2.

Figure 1: Size-rank relationships, ranked separately by size measure



Notes: Author's calculations of LEHD microdata. Percentile ranks for the y-axis can be recovered by multiplying the size rank by 100, and note that the lowest ranked firm is assigned the value $10^0 = 1$. To limit the disclosure risk, each series shows the predicted value of employment from a regression of the log size measure on a fifth order polynomial of the log percentile rank of the size measure in March of 2005, rounded to the nearest integer. See Appendix C.2 for additional details.

have a right tail that can be approximated by a straight line. This fact implies all three distributions have Pareto tails.[6] To our knowledge, out paper is the first in the literature to document that both the extensive margin and intensive margin have Pareto tails.

[6] While Kondo et al. (2019) show that a log normal distribution can obtain a good fit in the upper tail of the distribution over total number of workers within a firm, their results do not reject the fit of a Pareto in the far upper tail. In fact, they find that distributions that feature asymptotic Pareto tails (e.g., convolutions of Pareto and log-normal distributions) best fit the entire firm size distribution, and such distributions arise endogenously in our model. In addition, Kondo et al. (2019) do not look at the intensive and extensive margin distributions that we consider here.

2.4 Time-series properties

Now we turn to the time-series changes in these distributions. We find notable changes in these distributions over our sample period. As depicted in Figure 2(a), average firm size increased from about 23 employees to over 25 employees since 1990. This fact is consistent with the rise in concentration in the U.S. economy documented by Autor et al. (2020).[7]

Figures 2(b) and 2(c) present the novel facts that we focus on in this paper. Figure 2(b) plots the average of the intensive margin and shows the intensive margin remained stationary (or somewhat declining) despite the increase in firm size over our sample period.[8] Therefore, our decomposition of firm size implies the average extensive margin must exhibit a strong upward trend, as is confirmed in Figure 2(c). The extensive margin grew from 1.2 in 1990 to over 1.5 in 2014. Accordingly, average firm size over 1990-2014 can be accounted for by the extensive-margin growth in the number of establishments.[9]

This contrasting behavior between the intensive and extensive margins implies different forces are at work for these different components of firm growth.[10] To inves-

[7]Choi and Spletzer (2012) and Hathaway and Litan (2014) also document trends in firm size and establishment size using publicly available data, but do not analyze the intensive and extensive margins which require restricted access microdata to construct. In particular, average establishment size, $\sum_{j=1}^{F_t} \frac{N_{jt}}{E_{jt}} \cdot E_{jt} / \sum_{j=1}^{F_t} E_{jt}$, can be calculated as the raio of total employment to the total number of establishments using publicly available data. In contrast, the average intensive margin across firms, $\frac{1}{F_t} \sum_{j=1}^{F_t} \frac{N_{jt}}{E_{jt}}$, requires restricted access microdata to calculate each firm's number of workers per establishment.

[8]Recent papers by Rinz (2018), Rossi-Hansberg et al. (2021), and Hershbein et al. (2019) document the diverging trends in national concentration and local concentration, which is analogous to the diverging trends in the average firm size and the intensive margin. Note, once again, that the average intensive margin is conceptually different from the average establishment size in the economy. One can compute the average establishment size from publicly-available data, but one needs to access the micro-level data to compute the average intensive margin.

[9]We have cross-checked the trends in the firm size and the number of establishments per firm in an alternative administrative (and publicly-available) dataset, Business Dynamics Statistics (BDS) at the Census Bureau. The underlying data for BDS comes from different sources from QCEW. We find that similar trends exist in BDS, but the magnitudes of the changes are smaller than in QCEW. Investigating the source of the quantitative discrepancy is an important future research topic.

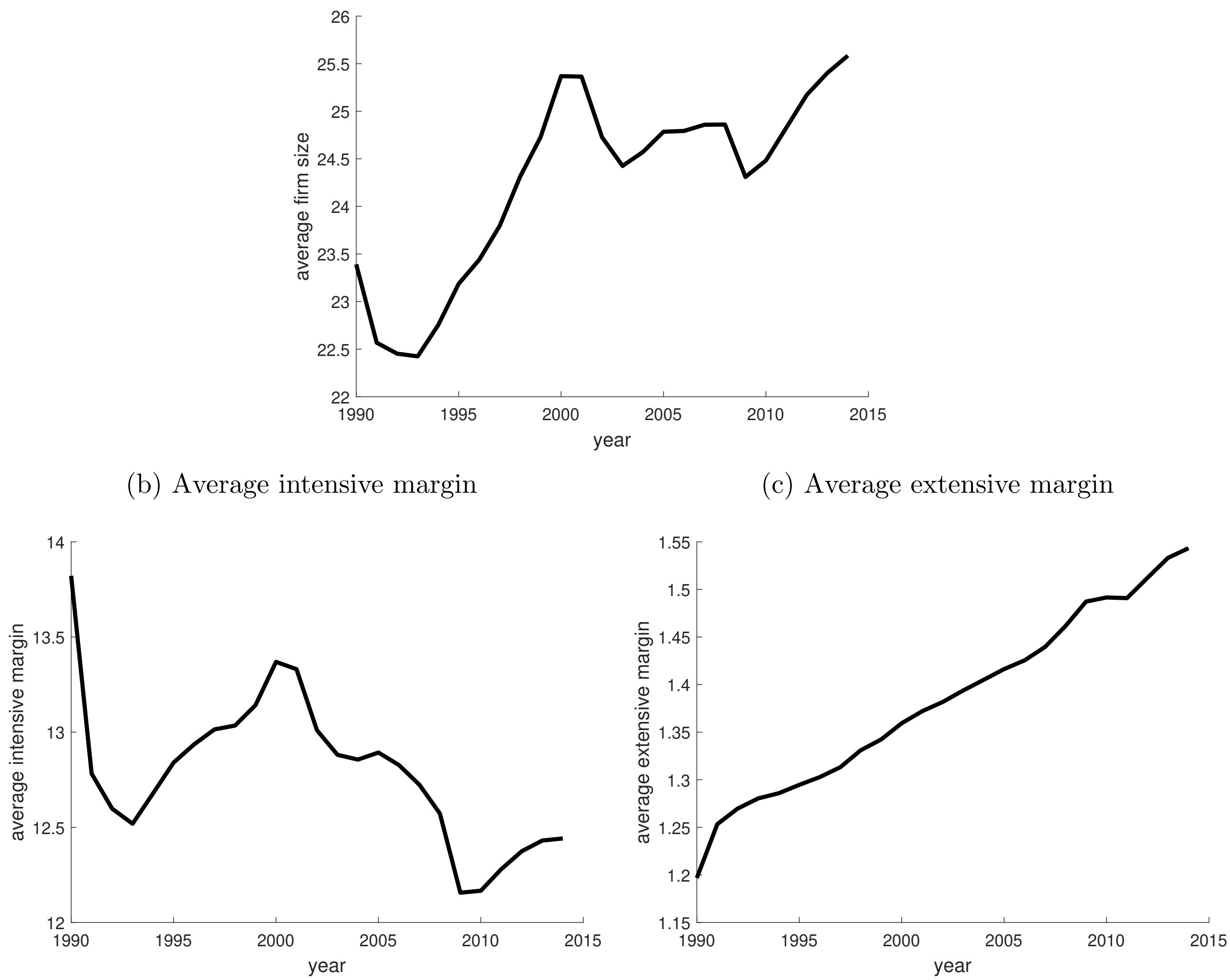
[10]A recent paper by Argente et al. (2019) document that the sales of the individual products decline over time, and emphasize the introduction of new products as a source of the firm sales growth. The expansion of firms by adding establishments are also analyzed in the industrial organization literature, such as Holmes (2011), and international trade literature, such as Garetto et al. (2019).

Figure 2: Time-series changes in average firm size, intensive and extensive margins

(a) Average firm size (number of workers)

(b) Average intensive margin

(c) Average extensive margin



Source: Author's calculations of Quarterly Census of Employment and Wages microdata.

tigate what drives the increase in firm size, in particular along the extensive margin, we consider disaggregations by sector and size bins in Appendix B. Overall, the preceding empirical documentation of firm growth shows the growth in average firm size between 1990 and 2014 is the result of high growth in creating new establishments, particularly by very large firms and firms in the service sector. Appendix B contains the estimation of the Pareto tail index. At the right tail, we observe increasing concentration in the firm-size distribution and the extensive-margin distribution between 1995 and 2014.[11]

3 Model

Motivated by the facts in the previous section, we construct a model of firm growth that matches the empirical patterns in Figure 1. In Section 5, we use this model to quantitatively analyze the fundamental causes of the growth in firm size along the extensive margin and the increasing concentration of the firm-size distribution.

3.1 Model setting

Time is continuous. The representative consumer provides labor and consumes a final good. The final good is produced by combining differentiated intermediate goods.

3.2 Representative consumer

The utility function of the representative households is

$$U = \int_0^\infty e^{-\tilde{\rho}t} L(t) u(C(t)/L(t)) dt,$$

where $u(C(t)/L(t)) = (C(t)/L(t))^{1-\sigma}/(1-\sigma)$ for $\sigma > 0$ and $\sigma \neq 1$ or $u(C(t)/L(t)) = \log(C(t)/L(t))$, corresponding to $\sigma = 1$. The consumer consumes, owns firms, and supplies labor inelastically. The labor supply is given exogenously and grows at the rate $\gamma \geq 0$. Denoting the real interest rate as r, the consumer's Euler equation is

$$\frac{\dot{C}(t)}{C(t)} = \frac{r - \rho}{\sigma}, \tag{1}$$

[11]The comparison in Table A.1, as well as the model estimation in Section 5, is between 1995 and 2014, instead of between 1990 and 2014, is due to the availability of the LEHD data.

where $\rho \equiv \tilde{\rho} - \gamma$. Final-good output $Y(t)$ is used for consumption, firm investments in innovative activities $R(t)$, and firm fixed entry costs $E(t)$, such that $Y(t) = C(t) + R(t) + E(t)$.

3.3 Final-good producers

The final-good sector is perfectly competitive. The final good is produced from differentiated intermediate goods. Intermediate goods have different qualities, and a high-quality intermediate good contributes more to the final-good production. The production function for the final good is

$$Y(t) = \left(\int_{\mathcal{N}(t)} q_i(t)^{\beta} x_i(t)^{1-\beta} di \right)^{\frac{1}{1-\beta}}, \tag{2}$$

where $x_i(t)$ is the quantity of intermediate good i, and $q_i(t)$ is its quality. $\mathcal{N}(t)$ is the set of actively-produced intermediate goods and $N(t) = |\mathcal{N}(t)|$ denotes the number of actively-produced intermediate goods. We assume $\beta \in (0, 1)$ so that the elasticity of substitution between differentiated goods $(1/\beta)$ is greater than one.

With the maximization problem

$$\max_{x_i(t)} \left(\int_{\mathcal{N}(t)} q_i(t)^{\beta} x_i(t)^{1-\beta} di \right)^{\frac{1}{1-\beta}} - \int_{\mathcal{N}(t)} p_i(t) x_i(t) di,$$

the inverse demand function for the intermediate good i is

$$p_i(t) = Y(t)^{\beta} \left(\frac{q_i(t)}{x_i(t)} \right)^{\beta}. \tag{3}$$

3.4 Intermediate-good producers

Production: The intermediate-good sector is monopolistically competitive. Each intermediate good is produced by one firm. A firm can potentially produce many intermediate goods. Below, especially when we compare the model to the data, we interpret one good as one establishment. In the existing literature, Luttmer (2011) and Acemoglu and Cao (2015) explicitly discuss this interpretation. This interpretation is particularly relevant for service sector firms, as a service establishment (such as a retail store) in one location and another establishment in another location provide different products (services) from the perspective of an Arrow-Debreu commodity space.[12]

A firm can add a new intermediate good (establishment) to its portfolio by investing in R&D that generates an *external innovation*. It can also increase the quality of the intermediate goods that it already produces by investing in R&D that generates an *internal innovation*. A new firm can enter the market by innovating its first product.[13]

We assume the intermediate goods are produced only by labor. This production process is the only place in the entire economy that uses labor as an input, thus allowing us to map the employment dynamics of the intermediate-good sector to our data analysis in Section 2.[14] The production function for intermediate good i is

$$x_i(t) = A(t)\ell_i(t), \tag{4}$$

where $A(t)$ is exogenous labor productivity that grows at rate θ.

Given the final-good producer's demand for its output (3), the intermediate goods producer's profit maximization results in standard optimal pricing with a markup over marginal cost:

$$p_i(t) = \frac{1}{1-\beta}\frac{w(t)}{A(t)}.$$

The optimal price, together with (3) and (4), implies that labor demand conditional

[12] In manufacturing, it is still an open question how products and establishments corresponds with each other. Bernard et al. (2010) find that 85% of product switching occurs within a plant, but this evidence does not necessarily contradict with our assumption as they consider a relatively narrow product category (SIC 5 digits). Also note that the existence of product switching itself does not affect our estimation directly, as long as the number of products within a plant is stable, because our estimation relies on the cross-sectional information.

[13] We refer to the two margins of firm growth as types of innovations, however this language is borrowed from the endogenous growth literature and is not necessary in our model environment. As we show below, the firm faces a convex adjustment cost of "innovation" along either margin of growth, which is isomorphic to a standard model of investment. We will discipline these model parameters using employment data instead of data on R&D (see section 5).

[14] A similar idea of mapping the employment process to the productivity process is employed by Hopenhayn and Rogerson (1993), Garca-Macia et al. (2019), and Mukoyama and Osotimehin (2019).

on aggregate variables is proportional to $q_i(t)$

$$\ell_i(t) = \frac{(1-\beta)^{\frac{1}{\beta}}\bar{w}(t)^{-\frac{1}{\beta}}}{A(t)Y(t)^{\frac{\beta}{1-\beta}}} q_i(t), \tag{5}$$

and is decreasing in the normalized wage defined by $\bar{w}(t) \equiv w(t)/(A(t)Y(t)^{\frac{\beta}{1-\beta}})$.

Because firms (and establishments) differ only in their level of quality $q_i(t)$, employment per establishment varies proportionally to $q_i(t)$ in the cross section. Similarly, profit is also proportional to $q_i(t)$:

$$\pi_i(t) = \underbrace{\left(\beta(1-\beta)^{\frac{1-\beta}{\beta}}\bar{w}(t)^{\frac{\beta-1}{\beta}}\right)}_{\equiv\bar{\pi}(t)} q_i(t). \tag{6}$$

Innovation: Innovations are carried out through R&D activity. Final goods are used as an input for R&D. For an existing intermediate-good firm, two kinds of innovations are possible: internal innovation and external innovation. We first note that the assumptions we make on innovations imply that all goods (establishments) within a firm has the same quality q. Therefore we will index the quality of each good with the firm index j.

Internal innovation raises the quality of the goods that a firm already produces. The firm-level total intensity of internal innovation is denoted by $Z_{I,j}(t)$ for firm j. The innovation intensity per good is $z_{I,j}(t) \equiv Z_{I,j}(t)/n_j(t)$, where $n_j(t)$ is the (discrete) number of goods firm j produces. We assume that the quality of each good in firm j improves according to the law of motion:

$$\frac{dq_j(t)}{dt} = z_{I,j}(t)q_j(t). \tag{7}$$

An implicit assumption here is that the total intensity $Z_{I,j}(t)$ contributes equally to the improvement of each good (and thus we divide by n_j in constructing the intensity for each good $z_{I,j}(t)$). As a consequence, the growth rate of the quality of each good is equal within a firm.

We assume different firms can have different costs for innovation. In particular, we partition firms into different (finite) types, and assume different types have different costs for innovation. We will detail later how types evolve over time. We denote the number of types by T and index the types by τ. The R&D cost for internal innovation is assumed to be $\mathbf{R}_I^\tau(Z_{I,j}(t), n_j(t), q_j(t))$. As in Klette and Kortum (2004), we assume the R&D cost function $\mathbf{R}_I^\tau(Z_{I,j}(t), n_j(t), q_j(t))$ exhibits constant returns to scale with respect to $Z_{I,j}(t)$ and $n_j(t)$. Then, the R&D cost per good can be denoted as

$$R_I^\tau(z_{I,j}(t), q_j(t)) \equiv \mathbf{R}_I^\tau(z_{I,j}(t), 1, q_j(t)) = \frac{\mathbf{R}_I^\tau(Z_{I,j}(t), n_j(t), q_j(t))}{n_j(t)}.$$

We further assume that,

$$R_I^\tau(z_{I,j}(t), q_j(t)) = h_I^\tau(z_{I,j}(t))q_j(t)$$

for a strictly convex function $h_I^\tau(\cdot)$. Assuming that R_I is proportional to q is essential in allowing the model to generate constant (although type-dependence) growth rate in the intensive margin and, hence, an establishment size distribution with Pareto tail, which is consistent with our data.

External innovation adds brand-new intermediate goods to the production portfolio of the firm. We assume the new good has the same quality as the average quality of the goods produced by that firm. As we noted above, assumption (7) implies that the growth rate of the quality is common within a firm. Given that any firm starts from one good, combined with the assumption on new goods here, the consequence of (7) is that at each point in time, all products that firm j produces always have the same quality. This property justifies our use of the firm index j on the quality of each good. The total intensity of external innovation is denoted by $Z_{X,j}(t)$. The innovation intensity per good (establishment) is $z_{X,j}(t) \equiv Z_{X,j}(t)/n_j(t)$. The Superposition Theorem of a Poisson process (Grimmett and Stirzaker (2001), p.283) implies that considering (i) the arrival rate of new goods to firm j with intensity $Z_{X,j}(t)$ and (ii) the sum of the (independent) arrival rates of new goods to each establishment in firm j with intensity $z_{X,j}(t)$ are equivalent. The R&D cost for external innovation is assumed to be $\mathbf{R}_X^\tau(Z_{X,j}(t), n_j(t), q_j(t))$, which is assumed to be constant returns to scale with respect to $Z_{X,j}(t)$ and $n_j(t)$. Once again, we can denote the cost per establishment as

$$R_X^\tau(z_{X,j}(t), q_j(t)) \equiv \mathbf{R}_X^\tau(z_{X,j}(t), 1, q_j(t)) = \frac{\mathbf{R}_X^\tau(Z_{X,j}(t), n_j(t), q_j(t))}{n_j(t)},$$

and we assume

$$R_X^\tau(z_{X,j}(t), q_j(t)) = h_X^\tau(z_{X,j}(t))q_j(t)$$

for a strictly convex function $h_X^\tau(\cdot)$.[15] Similarly to the intensive margin innovation, we assume that R_X is proportional to q to guarantee a constant (type-dependent) growth rate in the extensive margin and a distribution of establishment number with a Pareto tail as observed in our data.

Dynamic programming problem: We assume firms transition between different types from τ to τ' with Poisson transition rates $\lambda_{\tau\tau'}$. Each establishment depreciates (is forced to exit) with the Poisson rate δ_τ. We also impose an exogenous exit shock at the firm level. Let d_τ be the Poisson rate of the firm exit shock for a type-τ firm. We omit time notation here, because all variables and functions are constant over time along the balanced-growth path (BGP) that we construct.

Each firm is a collection of n goods (establishments) that are each characterized by a quality level, q. (Here, we omit the firm index j when there is no risk of confusion.) Note that, as we described above, our setting implies that all goods in one firm have the same quality at each point in time. Let $\mathbf{V}_\tau(q, n)$ denote the value function of the firm, such that the Hamilton-Jacobi-Bellman (HJB) equation for the firm is

$$\begin{aligned} r\mathbf{V}_\tau(q, n) = \max_{z_I, z_X} \ & n\Pi_\tau(q, n, z_I, z_X) + z_I \frac{\partial \mathbf{V}_\tau(q, n)}{\partial q} q \\ & - d_\tau \mathbf{V}_\tau(q, n) + \sum_{\tau'} \lambda_{\tau\tau'}(\mathbf{V}_{\tau'}(q, n) - \mathbf{V}_\tau(q, n)), \end{aligned}$$

[15] A major departure from Luttmer (2011) is that we allow for internal innovation. Internal innovation enables us to capture the characteristics of intensive-margin growth. In an earlier working paper version of this paper (Cao et al., 2020), we have shown that (i) a larger firm tends to have a larger intensive margin (Figure B.1) and (ii) the average establishment size is larger for older firms (Figure B.6). These facts indicate that intensive margin growth is an essential component of firm growth over time.

where the return function is

$$\Pi_\tau(q,n,z_I,z_X) \equiv \bar{\pi}q - R_I^\tau(z_I,q) - R_X^\tau(z_X,q) \\ + z_X(\mathbf{V}_\tau(q,n+1) - \mathbf{V}_\tau(q,n)) - \delta_\tau(\mathbf{V}_\tau(q,n) - \mathbf{V}_\tau(q,n-1)).$$

In this expression, $\bar{\pi}q$ is the total profit from a good, and $R_I^\tau(\cdot), R_X^\tau(\cdot)$ are the previously discussed functions governing investment in internal and external innovations, that yield the intensive and extensive innovation rates, (z_I, z_X), respectively.[16]

Because of separability, the value function for the firm is the sum of the value functions across establishments,

$$\mathbf{V}_\tau(q,n) = nV_\tau(q),$$

and the establishment-level HJB equation of a type-τ establishment is

$$rV_\tau(q) = \max_{z_I,z_X} \left[\begin{array}{l} \bar{\pi}q - R_I^\tau(z_I,q) - R_X^\tau(z_X,q) + z_I \frac{\partial V_\tau(q)}{\partial q} q \\ +z_X V_\tau(q) - (\delta_\tau + d_\tau)V_\tau(q) + \sum_{\tau'} \lambda_{\tau\tau'}(V_{\tau'}(q) - V_\tau(q)) \end{array} \right],$$

where $V_\tau(q)$ is the value of type-τ establishment with quality q.

As in Mukoyama and Osotimehin (2019), $V_\tau(q)$ can be shown to be linearly homogeneous in q along the BGP. That is, $V_\tau(q) = v_\tau q$ for a constant v_τ. The HJB

[16] From (3) and (4), the firm-level revenue can be written as (given a firm has n identical establishments with identical product quality $q(t)$ and total labor $\ell(t)$) $Y(t)^\beta \left(q(t)\right)^\beta \left(A(t)\ell(t)\right)^{1-\beta} n^\beta$. For a given n, the revenue function exhibits decreasing returns to ℓ. For firms with small values of n, such as $n=1$ or $n=2$, we expect the firm to behave as a decreasing-returns producer, given that n tend to remain fixed. When n is large, however, a large firm has both large ℓ and n, because q and n tend to grow together. For these firms, with ℓ and n both interpreted as production factors, the revenue function can be viewed to have higher returns to scale. Consistent with this intuition, Dinlersoz et al. (2018) find privately-owned firms reduce their debt over time, whereas listed firms carry large amounts of debt indefinitely. They interpret this financing behavior as consistent with privately owned firms operating under a decreasing-returns-to-scale production functions, and publicly-listed firms under constant returns to scale. Since listed firms typically have more establishments, our characterization of multi-establishment firms is consistent with this empirical evidence.

equation above can be normalized to

$$rv_\tau = \max_{z_I, z_X} \left[\bar{\pi} - h_I^\tau(z_I) - h_X^\tau(z_X) + (z_I + z_X - \delta_\tau - d_\tau)v_\tau + \sum_{\tau'} \lambda_{\tau\tau'}(v_{\tau'} - v_\tau) \right], \quad (8)$$

where $\bar{\pi}$ is given by (6). Accordingly, the HJB equation (8) implies that the choice of innovation intensities (z_I, z_X) is a function of the firm type only. We denote the decision rules as (z_I^τ, z_X^τ).[17]

Entry: An intermediate firm can enter the market by creating a new good (establishment). A new firm draws its type from an exogenous distribution, where m_τ is the probability that an entrant draws the type τ. Given a type τ, the entrant draws its initial *relative quality* $\hat{q}$ from a distribution $\Phi_\tau(\hat{q})$. We assume this relative quality $\hat{q}$ is equal to $q(t)/Q(t)$, where

$$Q(t) \equiv \frac{1}{N(t)} \int_{\mathcal{N}(t)} q_i(t) di$$

is the average quality of intermediate goods. The firm's value of entry $V^e(t)$ is thus

$$V^e(t) = \sum_\tau m_\tau \int V_\tau(\hat{q}Q(t)) d\Phi_\tau(\hat{q}).$$

We assume that any potential entrant can pay a cost $\phi Q(t)$, denominated in final goods, to begin production. Therefore, the free-entry condition is: $V^e(t) = \phi Q(t)$. By defining the value of entry relative to average product quality, $v^e \equiv V^e(t)/Q(t)$, we can rewrite the value of entry as

$$v^e = \sum_\tau m_\tau v_\tau \int \hat{q} d\Phi_\tau(\hat{q}) \quad (9)$$

and we can rewrite the free-entry condition as, $v^e = \phi$.[18] Let the number of entrants at time t be $\mu_e N(t)$, where μ_e is a constant along the balanced-growth path.

[17] Recall that, from (6), $\bar{\pi}$ is a function of $\bar{w}$ only. Thus, given r and $\bar{w}$, equation (8) and the first-order conditions can solve for v_τ, z_I^τ, and z_X^τ.

[18] Note that once r is given, we can find a value of $\bar{w}$ that satisfies the free entry condition.

3.5 Balanced-growth equilibrium

A *competitive equilibrium* of this economy is a wage $w(t)$, a consumer allocation $(C(t), R(t), E(t))$, a final-good-producer allocation $(Y(t), \{x_i(t)\}_{i\in\mathcal{N}(t)})$, an allocation for intermediate-good producers $\{\ell_j(t), q_j(t), p_j(t), z_{I,j}(t), z_{X,j}(t), n_j(t)\}$ for all active producers j, and a value of entry $V_e(t)$ such that at each instant, (i) consumers optimize, (ii) the final-good producers' allocation solves its profit maximization problem, (iii) the intermediate-good producers' allocations solve their profit-maximization problem, (iv) the free-entry condition holds, (v) the final-good market clears: $Y(t) = C(t)+R(t)+E(t)$, and (vi) the labor market clears: $L(t) = \int_{\mathcal{N}_t} \ell_i(t)di$.

We now construct a *balanced-growth equilibrium* of this economy. Assume the population $L(t)$ grows at an exogenous rate γ. Furthermore, let aggregate quality $Q(t)$ grow at a constant rate ζ and the number of establishments $N(t)$ (and the number of firms) grow at a constant rate η. Denote the growth rate of final output $Y(t)$ by g. Along a balanced-growth path (BGP), the growth rates of $Y(t)$, $C(t)$, $R(t)$, and $E(t)$ must all be equal. Thus, the Euler equation (1) requires $\dot{C}(t)/C(t) = g$ because $C(t)$ grows at the same rate as $Y(t)$. This fact implies $r = \rho + \sigma g$ along the BGP.

The quality-invariant component of profit $\bar{\pi}(t)$ in (6) is constant along the BGP. Therefore, $w(t)$ must grow at the same rate as $A(t)Y(t)^{\frac{\beta}{1-\beta}}$. Given that $Y(t)$ grows at rate g and $A(t)$ grows at rate θ, $w(t)$ must grow at the rate $\beta g/(1-\beta) + \theta$. Because labor income of the representative consumer must grow at the same rate as consumption, and because population growth is γ, the following relationship must hold:

$$g = \gamma + \theta + \frac{\beta}{1-\beta} g, \tag{10}$$

which implies output growth of $g = (1 - \beta/(1-\beta))^{-1}(\gamma + \theta)$. Similarly to Luttmer (2011), the output growth rate is dictated by the population growth rate γ and the intermediate-good productivity growth rate θ.[19] Lastly, we can also decompose aggregate growth into extensive and intensive margins: $g = \eta + \zeta$.[20]

[19] In the terminology of Jones (1995), our model exhibits "semi-endogenous" growth. It is straightforward to extend the model to exhibit fully endogenous growth.

[20] The derivation is contained in Online Appendix E.

4 Distributions of firm sizes and establishment sizes

The properties of our model allow us to characterize firm-size distributions from both margins of firm growth. In the next section, we will estimate model parameters by mapping the theoretical intensive and extensive margin size distributions to the analogous empirical distributions that we documented in Section 2.

Before characterizing distributions, note several relevant features of our model. First, a general property of our model is that establishments are homogeneous within a firm. This property implies that all establishments within a firm share a common level of quality, because (i) a firm enters with one establishment and each of the firm's new establishments inherit the same quality as its existing ones, and (ii) the intensity of internal innovation z_I^τ that determines the evolution of quality is common across all establishments within a firm (although it may change over time). This property also implies that establishment sizes are also common within a firm. Although in reality establishment sizes are not the same within a firm, we view this assumption as a useful simplification that affords us sharp analytical characterizations.

Second, note that the distribution over the number of establishments per firm evolves through external innovation and exit shocks, and thus corresponds to the extensive margin in Section 2. In contrast, the size of each establishment evolves through internal innovation but does not directly correspond to the intensive margin in Section 2 (because economy-wide establishment size distribution, across establishments, is not the same as economy-wide intensive margin distribution, across firms), but is closely related.[21]

Finally, we assume that the economy is on a stationary BGP. In a balanced-growth equilibrium, the number of firms grows at the same rate as the number of establishments $N(t)$. The distribution of the number of establishments per firm, $n(t)$, is stationary (in shares), and the distribution of establishment's relative quality, $q(t)/Q(t)$, or size, is also stationary, despite the fact that $Q(t)$ grows exponentially over time.

[21]Note the existence of these two margins is the major departure from Klette and Kortum (2004) and Luttmer (2011). In these papers, establishments are homogeneous and each establishment does not grow, so the only relevant innovation is external innovation.

In the remaining analyses, we use the following formal definition of Pareto tail. A random variable $\mathbf{X}$ defined over $\mathbb{R}^*$ has a Pareto tail with tail index $\xi > 0$ if $\lim_{x\to\infty} x^\xi \Pr(\mathbf{X} > x) = a$ for some $a > 0$. The distribution has a thin tail if $\lim_{x\to\infty} x^\xi \Pr(\mathbf{X} > x) = 0$ for any $\xi > 0$. For notational convenience, we assign tail index ∞ to distributions with a thin tail.

4.1 General characterizations

Our approach is to look at two separate margins: the distribution of the number of establishments per firm, summarized by $\bar{\mathcal{M}}_\tau(n)$, which is the measure of type-τ firms with n establishments divided by $N(t)$ (i.e., $N(t)$-normalized measure), and the distribution of establishment quality relative to average quality, $q(t)/Q(t)$, a measure of establishment size.[22] We denote the fraction of type-τ establishments with $q(t)/Q(t) \geq \hat{q}$ as $\bar{\mathcal{H}}_\tau(\hat{q})$.

The distribution of the number of establishment per firm, $\{\bar{\mathcal{M}}_\tau(n)\}_{n=1}^\infty$, can be characterized by the following difference equations:

$$\begin{aligned} 0 = &-(n(z_X^\tau + \delta_\tau) + d_\tau + \eta)\bar{\mathcal{M}}_\tau(n) + (n+1)\delta_\tau \bar{\mathcal{M}}_\tau(n+1) + (n-1) z_X^\tau \bar{\mathcal{M}}_\tau(n-1) \\ &- \sum_{\tau' \neq \tau} \lambda_{\tau\tau'} \bar{\mathcal{M}}_\tau(n) + \sum_{\tau' \neq \tau} \lambda_{\tau'\tau} \bar{\mathcal{M}}_{\tau'}(n) + \mu_e m_\tau \mathbf{1}_{\{n=1\}} \end{aligned} \tag{11}$$

for each τ and $n \geq 1$ (with the convention that $\bar{\mathcal{M}}_\tau(0) = 0$).

The distribution of establishment-level relative quality, which is proportional to establishment size, $\bar{\mathcal{H}}_\tau(\hat{q})$, is governed by the following Kolmogorov forward equation:

$$\begin{aligned} (z_I^\tau - \zeta)\hat{q}\frac{d\bar{\mathcal{H}}_\tau(\hat{q})}{d\hat{q}} = &-(\delta_\tau + d_\tau + \eta - z_X^\tau)\bar{\mathcal{H}}_\tau(\hat{q}) + \mu_e \frac{m_\tau}{M_\tau}(1 - \Phi(\hat{q})) \\ &- \sum_{\tau' \neq \tau} \lambda_{\tau\tau'} \bar{\mathcal{H}}_\tau(\hat{q}) + \sum_{\tau' \neq \tau} \lambda_{\tau'\tau} \frac{M_{\tau'}}{M_\tau} \bar{\mathcal{H}}_{\tau'}(\hat{q}). \end{aligned} \tag{12}$$

Lastly, for the distribution of firm size, recall all establishments in a firm grow at the same rate and hence have the same size. Therefore, we just need to keep

[22] On a BGP, these normalized measures are constant and $\sum_{\tau,n} n\bar{\mathcal{M}}_\tau(n) = 1$.

track of the joint distribution of establishment number and establishment size in order to study the firm-size distribution. That is, the distribution of firm size is some "convolution" of the distribution of the number of establishments per firm and the distribution of establishment size and can be derived as follows. Let $\mathcal{M}_\tau(n,\hat{q})$ be the normalized measure of type-τ firms with n establishments and $q(t) \geq \hat{q}Q(t)$. Then

$$\begin{aligned}(z_I^\tau - \zeta)\hat{q}\frac{d\mathcal{M}_\tau(n,\hat{q})}{d\hat{q}} = &-(n(z_X^\tau+\delta_\tau)+d_\tau+\eta)\mathcal{M}_\tau(n,\hat{q}) \\ &+(n+1)\delta_\tau\mathcal{M}_\tau(n+1,\hat{q};t)+(n-1)z_X^\tau\mathcal{M}_\tau(n-1,\hat{q}) \\ &+\sum_{\tau'\neq\tau}\lambda_{\tau'\tau}\mathcal{M}_{\tau'}(n,\hat{q})-\sum_{\tau'\neq\tau}\lambda_{\tau\tau'}\mathcal{M}_\tau(n,\hat{q})+\mu_e m_\tau(1-\Phi_\tau(\hat{q}))\mathbf{1}_{\{n=1\}}\end{aligned} \tag{13}$$

for each τ and $n \geq 1$ (with the convention that $\mathcal{M}_\tau(0,\hat{q})=0$). The detailed derivations of these equations are presented in Online Appendix D.1.

4.2 Distributions for one-type economy

When only one firm type exists, firm growth is governed by three endogenous variables: z_I, z_X, and μ_e. Denoting $p \equiv \log(\hat{q})$ and $\tilde{\mathcal{H}}(p) \equiv \bar{\mathcal{H}}(\exp(p))$, Online Appendix D.2 shows that

$$\tilde{\mathcal{H}}(p) = \int_{-\infty}^{p} e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}(\tilde{p}-p)}\frac{\delta+d+\eta-z_X}{z_I-\zeta}(1-\Phi(\exp(\tilde{p})))d\tilde{p} \tag{14}$$

holds, and this equation characterizes the distribution of establishment sizes. This expression implies $\tilde{\mathcal{H}}(\log y)$ is the complementary cumulative distribution function of a random variable $\mathbf{Y^e}$ defined by a convolution between a Pareto distribution with scale parameter 1 and tail index $(\delta+d+\eta-z_X)/(z_I-\zeta)$ and a distribution with cumulative distribution function (CDF) Φ. That is, $\mathbf{Y}^e$ is expressed as $\mathbf{Y}^e = \mathbf{Y}_1^e\mathbf{Y}_2^e$, where $\mathbf{Y}_1^e \sim \mathit{Pareto}(1,(\delta+d+\eta-z_X)/(z_I-\zeta))$ and $\mathbf{Y}_2^e \sim \Phi$.

Notice also that, when Φ is a log-normal distribution, $\tilde{\mathcal{H}}$ is a convolution of a Pareto distribution and a log-normal distribution analyzed in Reed (2001), and more recently, Cao and Luo (2017) and Sager and Timoshenko (2019). Therefore, we offer an alternative micro-foundation of this convolution distribution with an endogenous establishment growth rate, relative to the micro-foundation in Reed (2001) with an exogenous growth rate. Our micro-foundation is also more general because it allows for any distribution of Φ, whereas Reed (2001) only allows Φ to be a log-normal distribution.

Using this explicit solution, we can easily show that if $z_I > \zeta$ and Φ has a thin right tail (for example, when Φ is a log-normal or left-truncated log-normal distribution), then $\bar{\mathcal{H}}(p)$ has a Pareto tail with the index given by

$$\lambda^e \equiv \frac{d + \eta - z_X + \delta}{z_I - \zeta}. \tag{15}$$

When $z_I < \zeta$, the distribution has a thin tail.

For the distribution of the number of establishments per firm, we show (see Online Appendix D.2) that when $z_X > \delta$, it has a Pareto tail with the tail index given by

$$\lambda^{ne} \equiv \frac{\eta + d}{z_X - \delta}. \tag{16}$$

The following proposition summarizes the last two results. The proof of this proposition is given in Online Appendix D.3 using a Karamata Tauberian theorem from the literature on regular varying sequences and functions (Bingham et al., 1987, Corollary 1.7.3).[23]

Proposition 1 *On a stationary BGP with $z_I > \zeta$ and $z_X > \delta$ and the distribution of entrant sizes Φ has a thin tail, the stationary distribution of establishment sizes (across establishments) and the stationary distribution of the number of establishments per firm (across firms) have Pareto right tails and the tail indexes are given by (15) and (16), respectively.*

[23]Notice this result on the distribution of establishment number is stronger than the one in Luttmer (2011). In particular, he shows that for any $\xi > \frac{\eta+d}{z_X-\delta}$, $\lim_{n\to\infty} n^{\xi} \Pr\{\mathbf{X} > n\} = \infty$ and for any $\xi < \frac{\eta+d}{z_X-\delta}$, $\lim_{n\to\infty} n^{\xi} \Pr\{\mathbf{X} > n\} = 0$, whereas we show

$$\lim_{n\to\infty} n^{\frac{\eta+d}{z_X-\delta}} \Pr(\mathbf{X} > n) = a$$

for some $a > 0$. The latter limit implies the previous two limits but not vice versa.

Now, we analyze the distribution of firm size, which is the combination of above two. In the special case where the initial draw satisfies $\int \hat{q} d\Phi(\hat{q}) = 1$ $z_I = \zeta$ holds. In this case, the random variable for firm size $\mathbf{Z}$ can be written as $\mathbf{Z} = \mathbf{XY}$, where $\mathbf{X}$ is the number of establishments and $\mathbf{Y}$ is establishment size in a firm. In the cross section, $\mathbf{X}$ and $\mathbf{Y}$ are independent and the cdf of $\mathbf{Y}$ is given by Φ. The pdf for $\mathbf{X}$ is given by (23). We show that, utilizing the Tauberian theorem in Mimica (2016, Corollary 1.3), the following proposition holds (see Appendix D.2 and the proof in Appendix D.3).

Proposition 2 *On a stationary BGP with $z_I = \zeta$ and $z_X > \delta$, and when the distribution of entry sizes Φ has a thin tail, the firm-size distribution has a Pareto tail with the tail index equal to the tail index of the distribution of the number of establishments per firm given by* (16).

When the distribution of entry sizes Φ has a Pareto tail, we can show the tail index of the firm size distribution is the minimum of the tail index of Φ and the tail index given by (16).

In the empirically more relevant the case in which $z_I \neq \zeta$, i.e. $\int \hat{q} d\Phi(\hat{q}) \neq 1$. This case is more challenging because the dynamics of firm size are driven both by the dynamics of the establishment number and of the dynamics of relative establishment size. This fact implies that when we write firm size as a product of the number of establishments and average establishment size, $\mathbf{Z} = \mathbf{XY}$, in the cross section, $\mathbf{X}$ and $\mathbf{Y}$ are correlated, instead of being independent when $z_I = \zeta$. For example, when $z_X > \delta$ and $z_I > \zeta$, over time surviving firms, on average, have both a higher number of establishments and larger establishments.

To tackle this case, we use the system of difference-differential equations (13). This system involves an infinite number of functional equations on $\{\mathcal{M}_\tau(n, \hat{q})\}$. We use the Laplace transformations of $\{\mathcal{M}_\tau(n, \hat{q})\}$ to transform (13) into a system of only difference equations. Then we use the Karamata Tauberian theorem (Bingham et al., 1987, Corollary 1.7.3) to characterize the asymptotic behavior of its solution. This characterization enables us to utilize the Tauberian theorem in Mimica (2016, Corollary 1.3) and show that, if $z_X > \delta$ and $z_X - \delta + z_I - \zeta > 0$, the firm-size distribution has a Pareto tail with the tail index

$$\lambda^f \equiv \frac{\eta + d}{z_X - \delta + z_I - \zeta}. \tag{17}$$

(see Appendix D.2). The following proposition, which also describes the relationship among λ^e, λ^{ne}, and $\lambda^f > 1$, summarizes the result. The detailed proof is given in Online Appendix D.3.

Proposition 3 *On a stationary BGP with $z_X > \delta$ and $z_X - \delta + z_I - \zeta > 0$, the firm-size distribution has a Pareto tail with the tail index given by (17). If, in addition, $z_I > \zeta$, all three distributions for establishment size, the number of establishments per firm, and firm size have a Pareto tail and*

$$\frac{1}{\lambda^f} = \frac{1}{\lambda^{ne}} + \frac{1}{\lambda^e} - \frac{1}{\lambda^{ne}\lambda^e} \tag{18}$$

where λ^e, λ^{ne}, and $\lambda^f > 1$ are respectively the Pareto tail indexes for the distributions for establishment size, the number of establishments per firm, and firm size.

Because $\lambda^e, \lambda^{ne} > 1$, equation (18) shows $\lambda^f > \lambda^e$ or $\lambda^f > \lambda^{ne}$; that is, the tail of the firm-size distribution is strictly fatter than the tails of either the establishment-size distribution or the number-of-establishments distribution. The formula can also be used to calculate the tail index of the firm-size distribution from the tail indexes for the establishment size and the number-of-establishments distributions.[24]

5 Quantitative analysis

In this section, we estimate our model using the cross-sectional information to quantitatively analyze the pattern of firm growth over the 1995-2014 period. We first

[24]Equation (18) approximates the relationships of the estimated tail indexes in the data very well. For example, for 1995, with the tail index of the number-of-establishments distribution of 1.25 in Table A.1 and the tail index of the establishment size distribution of 1.55 from the estimation in Online Appendix F.1 (which is conceptually different from the tail index for the intensive margin in Table A.1), equation (18) implies a tail index of 1.07 for firm-size distribution, which lies between the estimates 0.99 and 1.10 for firm size distribution in Table A.1. Similarly, for 2015, with the tails parameters of 1.21 and 1.75, equation (18) implies a tail index of 1.08 for firm size distribution, which also lies between the estimates of 0.99 and 1.17 in Table A.1.

confirm the model is able to match the firm-size, the number-of-establishments, and establishment-size distributions well. We estimate the model for both 1995 and 2014 data and show how the model informs us about fundamental economic forces that changed with average firm-size growth over these years.

5.1 Model estimation

Firm Types: We find that at least two types are needed to match both the extensive- and intensive-margin distributions. The main reason is there are many single-unit firms that do not expand in extensive margin whereas there are also many firms that end up being at the right-tail of the extensive-margin distribution. This difference cannot be justified by the same innovation cost functions. We estimate a simple version of the model with two types, denoted $\tau \in \{L, H\}$. H-type firms are characterized by a *high* intensity of external innovation while L-type firms exhibit a *low* intensity of external innovation. When we calibrate the model to data, H-type firms turn out to have a lower cost of external investment, and expand their number of establishments faster than L-type firms. We assume H-type firms transition to L-type firms at the rate $\lambda_{HL} > 0$, whereas L-type firms do not transition to H-type firms, that is, $\lambda_{LH} = 0$. Thus, the L-type is the absorbing state.[25]

These assumptions on types allow us to easily obtain closed-form solutions for the tail indexes of the distributions of establishment number and establishment size, which we use to estimate the model. Using the results for two types in Luttmer (2011) and the derivations in Subsection 4.2, we can show that, under some parameter restrictions, the stationary distribution of the number of establishments per firm has a Pareto tail with the tail index given by:

$$\min\left\{\frac{\eta + \lambda_{HL} + d_H}{[z_X^H - \delta_H]_+}, \frac{\eta + d_L}{[z_X^L - \delta_L]_+}\right\}, \tag{19}$$

where $[x]_+ \equiv max(x, 0)$. The formula corresponds to (16) in the case of a single type.

[25] These assumptions are similar to those in Luttmer (2011). Under these assumptions, Luttmer (2011) also provides analytical solutions for the distribution of establishment numbers, which can be used to verify the accuracy of numerical solutions.

Similarly, using the procedure from Gabaix et al. (2016) and Cao and Luo (2017), the Pareto-tail index of the distribution of establishment size is given by

$$\min\left\{\frac{\eta+\delta_H+\lambda_{HL}+d_H-z_X^H}{[z_I^H-\zeta]_+},\frac{\eta+\delta_L+d_L-z_X^L}{[z_I^L-\zeta]_+}\right\}. \tag{20}$$

Estimation: We estimate the model in two steps. In the first step, we estimate $(z_X^H, z_X^L, \lambda_{HL}, \mu_e, m_H, m_L)$ using moments related to the number of establishments per firm, and then we estimate $(z_I^H, z_I^L, \Phi_H(\cdot), \Phi_L(\cdot))$ using moments related to the number of employees per establishment. This step is an intermediate step, as many of the estimated variables are endogenous variables. In the second step, we map these estimates to the fundamental parameters. In particular, as is common in the literature, we assume quadratic innovation costs of the form,

$$h_X^\tau(z_X^\tau) \equiv \chi_X^\tau\left(z_X^\tau\right)^2 \quad , \quad h_I^\tau(z_I^\tau) \equiv \chi_I^\tau\left(z_I^\tau\right)^2 \qquad \text{for } \tau \in \{L, H\}.$$

We assume log-normal distribution for initial relative quality: $\Phi_\tau(\cdot) \sim \exp(\mathcal{N}(\varrho_\tau, \varsigma_\tau))$. We then estimate the additional parameters using the estimated values recovered from the first step.[26]

A set of parameters are assigned in advance of estimation, whereas the remaining parameters are estimated to match empirical moments of the establishment number and establishment size distributions. Table 1 summarizes parameter estimates and targets.

The unit of time is set as a year. For preferences, we assume log utility and an effective discount rate of $\rho = 0.01$. We choose the elasticity of demand for the final-good producer to $\beta = 0.091$, which is consistent with a markup of 10% (Basu and Fernald (1995)). With respect to exogenous firm and establishment exit rates, we set $d_H = 0\%$ and $d_L = 0.4\%$ at the firm level and $\delta_H = \delta_L = 12\%$ in 1995 and $\delta_H = \delta_L = 10\%$ in 2015. These values amount to around a 3% quarterly exit rate for establishments and a 0.1% quarterly exogenous exit rate for firms, roughly consistent

[26] The targeted moments include several percentiles of the distributions and the slopes of the right tail of the distributions. More details can be found in Appendix F.2.

Table 1: Parameter Values and Targets

Concept	Parameter	Value	Target/Source
Parameters Set in Advance			
Elasticity of Demand	β	$1-\frac{1}{1.10}$	10% Markup
Intertemporal Elasticity	σ	1	Log utility
Discount Rate	ρ	0.01	Standard value
Population Growth Rate	γ	0.01	Census Bureau
Firm Exit Rates (Exogenous Component)	d_L, d_H	0.4%, 0%	BLS
Establishment Exit Rates (1995)	δ_L, δ_H	12%, 12%	BLS
Establishment Exit Rates (2014)	δ_L, δ_H	10%, 10%	BLS
Estimated Parameter Targets			
Entry Cost	ϕ		Normalized wage, $\bar{w}=1$
Growth Types	λ_{HL}, m_H		Establishment number distribution
Extensive Margin Costs	χ_X^H, χ_X^L		Establishment number distribution
Intensive Margin Costs	χ_X^H, χ_X^L		Establishment size distribution
Entrant Size	$\varrho_H, \varsigma_H, \varrho_L, \varsigma_L$		Establishment size distribution

with the establishment exit rate and the exit rate of very large firms in our data.[27] We set the population growth rate to the post-1960 average of $\gamma = 0.01$.

All other parameters are estimated to match empirical moments. We target $\eta = 0.01$ and $\zeta = 0.021$ to imply a growth rate of final output of 3.1% for 1995 and $\eta = 0.01$ and $\zeta = 0.013$, implying a final output growth rate of 2.3% for 2014. Given the log-utility specification and the value $\rho = 0.01$, the Euler equation in (1) implies an interest rate of $r = 0.04$ for 1995 and $r = 0.03$ for 2014, which are within the range of values that are standard in the literature.

Finally, we estimate the remaining parameters twice—once for 1995 moments of the establishment size and number distributions, and once for 2014 empirical moments. Hence, the resulting differences in parameter estimates between 1995 and

[27]Notice that overall quarterly firm exit rate in our model should be larger than 0.1% because firms also exit endogenously when they lose all establishments. This rate depends on other model ingredients such as the endogenous establishment creation rates z_X^τ. From model simulations, we find that our estimation implies an average quarterly firm exit rate of 2.75% in 1995 and 2.32% in 2014.

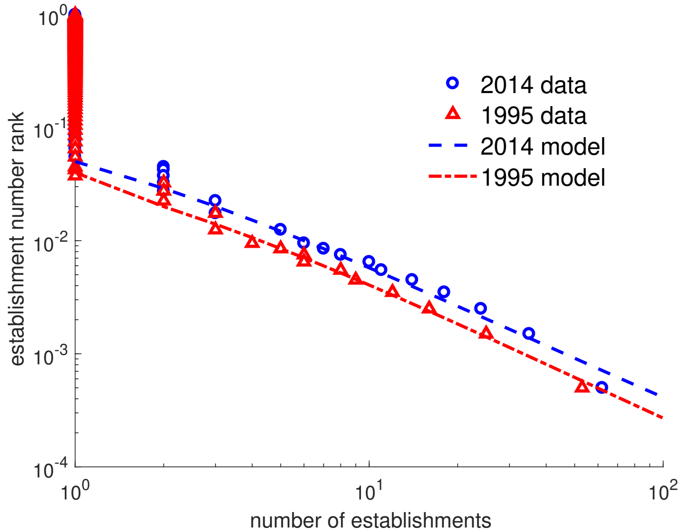


Figure 3: Distribution of number of establishments per firm, Data and Model

2014 are due to changes in the empirical distributions. For example, the distribution of establishment number changes shifts significantly from 1995 to 2014: the fraction of single establishment firms is 96% in 1995 and 95% in 2014; at the 99th percentile of the establishment number distribution, firms have only 5 establishment in 1995, while in 2014, they have 7 establishments; and the Pareto tail index decreases from 1.25 in 1995 to 1.21 in 2014 (Table A.1 in Appendix).

Figure 3 shows the model distributions of the number of establishments per firm match the empirical distributions from the 1995 and 2014 data very well. Figure 4 shows the model distributions of establishment size closely match the parametrized empirical distributions for 1995 and 2014 data described in Online Appendix F.1 (blue solid line), as well as publicly available BLS tabulations of establishment sizes (red circles).

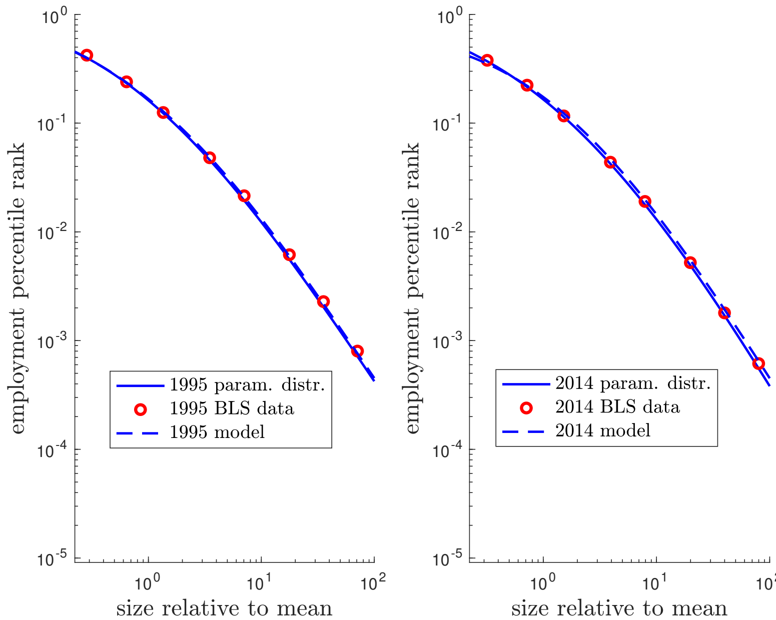


Figure 4: Distribution of number of employees per establishment, Data and Model

5.2 Quantitative changes in parameter values from 1995 to 2014

We compare the results of model estimation for 1995 with that in 2014. Differences in model outcomes over time inform us about the underlying economic mechanisms that generate the observed changes in the distributions over the number of establishments and average establishment size.

Table 2 shows the H-type firm's extensive-margin investment rate increased from 32.81% to 51.20% as its investment-cost coefficients decreased about 21% (from 0.6149 to 0.4797). In contrast, the L-type firms invest significantly less in 2014 than in 1995. As for entrants, in 2014, a larger fraction of entrants are H-type firms (5.23% in 1995 and 8.78% in 2014), yet the average duration of being a H-type shortened from around four years (= 1/0.2523) in 1995 to around two years (= 1/0.4900) in 2014. The estimation provides estimates of μ_e, which is the entry rate of firms relative to the total number of establishments. To recover the firm entry rate (relative to the

Table 2: Parameter Estimates and Model Outcomes, 1995 versus 2014

Parameter	Description	Value (1995)	Value (2014)
Innovation Investments			
z_X^H	H-type external innovation	0.3281	0.5120
z_X^L	L-type external innovation	0.0019	0.0002
z_I^H	H-type internal innovation	0.0559	0.0637
z_I^L	L-type internal innovation	0.0000	0.0087
Innovation Costs			
χ_X^H	H-type external innovation cost	0.6149	0.4797
χ_X^L	L-type external innovation cost	57.400	610.12
χ_I^H	H-type internal innovation cost	3.6056	3.8547
χ_I^L	L-type internal innovation cost	∞	15.134
Firm Entry			
μ_e	Entry rate	0.0981	0.0740
ϕ	Entry fixed cost	0.2009	0.1671
$\int \hat{q} d\Phi_H(\hat{q})$	H-entrant size relative to mean	0.6016	0.7595
ϱ_H	Mean of $\Phi_H(\cdot)$	−2.4909	−1.1511
ς_H	Standard deviation of $\Phi_H(\cdot)$	1.9914	1.3236
$\int \hat{q} d\Phi_L(\hat{q})$	L-entrant size relative to mean	0.9271	0.5567
ϱ_L	Mean of $\Phi_L(\cdot)$	−1.472	−3.4660
ς_L	Standard deviation of $\Phi_L(\cdot)$	1.6711	2.4002
Firm Types			
λ_{HL}	H to L transition rate	0.2524	0.4901
m_H	Fraction of H-type at entry	0.0523	0.0878
m_L	Fraction of L-type at entry	0.9477	0.9122

Notes: Estimates using Least-Square Minimization

number of firms), we need to divide μ_e by the average number of establishments per firm. The model recovers a declining firm entry rate of 7.55% in 1995 to 5.08% in 2014.[28] Notice that this decline is accompanied by a decrease in the estimated entry

[28] The level of the entry rates in the model are lower than the ones in the data but the magnitude of the decline is similar. We can bring the model entry rates closer to the data if we allow for type-dependent exit rates.

cost (from 0.2009 to 0.1617), in contrast to the recent literature that emphasizes increases in entry cost or entry barriers such as Decker et al. (2014). As we discuss below, while a decrease in entry cost would lead to an increase in entry in our model, the declines in external innovation cost, establishment exit rates, and growth rate reduce the entry rate by more.

5.3 Decomposing the changes in the entry rate and extensive margin

Using our estimated structural model, we conduct a series counterfactual experiments that decompose the changes in the firm entry rate and the number of establishments per firm into their constituent causes. We focus on these particular changes because they highlight (i) the tradeoff between incumbent innovation and entry and (ii) its implications for firm size distribution. Our decomposition procedure is as follows. We start with all parameter values (estimated and set in advance) for the year 1995. Then we incrementally switch each parameter, one by one, to the estimates from the year 2014. This procedure provides counterfactuals where only a subset of parameter values are changed to the 2014 values. Note that, while the order of switching matters for the quantitative results, the qualitative results and general intuition are robust to the order of switches.

Table 3 shows the result. The table, starting from the top row marked "Decomposition," begins from the 1995 estimated parameters and "turns on" the 2014 parameters as we go down each row. At the end of the final row, all parameters are switched to the ones in 2014, and therefore the sum of all rows is equal to the total change.

The first column conducts this exercise for the entry rate. In the estimated model, the entry rate decreases by 2.47 percentage points in total. Observing the parameters that generate a decline in the entry rate, this total turns out to be primarily driven by (i) the changes in the external innovation costs, (ii) the decline in the establishment exit rate, and (iii) the decline in the aggregate growth rate (as γ remains unchanged from 1995 to 2014, (10) implies that labor productivity growth rate θ declines). The second column is the decomposition for the number of establishments per firm (the

Table 3: Total Changes (1995-2014) and Decomposition

		Change in Firm Entry Rate	Percent Change in #Establishments/Firm
Decomposition:			
type fraction and persistence	(m_H, λ_{HL})	8.14	−20.04
entrant quality distribution	$(\varrho_H, \varsigma_H, \varrho_L, \varsigma_L)$	0.19	9.24
fixed entry cost	(ϕ)	5.59	−6.50
external innovation cost	(χ_X^H, χ_X^L)	−8.53	12.79
internal innovation cost	(χ_I^H, χ_I^L)	0.61	−1.60
establishment exit rates	(δ_H, δ_L)	−5.43	10.37
growth rate	(g)	−3.06	12.09
Total Change (1995-2014):		−2.47	12.15

extensive margin), which is the main focus of this paper. In the model, the average number of establishments per firm have increased by 12.14% between 1995 and 2014. Observing the parameters that increase the average number of establishments per firm, our decomposition reveals that the increase is primarily driven by the same three factors that reduce the entry rate.

This second decomposition provides an important insight in looking at the empirical facts in Section 2 through the lens of the model. When we think of the dominance of large firms in the recent years, our model indicates that the mechanism is associated with (i) the reduction of the cost for expanding through new establishments, (ii) the decline in the establishment exit rate, and (iii) the decline in the aggregate productivity growth rate.

To highlight the mechanisms underlying the model decomposition, we consider a version of the model with one type (omitting the τ subscript) that delivers essentially the same qualitative relationships yet provides a sharp characterization of firm-level outcomes (see Section 4.2) and general equilibrium outcomes. The one-type environment can be characterized by the firm's value function,

$$(\rho + \sigma g)v = \bar{\pi} - \chi_I z_I^2 - \chi_X z_X^2 + (z_I + z_X - \delta - d)v,$$

whose first-order conditions imply innovation intensities $z_I = v/(2\chi_I)$ and $z_X = v/(2\chi_X)$; the firm's value v which is pinned down by the free entry condition (9) $v = \phi / \int \hat{q} d\Phi(\hat{q})$; the value of $\bar{\pi}$ which is pinned down by the value function after obtaining values for (v, z_I, z_X); and the equilibrium entry rate,

$$\mu_e = \frac{1}{\int \hat{q} d\Phi(\hat{q})} \left(\frac{\gamma + \theta}{1 - (\beta/(1-\beta))} - \Big(z_I + z_X - \delta - d \Big) \right),$$

where output growth is given by equation (10), and we assume $\beta < 1/2$ to ensure positive entry in equilibrium ($\mu_e > 0$).[29] Given this characterization, the following comparative static results are immediate.

Proposition 4 *Consider a BGP of the one-type economy. The following comparative statics hold:*

(a) ***Entry costs:*** *An increase in the entry cost, ϕ, generates an increase in innovation intensities (z_I, z_X) and a decrease in the entry rate (μ_e).*

(b) ***Innovation costs:*** *A decrease in χ_I increases z_I but does not directly affect z_X. A decrease in χ_X increases z_X but does not directly affect z_I. The entry rate (μ_e) decreases when either cost χ_I or χ_X decreases.*

(c) ***Growth and exit rates:*** *An increase in population growth γ, productivity growth θ, or exit rates (δ, d) does not directly affect innovation intensities (z_I, z_X) but does generate higher output growth g and entry rate μ_e.*

From Proposition 4, one key prediction of the model is a trade-off between firm entry and incumbents' innovative activity (e.g., that leads to growth in the number of establishments per firm). From part (a) of the proposition, if the entry cost ϕ is higher, the free-entry condition implies that a firm needs to receive higher lifetime

[29] It is straightforward to show that $g = z_I + z_X - (\delta + d) + \mu_e \int \hat{q} d\Phi(\hat{q})$ along the balanced growth path. Rearranging and substituting equation (10) yields the expression for μ_e.

compensation for incurring the increased initial cost of creating a new product for the market. Therefore, an entering firm innovates at a higher rate and grows to a larger size along both the intensive and extensive margins, compared to an economy with a lower entry cost. Hence a higher entry cost is associated with both less entry and a greater number of establishments per firm, on average.[30]

However, our estimation infers that entry costs in fact declined over recent decades (see Table 2) and, therefore, the mechanisms associated with parts (b) and (c) of the proposition dominate the mechanism associated with a change in entry costs. First, declining innovation costs lead to higher innovation rates and less firm entry, as in part (b) of the proposition, which implies that there are more establishments through external innovation and fewer firms due to lower entry. As a result, there are more establishments per firm in an economy with lower innovation costs. Part (b) is important for our estimation results as Table 2 indicates that the external innovation cost for H-type firms indeed declined. Part (c) reflects the limits to output growth due to population and technology dynamics, as seen from the balanced-growth condition in equation (10). With lower labor productivity growth, incumbent firms' increased innovative activity diminishes the incentive of new firms to enter the market due to greater scarcity of inputs to production. Similarly, part (c) also states that a decline in exit rates discourages firm entry, because incumbent firms employ more resources for innovative activity and less frequently free up those resources through exit. Entrants have less incentive to enter the market if there is a high cost of employing resources to grow and recover their entry cost. As we see in Table 3, a decline in both aggregate productivity growth, θ, and exit rates were indeed large contributors to the reduction in the entry rate. Overall, parts (b) and (c) imply reduced entry rates and increased growth in the number of establishments per firm in recent decades, and the associated mechanisms quantitatively dominate the opposing force of the effect in part (a).

[30]Notice that the entry rate μ_e is defined as the ratio of the flow of new firms over the stock of establishments. The effective firm entry rate is the ratio of the flow of new firms over the stock of firms, i.e., μ_e divided by the average number of establishments per firm, as in Subsection 5.2. As a higher entry cost leads to a greater number of establishment per firms, firm entry rate decreases more than μ_e does. Similar to Hopenhayn et al. (2018), because population growth is held constant, the decrease in firm entry rate is equal to the sum of the decrease in firm exit rate, which results from the increase in extensive margin growth, and the increase in average firm size.

For the number of establishments per firm, recall that, in Part (b), a decline in external innovation cost increases z_X and reduces μ_e. In Part (c), declines in establishment exit rates and the aggregate growth rate reduce μ_e without directly affecting z_X. From our characterization of the tail of the extensive margin distribution,

$$\lambda^{ne} = 1 + \frac{\mu_e}{z_X - \delta}$$

holds. This equation can be derived from (16) and the fact that $\eta = z_X - d - \delta + \mu_e$ (from the law of motion for the aggregate number of establishments). The changes in z_X and μ_e lead to a fatter tail of the extensive margin distribution (a smaller λ^{ne}), because λ^{ne} is smaller when $\mu_e/(z_X - \delta)$ is smaller. This mechanism is consistent with our estimated outcome. Moreover, because the establishment number distribution has a lower Pareto tail index in 2014 compared to 1995, and a lower tail index is associated with a higher average (for example, when the tail index approaches 1, the average tends to ∞), the decline in λ^{ne} during this period implies that the average number of establishments per firm is higher in 2014. This overall intuition is consistent with the decomposition documented in Table 3.

6 Conclusion

In this paper, we decomposed firm growth into two margins: an extensive margin of building new establishments and an intensive margin of adding workers to existing establishments. We documented the patterns of extensive- and intensive-margin firm growth in the U.S. from 1990-2014 and found that U.S. growth is predominantly generated by the addition of new establishments in very large firms. We developed a model of firm growth that incorporates both the extensive and intensive margins as separate types of firm innovation and showed the model can generate a fat tail of large firms, both in terms of the number of establishments and the number of workers. We analytically characterized the Pareto tails of the firm-size distribution and derive a formula that characterizes how they are linked to the forces that determine innovation, firm entry, and aggregate growth.

We estimated the model parameters for 1995 and 2014 and interpreted the increase in firm size along both the intensive and extensive margins in the data as reflecting fundamental economic changes. We found that the cost for external innovation declined for firms that are actively expanding with new establishments. Moreover, the model infers that the cost of entry has also declined. We decomposed the effect of model parameters on model outcomes to reveal that the largest contributors to the recent dominance of large firms with many establishments are the decline in the external innovation cost, decline in the establishment exit rate, and the slowdown of the aggregate productivity growth for production.

An important future research agenda is to explain why these changes occurred during this time period. Numerous anecdotes indicate finding new locations for stores and restaurants became easier due to increasing availability of ICT and various "big data."[31] This change in technology may have contributed to the lower cost of external innovation. Faster information flows better enables a new business to succeed early, but also allows business models to be imitated more easily, creating a faster obsolescence and hampering the ability to grow quickly through the extensive margin. Our empirical and theoretical results can serve as a starting point for further investigations into these recent changes in the economic environment.

[31] Brynjolfsson and McElheran (2016) show that, in the U.S. manufacturing sector, larger plants and plants that belong to multi-unit firms have higher tendencies to adopt data-driven decision making. Ganapati (2018) highlights the surge of IT investment in the wholesale industry, which also experienced the increase in the market share of the largest 1% firms. Begenau et al. (2018) argue that the use of big data in financial markets has lowered the cost of capital for large firms relative to small firms. The dominance of large firms with big data raises a new set of normative and policy questions. Jones and Tonetti (2020) discusses various issues that arises in the "data economy" using a macroeconomic model that explicitly considers information flow within and across firms.

References

Acemoglu, D. and D. Cao (2015). Innovation by Entrants and Incumbents. *Journal of Economic Theory 157*, 255–294.

Aghion, P., A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li (2019). A Theory of Falling Growth and Rising Rents. NBER Working Paper 26448.

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

Argente, D., M. Lee, and S. Moreira (2019). The Life Cycle of Products: Evidence and Implications. Working paper.

Atalay, E., A. H. csu, M. J. Li, and C. Syverson (2019). How Wide Is the Firm Border? *Quarterly Journal of Economics 134*, 1845–1882.

Atkinson, A. B., T. Piketty, and E. Saez (2011, March). Top incomes in the long run of history. *Journal of Economic Literature 49*(1), 3–71.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020, 02). The Fall of the Labor Share and the Rise of Superstar Firms*. *The Quarterly Journal of Economics 135*(2), 645–709.

Axtell, R. L. (2001). Zipf Distribution of U.S. Firm Sizes. *Science 293*(5536), 1818–1820.

Begenau, J., M. Farboodi, and L. Veldkamp (2018). Big data in finance and the growth of large firms. NBER Working Paper 24550.

Bernard, A. B., S. J. Redding, and P. K. Schott (2010). Multi-Product Firms and Product Switching. *American Economic Review 100*, 70–97.

Bingham, N., C. Goldie, and J. Teugels (1987). *Regular Variation.* Cambridge University Press.

Brynjolfsson, E. and K. McElheran (2016). The Rapid Adoption of Data-Driven Decision-Making. *American Economic Review 106*, 133–139.

Cao, D., H. Hyatt, T. Mukoyama, and E. Sager (2020). Firm Growth through New Establishments. GCER Working Paper 20-20-02.

Cao, D. and W. Luo (2017). Persistent Heterogeneous Returns and Top End Wealth Inequality. *Review of Economic Dynamics 26*, 301 – 326.

Choi, E. J. and J. R. Spletzer (2012). The Declining Average Size of Establishments: Evidence and Explanations. *Monthly Labor Review*, 50–65.

De Loecker, J., J. Eeckhout, and G. Unger (2020, 01). The Rise of Market Power and the Macroeconomic Implications. *The Quarterly Journal of Economics 135*(2), 561–644.

Decker, R., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2014). The Role of Entrepreneurship in U.S. Job Creation and Economic Dynamism. *Journal of Economic Perspectives 28*, 3–24.

Dinlersoz, E., S. Kalemli-Ozcan, H. Hyatt, and V. Penciakova (2018). Leverage over the life cycle and implications for firm growth and shock responsiveness. NBER Working Paper 25226.

Fattal Jaef, R. N. (2018, April). Entry and exit, multiproduct firms, and allocative distortions. *American Economic Journal: Macroeconomics 10*(2), 86–112.

Gabaix, X. (2009). Power Laws in Economics and Finance. *Annual Review of Economics 1*(1), 255–294.

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016). The Dynamics of Inequality. *Econometrica 84*(6), 2071–2111.

Ganapati, S. (2018). The Modern Wholesaler: Global Sourcing, Domestic Distribution, and Scale Economies. Working paper.

Garca-Macia, D., C.-T. Hsieh, and P. J. Klenow (2019). How Destructive is Innovation? *Econometrica 87*, 1507–1541.

Garetto, S., L. Oldenski, and N. Ramondo (2019). Multinational Expansion in Time and Space. Working paper.

Grimmett, G. and D. Stirzaker (2001). *Probability and Random Processes, Third Edition*. Oxford University Press.

Gumpert, A., H. Steimer, and M. Antoni (2019). Firm organization with multiple establishments. CESifo Working Paper 7435.

Gutiérrez, G. and T. Philippon (2017). Declining Competition and Investment in the U.S. NBER Working Paper 23583.

Hathaway, I. and R. E. Litan (2014). Declining Business Dynamism in the United States: A Look at States and Metros. Brookings Institute.

Hershbein, B., C. Macaluso, and C. Yeh (2019). Concentration in U.S. Local Labor Markets: Evidence from Vacancy and Employment Data. Working paper.

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

Hopenhayn, H., J. Neira, and R. Singhania (2018, December). From population growth to firm demographics: Implications for concentration, entrepreneurship and the labor share. Working Paper 25382, National Bureau of Economic Research.

Hopenhayn, H. and R. Rogerson (1993). Job Turnover and Policy Evaluation: A General Equilibrium Analysis. *Journal of Political Economy 101*(5), 915–938.

Hsieh, C.-T. and E. Rossi-Hansberg (2019). The Industrial Revolution of Services. NBER Working Paper 25968.

Jones, C. I. (1995). R&D-Based Models of Economic Growth. *Journal of Political Economy*, 759–784.

Jones, C. I. and J. Kim (2018). A schumpeterian model of top income inequality. *Journal of Political Economy 126*(5), 1785–1826.

Jones, C. I. and C. Tonetti (2020, September). Nonrivalry and the economics of data. *American Economic Review 110*(9), 2819–58.

Kehrig, M. and N. Vincent (2019). Good Dispersion, Bad Dispersion. NBER Working Paper 25923.

Klette, T. J. and S. Kortum (2004). Innovating Firms and Aggregate Innovation. *Journal of Political Economy 112*(5), 986–1018.

Kondo, I. O., L. Lewis, and A. Stella (2019). On the U.S. Firm and Establishment Size Distributions. Working paper.

Luttmer, E. G. J. (2011). On the Mechanics of Firm Growth. *The Review of Economic Studies 78*(3), 1042–1068.

Mimica, A. (2016). Exponential Decay of Measures and Tauberian Theorems. *Journal of Mathematical Analysis and Applications 440*(1), 266 – 285.

Moreira, S., D. Argente, E. Oberfield, and V. Venkateswaran (2021). Scalable Expertise. Working paper.

Mukoyama, T. and S. Osotimehin (2019). Barriers to Reallocation and Economic Growth: The Effects of Firing Costs. *American Economic Journal: Macroeconomics 11*, 235–270.

Oberfield, E., E. Rossi-Hansberg, P.-D. Sarte, and N. Trachter (2020). Plants in Space. Working paper.

Peters, M. (2019). Heterogeneous Markups, Growth and Endogenous Misallocation. Econometrica forthcoming.

Reed, W. J. (2001). The Pareto, Zipf and Other Power Laws. *Economics Letters 74*(1), 15 – 19.

Rinz, K. (2018). Labor Market Concentration, Earnings Inequality, and Earnings Mobility. CARRA Working Paper 2018-10.

Rossi-Hansberg, E., P.-D. Sarte, and N. Trachter (2021). Diverging trends in national and local concentration. *NBER Macroeconomics Annual 35*, 115–150.

Sager, E. and O. A. Timoshenko (2019). The Double EMG Distribution and Trade Elasticities. *Canadian Journal of Economics 52*(4), 1523–1557.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2018, 10). Firming Up Inequality. *The Quarterly Journal of Economics 134*(1), 1–50.

Appendix

A Data and empirical documentation

A.1 Quarterly Census of Employment and Wages

This data appendix describes the Quarterly Census of Employment and Wages (QCEW) and draws heavily from the BLS Handbook of Methods.[32]

A.1.1 Definitions

The Quarterly Census of Employment and Wages (QCEW) is a count of employment and wages obtained from quarterly reports filed by almost every employer in the U.S., Puerto Rico and the U.S. Virgin Islands, for the purpose of administering state unemployment insurance programs. These reports are compiled by the Bureau of Labor Statistics (BLS) and supplemented with the Annual Refiling Survey and the Multiple Worksite Report for the purpose of validation and accuracy. The reports include an establishment's monthly employment level upon the twelfth of each month and counts any employed worker, whether their position is full time, part time, permanent or temporary. Counted employees include most corporate officials, all executives, all supervisory personnel, all professionals, all clerical workers, many farmworkers, all wage earners and all piece workers. Employees are counted if on paid sick leave, paid holiday or paid vacation. Employees are not counted if they did not earn wages during the pay period covering the 12th of the month, because of work stoppages, temporary layoffs, illness, or unpaid vacations. The QCEW does not count proprietors, the unincorporated self-employed, unpaid family members, certain farm and domestic workers that are exempt from reporting employment data, railroad workers covered by the railroad unemployment insurance system, all members of the Armed Forces, and most student workers at schools. If a worker holds multiple jobs across multiple firms, then that worker may be counted more than once in the QCEW.

[32] See `https://www.bls.gov/opub/hom/cew/home.htm` for the complete BLS Handbook of Methods.

A.1.2 BLS Sample

A sample we used as part of the BLS visiting researcher program provided data from 1990 to 2016 and covers thirty-eight states: Alaska, Alabama, Arkansas, Arizona, California, Colorado, Connecticut, Delaware, Georgia, Hawaii, Iowa, Idaho, Indiana, Kansas, Louisiana, Maryland, Maine, Minnesota, Montana, North Dakota, New Jersey, New Mexico, Nevada, Ohio, Oklahoma, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, Vermont, Washington, West Virginia, as well as the District of Columbia, Puerto Rico and the U.S. Virgin Islands.

A.1.3 LEHD Sample

The Employer Characteristics File maintained by the Longitudinal Employer-Household Dynamics program provided data for twenty-eight states: Alaska, Arizona, California, Colorado, Florida, Georgia, Iowa, Idaho, Illinois, Indiana, Kansas, Louisiana, Maryland, Missouri, Montana, North Carolina, New Mexico, New York, Pennsylvania, Oregon, Rhode Island, South Dakota, Texas, Utah, Washington, Wisconsin, West Virginia, and Wyoming.

A.1.4 Data cleaning and variable construction

To conform to official statistics, we clean the data in accordance with BLS procedure. First, while the QCEW contains monthly data as of the 12th of each month, we follow BLS convention by only using data from the final month within a quarter. As a result, our sample does not capture establishments that enter and exit within the same quarter. We additionally exclude firms from calculations in a given quarter if the absolute change in employment from the previous quarter exceeds 10 times the average employment between the two quarters. Statistics within this paper are not sensitive to the choice of multiple being 10.

We construct firms as collections of establishment with the same employment identification numbers (EINs). Firm-level employment is the sum of all employment in establishments associated with the same EIN and the number of establishments within a firm as the number of establishments that report using a common EIN.

To classify a firm's industry, we assign to a firm the average self-reported, 6-digit NAICS code of its establishments so that the firm is classified in the same way as its establishments are on average.

A firm's entry date is measured as the date at which the QCEW records a non-zero number of workers associated with a particular EIN after four consecutive quarters of recording zero workers. A firm's exit date is measured as the last date at which the QCEW records a non-zero number of workers associated with a particular EIN prior to four consecutive quarters of recording zero workers. A firm's age is measured by tracking firms after entering. Upon entry, the firm is assigned an age of 1 quarter and the firm's age is incremented by 1 quarter for each period that it does not exit.

B Further decomposition of time-series of margins

First, we examine the firm-size decomposition in the manufacturing, service, and agricultural sectors, and find that a significant increase in the number of service-sector establishments per firm is the driving force for the economy-wide increase in average firm size. Figure A.1(a) plots the evolution of average firm size in each sector, compared to the average firm size in 1990. We observe that all sectors excluding agriculture experienced an increase in average firm size over the sample period (1990-2014), but the service sector experienced the largest size increase. This observation is notable, because the service sector employs the majority of U.S. workers over this period. To account for sectoral firm size growth, we turn to the intensive and extensive margins plotted in Figures A.1(c) and A.1(e), respectively. Each sector's intensive margin exhibits a flat or slight downward trend similar to that in the overall economy. By strong contrast, the extensive margin for different sectors delivers the same message as the extensive margin in the overall economy, namely, that the growth in the number of establishments per firm accounts for the overall increase in average firm size across sectors and, by extension, the overall economy.

Next, we examine the firm-size decomposition conditional on firm size and find the establishment-driven growth in average firm size is concentrated in the economy's largest firms. Figure A.1(b) calculates the average size within size bins. We see a

Table A.1: Slope of the Size-Rank Relationship

	Firm size	**Extensive**	**Intensive**
	95th percentile and above		
1995	−1.10	−1.20	−1.35
2014	−0.99	−1.17	−1.32
	99th percentile and above		
1995	−1.17	−1.25	−1.39
2014	−0.99	−1.21	−1.24

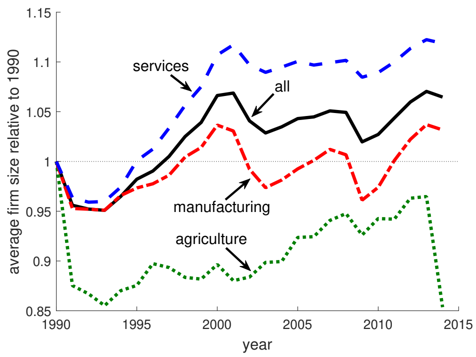
Notes: Authors' calculations of of LEHD microdata. Linear regression of log outcome on log rank for fitted values at or above the 95th percentile from a polynomial approximation of microdata. Slopes correspond to March of the respective year.

pattern of spreading out: very small firms with one to four employees have tended to become smaller, whereas the average size of larger firms with 100 employees increased over time. If we examine the very right tail of firms with 5,000 workers or more, firm size has been increasing over time since 1997, with a similar increase in firms that have 100 employees or more. The intensive margin does not exhibit an obvious relationship with firm size, as seen in Figure A.1(d). None of the series have an increasing trend, and in fact, the overall time-series pattern looks similar between very small firms (1 to 4 employees) and very large firms (5,000 or more employees) except for a spike for very large firms in the early 2000s. By contrast, growth in the average number of establishments per firm exhibits very different trends between small and large firms, as shown in Figure A.1(f). Very small firms are predominantly single-establishment firms over the entire sample period. Medium-size firms with 5 to 99 employees have had a modest increase in the number of establishments. Larger firms have had a startling increase in the number of establishments. On average, the firms with 5,000 or more employees had about four times more establishments in 2014 than in 1990. Thus, we conclude that a key mechanism that has generated the increase in firm size in recent years is expansion through the number of establishments in very large firms.

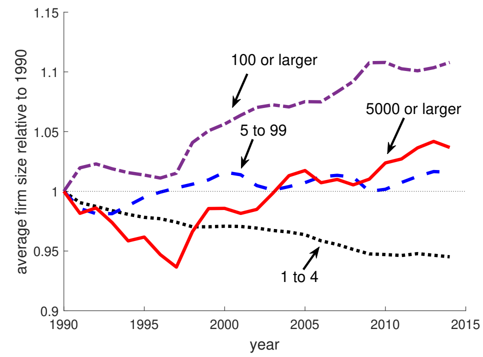
To see the behavior of the distribution at the right tail, we measure the slope of the upper percentiles of the firm-size distribution in Table A.1. Here, we include

Figure A.1: Average Firm Size, Intensive and Extensive Margins by Sector and Size Bins

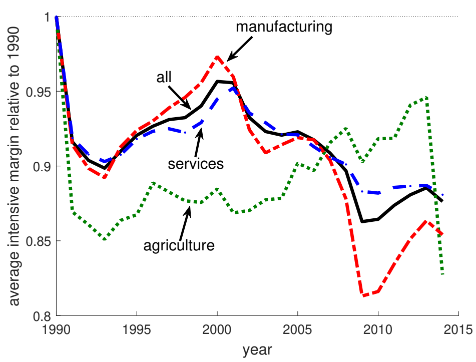
(a) Average firm size by sector (number of workers)



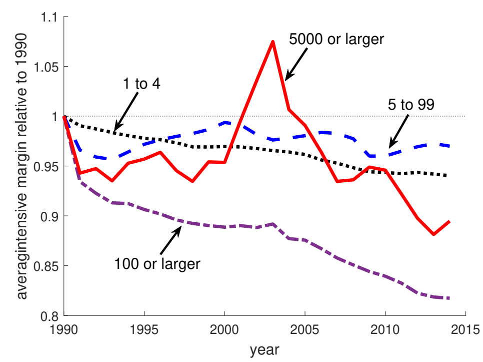
(b) Average firm size by size bin (number of workers)



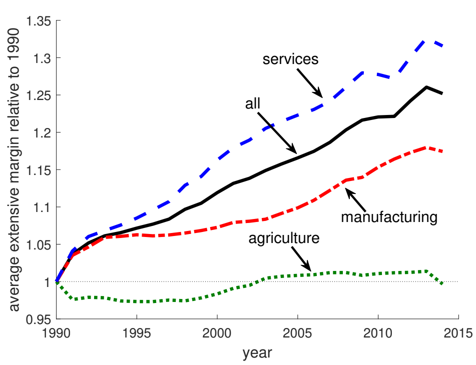
(c) Average intensive margin by sector



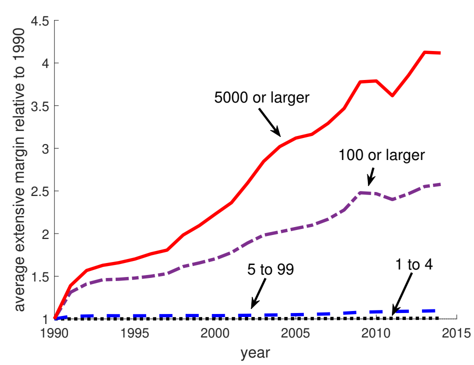
(d) Average intensive margin by size bin



(e) Average extensive margin by sector



(f) Average extensive margin by size bin



Source: Author's calculations of Quarterly Census of Employment and Wages microdata.

predicted values that are at or above the 95th percentile. The firm size distribution in 2014 has a slope that is close to (negative) 1, which indicates it has a very fat tail. Both the extensive margin and the intensive margin have steeper slopes than the employment distribution, which implies thinner tails. For the overall firm size, the right tail became thicker—looking at the 99 percentile and above, the tail index changed from -1.17 to -0.99 between 1995 and 2014. Table A.1 indicates both the extensive and intensive margins contributed to this thickening of the slope over time.

Online Appendix

C LEHD measures of size-rank relationships

C.1 Confidentiality protection for size-rank statistics in LEHD

Characterizing the employment distribution by firm rank, we used the Employer Characteristics File maintained by the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program. A number of steps were taken to minimize the disclosure risk associated with the release of statistics on the upper ranks of the firm size distribution. First, instead of a direct size-rank regression, we coarsened the underlying distribution, employing finer categories the closer we are at the firm size distribution. Coarsening ensures that there are a large number of observations in each cell, even at the upper range of the distribution in which we use the finest categories. To limit disclosure risk further, we estimated a fifth-order polynomial (plus a constant) on the (average) percentile rank associated with each rank category. To conform to U.S. Census Bureau disclosure requirements, all point estimates and standard errors were rounded to four significant digits.

Table A.2: Fifth-order polynomial approximations of size-rank relationships

	Employment			Establishments			Establishment Size		
	1995	2005	2014	1995	2005	2014	1995	2005	2014
Intercept	16.33	9.586	4.932	11.55	14.11	17.37	13.87	7.406	2.725
	(0.04863)	(0.04889)	(0.04988)	(0.06931)	(0.06404)	(0.06247)	(0.04857)	(0.04882)	(0.04979)
$ln(rank)$	-6.478	0.4619	5.243	-6.302	-8.406	-11.36	-5.143	1.333	6.056
	(0.04041)	(0.4060)	(0.04145)	(0.05760)	(0.5319)	(0.05191)	(0.04036)	(0.04055)	(0.04137)
$ln(rank)^2$	2.224	-0.2067	-1.888	2.274	2.961	3.919	1.723	-0.5921	-2.268
	(0.01241)	(0.1247)	(0.01273)	(0.01769)	(0.01633)	(0.01594)	(0.01239)	(0.01245)	(0.01270)
$ln(rank)^3 \times 10$	-3.936	-0.09348	2.583	-4.121	-5.157	-6.550	-3.055	0.6863	3.381
	(0.01785)	(0.01793)	(0.0183)	(0.02544)	(0.02348)	(0.02292)	(0.01782)	(0.01790)	(0.01827)
$ln(rank)^4 \times 10^2$	3.206	0.3817	-1.605	3.354	4.078	5.012	2.528	0.2700	-2.283
	(0.01219)	(0.01225)	(0.01250)	(0.0174)	(0.01604)	(0.0157)	(0.01218)	(0.01223)	(0.01248)
$ln(rank)^5 \times 10^3$	-9.863	-0.2010	0.3571	-0.9933	-1.184	-1.419	0.7974	0.009459	0.5591
	(0.003199)	(0.003213)	(0.003280)	(0.00456)	(0.004209)	(0.004108)	(0.003195)	(0.003208)	(0.003274)

Notes: Authors' calculations of LEHD microdata. Standard errors are in parentheses. Estimates characterize March of each respective year. All point estimates and standard errors were rounded to four significant digits to conform to U.S. Census Bureau discloure requirements. The dependent variables employment, establishments, and average establishment size are in logs.

C.2 Polynomial estimation procedure and results

We now describe the polynomial approximations that allow us to characterize the distributions of three measures of firm size: total employment, the number of establishments, and average establishment size. We first ranked firm-level data by each of these different size measures. Using these ranks, we started with the smallest rankings, and assigned categories based an observation being within percentile ranges. The ranges are defined as follows.

1. Starting with the lowest, group observations into 1% bins until the 95th percentile is reached, for a total of 95 categories.
2. Group observations into 0.5% bins until the 99th percentile is reached, for a total of 8 categories.
3. Group observations into 0.1% bins until the 99.9th percentile is reached, for a total of 9 categories.
4. Group the remaining observations into 0.01% bins, for a total of 10 categories.

Using this method creates a total of 122 categories. This method of grouping the data was meant to provide a balance between generating information that can be informative about the tails of the distributions that we are interested in, which protecting the confidentiality of the underlying microdata: even the finest cells have a relatively large number of observations (e.g., 0.01% $\times$ 5 million $=$ 500 observations). Each bin was assigned its average percentile rank (e.g., the lowest bin has an average percentile rank of 0.5, the next has an average percentile rank of 1.5, etc.). Polynomial approximations of our size measures use a transformation of this: $log((100 - \text{average percentile}) \times 1000)$. The transformation times 1000 was done for computational reasons, but conceptually is just a simple shift of the intercept because $log((100 - \text{average percentile}) \times 1000) = log(100 - \text{average percentile}) + log(1000)$. Fifth-order polynomials of this transformation of the average percentile rank (plus a constant) serve as regressors for each size measure.

Our dependent variables consist of the lograrithm of each size measure for size in March of 1995, 2005, and 2014. To avoid approximating the discrete jumps in the distribution between small values (1, 2, 3, etc.) a random draw form the interval $[-0.5\ 0.5]$ is applied to each observation. Results of these regressions are shown in Appendix Table A.2.

C.3 Recovering the size distribution from the polynomial approximations

While our polynomial approximations capture much of the rich features of the underlying microdata, they are difficult to immediately interpret. We therefore transform these polynomial estimates to create tables and figures that highlight important features of the underlying microdata. One feature of the data that we wish to highlight is that our estimation was done on coarsened (discritized) data. There are maximum and minimum values, and the are discrete values that the independent variable takes. Another feature of the underlying microdata for our size distribution is that total employment and the number of establishments are discrete, while our polynomials are of course continuous. While in practice average establishment size takes values other than integers, at most points in the distribution there are almost no multi-establishment firms and so the average establishment size is approximately a step function, especially for low values. To highlight these features of the data, we round each size measure after an exponential transformation.

We consider the relationships between employment rank and the number of establishments and establishment size. In this case, the size measure is not related to its rank, but to the rank of total employment. This is useful because when the data are so ranked, $log(\text{employment}) = log(\text{establishments}) + log(\text{establishment size})$. However, in this case, rounding no longer captures the salient features of moderate levels of employment (around employment of 10) where the number of establishments becomes distinct from zero. To capture these relationships, we fix total employment as the rounded value of total employment, and we estimate the number of establishments and average establishment size using Kuhn-Tucker optimization.

Let the log of the unrounded predicted value of total employment be e, the log number of establishments be p (plants) and the log average size be w (workers per establishment), both of which will be estimated based on unrounded predicted values

$\hat{p}$ and $\hat{w}$, respectively. Specifically, we minimize the squared distance between the estimated value and the polynomial approximation subject to three constraints. First, total log employment is the sum of the logs of the number of establishments and workers per establishment, and so $e = p + w$, and the value of this constraint is μ. Second, the total log number of plants must be at least one and so $p \geq 0$, which has value λ_1. Third, the total log number of workers per plant must be at least one and so $w \geq 0$, which has value λ_2. The Kuhn-Tucker problem is now:

$$\min_{p,w,\mu,\lambda_1,\lambda_2} (p - \hat{p})^2 + (w - \hat{w})^2 + \mu(e - p - w) + \lambda_1 p + \lambda_2 w$$

subject to the inequality constraints $p \geq 0$ and $e \geq 0$.

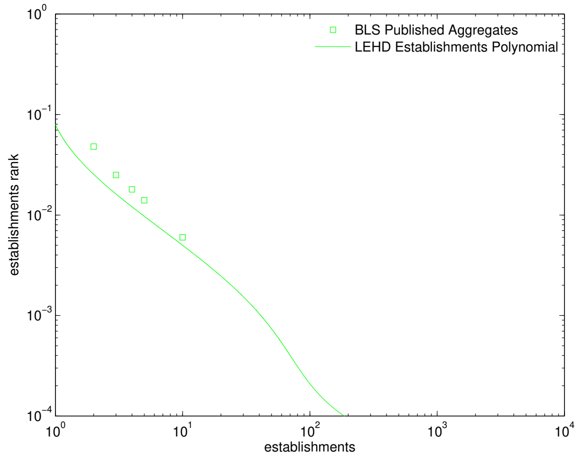
At an interior solution, $p > 0$ and $w > 0$, and so $\lambda_1 = \lambda_2 = 0$. In this case, $2(p - \hat{p}) = 2(w - \hat{w})$ and we can substitute the constraint $e = p + w$ to recover

$$\begin{aligned} p &= \frac{e + \hat{p} - \hat{w}}{2} \\ w &= \frac{e - \hat{p} + \hat{w}}{2}. \end{aligned}$$

Otherwise, at least one of the inequality constraints is binding and the solution is set to minimize the criterion function. In practice, this means that the number of establishments is set to zero when log total employment is greater than zero and an interior solution does not hold. The rounded employment counts e^* are then used to generate the final estimated number of establishments p^* and workers per establishment w^* according to:

$$\begin{aligned} p^* =& e^* \times \frac{p}{p + w} \\ w^* =& e^* \times \frac{w}{p + w}. \end{aligned}$$

Figure A.2: Comparison of Polynomial Approximation to Published Establishment Number Totals for 2005



Notes: Author's calculations of Longitudinal Employer-Household Dynamics microdata and Bureau of Labor Statistics published aggregates.

C.4 Comparison to BLS number of establishments distribution

Published aggregates from the Bureau of Labor Statistics (BLS) provide an opportunity to benchmark the number of establishments distribution for 2005.[33] Figure A.2 compare the approximated values from the LEHD estimation with the BLS published aggregates on a log-size, log-percentile rank scale. The BLS published aggregates indicate that 95.2% of firms have only one establishment. This implies that the upper 4.8% of the number of establishments distribution has two or more establishments. The LEHD data suggest that a smaller share of firms, 3.7%, have multiple establishments. The BLS published aggregates indicate that 2.5% of firms have three or more establishments, while the LEHD data indicate a share of 2.0%. The BLS published aggregates indicate that 1.8% of firms have four or more establishments, while the LEHD data indicate a share of 1.4%. The BLS published aggregates indicate that 1.4% of firms have four or more establishments, while the LEHD data indicate a

[33]Numbers for 2005 Q1 are taken from `https://www.bls.gov/bdm/sizeclassqanda.htm` (last accessed: October 4, 2019).

share of 1.1%. The BLS aggregates also indicate that 0.6% of firms have ten or more establishments, while the LEHD data indicate a share of 0.5%. Fitting a line to these five data points provide somewhat different slopes for the log-size, log-rank relationship. The natural explanation for these differences is that our LEHD microdata is for a 28-state subset, while the BLS published aggregates are national. Despite a level difference, the slopes are similar. The slope of this relationship is -1.27, while that of the LEHD using these same five data points is -1.23. The steeper slope of the BLS published aggregates suggests a somewhat thinner tail. Yet both of these estimated slopes are close to the slope estimates of -1.15 and -1.21 for the 95th, and 99th percentiles and above, respectively. Results using LEHD data include many more percentiles for the upper percentiles of the distribution, and so we naturally prefer these when we target moments for estimation.

D Distributional analyses

D.1 Derivations of Kolmogorov equations in Section 4.1

First note that $\bar{\mathcal{M}}_\tau(n) = \mathcal{M}_\tau(n, 0)$, therefore (11) is a special case of (13) with $\hat{q} = 0$. To derive the latter, let $\hat{\mathcal{M}}_{n,t,q}$ denote the measure of firms with n establishments and with each establishment having quality of at least $\hat{q}Q_t$. Then

$$\begin{aligned}
\hat{\mathcal{M}}_\tau\left(1, \hat{q}; t + \Delta t\right) = \hat{\mathcal{M}}_\tau &\left(1, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right) - (z_X^\tau + \delta_\tau + d_\tau)\Delta t \hat{\mathcal{M}}_\tau \left(1, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right) \\
&+ 2\delta_\tau \Delta t \hat{\mathcal{M}}_\tau \left(2, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right) + \mu_e m_\tau \Delta t N_t \left(1 - \Phi\left(\frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}\right)\right) \\
&+ \sum_{\tau' \neq \tau} \lambda_{\tau'\tau} \Delta t \hat{\mathcal{M}}_{\tau'} \left(1, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right) - \sum_{\tau' \neq \tau} \lambda_{\tau\tau'} \Delta t \hat{\mathcal{M}}_\tau \left(1, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right)
\end{aligned}$$

holds. Subtracting $\hat{\mathcal{M}}_\tau(n,\hat{q};t)$ from both sides, dividing by Δt, and take the limit $\Delta t \to 0$, we obtain

$$\begin{aligned}\frac{\partial \hat{\mathcal{M}}_\tau(1,\hat{q};t)}{\partial t} &= -\hat{q}(z_I^\tau - \zeta)\frac{\partial \hat{\mathcal{M}}_\tau(1,\hat{q};t)}{\partial \hat{q}} - (z_X^\tau + \delta_\tau + d_\tau)\hat{\mathcal{M}}_\tau(1,\hat{q};t)\\ &+ 2\delta_\tau \hat{\mathcal{M}}_\tau\left(2, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right) + \mu_e m_\tau N_t\left(1 - \Phi(\hat{q})\right)\\ &+ \sum_{\tau'\neq\tau}\lambda_{\tau'\tau}\hat{\mathcal{M}}_{\tau'}(1,\hat{q};t) - \sum_{\tau'\neq\tau}\lambda_{\tau\tau'}\hat{\mathcal{M}}_\tau(1,\hat{q};t).\end{aligned}$$

Now $\mathcal{M}_\tau(1,\hat{q};t) = \frac{\hat{\mathcal{M}}_\tau(1,\hat{q};t+\Delta t)}{N_t}$, thus

$$\begin{aligned}\frac{\partial \mathcal{M}_\tau(1,\hat{q};t)}{\partial t} + \eta\mathcal{M}_\tau(1,\hat{q};t) &= -\hat{q}(z_I^\tau - \zeta)\frac{\partial \mathcal{M}_\tau(1,\hat{q};t)}{\partial \hat{q}} - (z_X^\tau + \delta_\tau + d_\tau)\mathcal{M}_\tau(1,\hat{q};t)\\ &+ 2\delta_\tau \mathcal{M}_\tau\left(2, \frac{\hat{q}}{\exp\left((z_I^\tau - \zeta)\Delta t\right)}; t\right) + \mu_e m_\tau\left(1 - \Phi(\hat{q})\right)\\ &+ \sum_{\tau'\neq\tau}\lambda_{\tau'\tau}\mathcal{M}_{\tau'}(1,\hat{q};t) - \sum_{\tau'\neq\tau}\lambda_{\tau\tau'}\mathcal{M}_\tau(1,\hat{q};t).\end{aligned}$$

On a stationary BGP, $\frac{\partial \mathcal{M}_\tau(1,\hat{q};t)}{\partial t} = 0$, so we obtain (13) for $n = 1$. The steps for deriving (13) for $n > 1$ is similar. Equation (12) can be derived with similar steps as Cao and Luo (2017).

D.2 Analysis of distribution in one-type economy in Section 4.2

For the characterization of the distributions, we use the following mathematical notations: for two strictly positive functions f, g defined over $(0, x^*)$, where $x^* \in \mathbb{R}^* \cup \{+\infty\}$,

$$f(x) \sim_{x\to x^*} g(x)$$

if $\lim_{x\to x^*}\frac{f(x)}{g(x)} = 1$;[34] and

$$f(x) \propto_{x\to x^*} g(x)$$

[34]This $\sim$ notation follows the regular variation literature (Bingham et al. (1987)).

if $f(x) \sim_{x \to x^*} ag(x)$ for some $a > 0$. Using this notation, a random variable $\mathbf{X}$ defined over $\mathbb{R}^*$ has a Pareto tail with tail index $\xi > 0$ if $\Pr(\mathbf{X} > x) \propto_{x \to \infty} x^{-\xi}$.

In the case with only one type, our model assumptions imply that for a given firm, the quality (and therefore the size) of each establishment grows at a deterministic rate z_I that is common across all firms. The average quality $Q(t)$ grows at the rate ζ. Thus, the quality of the establishments in a firm that entered at time t_0 and whose initial draw of the normalized quality is $\hat{q}Q(t_0)$ can be represented as (denoting it by $q_{t_0}(t)$)

$$q_{t_0}(t) = \hat{q}Q(t_0)e^{z_I(t-t_0)} = \hat{q}Q(t)e^{(z_I-\zeta)(t-t_0)}.$$

From the labor demand (5) and the labor market equilibrium condition, it is straightforward to show that along the balanced-growth equilibrium, the relative labor demand $\ell(t)/L(t)$ of a particular establishment with quality $q_{t_0}(t)$ is equal to $q_{t_0}(t)/(N(t)Q(t))$. Therefore, the cross-sectional distribution of establishment size at a given time t is the same as the distribution of $\hat{q}e^{(z_I-\zeta)(t-t_0)}$. Denoting the time-$t$ number of establishments for a firm that starts at time t_0 as $n_{t_0}(t)$ (note that $n_{t_0}(t)$ is stochastic as the external innovation is random), the (relative) firm-size distribution follows the distribution of $n_{t_0}(t)\hat{q}e^{(z_I-\zeta)(t-t_0)}/N(t)$.

For the distribution of establishment sizes, equation (12) becomes

$$(z_I - \zeta)\hat{q}\frac{d\bar{\mathcal{H}}(\hat{q})}{d\hat{q}} = -(\delta + d + \eta - z_X)\bar{\mathcal{H}}(\hat{q}) + \mu_e(1 - \Phi(\hat{q})).$$

Let us use the change of variables $p \equiv \log(\hat{q})$ and $\tilde{\mathcal{H}}(p) \equiv \bar{\mathcal{H}}(\exp(p))$ to rewrite this equation as

$$(z_I - \zeta)\frac{d\tilde{\mathcal{H}}(p)}{dp} = -(\delta + d + \eta - z_X)\tilde{\mathcal{H}}(p) + \mu_e(1 - \Phi(\exp(p))).$$

This equation is a first-order ODE that has a general solution:

$$\tilde{\mathcal{H}}(p) = e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}(\underline{p}-p)}\tilde{\mathcal{H}}(\underline{p}) + \int_{\underline{p}}^{p} e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}(\tilde{p}-p)}\frac{\mu_e}{z_I-\zeta}(1 - \Phi(\exp(\tilde{p})))d\tilde{p},$$

for each $\underline{p}$. Taking the limit $\underline{p} \to -\infty$, and replacing μ_e with $\delta + d + \eta - z_X$,

$$\tilde{\mathcal{H}}(p) = \int_{-\infty}^{p} e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}(\tilde{p}-p)} \frac{\delta + d + \eta - z_X}{z_I - \zeta}(1 - \Phi(\exp(\tilde{p})))d\tilde{p},$$

which is (14) in the main text.

For the distribution of the number of establishments per firm, (11) becomes

$$0 = -(z_X + \delta + d + \eta)\bar{\mathcal{M}}(1) + 2\delta\bar{\mathcal{M}}(2) + \mu_e \tag{21a}$$

and

$$0 = -(n(z_X + \delta) + d + \eta)\bar{\mathcal{M}}(n) + (n+1)\delta\bar{\mathcal{M}}(n+1) + (n-1)z_X\bar{\mathcal{M}}(n-1) \tag{21b}$$

for $n > 1$. Luttmer (2011) provides a closed-form solution for $\{\bar{\mathcal{M}}(n)\}_{n=1}^{\infty}$:

$$\bar{\mathcal{M}}(n) = \frac{1}{n}\frac{\mu_e}{z_X}\sum_{k=0}^{\infty}\frac{1}{\upsilon_{n+k}}\left(\prod_{m=n}^{n+k}\upsilon_m\right)\prod_{m=1}^{n+k}\frac{z_X\upsilon_m}{\delta}, \tag{22}$$

where the sequence $\{\upsilon_n\}_{n=0}^{\infty}$ is defined recursively by $\upsilon_0 = 0$ and

$$\frac{1}{\upsilon_{n+1}} = 1 - \frac{z_X\upsilon_n}{\delta} + \frac{\eta + d + z_X n}{\delta n}.$$

The distribution of establishment number is given by a discrete random variable $\mathbf{X}$ with pdf

$$\Pr(\mathbf{X} = n) = \frac{\bar{\mathcal{M}}(n)}{\sum_{n'}\bar{\mathcal{M}}(n')}. \tag{23}$$

Because we normalize the measure of n-establishment firms by the total number of establishments $N(t)$, $\sum_n n\bar{\mathcal{M}}(n) = 1$, and thus $\sum_n \bar{\mathcal{M}}(n) < 1$. In Appendix D.3, we show that when $z_X > \delta$, it has a Pareto tail with the tail index given by $\frac{\eta+d}{z_X-\delta}$, which is the expression in (16).

We consider the firm size distribution in two cases. In the case where the initial draw satisfies $\int \hat{q}d\Phi(\hat{q}) = 1$, $z_I = \zeta$ holds. In this case, the random variable for firm size $\mathbf{Z}$ can be written as $\mathbf{Z} = \mathbf{XY}$, where $\mathbf{X}$ is the number of establishments and $\mathbf{Y}$

is establishment size in a firm. In the cross section, $\mathbf{X}$ and $\mathbf{Y}$ are independent and the cdf of $\mathbf{Y}$ is given by Φ. The pdf for $\mathbf{X}$ is given by (23).

Therefore, the fraction of firms with size $\mathbf{Z}(t) \geq \hat{\mathbf{z}}$, denoted by $\mathbf{M}(\hat{\mathbf{z}})$, can be computed as[35]

$$\mathbf{M}(\hat{\mathbf{z}}) = \frac{\sum_n \bar{\mathcal{M}}(n)(1 - \Phi(\hat{\mathbf{z}}/n))}{\sum_n \bar{\mathcal{M}}(n)}.$$

To determine the tail index of $\mathbf{M}(\cdot)$, we consider the Laplace transformation:[36]

$$\varphi(s) = \int_0^\infty \hat{\mathbf{z}}^s \left(-d\mathbf{M}(\hat{\mathbf{z}})\right). \tag{24}$$

Using the expression for $\mathbf{M}$ above, we show in Appendix D.3 that

$$\varphi(s) = \left\{\int_0^\infty \hat{\mathbf{z}}^s d\Phi(\hat{\mathbf{z}})\right\} \left\{\frac{\sum_n \bar{\mathcal{M}}(n) n^s}{\sum_n \bar{\mathcal{M}}(n)}\right\}.$$

Assuming that the entry distribution Φ has a thin tail and using the characterization in Proposition 1, we show in Appendix D.3 that

$$\varphi(s) \propto \frac{1}{\frac{\eta+d}{z_X-\delta} - s}$$

as $s \uparrow \frac{\eta+d}{z_X-\delta}$. Therefore, by the Tauberian theorem in Mimica (2016, Corollary 1.3), $\mathbf{M}$ has a Pareto tail with the tail index given by (16).

The case where $z_I \neq \zeta$, i.e. $\int \hat{q} d\Phi(\hat{q}) \neq 1$, is more challenging because the dynamics of firm size are driven both by the dynamics of the establishment number and of the dynamics of relative establishment size. This fact implies that when we write firm size as a product of the number of establishments and average establishment size, $\mathbf{Z} = \mathbf{X}\mathbf{Y}$, in the cross section, $\mathbf{X}$ and $\mathbf{Y}$ are correlated, instead of being independent when $z_I = \zeta$. For example, when $z_X > \delta$ and $z_I > \zeta$, over time surviving firms, on average, have both a higher number of establishments and larger establishments.

To tackle this case, we use the system of differential equations (13). The system

[35] Another way to obtain this result is to notice that $\mathcal{M}(n, \hat{q}) = \bar{\mathcal{M}}(n)(1 - \Phi(\hat{q}))$ solves (13).

[36] After a change of variable $\hat{\mathbf{z}} = \exp(p)$, the transformation can be re-written in its more familiar form: $\int_{-\infty}^{\infty} e^{sp} \left(-d\mathbf{M}(\exp(p))\right)$.

of differential equations (13) for $\mathcal{M}(n, \hat{q})$ simplifies to

$$(z_I - \zeta)\hat{q}\frac{d\mathcal{M}(1, \hat{q})}{d\hat{q}} = -(z_X + \delta + d + \eta)\mathcal{M}(1, \hat{q}) + 2\delta\mathcal{M}(2, \hat{q}) + \mu_e(1 - \Phi(\hat{q}))$$

(for $n = 1$) and

$$(z_I - \zeta)\hat{q}\frac{d\mathcal{M}(n, \hat{q})}{d\hat{q}} = -(n(z_X + \delta) + d + \eta)\mathcal{M}(n, \hat{q}) + (n+1)\delta\mathcal{M}(n+1, \hat{q}) + (n-1)z_X\mathcal{M}(n-1, \hat{q})$$

for $n > 1$. Multiplying both sides of these equations by $\hat{q}^{s-1}$ and integrating by parts from 0 to ∞,

$$\int_0^\infty \hat{q}^{s-1}\mathcal{M}(n, \hat{q})d\hat{q} = -\frac{1}{s}\int_0^\infty \hat{q}^s d\mathcal{M}(n, \hat{q}),$$

we obtain:

$$-(z_I - \zeta)s\hat{\varphi}(1, s) = -(z_X + \delta + d + \eta)\hat{\varphi}(1, s) - 2\delta\hat{\varphi}(2, s) + \int_0^\infty \hat{q}^{s-1}\mu_e(1 - \Phi(\hat{q}))$$

(for $n = 1$) and

$$-(z_I - \zeta)s\hat{\varphi}(n, s) = -(n(z_X + \delta) + d + \eta)\hat{\varphi}(n, s) + (n+1)\delta\hat{\varphi}(n+1, s) + (n-1)z_X\hat{\varphi}(n-1, s)$$

for $n > 1$, where

$$\hat{\varphi}(n, s) \equiv \int_0^\infty \hat{q}^s(-d\mathcal{M}(n, \hat{q})).$$

For each $s \geq 0$, the equations form a system of difference equations and allow us to solve for $\hat{\varphi}(n, s)$ for all $n \geq 1$ using the closed-form solution from Luttmer (2011) (with η being replaced by $\eta - (z_I - \zeta)s$). We show in Appendix D.3 that

$$\hat{\varphi}(n, s) \propto_{n\to\infty} n^{-\frac{d+\eta-(z_I-\zeta)s}{z_X-\delta}-1}.$$

Now, with the solution for $\hat{\varphi}(n, s)$, we can calculate the Laplace transform (24) as follows:

$$\varphi(s) = \frac{1}{\sum_n \bar{\mathcal{M}}(n)}\sum_n n^s \int_0^\infty (\hat{z}/n)^s(-d\mathcal{M}(n, \hat{z}/n)) = \frac{\sum_n n^s\hat{\varphi}(n, s)}{\sum_n \bar{\mathcal{M}}(n)}.$$

Using the asymptotic property of $\hat{\varphi}(n,s)$ above, we show in Appendix D.3 that if $z_X > \delta$ and $z_X - \delta + z_I - \zeta > 0$, $\varphi(s)$ is finite up to s^* determined by

$$\frac{d+\eta-(z_I-\zeta)s}{z_X-\delta} = s,$$

or, equivalently,

$$s^* = \lambda^f \equiv \frac{\eta+d}{z_X-\delta+z_I-\zeta}.$$

In addition, we can show that

$$\varphi(s) \propto_{s\uparrow s^*} \frac{1}{s^*-s}.$$

Therefore, by the Tauberian theorem in Mimica (2016, Corollary 1.3), $\mathbf{M}$ has a Pareto tail with the tail index given by s^*. See Appendix D.3 for further details.

D.3 Proofs

Proof of Proposition 1.

First, we show that the distribution of establishment sizes has Pareto tail with the tail index given by (15). We rewrite (14) as

$$e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}p}\tilde{\mathcal{H}}(p) = \int_{-\infty}^{p} e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}\tilde{p}}\frac{\delta+d+\eta-z_X}{z_I-\zeta}(1-\Phi(\exp(\tilde{p})))d\tilde{p}.$$

Because Φ has thin tail,

$$1-\Phi(\exp(\tilde{p})) < Ae^{-2\frac{\delta+d+\eta-z_X}{z_I-\zeta}\tilde{p}}$$

for some $A > 0$ and for all $\tilde{p} > 0$, which implies

$$a = \int_{-\infty}^{\infty} e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}\tilde{p}}\frac{\delta+d+\eta-z_X}{z_I-\zeta}(1-\Phi(\exp(\tilde{p})))d\tilde{p} < \infty.$$

Therefore

$$\lim_{p\to\infty} e^{\frac{\delta+d+\eta-z_X}{z_I-\zeta}p}\tilde{\mathcal{H}}(p) = a.$$

That is, $\mathcal{H}$ has Pareto tail with tail index given by (15).

The proof for the distribution of the number of establishments per firm is substantially more complicated. Here, $\frac{d+\eta}{z_X-\delta} > 1$ holds. In this Appendix, we provide the proof for the case $\frac{d+\eta}{z_X-\delta} < 2$. The proof for the case with higher $\frac{d+\eta}{z_X-\delta}$ is similar but with much more algebra.

To prove the result, we show a slightly stronger one:

$$\bar{\mathcal{M}}(n) \propto_{n\to\infty} n^{-\frac{d+\eta}{z_X-\delta}-1}. \tag{25}$$

To do so we use the probability generating function

$$\mathcal{P}(\omega) = \sum_{n=1}^{\infty} \bar{\mathcal{M}}(n)\omega^n.$$

The Karamata Tauberian theorem for power series Bingham et al. (1987) allows us to establish the asymptotic behavior of the cumulative sum of $\bar{\mathcal{M}}(n)$ $(n \to \infty)$ from the asymptotic behavior of $\mathcal{P}(\omega)$ $(\omega \to 1)$ if the latter diverges. However, $\mathcal{P}(1) = 1$ so the theorem does not directly apply. In order to apply the theorem, we need to work with $\mathcal{P}''(\omega)$. Lemma 1 below provides us with the asymptotic behavior of $\mathcal{P}''(\omega)$ $(\omega \to 1)$. By the Karamata Tauberian theorem for power series (Bingham et al., 1987, Corollary 1.7.3),

$$\sum_{k=0}^{n}(k+2)(k+1)\bar{\mathcal{M}}(k+2) \propto_{n\to\infty} n^{2-\frac{d+\eta}{z_X-\delta}}. \tag{26}$$

Now we use this result to prove (25).

Differentiating $\mathcal{P}(\omega)$ with respect to ω, we obtain

$$\mathcal{P}'(\omega) = \sum_{n=1}^{\infty} \bar{\mathcal{M}}(n)n\omega^{n-1} = \sum_{n=0}^{\infty} \bar{\mathcal{M}}(n+1)(n+1)\omega^n.$$

This implies

$$\omega\mathcal{P}'(\omega) = \sum_{n=1}^{\infty} \bar{\mathcal{M}}(n)n\omega^n$$

and

$$\omega^2\mathcal{P}'(\omega) = \sum_{n=1}^{\infty} \bar{\mathcal{M}}(n)n\omega^{n+1} = \sum_{n=2}^{\infty} \bar{\mathcal{M}}(n-1)(n-1)\omega^n.$$

Therefore

$$\begin{aligned}
&z_X\mathcal{P}'(\omega)\omega^2 - (z_X+\delta)\omega\mathcal{P}'(\omega) + \delta\mathcal{P}'(\omega) \\
&= \sum_{n=2}^{\infty} \left(z_X\bar{\mathcal{M}}(n-1)(n-1) - n(z_X+\delta)\bar{\mathcal{M}}(n)n + \delta\bar{\mathcal{M}}(n+1)(n+1)\right)\omega^n \\
&- (z_X+\delta)\bar{\mathcal{M}}(1)\omega + \delta\bar{\mathcal{M}}(2)2\omega + \delta\bar{\mathcal{M}}(1).
\end{aligned}$$

Using equalities (21b), the last equation is equivalent to

$$\begin{aligned}
&z_X\mathcal{P}'(\omega)\omega^2 - (z_X+\delta)\omega\mathcal{P}'(\omega) + \delta\mathcal{P}'(\omega) \\
&= \sum_{n=2}^{\infty} (d+\eta)\,\bar{\mathcal{M}}(n)\omega^n - (z_X+\delta)\bar{\mathcal{M}}(1)\omega + \delta\bar{\mathcal{M}}(2)2\omega + \delta\bar{\mathcal{M}}(1) \\
&= (d+\eta)\mathcal{P}(\omega) - \left((d+\eta+z_X+\delta)\bar{\mathcal{M}}(1) - 2\delta\bar{\mathcal{M}}(2)\right)\omega + \delta\bar{\mathcal{M}}(1).
\end{aligned}$$

Rearranging and regrouping different terms and using (21a), we arrive at

$$\mathcal{P}'(\omega)\left(\delta + z_X\omega^2 - (z_X+\delta)\omega\right) = (d+\eta)\mathcal{P}(\omega) - \mu_e\omega + \delta\bar{\mathcal{M}}(1). \tag{27}$$

Differentiating both sides twice and rearranging terms we obtain

$$\mathcal{P}'''(\omega)\left(\delta + z_X\omega^2 - (z_X+\delta)\omega\right) = \left(d+\eta+2(z_X+\delta) - 4z_X\omega\right)\mathcal{P}''(\omega) - \mathcal{P}'(\omega)2z_X - \mu_e. \tag{28}$$

Dividing both sides by $\delta + z_X\omega^2 - (z_X+\delta)\omega$ and observing that

$$\frac{1}{\delta + z_X\omega^2 - (z_X+\delta)\omega} = -\frac{1}{z_X-\delta}\left(\frac{1}{\omega - \frac{\delta}{z_X}} + \frac{1}{1-\omega}\right), \tag{29}$$

we rewrite (28) as

$$
\begin{aligned}
\mathcal{P}'''(\omega) = &- \left(d + \eta + 2(z_X + \delta) - 4z_X\omega\right)\mathcal{P}''(\omega)\frac{1}{z_X - \delta}\left(\frac{1}{\omega - \delta/z_X} + \frac{1}{1-\omega}\right)\\
&+ \mathcal{P}'(\omega)2z_X\frac{1}{z_X - \delta}\left(\frac{1}{\omega - \delta/z_X} + \frac{1}{1-\omega}\right)\\
&+ \mu_e\frac{1}{z_X - \delta}\left(\frac{1}{\omega - \delta/z_X} + \frac{1}{1-\omega}\right).
\end{aligned}
$$

Equivalently,

$$
\begin{aligned}
\mathcal{P}'''(\omega) = &\left\{- \left(d + \eta + 2(z_X + \delta) - 4z_X\omega\right)\mathcal{P}''(\omega) + \mathcal{P}'(\omega)2z_X + \mu_e\right\}\frac{1}{z_X - \delta}\left(\frac{1}{\omega - \delta/z_X} - \frac{1}{1-\delta/z_X}\right)\\
&+ \left\{- \left(d + \eta + 2(z_X + \delta) - 4z_X\omega\right)\mathcal{P}''(\omega) + \mathcal{P}'(\omega)2z_X + \mu_e\right\}\frac{1}{z_X}\\
&+ \mathcal{P}'(\omega)2z_X\frac{1}{z_X - \delta}\frac{1}{1-\omega} + \mu_e\frac{1}{z_X - \delta}\frac{1}{1-\omega} - 4z_X\mathcal{P}''(\omega)\frac{1}{z_X - \delta}\\
&+ \left(2 - \frac{d+\eta}{z_X - \delta}\right)\mathcal{P}''(\omega)\frac{1}{1-\omega}.
\end{aligned} \tag{30}
$$

Let $\mathcal{Q}$ be defined by

$$
\mathcal{Q}(\omega) \equiv \left\{- \left(d + \eta + 2(z_X + \delta) - 4z_X\omega\right)\mathcal{P}''(\omega) + \mathcal{P}'(\omega)2z_X + \mu_e\right\}\frac{1}{z_X - \delta}\left(\frac{1}{\omega - \delta/z_X} - \frac{1}{1-\delta/z_X}\right).
$$

It follows that $\mathcal{Q}(\omega)$ is finite for all $\omega < 1$. Lemma 1 and the fact that

$$
\frac{1}{\omega - \delta/z_X} - \frac{1}{1-\delta/z_X} = O(1-\omega)
$$

imply

$$
\lim_{\omega\to 1}\mathcal{Q}(\omega) = 0,
$$

when $\omega \to 1$. Therefore, by the Riemann-Lebsegue lemma, the Taylor expansion of $\mathcal{Q}(\omega)$

$$
\mathcal{Q}(\omega) = \sum_{n=0}^{\infty} q_n\omega^n
$$

satisfies

$$\lim_{n\to\infty} q_n = 0.$$

By comparing the power series for both sides of (30), coefficient by coefficient, we obtain

$$\begin{aligned}
&(n+3)(n+2)(n+1)\bar{\mathcal{M}}(n+3) \\
&= q_n + (d+\eta+2(z_X+\delta))\,(n+2)(n+1)\bar{\mathcal{M}}(n+2) - 4z_X(n+1)\bar{\mathcal{M}}(n+1) + 2z_X(n+1)\bar{\mathcal{M}}(n+1) \\
&+ \frac{2z_X}{z_X-\delta}\sum_{k=0}^{n}(k+1)\bar{\mathcal{M}}(k+1) + \frac{\mu_e}{z_X-\delta} - \frac{4z_X}{z_X-\delta}(n+2)(n+1)\bar{\mathcal{M}}(n) \\
&+ \left(2 - \frac{d+\eta}{z_X-\delta}\right)\sum_{k=0}^{n}(k+2)(k+1)\bar{\mathcal{M}}(k+2)
\end{aligned}$$

Observing that $(n+3)(n+2)(n+1)\bar{\mathcal{M}}(n)$ is the leading term on the right hand side of the last expression, so (26) implies

$$(n+3)(n+2)(n+1)\bar{\mathcal{M}}(n+3) \propto_{n\to\infty} n^{2-\frac{d+\eta}{z_X-\delta}}.$$

This limit is equivalent to (25).

Now (25) together with Lemma 2 yields

$$\Pr(\mathbf{X} \geq n) = \sum_{k\geq n}\bar{\mathcal{M}}(k) \propto n^{-\frac{d+\eta}{z_X-\delta}},$$

as $n \to \infty$. ■

Lemma 1 *Assume* $\frac{d+\eta}{z_X-\delta} \in (1,2)$, *then the second derivative of the probability generating function satisfies*

$$\mathcal{P}''(\omega) \propto (1-\omega)^{\frac{d+\eta}{z_X-\delta}-2}$$

as $\omega \to 1$.

Proof. Dividing both sides of (27) and using identity (29), we arrive at

$$\mathcal{P}'(\omega) + \mathcal{P}(\omega)\frac{d+\eta}{z_X-\delta}\left(\frac{1}{\omega-\frac{\delta}{z_X}} + \frac{1}{1-\omega}\right) = \frac{\mu_e\omega - \delta\bar{\mathcal{M}}(1)}{z_X-\delta}\left(\frac{1}{\omega-\frac{\delta}{z_X}} + \frac{1}{1-\omega}\right).$$

Let

$$\psi(\omega) = \left(\frac{\omega - \frac{\delta}{z_X}}{1-\omega}\right)^{\frac{d+\eta}{z_X-\delta}}$$

which satisfies

$$\psi'(\omega) = \frac{d+\eta}{z_X-\delta}\left(\frac{1}{\omega-\frac{\delta}{z_X}} + \frac{1}{1-\omega}\right)\psi(\omega). \tag{31}$$

Then the differential equation for $\mathcal{P}(\omega)$ above can be rewritten as

$$\frac{d}{d\omega}\left(\mathcal{P}(\omega)\psi(\omega)\right) = \mathcal{P}'(\omega)\psi(\omega) + \mathcal{P}(\omega)\psi'(\omega) = \psi(\omega)\frac{\mu_e\omega - \delta\bar{\mathcal{M}}(1)}{z_X-\delta}\left(\frac{1}{\omega-\frac{\delta}{z_X}} + \frac{1}{1-\omega}\right).$$

Integrating both sides from some $\underline{\omega} > \delta/z_X$ up to any $\omega \in (\underline{\omega}, 1)$:

$$\begin{aligned}
\mathcal{P}(\omega)\psi(\omega) &= \mathcal{P}(\underline{\omega})\psi(\underline{\omega}) + \int_{\underline{\omega}}^{\omega} \psi(\tilde{\omega})\frac{\mu_e\tilde{\omega} - \delta\bar{\mathcal{M}}(1)}{z_X-\delta}\left(\frac{1}{\tilde{\omega}-\frac{\delta}{z_X}} + \frac{1}{1-\tilde{\omega}}\right)d\tilde{\omega} \\
&= \mathcal{P}(\underline{\omega})\psi(\underline{\omega}) + \int_{\underline{\omega}}^{\omega} \psi'(\tilde{\omega})\frac{\mu_e\tilde{\omega} - \delta\bar{\mathcal{M}}(1)}{d+\eta}d\tilde{\omega},
\end{aligned}$$

where the second equality is due to (31). Equivalently,

$$\mathcal{P}(\omega) = \frac{c}{\psi(\omega)} + \frac{1}{\psi(\omega)}\int_{\underline{\omega}}^{\omega} \psi'(\tilde{\omega})\frac{\mu_e\tilde{\omega} - \delta\bar{\mathcal{M}}(1)}{d+\eta}d\tilde{\omega},$$

where $c = \mathcal{P}(\underline{\omega})\psi(\underline{\omega}) > 0$. Integrating by parts, we obtain

$$\int_{\underline{\omega}}^{\omega} \psi'(\tilde{\omega})\frac{\mu_e\tilde{\omega} - \delta\bar{\mathcal{M}}(1)}{d+\eta}d\tilde{\omega} = \psi(\omega)\frac{\mu_e\omega - \delta\bar{\mathcal{M}}(1)}{d+\eta} - \psi(\underline{\omega})\frac{\mu_e\underline{\omega} - \delta\bar{\mathcal{M}}(1)}{d+\eta} - \int_{\underline{\omega}}^{\omega}\psi(\tilde{\omega})\frac{\mu_e}{d+\eta}d\tilde{\omega}.$$

Therefore

$$\mathcal{P}(\omega) = \frac{c - \psi(\underline{\omega})\frac{\mu_e\underline{\omega} - \delta\bar{\mathcal{M}}(1)}{d+\eta}}{\psi(\omega)} + \frac{\mu_e\omega - \delta\bar{\mathcal{M}}(1)}{d+\eta} - \frac{\int_{\underline{\omega}}^{\omega}\psi(\tilde{\omega})\frac{\mu_e}{d+\eta}d\tilde{\omega}}{\psi(\omega)}.$$

The derivatives can be computed explicitly:

$$\mathcal{P}'(\omega) = -\frac{c-\psi(\underline{\omega})\frac{\mu_e\underline{\omega}-\delta\bar{\mathcal{M}}(1)}{d+\eta}}{\psi(\omega)}\frac{\psi'(\omega)}{\psi(\omega)} + \frac{\int_{\underline{\omega}}^{\omega}\psi(\tilde{\omega})\frac{\mu_e}{d+\eta}d\tilde{\omega}}{\psi(\omega)}\frac{\psi'(\omega)}{\psi(\omega)}$$
$$= -\frac{c-\psi(\underline{\omega})\frac{\mu_e\underline{\omega}-\delta\bar{\mathcal{M}}(1)}{d+\eta}}{\psi(\omega)}\frac{d+\eta}{z_X-\delta}\left(\frac{1}{\omega-\frac{\delta}{z_X}}+\frac{1}{1-\omega}\right) + \frac{\int_{\underline{\omega}}^{\omega}\psi(\tilde{\omega})\frac{\mu_e}{d+\eta}d\tilde{\omega}}{\psi(\omega)}\frac{d+\eta}{z_X-\delta}\left(\frac{1}{\omega-\frac{\delta}{z_X}}+\frac{1}{1-\omega}\right)$$

and

$$\mathcal{P}''(\omega) = \frac{c-\psi(\underline{\omega})\frac{\mu_e\underline{\omega}-\delta\bar{\mathcal{M}}(1)}{d+\eta}}{\psi(\omega)}\left\{\left(\frac{d+\eta}{z_X-\delta}\right)^2\left(\frac{1}{\omega-\frac{\delta}{z_X}}+\frac{1}{1-\omega}\right)^2 - \frac{d+\eta}{z_X-\delta}\left(-\frac{1}{(\omega-\frac{\delta}{z_X})^2}+\frac{1}{(1-\omega)^2}\right)\right\}$$

where

$$\mathcal{R}(\omega) = \frac{d+\eta}{z_X-\delta}\left(-\frac{1}{(\omega-\frac{\delta}{z_X})^2}+\frac{1}{(1-\omega)^2}\right)\frac{\int_{\underline{\omega}}^{\omega}\psi(\tilde{\omega})\frac{\mu_e}{d+\eta}d\tilde{\omega}}{\psi(\omega)}$$
$$-\left(\frac{d+\eta}{z_X-\delta}\right)^2\left(\frac{1}{(\omega-\frac{\delta}{z_X})^2}+\frac{2}{(\omega-\frac{\delta}{z_X})(1-\omega)}+\frac{1}{(1-\omega)^2}\right)\frac{\int_{\underline{\omega}}^{\omega}\psi(\tilde{\omega})\frac{\mu_e}{d+\eta}d\tilde{\omega}}{\psi(\omega)}$$
$$+\frac{d+\eta}{z_X-\delta}\left(\frac{1}{\omega-\frac{\delta}{z_X}}+\frac{1}{1-\omega}\right)\frac{\mu_e}{d+\eta}.$$

Now,

$$\int_{\underline{\omega}}^{\omega}\psi(\tilde{\omega})d\tilde{\omega} = \int_{\underline{\omega}}^{\omega}\left(\tilde{\omega}-\frac{\delta}{z_X}\right)^{\frac{d+\eta}{z_X-\delta}}(1-\tilde{\omega})^{-\frac{d+\eta}{z_X-\delta}}\,d\tilde{\omega} = \int_{\underline{\omega}}^{\omega}\left(\tilde{\omega}-\frac{\delta}{z_X}\right)^{\frac{d+\eta}{z_X-\delta}}d\left(\frac{(1-\tilde{\omega})^{-\frac{d+\eta}{z_X-\delta}+1}}{\frac{d+\eta}{z_X-\delta}-1}\right)$$
$$= \left.\frac{\left(\tilde{\omega}-\frac{\delta}{z_X}\right)^{\frac{d+\eta}{z_X-\delta}}(1-\tilde{\omega})^{-\frac{d+\eta}{z_X-\delta}+1}}{\frac{d+\eta}{z_X-\delta}-1}\right|_{\underline{\omega}}^{\omega} - \int_{\underline{\omega}}^{\omega}\frac{d+\eta}{z_X-\delta}\left(\tilde{\omega}-\frac{\delta}{z_X}\right)^{\frac{d+\eta}{z_X-\delta}-1}\frac{1}{\frac{d+\eta}{z_X-\delta}-1}(1-\tilde{\omega})^{-\frac{d+\eta}{z_X-\delta}+1}d\tilde{\omega}$$
$$= \frac{\left(\omega-\frac{\delta}{z_X}\right)^{\frac{d+\eta}{z_X-\delta}}(1-\omega)^{-\frac{d+\eta}{z_X-\delta}+1}}{\frac{d+\eta}{z_X-\delta}-1} + c_\psi + O(1-\omega)$$

where

$$c_\psi = -\frac{\psi(\underline{\omega})(1-\underline{\omega})}{\frac{d+\eta}{z_X-\delta}-1} - \int_{\underline{\omega}}^{1} \frac{d+\eta}{z_X-\delta}\left(\tilde{\omega}-\frac{\delta}{z_X}\right)^{\frac{d+\eta}{z_X-\delta}-1} \frac{1}{\frac{d+\eta}{z_X-\delta}-1}(1-\tilde{\omega})^{-\frac{d+\eta}{z_X-\delta}+1}\,d\tilde{\omega}.$$

So

$$\frac{\int_{\underline{\omega}}^{\omega}\psi(\tilde{\omega})d\tilde{\omega}}{\psi(\omega)} = \frac{1-\omega}{\frac{d+\eta}{z_X-\delta}-1} + \frac{c_\psi}{\psi(\omega)} + o(1-\omega)$$

Therefore the factor associated with $\frac{1}{1-\omega}$ in $\mathcal{R}(\omega)$ is

$$\left(\frac{d+\eta}{z_X-\delta} - \left(\frac{d+\eta}{z_X-\delta}\right)^2\right)\frac{1}{\frac{d+\eta}{z_X-\delta}-1}\frac{\mu_e}{d+\eta} + \frac{d+\eta}{z_X-\delta}\frac{\mu_e}{d+\eta} = 0.$$

which implies

$$\mathcal{R}(\omega) = O(1) + \frac{c_\psi}{\psi(\omega)(1-\omega)^2}\frac{d+\eta}{z_X-\delta}\left(1-\frac{d+\eta}{z_X-\delta}\right)$$

as $\omega \to 1$. So

$$\mathcal{P}''(\omega) \sim_{\omega\to 1} \left(c - \psi(\underline{\omega})\frac{\mu_e\underline{\omega} - \delta\bar{\mathcal{M}}(1)}{d+\eta} - c_\psi\right)\frac{d+\eta}{z_X-\delta}\left(\frac{d+\eta}{z_X-\delta}-1\right)\frac{1}{\psi(\omega)(1-\omega)^2}$$

Notice that

$$c - \psi(\underline{\omega})\frac{\mu_e\underline{\omega} - \delta\bar{\mathcal{M}}(1)}{d+\eta} - c_\psi > \psi(\underline{\omega})\left(\mathcal{P}(\underline{\omega}) - \frac{\mu_e\underline{\omega} - \delta\bar{\mathcal{M}}(1)}{d+\eta} + \frac{1-\underline{\omega}}{\frac{d+\eta}{z_X-\delta}-1}\right)$$

and

$$\begin{aligned}\mathcal{P}(\underline{\omega}) - \frac{\mu_e\underline{\omega} - \delta\bar{\mathcal{M}}(1)}{d+\eta} + \frac{1-\underline{\omega}}{\frac{d+\eta}{z_X-\delta}-1} &= \frac{\mathcal{P}'(\underline{\omega})\left(\delta + z_X\underline{\omega}^2 - (z_X+\delta)\underline{\omega}\right)}{d+\eta} + \frac{1-\underline{\omega}}{\frac{d+\eta}{z_X-\delta}-1} \\ &= \left\{\frac{\mathcal{P}'(\underline{\omega})(\delta - z_X\underline{\omega})}{d+\eta} + \frac{1}{\frac{d+\eta}{z_X-\delta}-1}\right\}(1-\underline{\omega}) > 0,\end{aligned}$$

when $\underline{\omega}$ is chosen sufficiently close to δ/z_X. Thus

$$c - \psi(\underline{\omega})\frac{\mu_e\underline{\omega} - \delta\bar{\mathcal{M}}(1)}{d+\eta} - c_\psi > 0.$$

■

Lemma 2 *Suppose that*

$$\bar{\mathcal{M}}(n) \propto_{n\to\infty} n^{-\frac{d+\eta}{z_X-\delta}-1},$$

then

$$\Pr(\mathbf{X} \geq n) = \sum_{k\geq n} \bar{\mathcal{M}}(k) \propto n^{-\frac{d+\eta}{z_X-\delta}}.$$

Proof. There exists $a > 0$ such that

$$\lim_{n\to\infty} \frac{\bar{\mathcal{M}}(n)}{n^{-\frac{d+\eta}{z_X-\delta}-1}.} = a.$$

Therefore, for any $\epsilon > 0$, there exists n^* such that, for all $n \geq n^*$

$$a - \epsilon < \frac{\bar{\mathcal{M}}(n)}{n^{-\frac{d+\eta}{z_X-\delta}-1}.} < a + \epsilon$$

Combining these inequalities with the definition of $\Pr(\mathbf{X} \geq n)$, we obtain, for all $n \geq n^*$

$$(a-\epsilon)\sum_{k\geq n} k^{-\frac{d+\eta}{z_X-\delta}-1} < \Pr(\mathbf{X} \geq n) = \sum_{k\geq n} \bar{\mathcal{M}}(k) < (a+\epsilon)\sum_{k\geq n} k^{-\frac{d+\eta}{z_X-\delta}-1}.$$

Notice that

$$\sum_{k\geq n} k^{-\frac{d+\eta}{z_X-\delta}-1} < \int_{k\geq n-1} k^{-\frac{d+\eta}{z_X-\delta}-1} dk = \frac{z_X-\delta}{d+\eta}(n-1)^{-\frac{d+\eta}{z_X-\delta}}$$

and

$$\sum_{k\geq n} k^{-\frac{d+\eta}{z_X-\delta}-1} > \int_{k\geq n} k^{-\frac{d+\eta}{z_X-\delta}-1} dk = \frac{z_X-\delta}{d+\eta} n^{-\frac{d+\eta}{z_X-\delta}}.$$

Therefore

$$(a-\epsilon)\frac{z_X-\delta}{d+\eta}n^{-\frac{d+\eta}{z_X-\delta}} < \Pr(\mathbf{X}\geq n) < (a+\epsilon)\frac{z_X-\delta}{d+\eta}(n-1)^{-\frac{d+\eta}{z_X-\delta}}$$

As this applies for any $\epsilon > 0$, we obtain

$$\lim_{n\to\infty}\frac{\Pr(\mathbf{X}\geq n)}{\frac{z_X-\delta}{d+\eta}n^{-\frac{d+\eta}{z_X-\delta}}} = a.$$

■

Proofs for Proposition 2. First we derive the expression for $\varphi(s)$ provide in the main text.

$$\begin{aligned}
\varphi(s) = \int_0^\infty \hat{\mathbf{z}}^s \frac{-d\sum_n \bar{\mathcal{M}}(n)(1-\Phi(\hat{\mathbf{z}}/n))}{\sum_n \bar{\mathcal{M}}(n)} &= \int_0^\infty \hat{\mathbf{z}}^s \frac{\sum_n \bar{\mathcal{M}}(n)d\Phi(\hat{\mathbf{z}}/n)}{\sum_n \bar{\mathcal{M}}(n)} \\
&= \int_0^\infty \frac{\sum_n \bar{\mathcal{M}}(n)n^s(\hat{\mathbf{z}}/n)^s d\Phi(\hat{\mathbf{z}}/n)}{\sum_n \bar{\mathcal{M}}(n)} \\
&= \left\{\int_0^\infty \hat{\mathbf{z}}^s d\Phi(\hat{\mathbf{z}})\right\}\left\{\frac{\sum_n \bar{\mathcal{M}}(n)n^s}{\sum_n \bar{\mathcal{M}}(n)}\right\}.
\end{aligned}$$

Now we prove that

$$\varphi(s) \propto \frac{1}{\frac{\eta+d}{z_X-\delta}-s} \tag{32}$$

as $s \uparrow \frac{\eta+d}{z_X-\delta}$. To do so, we use (25) which characterizes the asymptotic behavior of $\bar{\mathcal{M}}(n)$. This result implies that, there exists $a>0$ such that: for any $\epsilon>0$, there exists n^* so that

$$(a-\epsilon)n^{-\frac{d+\eta}{z_X-\delta}-1} < \bar{\mathcal{M}}(n) < (a+\epsilon)n^{-\frac{d+\eta}{z_X-\delta}-1}$$

for all $n \geq n^*$. Therefore

$$\left(\frac{\eta+d}{z_X-\delta}-s\right)\sum_{n\geq n^*}(a-\epsilon)n^{s-\frac{d+\eta}{z_X-\delta}-1} < \left(\frac{\eta+d}{z_X-\delta}-s\right)\sum_{n\geq n^*}\bar{\mathcal{M}}(n)n^s < \left(\frac{\eta+d}{z_X-\delta}-s\right)\sum_{n\geq n^*}(a+\epsilon)n^{s-\frac{d+\eta}{z_X-\delta}-1}.$$

Notice that

$$\left(\frac{\eta+d}{z_X-\delta}-s\right)\sum_{n\geq n^*} n^{s-\frac{d+\eta}{z_X-\delta}-1} < \left(\frac{\eta+d}{z_X-\delta}-s\right)\int_{n^*-1}^{\infty} x^{s-\frac{d+\eta}{z_X-\delta}-1}dx = (n^*-1)^{s-\frac{d+\eta}{z_X-\delta}}$$

and

$$\left(\frac{\eta+d}{z_X-\delta}-s\right)\sum_{n\geq n^*} n^{s-\frac{d+\eta}{z_X-\delta}-1} > \left(\frac{\eta+d}{z_X-\delta}-s\right)\int_{n^*}^{\infty} x^{s-\frac{d+\eta}{z_X-\delta}-1}dx = (n^*)^{s-\frac{d+\eta}{z_X-\delta}}$$

Since

$$\left(\frac{\eta+d}{z_X-\delta}-s\right)\sum_{n}\bar{\mathcal{M}}(n)n^s = \left(\frac{\eta+d}{z_X-\delta}-s\right)\sum_{n<n^*}\bar{\mathcal{M}}(n)n^s + \left(\frac{\eta+d}{z_X-\delta}-s\right)\sum_{n\geq n^*}\bar{\mathcal{M}}(n)n^s,$$

using the inequalities above and take the limit $s \uparrow \frac{\eta+d}{z_X-\delta}$, we obtain

$$a-2\epsilon < \left(\frac{\eta+d}{z_X-\delta}-s\right)\sum_{n}\bar{\mathcal{M}}(n)n^s < a+2\epsilon$$

for all $s \in (s^*, \frac{\eta+d}{z_X-\delta})$ with s^* sufficiently close to $\frac{\eta+d}{z_X-\delta}$. In other words,

$$\sum_{n}\bar{\mathcal{M}}(n)n^s \propto_{s\uparrow\frac{\eta+d}{z_X-\delta}} \frac{1}{\frac{\eta+d}{z_X-\delta}-s}.$$

Because Φ has thin tail $\int_0^\infty \hat{\mathbf{z}}^s d\Phi(\hat{\mathbf{z}})$ is finite and continuous in s which implies (32).
■

Proof of Proposition 3. Let $\hat{\varphi}(n,s)$ denote the Laplace transform of $\mathcal{M}(n,q)$. In Subsection 4.2, we derived the difference equations satisfied by $\hat{\varphi}(n,s)$ similar to the difference equations for $\bar{\mathcal{M}}(n)$. Using these difference equations and following the steps in the proof of Proposition 1, we can show that, there exists $a(s)$ such that

$$\lim_{n\to\infty} n^{\frac{d+\eta-(z_I-\zeta)s}{z_X-\delta}+1}\hat{\varphi}(n,s) = a(s)$$

and the convergence is uniform in s. Recall that the Laplace transformation for firm size distribution can be written as

$$\varphi(s) = \frac{\sum_n n^s \hat{\varphi}(n,s)}{\sum_n \bar{\mathcal{M}}(n)},$$

Armed with this result, we can follow the steps in the proof of Proposition 2 to show that

$$\sum_n \hat{\varphi}(n,s) n^s \propto \frac{1}{s^* - s}$$

Using this property and applying Mimica (2016, Corollary 1.3), we obtain the tail result stated in the proposition.

To derive (18), notice that, by (16),

$$d + \eta = (z_X - \delta)\lambda^{ne}$$

and by (15)

$$z_I - \zeta = \frac{d + \eta - (z_X - \delta)}{\lambda^e} = \frac{(z_X - \delta)(\lambda^{ne} - 1)}{\lambda^e}.$$

Plugging these expression in (17), we obtain

$$\lambda^f = \frac{(z_X - \delta)\lambda^{ne}}{z_X - \delta + \frac{(z_X-\delta)(\lambda^{ne}-1)}{\lambda^e}} = \frac{\lambda^{ne}\lambda^e}{\lambda^{ne} + \lambda^e - 1}.$$

Inverting the first and the last items, we arrive at (18). ■

E Derivations

E.1 Output growth

Output growth derives from firm-level investments in internal and external innovations. Thus we examine a decomposition of output growth into the extensive- and intensive-margin growth rates, η and ζ, at an aggregate level. Using the labor-market-clearing condition $L(t) = \int_{\mathcal{N}_t} \ell_i(t)di$ along with firm labor demand in equation (5) yields the following expression:

$$\bar{w}(t)L(t) = (1-\beta)\left(\frac{N(t)Q(t)}{A(t)Y(t)^{\frac{\beta}{1-\beta}}}\right)^{\beta} L(t)^{1-\beta}. \tag{33}$$

Because the normalized wage $\bar{w}(t)$ does not grow along the BGP, equation (33) implies a decomposition of output growth into the growth of the intensive and extensive margins,[37]

$$g = \eta + \zeta. \tag{34}$$

Because final output growth can be decomposed into the extensive and intensive margins of firm growth, it has a natural interpretation as the aggregate outcome of disaggregated firm behavior. First, consider the extensive margin. The total number of establishments of type τ is $N_\tau(t) \equiv M_\tau N(t)$, where $M_\tau \in [0,1]$ is the share of type τ establishments, and satisfies $\sum_\tau M_\tau = 1$. The law of motion for $N_\tau(t)$ is

$$\dot{N}_\tau(t) = z_X^\tau N_\tau(t) - (\delta_\tau + d_\tau)N_\tau(t) + \mu_e m_\tau N(t) - \sum_{\tau' \neq \tau} \lambda_{\tau\tau'} N_\tau(t) + \sum_{\tau' \neq \tau} \lambda_{\tau'\tau} N_{\tau'}(t).$$

The first term is the increase in the number of establishments due to external innovation, the second term is the loss of establishments due to exit, the third term accounts for product entry, and the fourth and fifth terms capture the product-number evolution due to changes in firm types. On the BGP, $N_\tau(t)$ grows at rate η for all τ, and thus this equation can be rewritten as:[38]

$$\eta = z_X^\tau - (\delta_\tau + d_\tau) + \mu_e \frac{m_\tau}{M_\tau} - \sum_{\tau' \neq \tau} \lambda_{\tau\tau'} + \sum_{\tau' \neq \tau} \lambda_{\tau'\tau} \frac{M_{\tau'}}{M_\tau}, \tag{35}$$

[37]Equation (33) implies the growth relationship $\gamma = \eta + \zeta - \theta - (\beta/(1-\beta))g$, which is combined with equation (10) to yield equation (34).

[38]The growth rate of the total number of establishments can also be written as the weighted sum of the growth rates of the number of type-τ establishments and the entry rate,

$$\eta = \sum_\tau M_\tau [z_X^\tau - (\delta_\tau + d_\tau)] + \mu_e,$$

which is found by multiplying M_τ to both sides of (35) and summing over τ.

Next, consider the intensive margin. Define the average quality of type τ firms as

$$Q_\tau(t) \equiv \frac{1}{N_\tau(t)} \int_{\mathcal{N}_\tau(t)} q_i(t)di,$$

where $\mathcal{N}_\tau(t)$ is the set of actively produced goods by type-τ firms. Further define s_τ as the quality share of type τ firms by

$$s_\tau \equiv \frac{N_\tau(t)Q_\tau(t)}{N(t)Q(t)}, \tag{36}$$

which satisfies

$$\sum_\tau s_\tau = 1. \tag{37}$$

On the BGP, s_τ is constant, which implies $Q_\tau(t)$ has to grow at the same rate as $Q(t)$ for all types τ.

Finally, using $g = \eta + \zeta$ defines aggregate output growth g as a function of firm-level innovations and shocks,

$$g = \left[z_I^\tau + z_X^\tau - (\delta_\tau + d_\tau)\right] + \mu_e \frac{m_\tau}{s_\tau} \int \hat{q} d\Phi_\tau(\hat{q}) - \sum_{\tau' \neq \tau} \lambda_{\tau\tau'} + \sum_{\tau' \neq \tau} \lambda_{\tau'\tau} \frac{s_{\tau'}}{s_\tau}. \tag{38}$$

The first term is the incumbent firms' contribution to g, the second term characterizes entrants' contribution to output growth, and the final terms capture the impact of changing firm types.[39] To derive (38) take the definition of $Q_\tau(t)$ and express as,

$$\frac{\dot{Q}_\tau(t)}{Q_\tau(t)} = -\frac{\dot{N}_\tau(t)}{N_\tau(t)} + \frac{d \int_{\mathcal{N}_\tau(t)} q_i(t)di/dt}{\int_{\mathcal{N}_\tau(t)} q_i(t)di}. \tag{39}$$

The first term of the right-hand side is $-\eta$. To compute the second term, consider a discrete time interval $\Delta t > 0$, compute $(\int_{\mathcal{N}_\tau(t+\Delta t)} q_i(t+\Delta t)di - \int_{\mathcal{N}_\tau(t)} q_i(t)di)/\Delta t$

[39] A simpler expression for g can be found by multiplying s_τ on both sides of (38) and summing across τ,

$$g = \sum_\tau s_\tau [z_I^\tau + z_X^\tau - (\delta_\tau + d_\tau)] + \mu_e \int \hat{q} d\Phi(\hat{q}).$$

and set $\Delta_t \to 0$. Note that the denominator of the second term is equal to $Q(t)N(t)$. Because

$$\begin{aligned}
&\int_{\mathcal{N}_\tau(t+\Delta t)} q_i(t+\Delta t)di - \int_{\mathcal{N}_\tau(t)} q_i(t)di \\
&\quad = [z_I^\tau + z_X^\tau]\Delta t Q_\tau(t)N_\tau(t) - (\delta_\tau + d_\tau)\Delta t Q_\tau(t)N_\tau(t) + \mu_e \Delta t m_\tau Q(t)N(t)\int \hat{q}d\Phi(\hat{q}) \\
&\qquad - \sum_{\tau'\neq\tau}\lambda_{\tau\tau'}\Delta t Q_\tau(t)N_\tau(t) + \sum_{\tau'\neq\tau}\lambda_{\tau'\tau}\Delta t Q_{\tau'}(t)N_{\tau'}(t) + o(\Delta t),
\end{aligned}$$

where the first term is the additional quality by internal and external innovation, the second term is the lost quality by exit, the third term is the gain from entry, and the fourth and the fifth terms are the loss and gain from the transitions of firm types. (The higher-order terms are omitted as $o(\Delta t)$.) Dividing by Δt and taking $\Delta t \to 0$,

$$\begin{aligned}
\frac{d\int_{\mathcal{N}_\tau(t)} q_i(t)di}{dt} &= Q_\tau(t)N_\tau(t)\Big(z_I^\tau + z_X^\tau - (\delta_\tau + d_\tau) + \mu_e m_\tau \tfrac{Q(t)N(t)}{Q_\tau(t)N_\tau(t)}\textstyle\int \hat{q}d\Phi(\hat{q}) \\
&\qquad - \textstyle\sum_{\tau'\neq\tau}\lambda_{\tau\tau'} + \sum_{\tau'\neq\tau}\lambda_{\tau'\tau}\tfrac{Q_{\tau'}(t)N_{\tau'}(t)}{Q_\tau(t)N_\tau(t)}\Big).
\end{aligned}$$

Therefore, (39) can be rewritten as

$$\zeta = -\eta + z_I^\tau + z_X^\tau - (\delta_\tau + d_\tau) + \mu_e \frac{m_\tau}{M_\tau}\frac{Q(t)}{Q_\tau(t)}\int \hat{q}d\Phi(\hat{q}) - \sum_{\tau'\neq\tau}\lambda_{\tau\tau'} + \sum_{\tau'\neq\tau}\lambda_{\tau'\tau}\frac{Q_{\tau'}(t)M_{\tau'}}{Q_\tau(t)M_\tau}.$$

Using the definition of s_τ and $g = \eta + \zeta$, we can obtain (38).

F Estimation procedures

F.1 Estimation of parametrized establishment size distributions

In this Appendix, we describe an estimation strategy for recovering model parameters from establishment size distributions. We assume that the data is drawn from simple parametric distributions that are known to fit the actual U.S. data from past studies. We first estimate these distributions using publicly available data on establishment size distributions. The data moments include the Pareto tail index of the establishment size distribution which cannot be inferred directly from publicly available data. The tail index is crucial for our estimation procedure.

We assume that in year t, the distribution of establishment size in number of employees (call it $\mathbf{l}$) takes the following form:

$$\Pr_t(\mathbf{l} \leq l) = G(\log l; \mu_t^e, \sigma_t^e, \lambda_t^e),$$

for establishment sizes $\ell = 1, 2, \ldots$ and in years $t = 1995, 2014$, where G is the CDF of the convolution between a normal distribution and an exponential distribution (see Sager and Timoshenko (2019) for more details on this type of distribution):

$$G(z; \mu, \sigma, \lambda) \equiv \Phi_n\left(\frac{z-\mu}{\sigma}\right) - e^{-\lambda(z-\mu)+\frac{\sigma^2}{2}\lambda^2}\Phi_n\left(\frac{z-\mu-\lambda\sigma^2}{\sigma}\right),$$

where Φ_n is the cdf of the standard normal distribution. This distribution flexibly nests both a normal distribution and an exponential distribution and conveniently allows for a thick right tail.

We estimate μ_t^e, σ_t^e, and λ_t^e by targeting the establishment size table published by the U.S. Bureau of Labor and Satistics' Quarterly Census of Employment and Wages (QCEW) for 1995 and 2014 using weighted least square minimization procedure in Sager and Timoshenko (2019). The estimation yields

$$\mu_{1995}^e = 0.3761, \quad \sigma_{1995}^e = 1.5745, \quad \lambda_{1995}^e = 1.5578$$

and

$$\mu_{2014}^e = 0.2823, \quad \sigma_{2014}^e = 1.6642, \quad \lambda_{2014}^e = 1.7531.$$

We have checked that outcome replicates the BLS-published distribution well. Note that the establishment size distribution has a tail index that increases over time ($\lambda_{2014}^e > \lambda_{1995}^e$), indicating that skewness in the establishment size distribution has declined over time.

F.2 Computational algorithm

We estimate the model in two steps. In Step 1, we estimate $(z_X^H, z_X^L, \lambda_{HL}, \mu_e, m_H, m_L)$ using moments related to the number of establishments per firm (Step 1a), and then we estimate $(z_I^H, z_I^L, \Phi(\cdot))$ using moments related to the number of employees per establishment (Step 1b). In Step 2, we assume functional forms for the cost functions $h_X^\tau(\cdot)$, $h_I^\tau(\cdot)$ and estimate the parameters of these functions using the estimates from Step 1.

Step 1a (Number of establishments per firm): In this step, we choose

$$(z_X^H, z_X^L, \lambda_{HL}, \mu_e, m_H, m_L)$$

parameters to target (i) percentiles of the distribution over the number of establishments per firm, (ii) the slope of the upper tail of the distribution, and (iii) the growth rate of the number of establishments $\eta \approx 1\%$. The empirical moments are described in Appendix C. With two types, (35) becomes

$$\begin{aligned}
\eta &= z_X^H - \delta_H - d_H + \mu_e \frac{m_H}{M_H} - \lambda_{HL} \\
\eta &= z_X^L - \delta_L - d_L + \mu_e \frac{m_L}{M_L} + \lambda_{HL} \frac{M_H}{M_L}.
\end{aligned}$$

Together with $M_H + M_L = 1$, we have a unique solution for M_H, M_L and η.[40]

Having η, we can use equation (19) to calculate the Pareto-tail index of the distribution of establishment number. We also use (11) to compute the whole distribution

[40]The solution is derived from the quadratic form,

$$M_L = \frac{-a_1 - \sqrt{a_1^2 - 4a_0a_2}}{2a_0},$$

where

$$\begin{aligned}
a_0 &= (z_X^H - \delta_H - d_H) - (z_X^L - \delta_L - d_L) \\
a_1 &= -(\mu_e(m_H + m_L) + \lambda_{HL}) + ((z_X^L - \delta_L - d_L) - (z_X^H - \delta_H - d_H)) \\
a_2 &= \mu_e m_L + \lambda_{HL},
\end{aligned}$$

and $M_H = 1 - M_L$. We then obtain η from either type's version of equation (35).

of establishment numbers including the fraction of single-establishment firms and several quantiles beyond the top 99%. Using these model moments, the estimation minimizes a weighted squared sum between the model and empirical moments.

Step 1b (Establishment size): In this step, we assume type τ entry-size distribution Φ_τ follows a log-normal distribution with mean ϱ_τ and variance ς_τ^2 so that $\Phi_\tau \sim \exp(\mathcal{N}(\varrho_\tau, \varsigma_\tau^2))$. We choose

$$(z_I^\tau, \varrho_\tau, \varsigma_\tau)_{\tau\in\{L,H\}}$$

and target the distribution of establishment size as well as the average growth rate of establishment size, $\zeta = g - \eta$. The empirical moments from the establishment size distribution include the Pareto tail index and several quantiles. These moments are computed from the estimated parameterized distribution described in Appendix F.1. The estimation of the parametrized distributions uses publicly-available BLS data.

In the model, ζ can be computed from (38). The tail index of the establishment size distribution in the model is given by (20). The whole distribution of establishment size can be computed by solving (12), and several quantiles are included in the list of targeted moments. Using these model moments, the estimation minimizes a weighted-squared-sum distance between the model and empirical moments as in Step 1a.

Step 2 (Recovering endogenous variables): In this step, we use the estimates from Step 1a and Step 1b, including $z_i^\tau, \tau \in \{H, L\}, i \in \{X, I\}$, to quantify the remaining model outcomes and allocations. To execute this step, we must parameterize the innovation-cost functions. We assume the innovation-cost functions take the form $h_i^\tau(z) = \chi_i^\tau z^\psi$, for $\tau \in \{L, H\}$, $i \in \{X, I\}$, where $\psi > 0$. The first-order condition in (8) implies $\psi\chi_i^\tau(z_i^\tau)^{\psi-1} = v_\tau$, and hence

$$-h_i^\tau(z_i^\tau) + z_i^\tau v_\tau = \left(1 - \frac{1}{\psi}\right) z_i^\tau v_\tau.$$

Substituting this expression in (8) and re-arranging, we arrive at:

$$\underbrace{\begin{bmatrix} A_{11} & -\lambda_{HL} \\ 0 & A_{22} \end{bmatrix}}_{\equiv \boldsymbol{A}} \begin{bmatrix} v_H \\ v_L \end{bmatrix} = \bar{\pi} \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

where

$$A_{11} = r - \left(1 - \frac{1}{\psi}\right)(z_X^H + z_I^H) + \delta_H + d_H + \lambda_{HL}$$

and

$$A_{22} = r - \left(1 - \frac{1}{\psi}\right)(z_X^L + z_I^L) + \delta_L + d_L + \lambda_{HL}.$$

From the estimates in Step 1a and Step 1b, all the elements of matrix $\boldsymbol{A}$ are known, including $r = \rho + \sigma g$ and $g = \eta + \zeta$. Therefore, we can then solve for v_H, v_L as functions of $\bar{\pi}$:

$$\begin{bmatrix} v_H \\ v_L \end{bmatrix} = \bar{\pi} \boldsymbol{A}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Now, combining this result with equation (9) and the free entry condition $v^e = \phi$, we obtain

$$\phi = \bar{\pi} \begin{bmatrix} m_H \exp\left(-\varrho_H + \frac{\varsigma_H^2}{2}\right) \\ m_L \exp\left(-\varrho_L + \frac{\varsigma_L^2}{2}\right) \end{bmatrix}' \boldsymbol{A}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

In other words, ϕ is uniquely determined as a function of $\bar{\pi}$. Given the functional form for investment cost functions, ϕ cannot be jointly identified with $\{\bar{\pi}, \chi_X^\tau, \chi_I^\tau\}_{\tau \in \{H,L\}}$ without information that can pin down the level of (appropriately normalized) $N(t)Q(t)$, because scaling up these parameters by the same proportion would lead to the same equilibrium investment policies $\{z_X^\tau, z_I^\tau\}_{\tau \in \{H,L\}}$ and μ_e. Due to the lack of such information, here, as in Hopenhayn and Rogerson (1993), we choose a value of ϕ for which the equilibrium normalized wage $\bar{w}$ equals 1. It follows from equation (6) that the equilibrium profit per unit of quality is given by $\bar{\pi} = \beta(1-\beta)^{(1-\beta)/\beta}$. Therefore, our identification assumption for ϕ is that the normalized flow profit is the same between 1995 and 2014, and thus all changes in estimated ϕ come from the changes in other estimated parameters of the model, such as z_I, z_X, λ_{HL}, and m_H, m_L.

Table A.3: Moment Fitness

Moments	1995 Data	1995 Model	2014 Data	2014 Model
Step 1a: Establishment Number Distribution				
Establishment growth rate (η)	0.010	0.010	0.010	0.010
Pareto tail index	1.25	1.25	1.21	1.21
Step 1b: Establishment Size Distribution				
Average growth rate (ζ)	0.021	0.022	0.013	0.014
Pareto tail index	1.56	1.56	1.75	1.75

Notes: Estimates using Least-Square Minimization.

Figure 3 shows the model distributions of the number of establishments per firm match the empirical distributions from the 1995 and 2014 data very well. Furthermore, Figure 4 shows the model distributions of establishment size closely match the parametrized empirical distributions for 1995 and 2014 data described in Appendix F.1 (blue solid line), as well as publicly available BLS tabulations of establishment sizes (red circles). The success of the model along these dimensions has to do with the existence of fat tails in the data. The model generates fat-tailed distributions endogenously, and therefore the estimation procedure is selecting parameters to match the general slopes of the Pareto tails in the data. Table A.3 shows the model distribution closely matches the remaining empirical targets including Pareto-tail estimates computed from the establishment size and number distributions.