# Heavy Tailed, but not Zipf: Firm and Establishment Size in the U.S.

Illenin O. Kondo*
Federal Reserve Bank of Minneapolis
kondo@illenin.com

Logan T. Lewis
Federal Reserve Board
logantlewis@ltlewis.net

Andrea Stella
Federal Reserve Board
andrea.stella@frb.gov

October 20, 2022

**Abstract**

Heavy tails play an important role in modern macroeconomics and international economics. Previous work often assumes a Pareto distribution for firm size, typically with a shape parameter approaching Zipf's law. This convenient approximation has dramatic consequences for the importance of large firms in the economy. But we show that a lognormal distribution, or better yet, a convolution of a lognormal and a non-Zipf Pareto distribution, provides a better description of the U.S. economy, using confidential Census Bureau data. These findings hold even far in the upper tail and suggest heterogeneous firm models should more systematically explore deviations from Zipf's law.

*JEL classifications*: L11, E24

*Keywords*: Firm size distribution, Lognormal, Pareto, Zipf's law, Granularity.

*This paper was previously circulated as, "On the U.S. Firm and Establishment Size Distributions". We thank Robert L. Axtell, Tim Dunne, Emmanuel Farhi, Andrew Figura, Doireann Fitzgerald, Xavier Gabaix, Colin Hottman, Erzo G.J. Luttmer, Erick Sager, Francis E. Warnock, and Chen Yeh for helpful comments and suggestions. We also thank seminar participants at the Federal Reserve Bank of Philadelphia, CEF 2018, IAAE 2018, NASMES 2018, CAED 2019, and RIDGE IT/GDM 2020 for their comments. Michael Kister provided excellent research assistance. Finally, we thank Foster, Grim and Haltiwanger for kindly sharing the TFP data from Foster et al. (2016). The views expressed here should not be interpreted as reflecting the views of the Federal Reserve Board of Governors, the Federal Reserve Bank of Minneapolis, or any other person associated with the Federal Reserve System. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

# 1 Introduction

Rich micro heterogeneity has become a key feature of modern macroeconomics. Important questions concerning growth, innovation, financial frictions, trade policy, misallocation, and monetary policy have been successfully explored using insights from heterogeneous firm models. The size distribution of firms and its determinants implicitly shape the source and magnitude of aggregate effects in these models. The size distribution of firms has therefore become a central observable outcome that these models aim to match or explain.

Using the Longitudinal Business Database, a confidential U.S. Census Bureau panel dataset of all non-farm private firms and establishments with at least one employee, we document several important facts about the U.S. firm and establishment employment size distributions. Statistically, we find that a lognormal fits the firm size distribution better than a Pareto. This finding holds even when we consider most cuts of the truncated upper tail of the firm size distribution: the Pareto distribution provides a better fit of the right tail for only a narrow range, and the far right tail is still better described by lognormal. These results overturn the best available evidence for the United States from Axtell (2001), the key reference in the literature.

Both lognormal and Pareto distributions, however, do not fit the entire size distribution well. We therefore move beyond these simple distributions and estimate the parameters of a combination of lognormal and Pareto: a convolution of a Pareto random variable multiplied by a lognormal random variable. Perhaps not surprisingly, we find that the convolution beats the fit of the Pareto or lognormal distributions alone. Moreover, when we consider a second distribution fit criterion that is crucial in granular economies—the fraction of employment accounted for by various bins of establishment or firm size—the convolution provides a markedly better fit. Economically, the convolution can arise in a heterogeneous firm model with two sources of firm-level shock, say a demand shock and a productivity shock. Using manufacturing (revenue) TFP data, we show that the empirical distribution is well described by a straightforward lognormal distribution. We also show that the revenue distribution in manufacturing is better described by lognormal than Pareto.[1]

Models in growth theory such as those by Luttmer (2007) and Luttmer (2011) explicitly aim to rationalize a stationary distribution with a right tail that is close to (but above) Zipf's law. Arkolakis (2016) presents a model that can generate both a limit Pareto right tail and a larger number of small establishments than a close-to-Zipf Pareto distribution: customer acquisition costs induce endogenous demand shifters that decline with productivity.[2] We show that, even far in the upper

[1] We also estimate the parameters of a statistical mixture of lognormal and Pareto. Even though we find that the mixture provides the best overall fit, the convolution also beats the fit of the Pareto or lognormal distributions alone. More importantly, the convolution provides a markedly better fit for the employment share distribution.

[2] Rossi-Hansberg and Wright (2007) offer an alternative growth theory in which thinner-than-Zipf size distributions and scale-dependence arise from random growth and industry-specific human capital accumulation. Therefore, to

tail, the lognormal can be a good approximation of the size distribution of firms. More importantly, these theories clarify how the stochastic properties of idiosyncratic shocks to firm dynamics pin down the tail properties of the limit firm size distribution. For example, in the canonical growth model that uses Brownian shocks to capital to generate Zipf's law for firm sizes, the tail index is a closed form function of the stochastic process linear drift, its variance, and the capital depreciation rate. In this context, we find a lack of stability of the truncated Pareto tail index, which suggests that the Pareto approximation may prove problematic for the robustness of quantitative applications. In contrast, we show the convolution of a lognormal and a Pareto distribution to be an even better approximation.[3] Together, we view these results as suggestive evidence that models with multiple sources of firm heterogeneity are a more promising direction to flexibly generate heavy-tailed distributions other than Zipf's law. In fact, we note that the customer accumulation technology in Arkolakis (2016) delivers a firm size distribution that is a convolution of a standard Melitz-style Pareto size distribution and an endogenous demand shifter.[4]

Similarly, the Zipf's law assumption plays a crucial role in the macroeconomics of granular economies. When an economy is dominated by large firms, idiosyncratic shocks to these firms may be an important source of aggregate fluctuations, depending on the heaviness of the right tail of the distribution (Gabaix 2011, di Giovanni and Levchenko 2012, Stella 2015). This insight shapes the results of Gaubert and Itskhoki (2020), who argue that Pareto-driven granularity is critical for matching the concentration of firm-level exports within and across sectors in France. While Gaubert and Itskhoki (2020) estimate an economy-wide Pareto shape parameter that is slightly thinner than Zipf (around 1.1), they also report a substantial fraction of sectors having heavier-than-Zipf sector-level Pareto estimates. We show that estimating a convolution can avoid such implausibly low Pareto shape parameters. Our results suggest that granularity would likely play a smaller role in determining comparative advantage in their framework.

The shape of the firm size distribution affects the welfare gains from reducing trade costs and the source of those gains. di Giovanni and Levchenko (2013) show that in a canonical heterogeneous firm model, as the firm size distribution approaches Zipf, welfare gains from typical trade liberalizations rise, but these gains come nearly entirely from the intensive margin, the change in trade flows from existing exporters. In this world, some firms with a productivity above which firms choose to export are very large; reductions in tariffs or transport costs affect these large firms directly, with large changes in trade flows. In addition, in a Zipf world, reductions in fixed or

contrast the role of physical capital intensity, we report our size distribution estimation results for both manufacturing and services in Appendix C.

[3] See Nair et al. (2013) for a primer on the mathematics and the estimation of alternative heavy-tailed phenomena.

[4] In Arkolakis (2016), the market penetration choice in effect yields a negative demand shift that converges to 0 when productivity goes to $+\infty$. Since the productivity follows a Pareto distribution, the firm size distribution converges to the 'unmodified' Melitz-style Pareto size distribution in the right tail limit.

sunk costs of exporting have minimal welfare gains; marginal exporters that make up the extensive margin are too small to be economically consequential. Our results imply that the U.S. economy is significantly away from Zipf's law and provide greater scope for the extensive margin to play a significant role in trade dynamics and the gains from trade.[5]

International trade models with a Pareto firm size distribution also struggle to simultaneously match direct estimates of the firm size distribution and other important moments of firm heterogeneity, such as the exporter size premium. Exporters tend to be about 4.5 times larger than non-exporters, but Armenter and Koren (2015) show that the shape parameter implied by the exporter size premium in the data would be about 1.65. If U.S. firms were near Zipf (with a shape parameter of 1.065), these models would imply that the average exporter would be 41 times larger than the average non-exporter. A second source of heterogeneity, such as heterogeneous fixed costs, is required to rationalize both the direct distributional estimate with the exporter size premium.

Other papers in the international trade literature have also explored the implications of deviations from Zipf's law and the Pareto distribution using, for example, a lognormal, a Truncated Pareto, or multidimensional sources of firm heterogeneity. Adão et al. (2020) review the different distributional assumptions that are common in the literature and document size differences in key elasticities governing the gains from trade. Head et al. (2014) examine the consequences of a lognormal distribution in trade. Nigai (2017) shows that the combination of a Pareto right tail and a lognormal left tail not only provides a better empirical fit but also generates sizeable differences in the gains from trade. The mixture we estimate in this paper has similarities with the two-piece distribution in Nigai (2017). However, motivated by our findings on the fit and stability of the Pareto in the upper tail, the mixture we consider deviates from a pure Pareto by also including a lognormal component in the upper tail. Sager and Timoshenko (2019) argue that the Double Exponentially Modified Gaussian provides a better fit of export data with quantitative implications for trade elasticity estimates. The convolution we estimate in this paper is a parsimonious special case that they also examine explicitly. It yields parameter estimates that are easily compared to simple Pareto and lognormal distributions and can be generated from two random variables, one distributed lognormal and the other Pareto, that are readily familiar to economic modelers. Despite having only three parameters, we find that the convolution is flexible enough to provide a good fit for the U.S. size distributions. Finally, Fernandes et al. (2015) model the firm productivity distribution with a lognormal distribution to match the empirical evidence on the importance of the intensive margin of trade.

The rest of the paper is organized as follows. Section 2 introduces the data we use in the paper. Section 3 explains the parametric distributions that we fit to the data. Section 4 presents our main

[5]di Giovanni et al. (2011) show how international trade outcomes affect estimates of the firm size distribution through the lens of standard trade models.

results on the employment size distributions and analyzes the TFP establishment distribution in manufacturing. Section 5 concludes.

## 2 Data Description

The Center for Economic Studies at the U.S. Census Bureau created and maintains a longitudinally-linked establishment-level database: the Longitudinal Business Database (LBD). The LBD covers the non-farm private economy of establishments with at least one employee. The wide coverage of the LBD comes at a cost, as it only provides number of employees, payroll, location, firm ID, and sectoral affiliation of each establishment; we do not observe revenues, intermediate inputs, capital investment, prices or other important information. See Jarmin and Miranda (2002) for additional details on the LBD. Table 1 shows the number of establishments and firms we use in each year. 2012 was the latest available Census year, and we chose 10-year intervals based on that, with the addition of 1997 for comparability to Axtell (2001).

Table 1: Number of observations

| Year | Est. | Firm |
|---|---|---|
| 1982 | 4,490,000 | 3,620,000 |
| 1992 | 5,580,000 | 4,390,000 |
| 1997 | 6,060,000 | 4,770,000 |
| 2002 | 6,290,000 | 4,900,000 |
| 2012 | 6,590,000 | 4,980,000 |

Note: Numbers are rounded. Source: LBD.

The LBD establishment is defined as a single physical location where business is conducted; this definition is not equivalent to the IRS Establishment Identification Number (EIN), which might be comprised of more than one LBD establishment. The LBD establishment is also not equivalent to a legal entity; the LBD includes a firm ID variable that groups together establishments owned by the same firm.[6]

In most of this paper, we measure the size of an establishment or firm with its number of employees. The analysis is conducted on the whole universe, but in Section 4.4 and Appendices C and D, we estimate size distributions on disaggregated data. We first consider two subsamples:

[6]The LBD firm ID was created using information from the quinquennial Economic Census and the annual Company Organization Survey. The Company Organization Survey is only submitted to large firms and a subset of small firms, so the firm ID is not entirely reliable outside of Census years; this does not affect our results as we only use Census years in our analysis.

manufacturing and services, where the latter excludes retail, wholesale, and FIRE (finance, insurance, and real estate).[7]. Then, we estimate distributions across two-digit sectors and report results for both the largest sectors and summary statistics across all sectors.

The LBD covers nearly the entire U.S. business population, but only provides us with limited information. Census surveys on manufacturing establishments include much richer detail on their operation. Foster et al. (2016) estimate Revenue Total Factor Productivity (TFPR)—specifically, establishment-level revenue total factor productivity—with data from the quinquennial Census of Manufactures (CM). Since the distribution of productivity shocks represents an important primitive assumption in many theories, we extend our analysis to the TFPR distribution of manufacturing establishments, discussed in Section 4.5. Table 2 shows the number of establishments and firms in the services and manufacturing sectors as well as the number of establishments in the TFPR dataset, which does not include 2012. Note that services make up the vast majority of the establishments and firms in the LBD and thus contribute a greater share to the distribution calculations for the universe of establishments and firms. Also note that the TFPR data is available for roughly half of manufacturing establishments.

Table 2: Number of observations by sector

| | Services | | Manufacturing | | |
|---|---|---|---|---|---|
| Year | Est. | Firm | Est. | Firm | TFPR est. |
| 1982 | 1,430,000 | 1,280,000 | 330,000 | 270,000 | 190,000 |
| 1992 | 2,000,000 | 1,730,000 | 350,000 | 290,000 | 190,000 |
| 1997 | 2,240,000 | 1,920,000 | 360,000 | 300,000 | 210,000 |
| 2002 | 3,070,000 | 2,500,000 | 330,000 | 280,000 | 180,000 |
| 2012 | 3,440,000 | 2,760,000 | 280,000 | 230,000 | |

Note: Numbers are rounded.

Source: LBD for employment and Foster et al. (2016) for TFPR.

Since the U.S. Census Bureau prefers researchers not to disclose too many tabulations of the raw data, we use the Business Dynamic Statistics (BDS) to complement the statistics we disclosed and produce some of the charts and tables in the paper. The BDS are publicly available on the U.S. Census Bureau's website and are drawn from the same underlying data that we use for our analysis. All of our distribution estimates, however, are obtained using the underlying confidential LBD.

[7]We define the manufacturing sector as all establishments with two-digit SIC codes $\in [20,40)$ for years 1982, 1992, and 1997. For 2002 and 2012, we define manufacturing as establishments with two-digit NAICS codes $\in [31,33]$. The services sector is defined as all two-digit SIC codes $\in [70,90)$ and two-digit NAICS codes $\in [54,81]$ and 51 for the same years. We assign firms to the sector where most of its employees work. FIRE stands for Finance, Insurance and Real Estate.

# 3 Parametric Distributions and Estimation Methods

Motivated by the existing literature, we fit four parametric distributions to the data. The first and most popular distribution is Pareto. Axtell (2001) provides the benchmark evidence that the employment and sales firm size distributions in the U.S. are well approximated by a Pareto close to Zipf's law. As a consequence, along with analytical tractability, Pareto is widely used in heterogeneous firm models that assume an exogenous distribution, and much of the endogenous growth literature focuses on generating a Pareto distribution. The CDF of a Pareto with scale parameter $x_m$ and shape parameter $\alpha$ is:

$$F_P(x) = 1 - \left(\frac{x_m}{x}\right)^{\alpha}, \tag{1}$$

with $x_m > 0$ and $\alpha > 0$. For this type of Pareto distribution, the mean is $\frac{\alpha}{\alpha-1}x_m$ for $\alpha > 1$ and the variance is $\frac{\alpha}{(\alpha-1)^2(\alpha-2)}x_m^2$ for $\alpha > 2$. When $\alpha \leq 2$, the variance is undefined, and when $\alpha \leq 1$, the mean and variance are undefined.[8] These properties are especially important given the range of shape parameter estimates we find in the data.

The lognormal distribution has frequently been considered as a possible alternative to the Pareto distribution. The log of a lognormal random variable follows a normal distribution. The CDF of a lognormal with parameters $\mu$ and $\sigma$ is:

$$F_L(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \tag{2}$$

with $\mu \in (-\infty, \infty)$, $\sigma > 0$, and where $\Phi(x)$ is the CDF at $x$ of a standard normal distribution. For the lognormal distribution, the mean is given by $e^{\mu+\sigma^2/2}$ and the variance $(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$.

Besides the two most popular parametric distributions in the literature, we consider two distributions that parsimoniously combine Pareto and lognormal with the hope to provide a better fit. One is a pure statistical mixture of the two distributions and the other a convolution of the two distributions.

Specifically, the CDF $F_M$ of the mixture of a Pareto and a lognormal using a mixing parameter $p$ is:

$$F_M(x) = pF_L(x) + (1-p)F_P(x), \tag{3}$$

[8] Despite the mean and variance being undefined, Pareto is not degenerate when $\alpha < 1$. However, the employment share distribution implied by a Pareto with shape parameter below one does become degenerate when the mean is undefined, as we show in Table 9. If a distribution mixes any Pareto with a shape parameter below one, some incredibly large firms would be generated in reasonably sized samples, leaving too many employees belonging to the largest bin of firms.

where $F_L$ is the CDF of a lognormal with parameters $\mu$ and $\sigma$ and $F_P$ is the CDF of a Pareto with scale parameter $x_m$ and shape parameter $\alpha$.[9]

Finally, we define the convolution as the product of a Pareto random variable with CDF $F_P$ and a lognormal random variable with CDF $F_L$. Equivalently, the log of such convolution is the sum of a normal distribution and an exponential distribution.[10] Thus, the CDF of the convolution is:

$$F_C(\ln x) = \Phi(\alpha(x-\mu);0,\alpha\sigma) - e^{-\alpha(x-\mu)+\frac{(\alpha\sigma)^2}{2}}\Phi(\alpha(x-\mu);(\alpha\sigma)^2,\alpha\sigma), \tag{4}$$

where $\Phi(x;\mu,\sigma)$ is the CDF at $x$ of a normal distribution with parameters $\mu$ and $\sigma$.

Most of the previous literature on the firm size distribution evaluated the fit of the Pareto distribution using regression analysis on binned data (Axtell 2001, Gabaix 2009). It has been widely documented in both the statistics and econometrics literature that regression analysis is not well suited to test the goodness of fit of a Pareto distribution.[11] Since we are estimating parametric models with a known and simple likelihood, we use maximum likelihood estimation for its excellent statistical properties, and the likelihood of a parametric distribution model is a simple function of its PDF.[12] To determine which distribution best fits the data, we rely on formal statistical testing.

For nested models, we use the popular likelihood ratio test. If $L_1$ is the maximum likelihood of a model, $L_0$ is the maximum likelihood of a reduced version of the model, and $k$ is difference between number of parameters, then $\Lambda = -2\ln(\frac{L_0}{L_1})$ is asymptotically distributed according to $\chi^2_k$.

For non-nested models, we use a test developed by Vuong (1989). This test is a function of the likelihood ratio test: $\tilde{\Lambda} = n^{\frac{1}{2}}\frac{\Lambda}{\omega_n}$, where $f$ is the model in the numerator of the ratio, $g$ the model in the denominator, and $\omega_n$ is the empirical counterpart of the asymptotic variance of the likelihood ratio statistic $\Lambda$. Under the null hypothesis $H_0$ that the two models are equivalent, $\tilde{\Lambda} \xrightarrow{D} N(0,1)$; under the first alternative $H_f$ that model $f$ is better, $\tilde{\Lambda} \xrightarrow{a.s} \infty$; finally, under the second alternative $H_g$ that model $g$ is better, $\tilde{\Lambda} \xrightarrow{a.s} -\infty$. We also use the Akaike information criterion (AIC), which has the attractive feature of penalizing models with a higher number of parameters.

Our preferred measure of establishment and firm size, the number of employees, is discrete, whereas all the distributions we described so far have a continuous support. We follow Buddana

[9]Our mixture specification mixes Pareto and lognormal even in the upper tail. The assumption that the upper tail is strictly Pareto is a question we seek to answer. See Figure 2 and Table 5 for evidence that the assumption might in fact be restrictive even in the upper tail.

[10]See Reed (2001) for an example of stochastic growth processes that can yield such distribution. Sager and Timoshenko (2019) also rationalize this distribution using a model with Pareto productivity shocks and lognormal demand shocks.

[11]Clauset et al. (2009) discuss this issue in detail and also provide a Monte Carlo simulation to show that the lognormal distribution can approximate a Pareto closely when evaluated with the regression analysis approach used in Axtell (2001) and Gabaix (2009). See also Eeckhout (2009).

[12]We ran Monte Carlo simulations to investigate the accuracy of the MLE in estimating our models and found it to be reliable. Results are in Appendix A.2.

and Kozubowski (2015) and discretize the distributions. In particular, if $F(\cdot)$ is the CDF of a continuous distribution, the PMF of the discretized distribution is defined as $\Pr(X = n) \equiv F(n+1) - F(n)$. In other words, the continuous distribution is discretized by creating a bin for each integer value.

Finally, our data includes only establishments and firms with at least one employee, but the lognormal and convolution have support starting at zero; to make the estimation possible, we shift the lognormal to the right by one unit, along with the lognormal components of the mixture and the convolution.

# 4 Employment Distribution Results in the United States

In this section, we highlight two stylized facts that emerge from the U.S. firm and establishment employment distributions in the census years 1982, 1992, 1997, 2002, and 2012.[13] We use data from the LBD, except for Section 4.5, which uses TFPR data from the Census of Manufactures.

## 4.1 Lognormal versus Pareto

**Stylized Fact 1** *A lognormal fits both establishment and firm size employment distributions better than the commonly used Pareto, even far in the truncated upper tail.*

Table 3: Pareto and lognormal estimates using the entire sample

| | Pareto | | Lognormal | | | |
|---|---|---|---|---|---|---|
| | $\alpha$ | | $\mu$ | | $\sigma$ | |
| Year | Est. | Firm | Est. | Firm | Est. | Firm |
| 1982 | 0.57 | 0.61 | 1.38 | 1.21 | 1.52 | 1.81 |
| 1992 | 0.56 | 0.61 | 1.40 | 1.21 | 1.53 | 1.71 |
| 1997 | 0.56 | 0.61 | 1.41 | 1.17 | 1.56 | 1.74 |
| 2002 | 0.55 | 0.60 | 1.44 | 1.15 | 1.57 | 1.80 |
| 2012 | 0.56 | 0.62 | 1.37 | 1.14 | 1.61 | 1.80 |

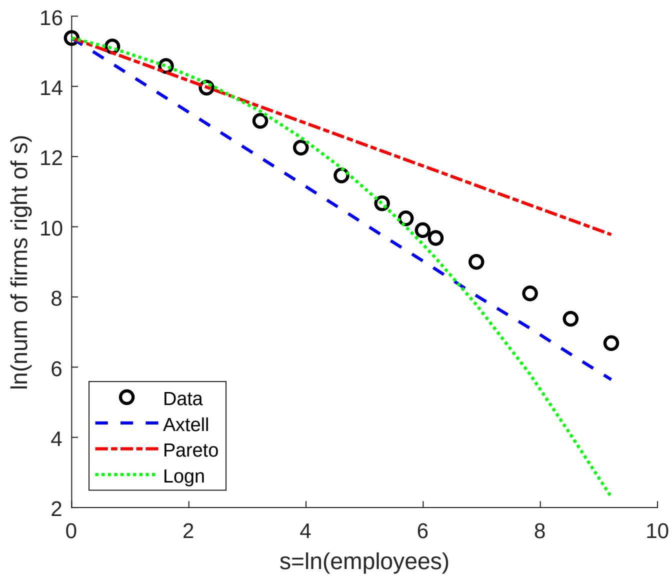
Table 3 shows the maximum likelihood estimates of the lognormal and Pareto distributions using the entire sample of establishments and firms. Both the Vuong test and AIC find that the

[13] To capture long-run trends we use 10-year intervals starting with 1982. We add 1997 for comparison to Axtell (2001).

lognormal distribution is preferred over Pareto for establishments and firms in all years, while the best-fitting Pareto consistently has a shape parameter significantly below one, around 0.6. These parameters are all tightly estimated, with standard errors and test statistics available upon request.

Figure 1 provides a graphical representation of the fit of the parametric distributions with 1997 binned data. It depicts the Complementary Cumulative Distribution Function (CCDF) in log space: for each logarithm of $s$ number of employees, it shows how many firms are larger than $s$. The black circles represent binned data; distributions with fits above the black circles represent too many large firms compared to the data, while distributions with fits below the black circles have too few large firms relative to the data.

Figure 1: Data vs. Estimated Distributions




Note: The black circles are computed using the number of observations in the last column of Table 4, and the lines depict the CCDF implied by the estimated parametric distributions.

Our estimate of the Pareto shape parameter (the dot-dashed line) fits the left tail but implies far too many large firms. Notably, the Axtell Pareto fit (the dashed line) with shape parameter 1.06 remains below the data, but, overall, it matches the slope of the black circles excluding the left tail, which is essentially by design of the regression-based estimation procedure on binned data. The lognormal fit is reasonable through about $e^5 \approx 150$ employees but produces too few very large firms. Nonetheless, its excellent fit in the left tail and the middle of the distribution underpins its better overall fit, compared to Pareto, using MLE.

As Figure 1 shows, in the log-log space of the CCDF, the curvature of the data cannot be matched by the straightness of a Pareto distribution. In other words, a Pareto distribution cannot fit the entire firm size distribution, including the left tail. Axtell's Pareto distribution essentially never touches the data, though it has a similar slope as the right tail of the data distribution. MLE reasonably tries to fit as much data as possible by flattening the Pareto line to fit the left tail, where most firms are. For this reason, MLE finds a Pareto shape parameter lower than one to be a better fit than Axtell's estimate. A shape parameter lower than one is not plausible, but it should not be viewed as evidence against MLE. Instead, it is evidence against a Pareto distribution.

In summary, Pareto fails at fitting the entire employment distribution, while lognormal does well in the left tail and middle of the empirical distribution, but fails in the right tail, when using the parameter estimates from the whole distribution. We investigate the right tail of the distribution next, but the results on the entire distribution are important, as many leading theories assume a Pareto distribution for the entire employment distribution and not just the right tail. Section 4.2 considers a mixture and a convolution of Pareto and lognormal that both fit well the entire distribution and not just a part of it.

**Right Tail Estimates** While the lognormal distribution might be a better fit overall, only the upper tail is the relevant portion of the distribution for some economic questions. Table 4 presents the parameter estimates for a Pareto distribution and a truncated lognormal distribution at various employment thresholds for the 1997 firm size distribution. We estimate the Pareto shape parameter using maximum likelihood (the second column) and using Axtell's empirical strategy (the fourth column).[14] Using the BDS, we calculated the share of total U.S. employment that is accounted for by the firms above some of the thresholds in Table 4: firms with at least 5 employees account for 87.8% of employment, firms with at least 50 employees account for 69.5%, firms with at least 100 employees account for 62.0%, firms with at least 500 employees account for 47.7%, firms with at least 1,000 employees account for 42.6%, firms with at least 5,000 employees account for 35.5%, and firms with at least 10,000 employees still account for 24.7% of total employment. Thus, fitting the right tail of the distribution can still fit a large fraction of U.S. employment.

[14] Estimates for the other Census years are available upon request. The main results in this paper are robust over time.

Table 4: Pareto and lognormal in the firm size right tail

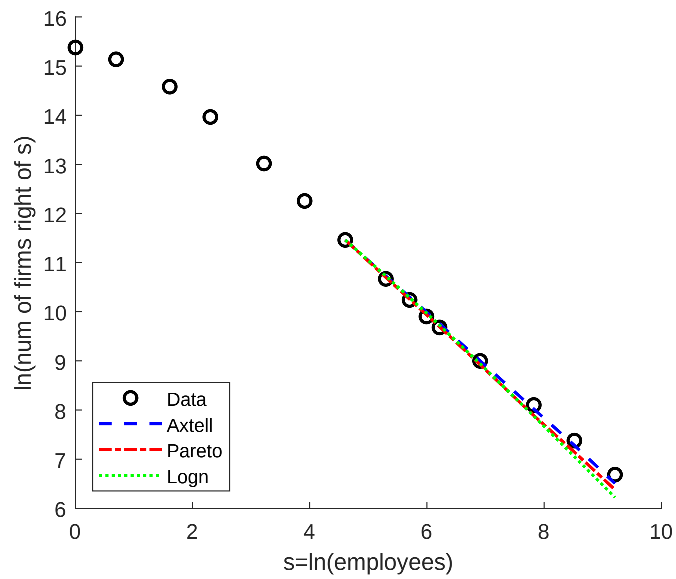
| | Pareto | lognormal | | Axtell | |
|---|---|---|---|---|---|
| Threshold | $\alpha$ | $\mu$ | $\sigma$ | $\alpha$ | N (rounded) |
| 2 | 0.76 | -0.02 | 1.93 | 1.08 | 3,750,000 |
| 5 | 0.97 | -5.88 | 3.11 | 1.11 | 2,150,000 |
| 10 | 1.05 | -14.97 | 4.26 | 1.12 | 1,160,000 |
| 25 | 1.11 | -14.79 | 4.22 | 1.12 | 450,000 |
| 50 | 1.12 | -16.44 | 4.46 | 1.13 | 210,000 |
| 100 | 1.10 | -19.10 | 4.85 | 1.14 | 95,000 |
| 200 | 1.05 | -22.01 | 5.28 | 1.15 | 43,000 |
| 300 | 1.02 | -23.80 | 5.56 | 1.16 | 28,000 |
| 400 | 1.01 | -24.73 | 5.72 | 1.17 | 20,000 |
| 500 | 1.01 | -21.29 | 5.40 | 1.18 | 16,000 |
| 1,000 | 1.01 | -5.32 | 3.73 | 1.24 | 8,100 |
| 2,500 | 1.05 | 1.99 | 2.68 | 1.33 | 3,300 |
| 5,000 | 1.11 | 5.70 | 2.00 | 1.36 | 1,600 |
| 10,000 | 1.23 | 7.21 | 1.66 | 1.48 | 800 |

Notes: The estimates are reported using the 1997 firm size sample in order to ensure consistency and comparability with Axtell (2001), the main benchmark in the literature. The Axtell column represents a binned regression adapted to the truncated sample.

In Table 4, the Pareto shape parameter is not stable as the upper tail threshold changes. At various thresholds the shape parameter is in fact near one, corresponding to Zipf's law. Note, however, that Zipf's law draws its power from the thickness of the right tail, and at the highest thresholds of 5,000 and 10,000, the shape parameter is increasingly above one. Indeed, the lack of stability of the shape parameter estimates across cutoffs suggests that the underlying distribution is not Pareto: a true Pareto distribution would have shape parameter estimates that are invariant to the cutoff or, at least, more stable further in the upper tail.[15]

[15] The well-known Hill plot for visually identifying power-law distributions is based on this stability argument. The lognormal distribution does not have the same property: if a sample is drawn from lognormal, estimating a truncated lognormal on a truncated sample would not deliver the same coefficients as on the whole sample. For this reason, the estimates of a truncated lognormal on the right tail do not translate to an overall distribution. Some theories of the size distribution such as Reed (2001) and Arkolakis (2016) generate a double Pareto size distribution as opposed to the simple Pareto distribution we consider. Note that, even in that case, the distribution should follow a simple Pareto distribution in the upper tail.

Figure 2: Data vs. Estimated Distributions Above 100 Employees




Note: The black circles are computed using the number of observations in the last column of Table 4, and the lines depict the CCDF implied by the estimated parametric distributions.

**Does the Pareto Provide a Better Fit?** In Figure 2, we show the fit of the parametric distributions when truncating firms below 100 employees. In contrast to Figure 1, the fit of each distribution is much closer to each other and to the few available data bins. Nonetheless, in Table 5, we use the Vuong statistic to formally test which distribution has a better fit for each truncation threshold, using all available confidential data from the LBD. The Vuong statistic uses simple standard normal critical values for a two-tailed test, and these p-values are provided in the third column of the table.

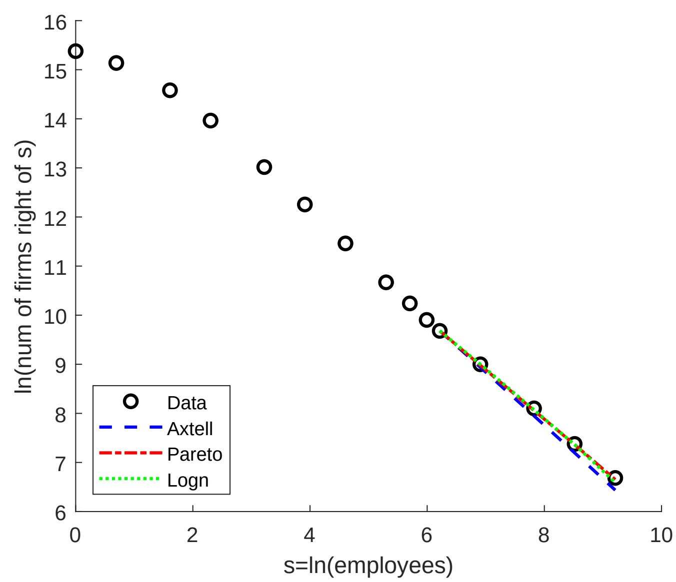
The lognormal provides the best fit both when the threshold is at or below 10 employees and when the threshold is at or above 400 employees. For a threshold of 300 employees, neither distribution is preferred. Thus, the lognormal fit typically dominates the Pareto fit even far in the upper tail, except in a narrow truncation window.

In some situations, however, the analytical tractability of Pareto may necessitate its use. For applications in which the right tail is crucial, Table 4 can provide useful estimates of the Pareto shape parameter. Since the far right tail (above 10,000 employees) has a shape parameter significantly away from Zipf (1.23), this cautions against using 1.06 in calibration exercises.

Table 5: Pareto vs lognormal (Vuong statistic)

| Threshold | Vuong statistic | p-value | Winner |
|---|---|---|---|
| 2 | -223.6 | 0.00 | lognormal |
| 5 | -58.4 | 0.00 | lognormal |
| 10 | -11.3 | 0.00 | lognormal |
| 25 | 18.3 | 0.00 | Pareto |
| 50 | 22.4 | 0.00 | Pareto |
| 100 | 18.4 | 0.00 | Pareto |
| 200 | 6.4 | 0.00 | Pareto |
| 300 | 1.0 | 0.32 | None |
| 400 | -2.0 | 0.05 | lognormal |
| 500 | -3.2 | 0.00 | lognormal |
| 1000 | -3.3 | 0.00 | lognormal |
| 2500 | -3.3 | 0.00 | lognormal |
| 5000 | -3.4 | 0.00 | lognormal |
| 10000 | -2.7 | 0.01 | lognormal |

Figure 3: Data vs. Estimated Distributions Above 500 Employees




Note: The black circles are computed using the number of observations in the last column of Table 4, and the lines depict the CCDF implied by the estimated parametric distributions.

While it is commonly assumed that the Pareto distribution can provide a better fit of the right tail, Figure 3, which shows firms with more than 500 employees, makes clear that the lognormal distribution can fit these data well. But these distributions still provide very different implications for the largest firms. Intuitively, this can be seen by extending Figure 3 further into the right tail, beyond the available data bins, in which Pareto produces far more extremely large firms. We show this in Figure 5 in the appendix.

**Revisiting the Zipf's Law Evidence** How can these results be reconciled with those of Axtell (2001)? Using a popular methodology, Axtell explores the fit of a Pareto distribution by running a regression of the logarithm of the frequency distribution on the logarithm of (binned) firm size. Axtell (2001, p. 1820) concludes: *"the Zipf distribution is an unambiguous target that any empirically accurate theory of the firm must hit."*

How well a line fits the log frequency plot is often interpreted as evidence of the fit of the Pareto distribution. As extensively explained by Clauset et al. (2009) and Eeckhout (2009), this method is ill-suited to determine how well the Pareto fits, as it generates significant systematic errors.[16]

Moreover, Axtell (2001) computes the frequency distribution using successive bins of increasing size in powers of three. This approach yields only 13 data points and the regression estimation on such a binned sample effectively gives more weight to the observations in the right tail.

We first replicated the Axtell (2001) procedure to ensure that differences in our results were not due to the underlying sample used. Axtell (2001) also uses the underlying data which became the LBD. In his regressions, the slope was 2.059 with a standard deviation of 0.054, which implies a Pareto shape parameter at 1.059. Our replication of his linear regression with our data produced a slope of 2.034 with a standard deviation of 0.045, which implies that the differences in results are not due to the underlying data, but to the methodologies used.

Visual inspection of Figure 1 from Axtell (2001, p. 1819) also provides some intuition for the difference in results. Axtell's first and last bins are both below the regression line. To fit the first bin, which contains a large portion of our sample, would require a shallower slope, or a Pareto shape parameter below one. To fit the last bin, containing the largest firms, would require a steeper slope, or a Pareto shape parameter above the 1.06 that Axtell found. This is consistent with the larger shape parameter for the far right tail in Table 4.

**Discussion** Overall, the evidence in this section indicates that the lognormal typically provides a better fit than the Pareto, even in the truncated right tail of the firm size distribution. In the binned right tail, the best-fitting lognormal may look quite similar to the best-fitting Pareto.

[16] In Appendix A.1, we further analyze the ability of the Axtell (2001) regression approach to recover the true parameters of a Pareto distribution and other distributions. By design, the regression approach yields a power law shape parameter. The parameter estimated on the full (non-binned) data has a bias towards Zipf's law when the true distribution is not Pareto.

Does this mean that, despite the better statistical fit of a lognormal, these different distributions are interchangeable in economic models? Existing models suggest the two distributions are not equivalent in their implications or origins. We think that, when the analytical convenience of the Pareto distribution or exact scale invariance is not a theoretical requirement, quantitative practitioners should be careful about choosing the Pareto distribution over the lognormal distribution, even far in the right tail.[17] This means that Pareto in general, whether our estimate or Axtell's, is likely not the more appropriate choice for most economic contexts.

## 4.2 Mixture and Convolution Distributions

We have so far compared how the Pareto and lognormal distributions fit the firm and establishment size distributions. It is evident from Figure 1 that neither fits the entire distribution well. Lognormal can provide a good fit of the left tail and middle of the distribution, at the cost of missing the right tail. Both distributions are shown in Figures 2 and 3 to be able to fit the right tail, as long as the rest of the distribution is ignored. However, recent papers have shown that for some economic theories the fit of the distribution outside of the right tail can be consequential. For this reason, we explore the fit of a statistical mixture of lognormal and Pareto as well as a convolution of a Pareto random variable multiplied by a lognormal random variable.

**Stylized Fact 2** *Both the mixture and the convolution of lognormal and Pareto distributions fit the size distributions better than lognormal alone. Statistically, a mixture provides the best size distribution fit, but economically, the convolution may be more suitable given its better employment share distribution fit.*

**Mixture Estimates** Table 6 provides the parameter estimates for the statistical mixture distribution of a lognormal and a Pareto. The parameters $\mu, \sigma$, and $\alpha$ have the same meanings as before (see Section 3). Now we also estimate $x_m$, the minimum of the Pareto distribution. This approach effectively means that the Pareto distribution is allowed to work on an endogenously chosen cutoff of the right tail of the distribution. In practice, this is approximately 3 employees and stable across both firms and establishments.

The mixture also has the mixing $p$ parameter, specifying the degree to which the estimated distribution is lognormal. For both establishments and firms, this lognormal mixing parameter $p$ is about 0.8 to 0.9, without much meaningful difference between establishments and firms across years. If anything, the distribution appears to be getting more lognormal over time, especially for establishments.

[17] Typically, the analytical tractability of the Pareto distribution come from its scale invariance, in the sense that the distribution is invariant to truncation of the left tail. See Rossi-Hansberg and Wright (2007) for related discussions on the apparent scale dependence of both the establishment size distribution and the firm size distribution.

Comparing the estimates between Tables 3 and 6, the Pareto shape parameter is systematically higher in the mixture. This is consistent with the estimates of the right tail in Table 4. Since the scale parameter $x_m \approx 3$, the mixture is mixing in a Pareto only above this threshold, analogous to the truncated estimations. This result also shows that the estimation did not favor a higher threshold $x_m$ yielding a Pareto component closer to Zipf's law (see Table 4). Over time, the estimated Pareto shape parameters also appear to be slightly declining for firms, but not for establishments.

Table 6: Parameter Estimates - Mixture

| | $\mu$ | | $\sigma$ | | p | | $x_m$ | | $\alpha$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| Year | Est. | Firm | Est. | Firm | Est. | Firm | Est. | Firm | Est. | Firm |
| 1982 | 1.18 | 1.02 | 1.50 | 1.43 | 0.83 | 0.86 | 3.39 | 3.37 | 0.80 | 0.75 |
| 1992 | 1.22 | 1.00 | 1.53 | 1.45 | 0.84 | 0.85 | 3.55 | 3.48 | 0.85 | 0.76 |
| 1997 | 1.25 | 1.00 | 1.58 | 1.49 | 0.86 | 0.86 | 3.57 | 3.47 | 0.85 | 0.74 |
| 2002 | 1.32 | 1.07 | 1.57 | 1.48 | 0.89 | 0.89 | 3.55 | 3.35 | 0.80 | 0.69 |
| 2012 | 1.24 | 0.97 | 1.60 | 1.52 | 0.91 | 0.89 | 4.47 | 3.62 | 0.83 | 0.68 |

**Convolution Estimates** The parameter estimates for the convolution of a lognormal and Pareto distribution are shown in Table 7. The lognormal $\mu$ and $\sigma$ parameters are systematically lower than their mixture counterparts, while most notably, the Pareto shape parameter $\alpha$ is higher, and always well above 1, a point that we discuss later in the contexts of the employment share distribution and granularity.

Table 7: Parameter Estimates - Convolution

| | $\mu$ | | $\sigma$ | | $\alpha$ | |
|---|---|---|---|---|---|---|
| Year | Est. | Firm | Est. | Firm | Est. | Firm |
| 1982 | 0.62 | 0.44 | 1.27 | 1.23 | 1.29 | 1.26 |
| 1992 | 0.70 | 0.45 | 1.32 | 1.26 | 1.39 | 1.26 |
| 1997 | 0.72 | 0.43 | 1.37 | 1.29 | 1.43 | 1.25 |
| 2002 | 0.76 | 0.44 | 1.39 | 1.28 | 1.46 | 1.23 |
| 2012 | 0.75 | 0.34 | 1.46 | 1.33 | 1.58 | 1.22 |

A lower shape parameter $\alpha$ for firms, relative to establishments, means that the right tail firm size is thicker than the establishment size right tail. The mixture also has a lower Pareto shape parameter for firms than establishments, but since both are below one, both distributions are economically implausible.[18]

The parameter estimates remain reasonably stable over time. If anything, the distribution appears to embed a more dispersed lognormal component over time, especially for establishments. The Pareto component for the establishment size distribution also appears to have a slightly less-heavy tail and it is becoming less heavy over time, while the tail of the Pareto component for the firm size distribution, though slightly declining, is more stable over time.

**Testing the Four Models** With estimates for four distributions in hand, we formally test which distribution fits the data best. Pareto and lognormal are both nested in a mixture and can therefore be tested with the likelihood ratio test. For non-nested models, we use the test developed by Vuong (1989) described earlier. For the 6 paired tests (testing each of 4 distributions against each other), the rankings are consistent across years and between establishments and firms: A mixture is always the most preferred distribution and a convolution the second most preferred. Test statistics and p-values are in Table 16 in Appendix B. As an alternative, we also computed the AIC for all the distributions and find identical rankings, in Tables 17 and 18.

While the statistical ranking is clear, it does not provide any feel for the nature of the fit. For this, we turn to a tabulation of the 1997 BDS and data simulated using the parameter estimates in Tables 3 to 7. The first column of Table 8 shows the employment size categories provided by the BDS; the second column shows the tabulation of the U.S. firms by size in 1997: for instance, 54.8% of firms have between 1 and 4 employees. Starting from the third column, we show the tabulation of simulated data.

Table 8 clearly indicates that the Pareto with the shape parameter from Axtell (2001) ($\alpha = 1.06$) provides a poor fit of the U.S. firm size distribution. Even the Pareto with our estimate of the shape parameter ($\alpha = 0.61$), the second column, does not fit the data well, putting too much weight on the left and right tails. Lognormal provides a better fit than the Pareto distributions, but its right tail is too thin as shown by the last two bins. Finally, the mixture and convolution both provide a very good fit of the U.S. firm size distribution.

Figure 4 depicts the implied CCDF in log space. It confirms that the mixture and convolution provide an improved fit. The convolution also has a slightly too-thin tail, while the mixture has a slightly too-thick tail. Still, by this criterion, both handily outperform the lognormal and Pareto distributions alone.

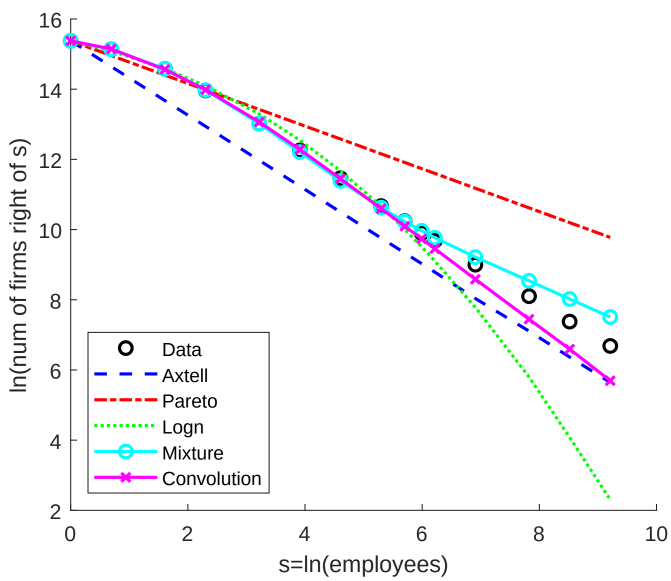
[18]The Pareto estimated on the entire sample actually has a lower shape parameter on establishments than firms. Since more flexible distributions show a thicker tail for firms, this further emphasizes the importance of using these more flexible distributions. See also Rossi-Hansberg and Wright (2007, figure 1 on p. 1649) for related empirical evidence and a model of larger deviations from Zipf's law in the size distribution of establishments compared to the firm size distribution.

Table 8: Tabulation of 1997 BDS data and simulated data

| | BDS | Pareto $\alpha = 0.61$ | Pareto $\alpha = 1.06$ | Pareto $\alpha = 1.23$ | Lognormal | Mixture | Convolution |
|---|---|---|---|---|---|---|---|
| 1 to 4 | 54.8 | 62.4 | 81.8 | 86.2 | 54.9 | 55.0 | 55.8 |
| 5 to 9 | 21.2 | 12.9 | 9.5 | 7.9 | 17.3 | 20.4 | 19.4 |
| 10 to 19 | 12.2 | 8.5 | 4.5 | 3.4 | 12.4 | 12.5 | 12.3 |
| 20 to 49 | 7.5 | 6.9 | 2.6 | 1.7 | 9.5 | 7.9 | 8.1 |
| 50 to 99 | 2.3 | 3.2 | 0.8 | 0.5 | 3.4 | 2.3 | 2.6 |
| 100 to 249 | 1.2 | 2.6 | 0.5 | 0.2 | 1.8 | 1.1 | 1.3 |
| 250 to 499 | 0.4 | 1.2 | 0.2 | 0.1 | 0.4 | 0.3 | 0.4 |
| 500 to 999 | 0.2 | 0.8 | 0.1 | 0.0 | 0.1 | 0.2 | 0.2 |
| $\geq$ 1,000 | 0.2 | 1.5 | 0.1 | 0.0 | 0.0 | 0.2 | 0.1 |

Note: The first column shows the BDS employment brackets. The second column shows the share of firms within each bracket in the BDS. The third to last columns show the share of firms within each bracket in simulated data drawn from parametric distributions.

Figure 4: Data vs. Estimated Distributions




Note: The black circles are computed using the number of observations in the last column of Table 4, and the lines depict the CCDF implied by the estimated parametric distributions.

**Implications for Employment Shares** In Table 9, we take a completely different cut of the data: the fraction of overall employment accounted for by firms in each bin.[19] Since these data moments are not explicitly targeted in the estimation, we can use them as an "out-of-sample" test of the fit of the estimated distributions. The second column shows, for example, that firms with 20 to 49 employees account for 10.6 percent of overall employment in the economy. Here, Axtell's estimate

[19] See Appendix B for confidence bands and a description of how this table was constructed.

of Pareto with a shape parameter of 1.06 fits the data somewhat better in absolute terms for each bin. Notably, however, too much employment is accounted for by very small firms (11.6 percent versus 5.7 percent in the data for firms with 1 to 4 employees) and very large firms (36.6 percent versus 24.7 percent for firms with more than 10,000 employees).

Table 9: Fraction of 1997 firm employment

| | | Pareto | | | | | |
|---|---|---|---|---|---|---|---|
| | BDS | $\alpha = 0.61$ | $\alpha = 1.06$ | $\alpha = 1.23$ | Lognormal | Mixture | Convolution |
| 1 to 4 | 5.65 | 0.00 | 11.57 | 27.25 | 7.04 | 0.46 | 6.57 |
| 5 to 9 | 6.53 | 0.00 | 5.44 | 10.51 | 7.52 | 0.53 | 7.14 |
| 10 to 19 | 7.73 | 0.00 | 5.41 | 9.28 | 11.06 | 0.67 | 9.26 |
| 20 to 49 | 10.62 | 0.00 | 6.95 | 10.41 | 19.07 | 0.94 | 13.56 |
| 50 to 99 | 7.52 | 0.00 | 5.06 | 6.60 | 15.56 | 0.64 | 9.80 |
| 100 to 249 | 8.72 | 0.01 | 6.41 | 7.28 | 18.00 | 0.69 | 11.22 |
| 250 to 499 | 5.54 | 0.01 | 4.63 | 4.58 | 9.76 | 0.44 | 7.05 |
| 500 to 999 | 5.09 | 0.01 | 4.45 | 3.91 | 6.18 | 0.44 | 5.94 |
| 1,000 to 2,499 | 7.07 | 0.02 | 5.61 | 4.30 | 4.00 | 0.67 | 6.43 |
| 2,500 to 4,999 | 5.40 | 0.02 | 4.05 | 2.70 | 1.18 | 0.61 | 3.98 |
| 5,000 to 9,999 | 5.46 | 0.03 | 3.89 | 2.30 | 0.44 | 0.73 | 3.34 |
| $\geq$ 10,000 | 24.68 | 99.90 | 36.54 | 10.87 | 0.19 | 93.18 | 15.70 |

Note: The first column shows the BDS employment brackets. The second column shows the employment share of firms within each bracket in the BDS. The third to last columns show the employment share of firms within each bracket in simulated data drawn from parametric distributions.

This table also makes clear the odd results of a Pareto distribution or mixture containing a Pareto with a shape parameter below one. Simulating such distributions eventually generate some very large firms, whose employment completely dominates the economy. The lognormal, as expected from Figure 4, generates too few large firms of too small a size. However, for most bins, the relative employment accounted for in each bin is comparable to the data (that is, the ratio of any two rows excluding the large firm sizes). Notably, by this metric, the convolution fits the data remarkably well. The left tail has a bit too much of total employment and the right tail has too little, but we show in Appendix B that in Monte Carlo simulations, the convolution is quite capable of reproducing the values we observe in the BDS. This table also highlights the greater flexibility of the convolution: Consider a Pareto distribution estimated on the right tail from Table 4 ($\alpha = 1.23$), that is estimated entirely *within* the largest bin of the table, with more than 10,000 employees. If this estimate is applied to the entire distribution, this bin accounts for too little of firm employment (10.9 percent), and the convolution demonstrates a relatively better fit across most bins, particularly in the far left tail as well.

## 4.3 Some Quantitative Implications in Granular Economies

In this section, we highlight quantitative implications for aggregate volatility in granular economies that sharply distinguish the distributions. This example illustrates a simple macroeconomic context in which distributional assumptions matter, and motivates the relevance of our estimates for macroeconomic models with cross-sectional firm heterogeneity.

Gabaix (2011) provides a useful framework to emphasize the potential of large firms to generate sizable aggregate shocks in a "granular" economy. He focuses on a Pareto distribution as it approaches Zipf's law, and shows that idiosyncratic shocks can be a more important source of aggregate volatility under Zipf's law, compared to a lognormal.

In this section, we quantify the contribution of idiosyncratic shocks to aggregate volatility for each estimated distribution; these visually similar distributions have vastly different aggregate volatility properties. Consider an economy with $N$ firms, each of which experience multiplicative shocks with fixed variance $\sigma^2$ to their size. Gabaix (2011) shows that aggregate volatility declines quickly with more firms (higher $N$) in many cases, but that as the firm size distribution approaches Zipf's law, aggregate volatility declines more slowly. Quantitatively, he calculates that with a firm volatility of $\sigma = 12\%$ and a Zipf distribution of firm sizes, GDP volatility is about 1.4%.

We replicate this quantitative exercise using our 1997 estimates for the firm size distribution; we simulated 100,000 samples taking $10^6$ draws from several distributions, including Zipf ($\alpha = 1.0$), Axtell's Pareto ($\alpha = 1.06$), our Pareto estimate ($\alpha = 0.61$) and our Pareto estimate of the right tail ($\alpha = 1.23$), along with our other parametric distributions.[20] We computed the GDP volatility for each simulation and we show the means in Table 10.

Table 10: Calibration of Aggregate Volatility under Granularity

| Pareto | | | | | | |
|---|---|---|---|---|---|---|
| $\alpha = 0.61$ | $\alpha = 1.0$ | $\alpha = 1.06$ | Pareto $\alpha = 1.23$ | Lognormal | Mixture | Convolution |
| 7.02% | 1.43% | 0.98% | 0.35% | 0.05% | 4.61% | 0.39% |

Note: We calculated $100,000$ times $10^6$ draws from each distribution, calculated the median of the Herfindahl index across the 100,000 draws, and computed the aggregate volatility using the standard deviation implied by equation (8) in Appendix E. Pareto with a shape parameter 0.61 is our maximum likelihood estimate. 1 is Zipf's law, 1.06 is Axtell's estimate, and 1.23 assumes the best fitting Pareto in the right tail as the entire distribution. Our estimates are parameterized using 1997 estimates for the firm size distribution from Tables 3, 6, and 7.

Gabaix compares his calibration with a U.S. aggregate volatility of 1.7% and makes the point

[20]Gabaix (2011) chose N=$10^6$ because it is in the same order of magnitude as the number of firms in the U.S. economy as reported by the LBD, Table 1.

that, with firms distributed according to Zipf's law or Axtell's Pareto, idiosyncratic shocks can explain a significant portion of aggregate volatility. Our estimates point at a different and more extreme picture: with a Pareto shape parameter below one, idiosyncratic shocks do not cancel out in the aggregate, producing too much aggregate volatility. In the model shown in Appendix E, we formally prove this unappealing economic implication of going *past* Zipf's law, with a shape parameter below one, and contrast it to the implications of a lognormal distribution.[21] When firms are lognormally distributed, the law of large numbers always kicks in and idiosyncratic shocks cancel out in the aggregate. Finally, the convolution—our preferred estimated distribution—allows for idiosyncratic shocks to matter in the aggregate, but with a much more diminished role compared to Zipf's law. Notably, the volatility generated by the convolution is very similar to a Pareto distribution with a shape parameter of 1.23, demonstrating that even with Pareto, moving plausibly away from Zipf dramatically reduces the quantitative importance of granularity. These results are in line with Stella (2015) and Yeh (2018), which also find a more limited contribution of firm-level shocks to aggregate volatility than Gabaix (2011).[22]

## 4.4 Results by Sector

So far, we have considered the firm and establishment distributions for the U.S. economy as a whole. Given the evidence on structural change in the U.S. economy and the significant differences among sectors, it is surprising that there is little evidence on the sectoral size distributions. In this section, we summarize a disaggregated approach by sector, leaving the detailed tables and discussion for Appendices C and D.

**Manufacturing vs. Services** We estimate each parametric distribution over time on two subsamples: manufacturing and services, where the latter excludes retail, wholesale, and FIRE. In both sectors and across time, formal statistical tests and the AIC provide a consistent ranking of distribution fit: the mixture fits the best, then convolution, then lognormal, and last Pareto. Therefore, the broad qualitative conclusions hold when considering, for example, only manufacturing establishments and firms. This is relevant for applications like international trade, in which goods are significantly more tradable than services.

Pareto estimates are fairly similar between manufacturing and services and also over time, robustly far below 1. Lognormal estimates provide a larger fitted mean and variance for manufacturing establishments and firms compared to their services' counterparts. Over time, the lognormal

[21] See Appendix E for model details.

[22] The values in Table 10 should be taken as upper bounds, as the calibration assumes that all firms have the same volatility. Declining volatility with firm size can significantly diminish the aggregate impact of idiosyncratic shocks. Yeh (2018) estimates the size-variance relationship in the LBD, and uses it to recalibrate the contribution of firm-level idiosyncratic shocks to aggregate fluctuations.

services parameters have trended upwards, but the same is not true for lognormal manufacturing parameters. Mixture estimates suggest that manufacturing establishments and firms have a higher share of lognormal than services, but only mixture estimates for establishments have plausible Pareto shape parameters above 1. Finally, the convolution tends to have larger lognormal and Pareto parameters for manufacturing establishments and firms than their services' counterparts. All Pareto shape parameters are well above 1.

**2-digit Sectors** We consider further disaggregated data by estimating each distribution on 2-digit SIC sectors for 1997. We report the mean and standard deviation of these statistics across all sectors, and we also explicitly report the largest 10 sectors in Appendix D. At this level of disaggregation, there is more heterogeneity across sectors, but the broad conclusions we find in the aggregate data hold: a lognormal almost always provides a better fit than Pareto, and convolution tends to outperform both. That is, our qualitative ranking is not driven by one or more idiosyncratic sectors.

## 4.5 Manufacturing revenue TFP and revenue

**Stylized Fact 3** *The distribution of establishment-level manufacturing revenue total factor productivity and revenue are also better described by lognormal than Pareto.*

Modern macroeconomic models prominently feature firm heterogeneity and therefore require assumptions on the distribution of productivity shocks. In particular, the analytical tractability of the Pareto distribution and its apparent good fit to the data have made it a common assumption. For instance, in their influential paper on gains from trade in new trade models, Arkolakis et al. (2012) assume Pareto. In standard monopolistic competition models, the firm size distribution reflects the productivity distribution, but given selection, demand functions, and potentially many sources of shocks, this is not always the case. Mrázová et al. (2021) characterize how the analytical properties of the demand function critically shape the implied distribution of sales and output using standard distributions for productivity such as lognormal or Pareto.[23] Therefore, in this section, we focus directly on estimates of the revenue and revenue TFP distributions available for the U.S. manufacturing sector. We use the TFPR measures as estimated for establishments by Foster et al. (2016) using the Census of Manufactures. This revenue TFP measure controls for only sectoral prices, and many heterogeneous firm models will imply substantial price heterogeneity within a

[23] Jones (2020) proposes an alternative growth model in which productivity draws from a Pareto distribution are not essential for generating exponential growth: combinatorially large samples from a thin-tailed distribution can yield exponential growth.

sector directly proportional to true productivity. This suggests caution in using these estimates as the literal parameters for the true TFP distribution in a model.[24]

Table 11: Manufacturing establishments

| | $x_m$ | $\alpha$ | $p$ | $\mu$ | $\sigma$ |
|---|---|---|---|---|---|
| | | *Pareto* | | | |
| TFP | | 0.12 | | | |
| Revenue | | 0.15 | | | |
| | | *Lognormal* | | | |
| TFP | | | | 1.93 | 0.54 |
| Revenue | | | | 6.88 | 1.99 |
| | | *Mixture* | | | |
| TFP | 19.65 | 1.83 | 0.98 | 1.90 | 0.50 |
| Revenue | 40.00 | 0.38 | 0.91 | 6.93 | 1.94 |
| | | *Convolution* | | | |
| TFP | | 3.57 | | 1.65 | 0.46 |
| Revenue | | 0.80 | | 5.64 | 1.58 |

Note: The scale parameter $x_m$ is not reported for the case of the simple Pareto due to Census disclosure restrictions. TFPR measures of TFP were constructed by Foster et al. (2016) using the Census of Manufactures.

Table 11 shows the results of the maximum likelihood estimation of the four parametric distributions we consider. Starting with the Pareto shape parameter fit in the top sub-panel, the estimate is around 0.15 for both revenue and revenue TFP, inheriting all of the economic issues inherent with an $\alpha$ below one. The second sub-panel provides the lognormal fit. The mixture estimates again show a very high proportion $p$ of lognormal. The lognormal parameters $\mu$ and $\sigma$ are similar to the estimates from the lognormal alone, while the shape parameter is well above 1 for TFPR and well below for revenues. Those lognormal parameters remain fairly similar in the convolution, with a high Pareto shape $\alpha$ for TFPR and a $\alpha$ below 1 for revenues.

Using the likelihood-based statistical tests on the TFPR distribution, we find that the mixture distribution generally outperforms lognormal, but, in 2002, the convolution outperforms both the mixture and lognormal, with the mixture still providing a better fit than the lognormal alone. The rankings are statistically significant typically at well beyond a 1 percent level. Our results are consistent with the evidence provided by Combes et al. (2012) and Nigai (2017) using French

[24] We present only estimates for 1997; estimates for the years 1982, 1992, and 2002 for the TFPR distribution are available upon request.

data. The statistical ranking of the distributions on the revenue sample is robustly in favor of convolution, ranking it as the best-fitting, followed by the mixture, the lognormal, and last the Pareto.

Taken together, the data strongly suggest that Pareto provides a poor fit and that lognormal is a more reasonable distribution for TFP than Pareto. Moreover, the parameter estimates for the lognormal distribution are remarkably similar between the standalone lognormal and its combinations with a Pareto. In contrast, such stability does not hold for the Pareto shape parameter estimates across the various estimated distributions. These findings suggest that given only one source of heterogeneity for TFP, a lognormal TFP distribution is more reasonable than Pareto. In addition, using Pareto-distributed demand shocks, this interpretation is consistent with our preferred distribution for the overall employment size distribution: a convolution of lognormal and Pareto.

# 5 Conclusion

In this paper, we use confidential microdata from the U.S. Census and maximum likelihood estimation to precisely characterize the U.S. firm and establishment size distributions, as measured by the number of employees and, for manufacturing, TFP and revenue. We think we provide the type of comprehensive empirical evidence that is missing in the literature, because of the nature of the data we use, the whole population of U.S. firms and establishments over a long period of time, and because of the rigorous statistical methods we employ. We establish three stylized facts about these distributions and provide guidance for researchers in parameterizing models that include firm heterogeneity.

The commonly used Pareto distribution does not provide a better fit for the size distribution of U.S. establishments or firms compared to the lognormal distribution, its less-popular alternative. In fact, the lognormal distribution even fits most truncations of the right tail better. For these reasons, we argue that if a single simple distribution is required, it is worth considering either the lognormal distribution or a calibration of Pareto that does not over-weight the right tail. That said, neither of these simple distributions is without serious trade-offs.

The firm and establishment size distributions are much better approximated with a combination of a lognormal distribution and a Pareto distribution. We find the convolution of a lognormal and a Pareto provides a much better fit along with a much-better match to the employment share distribution. While the mixture has the best statistical fit, its Pareto shape parameter below one implies that it inherits the unappealing characteristics of the Pareto distribution in that range. Thus, when evaluating distribution fit by the fraction of employment accounted for by firms of each size, the convolution provides a clearly superior fit to the mixture. As such, the convolution provides a more suitable choice economically for use in other applications. It can also be generated in a reasonable way as the product of a Pareto-distributed random variable and a lognormally distributed one; thus, models incorporating both productivity shocks and taste shocks could easily generate such a distribution.

Finally, we make use of manufacturing TFP estimates to consider the distribution of this alternative measure of firm heterogeneity. We find that the better fit of lognormal relative to Pareto is, if anything, even greater for TFP than for employment size. Given only one source of heterogeneity, lognormal TFP draws are more reasonable than Pareto. Adding a second source of heterogeneity with Pareto-distributed draws would further help fit the overall employment size distribution. That said, given the difficulties with measuring revenue TFP and their imperfect correspondence to a model, these results are not definitive about the fundamental sources of heterogeneity (productivity, taste, etc) and their distributions.

Aside from providing guidance for calibrating models with exogenously defined firm heterogeneity, our results also highlight that future models of endogenous growth should rationalize appreciable deviations from straight Pareto size distributions, including in the upper tail.

# References

**Adão, Rodrigo, Costas Arkolakis, and Sharat Ganapati**, "Measuring the Aggregate Implications of Firm Heterogeneity," 2020, p. 85.

**Arkolakis, Costas**, "A Unified Theory of Firm Selection and Growth," *The Quarterly Journal of Economics*, February 2016, *131* (1), 89–155.

____ **, Arnaud Costinot, and Andrés Rodríguez-Clare**, "New Trade Models, Same Old Gains?," *American Economic Review*, February 2012, *102* (1), 94–130.

**Armenter, Roc and Miklós Koren**, "Economies of Scale and the Size of Exporters," *Journal of the European Economic Association*, June 2015, *13* (3), 482–511.

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

**Buddana, Amrutha and Tomasz J. Kozubowski**, "Discrete Pareto Distributions," *Economic Quality Control*, 2015, *29* (2), 143–156.

**Clauset, Aaron, Cosma Rohilla Shalizi, and M. E. J. Newman**, "Power-law distributions in empirical data," *SIAM Review*, November 2009, *51* (4), 661–703. arXiv: 0706.1062.

**Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux**, "The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection," *Econometrica*, November 2012, *80* (6), 2543–2594.

**di Giovanni, Julian and Andrei A. Levchenko**, "Country Size, International Trade, and Aggregate Fluctuations in Granular Economies," *Journal of Political Economy*, December 2012, *120* (6), 1083–1132.

____ **and** ____ , "Firm entry, trade, and welfare in Zipf's world," *Journal of International Economics*, March 2013, *89* (2), 283–296.

____ **,** ____ **, and Romain Rancière**, "Power laws in firm size and openness to trade: Measurement and implications," *Journal of International Economics*, September 2011, *85* (1), 42–52.

**Durrett, Rick**, "Probability: Theory and Examples 5th Edition," 2017.

**Eeckhout, Jan**, "Gibrat's Law for (All) Cities: Reply," *American Economic Review*, August 2009, *99* (4), 1676–1683.

**Fernandes, Ana M., Peter J. Klenow, Sergii Meleshchuk, Martha Denisse Pierola, and Andrés Rodrıguez-Clare**, "The Intensive Margin in Trade: Moving Beyond Pareto," 2015.

**Foster, Lucia, Cheryl Grim, and John Haltiwanger**, "Reallocation in the Great Recession: Cleansing or Not?," *Journal of Labor Economics*, 2016, *34* (S1), S293–S331.

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

——, "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 2011, *79* (3), 733–772.

**Gaubert, Cecile and Oleg Itskhoki**, "Granular Comparative Advantage," *Journal of Political Economy*, June 2020, (Forthcoming).

**Head, Keith, Thierry Mayer, and Mathias Thoenig**, "Welfare and Trade without Pareto," *American Economic Review*, May 2014, *104* (5), 310–316.

**Jones, Charles I**, "Recipes and Economic Growth: A Combinatorial March Down an Exponential Tail," 2020.

**Luttmer, Erzo G. J.**, "Selection, Growth, and the Size Distribution of Firms," *The Quarterly Journal of Economics*, August 2007, *122* (3), 1103–1144.

——, "On the Mechanics of Firm Growth," *The Review of Economic Studies*, July 2011, *78* (3), 1042–1068.

**Mrázová, Monika, J Peter Neary, and Mathieu Parenti**, "Sales and markup dispersion: theory and empirics," *Econometrica*, 2021, *89* (4), 1753–1788.

**Nair, Jayakrishnan, Adam Wierman, and Bert Zwart**, "The Fundamentals of Heavy Tails," 2013.

**Nigai, Sergey**, "A tale of two tails: Productivity distribution and the gains from trade," *Journal of International Economics*, January 2017, *104*, 44–62.

**Quandt, Richard E.**, "On the Size Distribution of Firms," *The American Economic Review*, 1966, *56* (3), 416–432.

**Reed, William J**, "The Pareto, Zipf and other power laws," *Economics Letters*, December 2001, *74* (1), 15–19.

**Rossi-Hansberg, Esteban and Mark LJ Wright**, "Establishment size dynamics in the aggregate economy," *American Economic Review*, 2007, *97* (5), 1639.

**Sager, Erick and Olga A. Timoshenko**, "The EMG Distribution and Trade Elasticities," *Canadian Journal of Economics/Revue canadienne d'économique*, 2019, *52* (4), 1523–1557.

**Stella, Andrea**, "Firm dynamics and the origins of aggregate fluctuations," *Journal of Economic Dynamics and Control*, June 2015, *55*, 71–88.

**Vuong, Quang H.**, "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, March 1989, *57* (2), 307.

**Yeh, Chen**, "Revisiting the origins of business cycles with the size-variance relationship," July 2018.

# A Monte Carlo experiments

## A.1 Axtell's OLS Regression Analysis on Binned Data

To investigate the performance of Axtell's methodology to determine the fit of a Pareto distribution, we generated 250,000 synthetic datasets for each of the following parametric distributions: Pareto, lognormal, mixture and convolution with our parameter estimates obtained with 1997 Census data, and Pareto with Axtell's estimated parameter (1.06). We then implemented Axtell's methodology to estimate the Pareto shape parameter on each of the four sets of 250,000 synthetic datasets.

Axtell (2001) explains in "References and Notes" 30 how he implements the regression in Figure 1; the exact methodology followed by Axtell is not entirely clear to us. Our interpretation is as follows. First, we compute the number of firms (establishments) in each bin, with bins having width increasing in powers of three (the first bin is between 1 and 3, the second between 4 and 12, and so on). Second, we divide the number of firms (establishments) by the width of the bin and then by the number of firms (establishments) in the bin. Third, we take the geometric mean of the size of the firms within each bin. Finally, we run the regression of the logarithm with base 10 of the adjusted bin frequency on the logarithm with base 10 of the geometric mean of sizes within bin. The results are qualitatively robust to changes in the implementation of the tail coefficient estimation algorithm described in Axtell's original paper such as changing the bin widths or constructing the average size for a bin using firms elements within the bin rather than endpoints.

Table 12: Axtell's regression analysis on synthetic data

| | Axtell's Pareto | Pareto | Lognormal | Mixture | Convolution |
|---|---|---|---|---|---|
| True $\alpha$ | 1.06 | 0.61 | | 0.74 | 1.25 |
| $\hat{\alpha}$ | 1.01 | 0.59 | 1.36 | 0.78 | 1.10 |
| c.i. | (0.92, 1.08) | (0.54,0.62) | (1.28,1.49) | (0.70,0.83) | (1.03,1.16) |
| $N$ | 4,770,000 | 4,770,000 | 4,770,000 | 4,770,000 | 4,770,000 |
| Num. sim. | 250,000 | 250,000 | 250,000 | 250,000 | 250,000 |

Note: $\hat{\alpha}$ is the mean of the distribution of OLS coefficients for each synthetic dataset. c.i. is the 95% confidence interval. $N$ is the number of observations. Each simulation is run with 4,770,000 observations, which is the number of firms in 1997 according to the BDS.

Table 12 shows that Axtell's methodology is able to reasonably uncover the Pareto shape parameter when the data is drawn from a Pareto, with some downward bias. However, Axtell's methodology produces a shape parameter close to but above one even when data is drawn from a convolution with our estimated parameters. In other words, if the true firm size distribution were to be drawn from a convolution, Axtell's methodology would incorrectly find empirical support for Zipf's law.

## A.2 MLE simulations

We now investigate the performance of Maximum Likelihood Estimation in estimating the parameters of the distributions of interest. We drew 1 million observations from the Pareto, lognormal, mixture and convolution using our estimated 1997 coefficients as true parameters.[25] We discretized the data and then estimated the parameters using the same MLE procedure used in the paper. Finally, we computed the distance between the true parameters and the estimated parameters, and the pair-wise likelihood ratio tests among all the distributions. We repeated this exercise 250 times for each distribution.

Table 13: RMSEs

| | Pareto | Lognormal | Mixture | Convolution |
|---|---|---|---|---|
| $\mu$ | | *1.17* | *1.00* | *0.49* |
| RMSEs | | 0.00 | 0.34 | 0.00 |
| $\sigma$ | | *1.74* | *1.49* | *1.29* |
| RMSEs | | 0.00 | 0.14 | 0.00 |
| $p$ | | | *0.86* | |
| RMSEs | | | 0.08 | |
| $\alpha$ | *0.61* | | *3.47* | *1.25* |
| RMSEs | 0.06 | | 0.13 | 0.01 |
| $x_m$ | | | *0.74* | |
| RMSEs | | | 0.15 | |
| $N$ | 1,000,000 | 1,000,000 | 1,000,000 | 1,000,000 |
| Num. sim. | 250 | 250 | 250 | 250 |

Note: For each distribution, we show the 1997 estimated coefficients and the RMSEs from the simulation. $N$ is the number of observations in each simulation.

[25]We used a million observations because using the number of observations in the 1997 LBD was not computationally feasible for this exercise.

Table 14: Likelihood-based ratio tests

| *True:* | Pareto | Lognormal | Mixture | Convolution |
|---|---|---|---|---|
| *Alternative:* | | | | |
| Pareto | | 100.0% | 97.6% | 100.0% |
| lognormal | 96.8% | | 97.6% | 100.0% |
| Mixture | 33.2% | 0.0% | | 100.0% |
| Convolution | 70.0% | 1.6% | 97.2% | |
| *N* | 1,000,000 | 1,000,000 | 1,000,000 | 1,000,000 |
| Num. sim. | 250 | 250 | 250 | 250 |

Note: For each distribution, we show the percentage of times that the likelihood-ratio test correctly picks the true distribution. The LRT test is used between mixture and Pareto and between mixture and lognormal. The Vuong test is used for all other pairs. $N$ is the number of observations in each simulation.

Table 13 shows the Root Mean Squared Errors (RMSEs) computed using the distances between the true parameters and their MLE estimates in the 250 simulations; the errors made by MLE are for the most part very small. Table 14 presents the percentage of times that the likelihood-based ratio tests with 95% confidence were able to pick the correct distribution, with the true distribution shown in the column header and the alternative distribution shown in the first column. The tests are able to almost always pick the correct distribution when the true distributions are mixture and convolution, and when testing between lognormal and Pareto. The tests have a harder time when the alternative distribution is a more flexible version of the true distribution, as the former has more coefficients and is able to approximate the latter. For instance, if the true distribution is Pareto and the alternative is mixture, the mixture can perfectly approximate a Pareto by having its mixture parameter $p$ equal to zero, which makes it very hard for the likelihood-ratio test to distinguish between them.

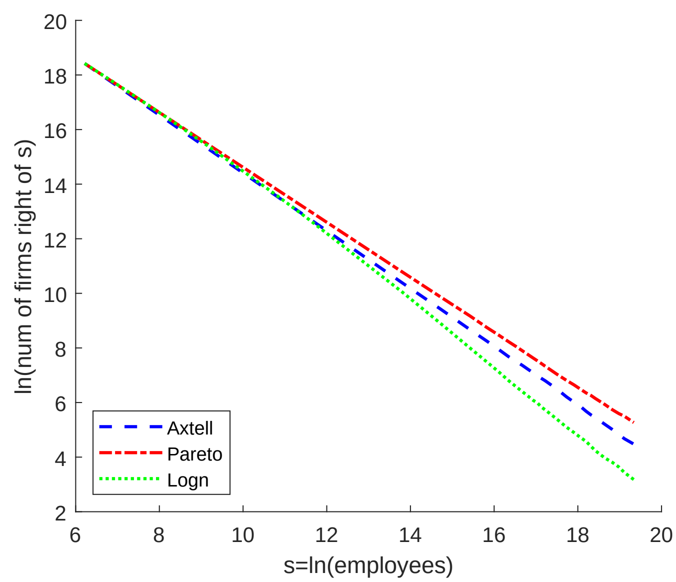
To help parse Table 14, consider a specific example. The first column shows the results of the tests when the sample is drawn from a Pareto distribution; the likelihood-based ratio tests correctly pick Pareto over lognormal 96.8% of the time; as explained earlier. The tests are not as good in distinguishing between Pareto and the more flexible mixture and convolution: the tests correctly pick Pareto only 33.2% of the samples when a mixture distribution is used as the alternative, and 70% when a convolution is the alternative. However, when the sample is drawn from Pareto, mixture and convolution are estimated to be very similar to Pareto, whereas the parameters of mixture and convolution estimated with Census data are not such that the two distributions look like either a straight Pareto or a straight lognormal. For this reason, it is unlikely that the weakness of likelihood-based tests in distinguishing nested or quasi-nested models applies to our empirical results.

# B Additional Empirical Results

## B.1 Visualizing the Estimated Upper Tails

We extende Figure 3 by increasing the number of observations drawn from the truncated parametric distributions so as to have enough observations in the largest bins. Figure 5 shows that the Pareto distributions produce a larger share of far more extremely large firms than the estimated truncated lognormal.

Figure 5: Estimated Distributions Above 500 Employees




## B.2 Confidence Intervals for Implied Employment Shares

In Table 15, we show the 95% confidence intervals for Table 9, obtained by drawing 4.77 million firms (as in the 1997 LBD) 100,000 times from each each distribution with the parameter values we estimate.

Here we see that for the Axtell calibration ($\alpha = 1.06$), the fraction of firms accounted for by the largest and smallest bins has a large variance, though the confidence intervals include the true data values. The lognormal distribution is much more consistently simulated across draws, with its thinner right tail. Our Pareto estimate with a shape parameter below one and the mixture, also incorporating a Pareto estimate with a shape parameter below one, consistently draws massive firms which account for nearly all of employment. Using our Pareto estimate from the right tail ($\alpha = 1.23$), the largest bin has a sizable variance, though it does encompass the BDS value; but this calibration also implies far too many small firms, well above the BDS, and this makes this

Pareto struggle to encompass the BDS value in the confidence bands for most bins.

Finally, the confidence bands for the convolution are fairly economically narrow but often include the true value, including both the small firm bin and the largest firm bin.

Table 15: Fraction of 1997 firm employment: 95% confidence intervals

| | | Pareto | | | | | |
|---|---|---|---|---|---|---|---|
| | BDS | $\alpha = 0.61$ | $\alpha = 1.06$ | $\alpha = 1.23$ | Lognormal | Mixture | Convolution |
| 1 to 4 | 5.65 | (0,0.01) | (5.23,13.94) | (22.09,28.97) | (7.01,7.07) | (0.01,1.1) | (5.37,6.97) |
| 5 to 9 | 6.53 | (0,0.01) | (2.46,6.56) | (8.52,11.17) | (7.49,7.56) | (0.02,1.27) | (5.84,7.58) |
| 10 to 19 | 7.73 | (0,0.01) | (2.44,6.52) | (7.52,9.87) | (11.01,11.11) | (0.02,1.59) | (7.58,9.83) |
| 20 to 49 | 10.62 | (0,0.01) | (3.14,8.38) | (8.43,11.07) | (18.98,19.15) | (0.03,2.24) | (11.09,14.39) |
| 50 to 99 | 7.52 | (0,0.01) | (2.29,6.1) | (5.35,7.03) | (15.48,15.65) | (0.02,1.51) | (8.02,10.4) |
| 100 to 249 | 8.72 | (0,0.03) | (2.9,7.73) | (5.9,7.77) | (17.88,18.12) | (0.02,1.62) | (9.18,11.92) |
| 250 to 499 | 5.54 | (0,0.03) | (2.09,5.59) | (3.71,4.94) | (9.63,9.89) | (0.01,1.05) | (5.76,7.5) |
| 500 to 999 | 5.09 | (0,0.04) | (2.01,5.39) | (3.16,4.27) | (6.04,6.32) | (0.01,1.05) | (4.86,6.35) |
| 1,000 to 2,499 | 7.07 | (0,0.07) | (2.53,6.82) | (3.45,4.78) | (3.83,4.17) | (0.02,1.58) | (5.26,6.9) |
| 2,500 to 4,999 | 5.40 | (0,0.07) | (1.82,4.99) | (2.11,3.17) | (1.03,1.32) | (0.02,1.44) | (3.23,4.36) |
| 5,000 to 9,999 | 5.46 | (0,0.09) | (1.75,4.89) | (1.68,2.88) | (0.32,0.57) | (0.02,1.72) | (2.7,3.76) |
| $\geq$ 10,000 | 24.68 | (99.63,100) | (23.54,71.3) | (5.38,27.74) | (0.07,0.35) | (83.83,99.8) | (10.6,31.1) |

## B.3 Test Statistics for Best Fitting Distribution

Table 16 presents the values of the likelihood-based test statistics used to determine which parametric distribution fits the data best in Section 4.2 for the 1997 firm distribution. Tables 17-20 show the AIC statistics.

Table 16: Likelihood-based tests for 1997 firm distribution

| $f/g$: | Convolution | Lognormal | Pareto |
|---|---|---|---|
| Mixture | 45.5 | 76,445.5 | 425,545.2 |
| | 0.00 | 0.00 | 0.00 |
| Convolution | | 181.6 | 493.8 |
| | | 0.00 | 0.00 |
| Lognormal | | | 495.8 |
| | | | 0.00 |

Notes: The row distributions are the models in the numerator of the ratio, $f$; the column distributions are the models in the denominator of the ratio, $g$. The first row for each pair shows either the LR test for nested models or the Vuong test for non-nested models; the second row shows the p-value.

Table 17: AIC statistics for establishment distribution

| | Lognormal | Pareto | Mixture | Convolution |
|---|---|---|---|---|
| 1982 | 28,898,876 | 29,918,494 | 28,842,314 | 28,851,956 |
| 1992 | 36,338,039 | 37,651,078 | 36,278,838 | 36,300,936 |
| 1997 | 39,785,313 | 41,113,303 | 39,736,610 | 39,756,014 |
| 2002 | 41,689,173 | 43,103,748 | 41,656,152 | 41,661,907 |
| 2012 | 43,192,318 | 44,464,030 | 43,161,973 | 43,177,085 |

Table 18: AIC statistics for firm distribution

| | Lognormal | Pareto | Mixture | Convolution |
|---|---|---|---|---|
| 1982 | 22,177,482 | 22,688,365 | 21,956,870 | 21,966,470 |
| 1992 | 27,019,369 | 27,727,694 | 26,867,381 | 26,883,931 |
| 1997 | 29,436,483 | 30,134,680 | 29,283,598 | 29,298,494 |
| 2002 | 30,496,847 | 31,182,538 | 30,283,990 | 30,288,029 |
| 2012 | 30,403,594 | 30,959,161 | 30,247,239 | 30,254,282 |

Table 19: AIC statistics for 1997 estab. distribution in the tail

| Threshold | Pareto | Lognormal |
|---|---|---|
| 2 | 24,626,695 | 24,393,693 |
| 5 | 15,832,300 | 15,816,451 |
| 10 | 9,727,014 | 9,725,856 |
| 25 | 4,531,605 | 4,532,775 |
| 50 | 2,380,765 | 2,381,625 |
| 100 | 1,220,572 | 1,220,996 |
| 200 | 619,642 | 619,719 |
| 300 | 423,432 | 423,443 |
| 400 | 324,446 | 324,431 |
| 500 | 265,706 | 265,688 |
| 1000 | 144,636 | 144,610 |
| 2500 | 63,908 | 63,879 |
| 5000 | 33,939 | 33,909 |
| 10000 | 17,752 | 17,734 |

Table 20: AIC statistics for 1997 firm distribution in the tail

| Threshold | Pareto | Lognormal |
|---|---|---|
| 2 | 34,535,926 | 34,024,886 |
| 5 | 23,484,232 | 23,416,630 |
| 10 | 14,979,496 | 14,958,621 |
| 25 | 7,263,395 | 7,255,654 |
| 50 | 3,850,413 | 3,847,002 |
| 100 | 1,902,059 | 1,901,481 |
| 200 | 827,330 | 827,160 |
| 300 | 494,318 | 494,188 |
| 400 | 337,863 | 337,696 |
| 500 | 252,919 | 252,778 |
| 1000 | 98,486 | 98,417 |
| 2500 | 23,327 | 23,310 |
| 5000 | 6,273 | 6,261 |
| 10000 | 1,333 | 1,321 |

## C Results by Manufacturing vs. Services

The employment size distributions in Section 4 were estimated using the entire U.S. population of firms and establishments. Over the past few decades, the U.S. economy has been going through a deep structural transformation that shifted employment away from manufacturing and towards the services sectors. In this appendix, we examine manufacturing and services separately. We find that the manufacturing and services sectors have notably different distribution estimates, but the distribution fit ranking is the same as in the aggregate data.

Table 21: Manufacturing vs Services

| | Fraction of employment | | Average firm size | |
|---|---|---|---|---|
| Year | Manufacturing | Services | Manufacturing | Services |
| 1977 | 29.2% | 22.9% | 74.3 | 13.5 |
| 1982 | 25.6% | 25.9% | 70.3 | 15.0 |
| 1987 | 21.7% | 29.9% | 64.2 | 16.1 |
| 1992 | 19.7% | 33.0% | 61.2 | 17.4 |
| 1997 | 17.9% | 35.1% | 61.3 | 18.8 |
| 2002 | 14.5% | 37.9% | 57.6 | 20.7 |
| 2007 | 12.3% | 40.1% | 55.1 | 20.5 |
| 2012 | 10.6% | 42.9% | 51.8 | 20.9 |

Source: BDS and authors' calculations. The second and third columns show the fraction of total employment in manufacturing and services, respectively; the fourth and fifth columns show the average employment size of manufacturing and services firms, respectively. The definition of manufacturing and services is in Section 2.

As shown in Table 21, the employment weight of manufacturing declined substantially over the period considered in this paper, whereas that of services almost doubled. Moreover, Table 21 shows substantial compositional differences between these two sectors, as the average firm size has been roughly three and a half times as large in manufacturing than in services, although the gap has been closing over time.

**Pareto and Lognormal Estimates** Table 22 presents the estimates for the lognormal distribution by sector. Focusing on the average line, manufacturing establishments have both a larger $\mu$ and $\sigma$ than services establishments, implying both a greater fitted mean and variance. A similar pattern holds for manufacturing firms relative to services firms.

Table 23 presents the Pareto shape parameter $\alpha$ by sector. Notably, manufacturing establishments and firms have an $\alpha$ even lower than for all firms, averaging about 0.4 compared to services' 0.6. These results suggest that in models focusing on manufacturing firms, such as standard models in international trade, the fit of a Pareto distribution on the overall empirical distribution is not more sensible than a lognormal.

Over time, both the mean and the variance of the estimated lognormal appear to be increasing for firms and establishments in the services sector, but not in the manufacturing sector.

Table 22: Lognormal by sector

| | Manufacturing | | Services | |
|---|---|---|---|---|
| Year | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ |
| | | *Establishments* | | |
| 1982 | 2.42 | 1.75 | 1.09 | 1.50 |
| 1992 | 2.32 | 1.76 | 1.21 | 1.53 |
| 1997 | 2.34 | 1.77 | 1.23 | 1.58 |
| 2002 | 2.21 | 1.75 | 1.43 | 1.60 |
| 2012 | 2.13 | 1.77 | 1.40 | 1.64 |
| Average | 2.28 | 1.76 | 1.27 | 1.57 |
| | | *Firms* | | |
| 1982 | 2.13 | 1.65 | 1.01 | 1.48 |
| 1992 | 2.05 | 1.69 | 1.11 | 1.54 |
| 1997 | 2.08 | 1.71 | 1.09 | 1.60 |
| 2002 | 1.98 | 1.68 | 1.25 | 1.60 |
| 2012 | 1.90 | 1.72 | 1.20 | 1.63 |
| Average | 2.03 | 1.69 | 1.13 | 1.57 |

Table 23: Pareto $\alpha$ by sector

| | *Establishments* | | *Firms* | |
|---|---|---|---|---|
| Year | Manufacturing | Services | Manufacturing | Services |
| 1982 | 0.38 | 0.65 | 0.42 | 0.67 |
| 1992 | 0.39 | 0.61 | 0.43 | 0.64 |
| 1997 | 0.39 | 0.60 | 0.43 | 0.64 |
| 2002 | 0.41 | 0.55 | 0.44 | 0.60 |
| 2012 | 0.42 | 0.55 | 0.45 | 0.60 |
| Average | 0.40 | 0.59 | 0.44 | 0.63 |

**Mixture Estimates** Table 24 reports the estimates for the mixture distribution by sector. The lognormal parameters remain remarkably similar to those from estimating lognormal alone, but there are notable differences in the Pareto shape parameter.

Table 24: Mixture by sector

| | Manufacturing | | | | | Services | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Year | $\mu$ | $\sigma$ | $p$ | $x_m$ | $\alpha$ | $\mu$ | $\sigma$ | $p$ | $x_m$ | $\alpha$ |
| | | | | | *Establishments* | | | | | |
| 1992 | 2.32 | 1.79 | 0.94 | 3.66 | 0.88 | 0.89 | 1.43 | 0.79 | 3.50 | 0.78 |
| 1997 | 2.34 | 1.80 | 0.96 | 4.49 | 1.12 | 0.96 | 1.51 | 0.81 | 3.51 | 0.77 |
| 2002 | 2.22 | 1.78 | 0.95 | 3.50 | 1.01 | 1.34 | 1.59 | 0.92 | 3.56 | 0.76 |
| 2012 | 2.13 | 1.79 | 0.98 | 4.48 | 1.96 | 1.29 | 1.63 | 0.92 | 4.48 | 0.81 |
| Average | 2.25 | 1.79 | 0.96 | 4.03 | 1.24 | 1.12 | 1.54 | 0.86 | 3.76 | 0.78 |
| | | | | | *Firms* | | | | | |
| 1992 | 1.90 | 1.63 | 0.88 | 4.48 | 0.59 | 0.74 | 1.35 | 0.78 | 3.49 | 0.78 |
| 1997 | 1.94 | 1.66 | 0.89 | 4.52 | 0.58 | 0.77 | 1.41 | 0.80 | 3.50 | 0.76 |
| 2002 | 1.85 | 1.62 | 0.90 | 4.53 | 0.57 | 1.06 | 1.48 | 0.87 | 3.48 | 0.71 |
| 2012 | 1.76 | 1.66 | 0.90 | 4.41 | 0.58 | 1.00 | 1.52 | 0.89 | 4.36 | 0.71 |
| Average | 1.86 | 1.64 | 0.89 | 4.49 | 0.58 | 0.89 | 1.44 | 0.83 | 3.71 | 0.74 |

Note: 1982 not reported as estimates have unusually large standard errors.

Starting with manufacturing establishments, the Pareto scale parameter, $x_m$, is estimated to take effect around 4 employees and the Pareto shape parameter, $\alpha$, is estimated to be a reasonable 1.24, but the mixing parameter $p = 0.96$ suggests that the manufacturing establishment size distribution is almost entirely lognormal. By contrast, the establishment size distribution in the services sector is somewhat less lognormal with $p = 0.86$ on average, but the corresponding Pareto shape parameter remains robustly below one. Manufacturing firms have a lower mix of lognormal ($p = 0.89$) but, unlike manufacturing establishments, feature a low Pareto shape parameter below one. The contrast seems less pronounced in the services sector: estimates for the size distribution of firms and establishments are quite similar.

The estimated parameters are relatively stable ovetime. The estimated distributions appear to be getting more lognormal over time, especially in the services sector. The estimated Pareto shape parameters also appear to be relatively stable and below one, except for manufacturing establishments.

**Convolution Estimates** Finally, Table 25 shows the sectoral parameter estimates for the convolution distribution. As with the previous distributions, manufacturing establishments and firms have much higher estimates of mean $\mu$ and standard deviation $\sigma$ than their services counterparts.

The Pareto shape parameter is well above 1 for all subsamples, which further suggests that the convolution yields reasonable parameter estimates and is a solid candidate for use in calibrated models.

Over time, for establishments in both sectors, there is a steady rise in the shape parameter $\alpha$ of the Pareto component. Consequently, there is a growing deviation from Zipf's law for establishments, especially in manufacturing, where the shape parameter is estimated to be above 2 since 1997. In contrast, the shape parameter of the convolution's Pareto component is much more stable and closer to 1 for firms in the manufacturing sector. The firm size distribution in services, however, features a slight upward trend in the Pareto shape parameter estimate.

These results are largely consistent with the findings in the aggregate sample. These findings also suggest that the manufacturing sector, unlike services, has experienced different dynamics at the firm level, relative to the establishments that comprise these firms.

Table 25: Convolution by sector

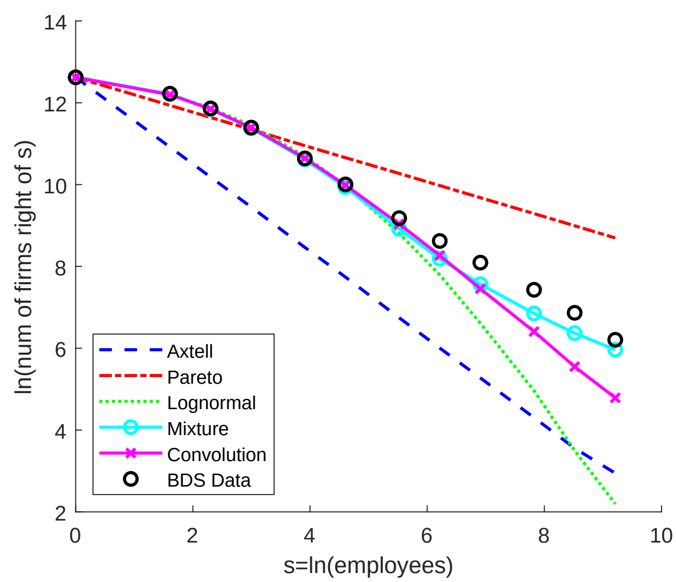
| | Manufacturing | | | Services | | |
|---|---|---|---|---|---|---|
| | $\mu$ | $\sigma$ | $\alpha$ | $\mu$ | $\sigma$ | $\alpha$ |
| | | | *Establishments* | | | |
| 1982 | 1.75 | 1.61 | 1.49 | 0.27 | 1.16 | 1.17 |
| 1992 | 1.75 | 1.66 | 1.75 | 0.40 | 1.23 | 1.20 |
| 1997 | 1.91 | 1.72 | 2.31 | 0.43 | 1.30 | 1.22 |
| 2002 | 1.73 | 1.68 | 2.08 | 0.80 | 1.44 | 1.55 |
| 2012 | 1.78 | 1.73 | 2.89 | 0.84 | 1.52 | 1.76 |
| Average | 1.78 | 1.68 | 2.10 | 0.55 | 1.33 | 1.38 |
| | | | *Firms* | | | |
| 1982 | 1.26 | 1.37 | 1.15 | 0.19 | 1.12 | 1.16 |
| 1992 | 1.18 | 1.41 | 1.14 | 0.26 | 1.18 | 1.15 |
| 1997 | 1.22 | 1.45 | 1.16 | 0.26 | 1.23 | 1.16 |
| 2002 | 1.15 | 1.42 | 1.19 | 0.46 | 1.29 | 1.23 |
| 2012 | 1.04 | 1.45 | 1.15 | 0.43 | 1.37 | 1.27 |
| Average | 1.17 | 1.42 | 1.16 | 0.32 | 1.24 | 1.19 |

**Visual Representations** Figure 6 shows a graphical representation of the fit of the parametric distributions of manufacturing with 1997 BDS data. Like the previous figures, it depicts the complementary cumulative distribution function in log space. Again, distributions above the black circles represent too many large firms compared to the data. Here we note that the convolution and mixture continue to fit the overall shape of the CCDF well, while our Pareto estimate generates far too many large firms. The lognormal distribution fits the left tail and middle of the distribution well, but generates too few large firms. We also plot Axtell's Pareto estimate of 1.06 for comparison, and it has a similar fit to the manufacturing distribution as it does for the overall distribution.

In Figure 7, we show the same CCDF plot for services with 1997 BDS data. Relative to manufacturing, the mixture provides an even better fit across all bins, while lognormal struggles more with the right tail. Our Pareto estimate remains systematically above the data, while Axtell's is below.

Finally, we formally test which distribution fits the data best by sector-year for establishments and for firms. For both establishments and firms, in both manufacturing and services, and across all the years 1982, 1992, 1997, 2002, and 2012, the formal statistical tests and the AIC all provide a consistent ranking of distribution fit: the mixture fits the best, then convolution, then lognormal, and last Pareto.[26] The rankings are statistically significant at least at a 5 percent level and typically at a much tighter level.[27]

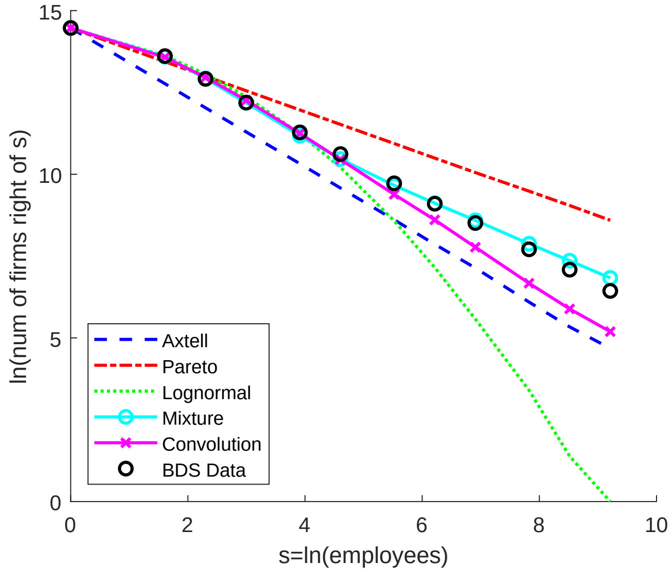
Figure 6: BDS Data vs. Estimated Distributions for Manufacturing




[26]Detailed results available upon request.

[27]Our results are not consistent with the findings by Quandt (1966), who estimates the firm size distribution in several sectors separately. His ranking of distribution fit varies from sector to sector. However, his sample only included very large firms, and he did not consider the lognormal distribution or mixtures of lognormal and Pareto.

Figure 7: BDS Data vs. Estimated Distributions for Services



# D Results by 2-digit Sector

In this appendix, we examine sectoral heterogeneity further by estimating distributions on 2-digit SIC sectors. While there is some interesting heterogeneity across sectors, our broad conclusions hold sector-by-sector: lognormal almost always provides a better fit than Pareto, and convolution tends to outperform both. That is, our conclusions are not driven by one or more idiosyncratic sectors.

To get a better sense of the heterogeneity across industries, we report all estimates for the largest 10 2-digit SIC sectors by employment in Tables 26 through 29 below. Concerns about disclosing too many statistics and for some smaller industries led us to focus on the largest 10. They include a nice mix of service and retail-oriented sectors. In the last two row of each table, we report the results for manufacturing as a whole and for all establishments/firms, both repeated from the other tables in the paper.

Table 26: Pareto and lognormal estimates for top 10 2-digit sectors, 1997

| | | Pareto | | Lognormal | | | |
|---|---|---|---|---|---|---|---|
| | | $\alpha$ | | $\mu$ | | $\sigma$ | |
| Sector | | Est. | Firm | Est. | Firm | Est. | Firm |
| 17 | Construction (Special Trade) | 0.64 | 0.64 | 1.14 | 1.13 | 1.42 | 1.41 |
| 50 | Wholesale Trade (Durable) | 0.55 | 0.57 | 1.49 | 1.39 | 1.41 | 1.49 |
| 51 | Wholesale Trade (Non-durable) | 0.54 | 0.58 | 1.48 | 1.3 | 1.55 | 1.59 |
| 53 | General Merchandise Stores | 0.33 | 0.63 | 2.81 | 1.08 | 1.93 | 1.78 |
| 54 | Food Stores | 0.48 | 0.58 | 1.81 | 1.35 | 1.56 | 1.53 |
| 58 | Eating and Drinking Places | 0.43 | 0.47 | 2.13 | 1.87 | 1.34 | 1.49 |
| 59 | Miscellaneous Retail | 0.61 | 0.69 | 1.29 | 1.09 | 1.26 | 1.5 |
| 73 | Business Services | 0.55 | 0.62 | 1.31 | 1.06 | 1.86 | 1.79 |
| 80 | Health Services | 0.55 | 0.59 | 1.46 | 1.29 | 1.55 | 1.55 |
| 87 | Services* | 0.67 | 0.71 | 0.96 | 0.83 | 1.63 | 1.59 |
| | Manufacturing | 0.39 | 0.43 | 2.34 | 2.08 | 1.77 | 1.71 |
| | All establishments/firms | 0.56 | 0.62 | 1.37 | 1.14 | 1.61 | 1.80 |

* This services category includes engineering, accounting, research, management, and related services.

In Table 26, the Pareto shape parameter is robustly below 1 and similar to our aggregate results in Table 3 of the main paper. Likewise, the lognormal estimates are similar to the aggregates, suggesting that outlier industries are not driving our main results.

Table 27: Mixture estimates for top 10 2-digit sectors, 1997

| | | $\mu$ | | $\sigma$ | | $p$ | | $x_m$ | | $\alpha$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Est. | Firm | Est. | Firm | Est. | Firm | Est. | Firm | Est. | Firm |
| 17 | Construction (Special Trade) | 1.07 | 1.04 | 1.45 | 1.43 | 0.93 | 0.92 | 3.58 | 3.59 | 1.21 | 1.11 |
| 50 | Wholesale Trade (Durable) | 1.79 | 1.66 | 1.20 | 1.27 | 0.77 | 0.74 | 1.00 | 1.00 | 0.85 | 0.78 |
| 51 | Wholesale Trade (Non-durable) | 1.42 | 1.21 | 1.57 | 1.56 | 0.94 | 0.93 | 4.47 | 3.57 | 1.04 | 0.73 |
| 53 | General Merchandise Stores | 5.05 | 1.24 | 0.51 | 1.06 | 0.27 | 0.63 | 1.71 | 1.00 | 0.58 | 0.62 |
| 54 | Food Stores | 1.80 | 1.22 | 1.66 | 1.35 | 0.89 | 0.90 | 4.62 | 4.58 | 2.20 | 0.72 |
| 58 | Eating and Drinking Places | 2.62 | 2.13 | 0.99 | 1.18 | 0.75 | 0.76 | 1.49 | 1.20 | 1.22 | 0.75 |
| 59 | Miscellaneous Retail | 1.14 | 1.45 | 1.29 | 0.88 | 0.86 | 0.54 | 4.48 | 1.16 | 1.32 | 0.96 |
| 73 | Business Services | 1.24 | 0.69 | 2.02 | 1.87 | 0.86 | 0.74 | 2.43 | 2.33 | 1.07 | 0.75 |
| 80 | Health Services | 1.69 | 1.60 | 0.60 | 0.62 | 0.33 | 0.38 | 1.32 | 1.25 | 0.65 | 0.70 |
| 87 | Services* | 0.92 | 1.39 | 1.54 | 1.11 | 0.99 | 0.33 | 77.43 | 1.15 | 0.98 | 0.87 |
| | Manufacturing | 2.34 | 1.94 | 1.80 | 1.66 | 0.96 | 0.89 | 4.49 | 4.52 | 1.12 | 0.58 |
| | All establishments/firms | 1.25 | 1.00 | 1.58 | 1.49 | 0.86 | 0.86 | 3.57 | 3.47 | 0.85 | 0.74 |

* This services category includes engineering, accounting, research, management, and related services.

Table 27 provides estimates of the mixture distribution for each of the 10 largest sectors. Most sectors have high estimates of $p$, the share of the distribution that is lognormal. The Pareto shape parameter $\alpha$ has estimates above and below 1 for establishments. The Pareto shape estimates for firms are more consistent and similar to the overall estimate of 0.74.

Table 28: Regression-based Pareto estimates for 2-digit NAICS sectors with BDS data, 1997

| | | Establishments | | | | Firms | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\alpha$ | s.e. | $N$ | $R^2$ | $\alpha$ | s.e. | $N$ | $R^2$ |
| 11 | Agriculture, forestry, fishing and hunting | 1.87 | 0.37 | 7 | 0.92 | 2.40 | 0.50 | 10 | 0.85 |
| 21 | Mining, quarrying, and oil and gas extraction | 1.47 | 0.40 | 7 | 0.88 | 1.45 | 0.23 | 10 | 0.93 |
| 22 | Utilities | 1.19 | 0.39 | 7 | 0.86 | 0.88 | 0.19 | 10 | 0.92 |
| 23 | Construction | 1.87 | 0.34 | 7 | 0.93 | 1.92 | 0.28 | 10 | 0.93 |
| 31-33 | Manufacturing | 1.22 | 0.39 | 7 | 0.87 | 1.09 | 0.15 | 10 | 0.96 |
| 42 | Wholesale trade | 1.80 | 0.41 | 7 | 0.90 | 1.75 | 0.20 | 10 | 0.96 |
| 44-45 | Retail trade | 2.09 | 0.49 | 7 | 0.89 | 1.33 | 0.06 | 10 | 1.00 |
| 48-49 | Transportation and warehousing | 1.38 | 0.28 | 7 | 0.94 | 1.47 | 0.13 | 10 | 0.98 |
| 51 | Information | 1.37 | 0.39 | 7 | 0.88 | 1.20 | 0.12 | 10 | 0.98 |
| 52 | Finance and insurance | 1.52 | 0.33 | 7 | 0.92 | 1.18 | 0.10 | 10 | 0.98 |
| 53 | Real estate and rental and leasing | 1.97 | 0.41 | 7 | 0.91 | 2.09 | 0.26 | 10 | 0.95 |
| 54 | Professional, scientific, and technical services | 1.63 | 0.32 | 7 | 0.93 | 1.71 | 0.18 | 10 | 0.97 |
| 55 | Management of companies and enterprises | 1.10 | 0.38 | 7 | 0.86 | 0.87 | 0.16 | 10 | 0.94 |
| 56 | Admin, support, waste, and remediation | 1.44 | 0.38 | 7 | 0.89 | 1.33 | 0.16 | 10 | 0.97 |
| 61 | Educational services | 1.22 | 0.29 | 7 | 0.92 | 1.13 | 0.21 | 10 | 0.93 |
| 62 | Health care and social assistance | 1.43 | 0.32 | 7 | 0.92 | 1.17 | 0.14 | 10 | 0.97 |
| 71 | Arts, entertainment, and recreation | 1.52 | 0.34 | 7 | 0.92 | 1.52 | 0.25 | 10 | 0.98 |
| 72 | Accommodation and food services | 1.82 | 0.40 | 7 | 0.91 | 1.30 | 0.11 | 10 | 0.93 |
| 81 | Other services (except public) | 1.99 | 0.30 | 7 | 0.95 | 2.09 | 0.25 | 10 | 0.95 |
| | All establishments/firms | 1.54 | 0.34 | 7 | 0.92 | 1.17 | 0.09 | 10 | 0.99 |

Pareto shape estimates obtained with an Axtell-style regression using BDS binned data. Bin values in the regression are the total number of employees in the bin divided by the number of firms or establishments in that bin.

Table 28 provides estimates from an Axtell-style regression using BDS binned data. The BDS reports sectors on a NAICS basis (rather than SIC), so the sectors are not directly comparable to the sectoral estimates obtained via MLE using the microdata in Table 27. The shape parameter $\alpha$ tends to be large for establishments in most sectors, and all sectors together yield an estimate of 1.54, well above Zipf. That said, with a sample size $N = 7$, standard errors are large. For firms, estimates vary from as low as 0.87 (management) to 2.40 (agriculture, forestry, fishing, and hunting). For all sectors, the estimate of 1.17 is notably higher than the traditional Axtell estimate. We see this as evidence that the choice of bins matters, and there is no clear rationale for one binning scheme over another. In addition, these estimates suffer from the statistical problems inherent in all such regressions, and should be taken with caution.

Table 29: Convolution estimates for top 10 2-digit sectors, 1997

| | | $\mu$ | | $\sigma$ | | $\alpha$ | |
|---|---|---|---|---|---|---|---|
| | | Est. | Firm | Est. | Firm | Est. | Firm |
| 17 | Construction (Special Trade) | 0.84 | 0.73 | 1.38 | 1.35 | 3.28 | 2.51 |
| 50 | Wholesale Trade (Durable) | 1.45 | 0.85 | 1.41 | 1.37 | 25.93 | 1.84 |
| 51 | Wholesale Trade (Non-durable) | 1.06 | 0.66 | 1.49 | 1.42 | 2.39 | 1.52 |
| 53 | General Merchandise Stores | 2.76 | 0.17 | 1.93 | 1.04 | 18.78 | 0.98 |
| 54 | Food Stores | 1.20 | 0.64 | 1.42 | 1.21 | 1.63 | 1.35 |
| 58 | Eating and Drinking Places | 2.09 | 1.38 | 1.34 | 1.33 | 26.94 | 2.10 |
| 59 | Miscellaneous Retail | 1.25 | 0.49 | 1.26 | 1.12 | 29.12 | 1.81 |
| 73 | Business Services | 0.51 | 0.10 | 1.64 | 1.38 | 1.22 | 0.99 |
| 80 | Health Services | 0.47 | 0.38 | 1.07 | 1.03 | 0.98 | 1.05 |
| 87 | Services* | 0.20 | 0.04 | 1.37 | 1.28 | 1.27 | 1.20 |
| | Manufacturing | 1.91 | 1.22 | 1.72 | 1.45 | 2.31 | 1.16 |
| | All establishments/firms | 0.72 | 0.43 | 1.37 | 1.29 | 1.43 | 1.25 |

* This services category includes engineering, accounting, research, management, and related services.

Table 29 provides estimates for the convolution for the same 10 sectors in the microdata. Here there is more heterogeneity at the establishment level: some sectors, especially retail sectors, have large $\alpha$ Pareto shape parameters, far from Zipf, and suggesting thinner right tails (the lognormal $\sigma$ parameters also tend to be smaller for these retail sectors). One standout is health services, which appears near-Zipf in the $\alpha$ parameter for establishments. At the firm level, most sectors remain far from Zipf.

Table 30: Distribution estimates across 2-digit sectors, 1997

| | | Establishments | | | Firms | | |
|---|---|---|---|---|---|---|---|
| Distribution | Parameter | Mean | St. Dev. | All | Mean | St. Dev. | All |
| Pareto | $\alpha$ | 0.52 | 0.15 | 0.56 | 0.56 | 0.15 | 0.62 |
| Lognormal | $\mu$ | 1.76 | 0.76 | 1.37 | 1.55 | 0.69 | 1.14 |
| Lognormal | $\sigma$ | 1.62 | 0.26 | 1.61 | 1.67 | 0.25 | 1.80 |
| Mixture | $\mu$ | 1.91 | 0.94 | 1.25 | 1.69 | 0.84 | 1.00 |
| Mixture | $\sigma$ | 1.47 | 0.43 | 1.58 | 1.41 | 0.51 | 0.86 |
| Mixture | $p$ | 0.80 | 0.19 | 0.86 | 0.69 | 0.21 | 0.86 |
| Mixture | $x_m$ | 4.90 | 10.49 | 3.57 | 4.08 | 6.93 | 3.47 |
| Mixture | $\alpha$ | 6.28 | 25.61 | 0.85 | 3.64 | 23.75 | 0.74 |
| Convolution | $\mu$ | 1.38 | 0.96 | 0.72 | 0.81 | 0.72 | 0.43 |
| Convolution | $\sigma$ | 1.51 | 0.29 | 1.37 | 1.39 | 0.25 | 1.29 |
| Convolution | $\alpha$ | 10.60 | 10.54 | 1.43 | 2.25 | 4.21 | 1.25 |

"All" columns show estimates for the entire all establishments/firms.

Going beyond these largest 10 sectors, in Table 30, we report the average and standard deviation from the estimations performed separately on every two-digit SIC sector. A few patterns are clear: the Pareto shape parameter is robustly below 1 for both establishments and firms, demonstrating that our benchmark MLE-based result is not driven by outlier industries. A mixture is overwhelmingly lognormal ($p$ is 0.8 for establishments and 0.69 for firms). However, for both the mixture and the convolution, there appears to be substantial heterogeneity away from Zipf's law in the estimates of the Pareto shape parameter $\alpha$.

Finally, we evaluate the relative statistical fit of each distribution among all 2-digit sectors. Our key findings for both firms and establishments hold broadly. We only report these results qualitatively due to Census disclosure restrictions.

- Starting with establishments, a lognormal distribution is statistically preferred to Pareto in a significant way in all but one SIC 2-digit sectors, and in that one sector neither distribution is statistically significantly better than the other.

- For the firm employment distribution, there is only one SIC 2-digit sector in which Pareto was estimated to be statistically significantly better than lognormal. In almost all of the other sectors, lognormal is estimated to have a statistically significantly better fit than Pareto.

- The convolution provides a statistically significantly better fit than Pareto in all but one SIC 2-digit sectors for both the establishment and firm distributions.

- The convolution is also estimated to have a statistically significantly better fit than lognormal in the majority of SIC 2-digit sectors for the firm distribution.
- For the establishment distribution, the convolution and the lognormal are estimated to have a statistically significantly better fit than the other in a similar number of SIC 2-digit sectors, but for a lot of sectors neither provides a better fit in a statistically significant way.

# E Theoretical Appendix

In this section, we show that correctly characterizing the firm size distribution has first-order implications for the effect of firm-level idiosyncratic shocks on aggregate activity, as seemingly innocuous statistical differences can lead to strikingly different economic implications. A Pareto shape parameter less than one implies an upper tail heavier than Zipf's law, and therefore leads to problematic theoretical implications, as the firm size distribution mean would not be well-defined. We extend the analysis of Gabaix (2011) to illustrate that with a Pareto shape parameter below one, aggregate volatility does not decrease in the number of firms, thus generating 'too much' aggregate volatility; in the paper we also showed that, if we allow idiosyncratic shocks to be drawn from our preferred convolution, the granular origins of aggregate fluctuations are significantly diminished. This theoretical and quantitative example illustrates the importance of accurately characterizing the size distribution in macroeconomic models

Let $\ell_t^i$ denote the employment at firm $i$ at time $t$. Aggregate employment is then simply $L_{N,t} = \sum_{i=1}^{N} \ell_t^i$, where $N$ denotes the number of firms.[28] Consider a set of multiplicative shocks $\varepsilon_{i,t}$ to the size of each firm such that $\varepsilon_{i,t}$ has mean 0 and variance $\varsigma_i$:

$$\Delta \ell_{i,t+1} \equiv \ell_{i,t} \varepsilon_{i,t+1}. \tag{5}$$

Then the aggregate employment growth rate is simply:

$$\frac{\Delta L_{N,t+1}}{L_{N,t}} = \sum_{i=1}^{N} \frac{\Delta \ell_{i,t+1}}{L_{N,t}} \tag{6}$$

and aggregate volatility, the variance of aggregate growth is:

$$\sigma_{N,t}^2 \equiv \mathbf{var}_t \left[ \sum_{i=1}^{N} \frac{\Delta \ell_{i,t+1}}{L_{N,t}} \right] = \sum_{i=1}^{N} \left( \frac{\ell_{i,t}}{L_{N,t}} \right)^2 \varsigma_i^2. \tag{7}$$

In the symmetric case where $\varsigma_i = \sigma \quad \forall i$, the Herfindahl index $h_{N,t}^2$ summarizes the aggregation of idiosyncratic shocks:

$$\sigma_{N,t}^2 = \sigma^2 \sum_{i=1}^{N} \left( \frac{\ell_{i,t}}{L_{N,t}} \right)^2 \equiv \sigma^2 h_{N,t}^2. \tag{8}$$

[28]For the rest of the theoretical exposition, we use "firm" to denote the individual economic entity.

The Herfindahl, in turn, can be rewritten as:

$$h_{N,t}^2 = \frac{\sum_{i=1}^N \ell_{i,t}^2}{\left(\sum_{i=1}^N \ell_{i,t}\right)^2} = \frac{\left(N^{-2}\sum_{i=1}^N \ell_{i,t}^2\right)}{\left(N^{-1}\sum_{i=1}^N \ell_{i,t}\right)^2} = N^{-1}\frac{\left(N^{-1}\sum_{i=1}^N \ell_{i,t}^2\right)}{\left(N^{-1}\sum_{i=1}^N \ell_{i,t}\right)^2}. \tag{9}$$

Therefore, when $\mathbb{E}\left[\ell\right]$ and $\mathbb{E}\left[\ell^2\right]$ are finite,

$$h_{N,t}^2 \times N \xrightarrow{a.s.} \frac{\mathbb{E}\left[\ell^2\right]}{\left(\mathbb{E}\left[\ell\right]\right)^2}. \tag{10}$$

**Definition 1** *Consider a random variable $Y$, a sequence of random variables $\zeta_N$, and a sequence of positive numbers $a_N$. Following Gabaix (2011), a convergence in distribution such that $\zeta_N / a_N \xrightarrow{d} Y$ as $N \to \infty$ is also denoted $\zeta_N \sim a_N Y$ and $\zeta_N$ is said to scale like $a_N$.*

Using equation 9 and the scaling definition above, we can characterize the scaling properties of granular shocks when the moments of the size distribution are finite.

**Proposition 1** *If the size distribution has finite mean and variance, then the size Herfindahl index is such that:*

$$h_N^2 \sim \frac{1}{N} \tag{11}$$

*and thus the macro variance $\sigma_N^2$ is decreasing in $\frac{1}{N}$.*

The proof is a straight application of the law of large numbers and is the same as the proof of Proposition 1 in Gabaix (2011). Proposition 1 implies that, in terms of the micro origins of macroeconomic volatility, there is no material difference between a lognormal size distribution and a Pareto size distribution with shape parameter when $\alpha > 2$.

However, when the Pareto shape parameter $\alpha$ is equal to or lower than 2, the distribution has undefined variance, and it has both undefined mean and variance when $\alpha$ is equal to or lower than 1. As we have shown, these are relevant regions of the Pareto parameter space, and so we describe the behavior of the Herfindahl index in the following proposition which extends Proposition 2 in Gabaix (2011).

**Proposition 2** *If firm size is distributed Pareto with shape parameter $\alpha$, then the size Herfindahl* *index is characterized by:*

$$h_N^2(\alpha) \propto \begin{cases} 1/N & \textit{when } \alpha > 2 \\ 1/\left(\frac{N}{\ln N}\right) & \textit{when } \alpha = 2 \\ 1/\left(N^{1-\frac{1}{\alpha}}\right)^2 & \textit{when } \alpha \in (1,2) \\ 1/(\ln N)^2 & \textit{when } \alpha = 1 \\ 1 & \textit{when } \alpha \in (0,1) \end{cases} \tag{12}$$

*and thus, in the case of $\alpha < 1$, the macro variance $\sigma_N^2$ no longer decays with N for N large enough.*

With $\alpha > 2$, we return to the world of Proposition 1. With a lower shape parameter, aggregate volatility declines more slowly with $N$. When $\alpha < 1$, the upper tail is heavier than Zipf's law ($\alpha = 1$) and aggregate volatility does not decline at all with $N$: the largest firms so disproportionately dominate that aggregate volatility *is* the idiosyncratic volatility of these firms. This is a starkly different context than the better-fitting lognormal, where volatility declines rapidly with $N$.

Proof: We prove Proposition 2 using Theorem 3.8.2 in Durrett (2017, p. 186-187), which is as follows:

**Theorem 1** *Suppose that $X_1$, $X_2$, ... are i.i.d. with a distribution that satisfies (i) $\lim_{x\to\infty} P(X_1 > x)/P(|X_1| > x) = \theta \in [0,1]$ and (ii) $P(|X_1| > x) = x^{-\alpha}L(x)$ with $\alpha \in (0,2)$ and $L(x)$ slowly varying. Let $s_N = \sum_{i=1}^N X_i$, $a_N = inf\{x : P(|X_1| > x) \leq 1/N\}$, and $b_N = N\mathbb{E}[X_1 \mathbb{1}_{|X_1| \leq a_N}]$. As $N \to \infty$, $(s_N - b_N)/a_N$ converges in distribution to a nondegenerate random variable Y. When $\alpha < 1$, we can let $b_N = 0$.*

Using equations (8) and (9), we can write

$$\sigma_N = \sigma \frac{(\sum_{i=1}^N \ell_i^2)^{1/2}}{(\sum_{i=1}^N \ell_i)} \tag{13}$$

When $\alpha > 2$, a Pareto random variable has finite mean and variance, and so we simply apply Proposition 1. When $\alpha = 2$, we must apply 1 to the numerator, $\sum_{i=1}^N \ell_i^2$. Here, $a_N = N$ and $b_N = N\int_1^N y \cdot y^{-(1+1)}dy = N\ln(N)$, and thus:

$$N^{-1}(\sum_{i=1}^N \ell_i^2 - N\ln(N)) \xrightarrow{d} u,$$

where $u$ is a random variable following a nondegenerate distribution that does not depend on $N$.

Thus:

$$\sum_{i=1}^{N} \ell_i^2 \sim N \ln(N).$$

It follows that:

$$h_N = \frac{(\sum_{i=1}^{N} \ell_i^2)^{1/2}}{\sum_{i=1}^{N} \ell_i} \xrightarrow{d} \frac{(N\ln(N))^{1/2} u^{1/2}}{N\mathbb{E}[\ell_i]} \propto \left(\frac{\ln(N)}{N}\right)^{\frac{1}{2}}$$

When $1 < \alpha < 2$, we again apply Theorem 1 to determine the numerator in equation 13. Since $l_i$ is Pareto distributed with scale parameter equal to 1 and shape parameter equal to $\alpha$, $l_i^2$ is Pareto distributed with the same scale parameter and shape parameter equal to $\frac{\alpha}{2}$. Since $\alpha < 2$, $\frac{\alpha}{2} < 1$ and $b_N = 0$.

$$P(\ell^2 > x) = P(\ell > x^{1/2}) = (x^{1/2})^{-\alpha} = x^{-\alpha/2},$$

which implies that $a_N = N^{2/\alpha}$. With $a_N$ and $b_N$ we apply Theorem 1:

$$N^{-2/\alpha} \sum_{i=1}^{N} \ell_i^2 \xrightarrow{d} u,$$

It follows that

$$h_N = \frac{(\sum_{i=1}^{N} \ell_i^2)^{1/2}}{\sum_{i=1}^{N} \ell_i} \xrightarrow{d} \frac{N^{1/\alpha} u^{1/2}}{N\mathbb{E}[\ell_i]} = \frac{u^{1/2}}{N^{1-\frac{1}{\alpha}}\mathbb{E}[\ell_i]} \propto 1/N^{1-\frac{1}{\alpha}}$$

When $\alpha = 1$, we have to apply Theorem 1 to both the numerator and denominator. For $\sum_{i=1}^{N} \ell_i^2$, $a_N = N^2$ and $b_N = 0$. For $\sum_{i=1}^{N} \ell_i$, $P(\ell > x) = x^{-1} \leq 1/N$ implies $a_N = N$; $b_N = N\int_1^N x \cdot x^{-(1+1)} dx = N\ln(N)$. We then have:

$$\frac{1}{N}(\sum_{i=1}^{N} \ell_i - N\ln(N)) \xrightarrow{d} g,$$

where $g$ is random variable following a nondegenerate distribution that does not depend on $N$. This implies that

$$\sum_{i=1}^{N} \ell_i \sim N \ln(N).$$

$$h_N = \frac{(\sum_{i=1}^{N} \ell_i^2)^{1/2}}{\sum_{i=1}^{N} \ell_i} \sim \frac{N}{N\ln(N)} \propto 1/\ln(N)$$

Finally, when $0 < \alpha < 1$, $a_N = N^{\frac{1}{\alpha}}$ and $b_N = 0$ for $\sum_{i=1}^{N} \ell_i$, and $a_N = N^{\frac{2}{\alpha}}$ and $b_N = 0$ for $\sum_{i=1}^{N} \ell_i^2$. This implies

$$\sum_{i=1}^{N} \ell_i \sim N^{1/\alpha},$$

and

$$\sum_{i=1}^{N} \ell_i^2 \sim N^{1/\alpha},$$

and thus,

$$h_N = \frac{(\sum_{i=1}^{N} \ell_i^2)^{1/2}}{\sum_{i=1}^{N} \ell_i} \sim \frac{N^{1/\alpha}}{N^{1/\alpha}} \propto 1.$$