

On the U.S. Firm and Establishment Size Distributions

Illenin O. Kondo*
University of Notre Dame

Logan T. Lewis
Federal Reserve Board

Andrea Stella
Federal Reserve Board

October 30, 2018

Abstract

This paper revisits the empirical evidence on the nature of firm and establishment size distributions in the United States using the Longitudinal Business Database (LBD), a confidential Census Bureau panel of all non-farm private firms and establishments with at least one employee. We establish five stylized facts that are relevant for the extent of granularity and the nature of growth in the U.S. economy: (1) with an estimated shape parameter significantly below 1, the best-fitting Pareto distribution substantially differs from Zipf's law for both firms and establishments; (2) a lognormal distribution fits both establishment and firm size distributions better than the commonly-used Pareto distribution, even far in the upper tail; (3) a convolution of lognormal and Pareto distributions fits both size distributions better than lognormal alone while also providing a better fit for the employment share distribution; (4) the estimated parameters are different across manufacturing and services sectors, but the distribution fit ranking remains unchanged in the sectoral subsamples. Finally, using the Census of Manufactures (CM), we find that (5) the distribution of establishment-level total factor productivity—a common theoretical primitive for size—is also better described by lognormal than Pareto.

JEL classifications: L11, E24

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

*The views expressed here should not be interpreted as reflecting the views of the Federal Reserve Board of Governors 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. We thank Colin Hottman for helpful comments and suggestions. We also thank seminar participants at the Federal Reserve Bank of Philadelphia, CEF 2018, IAAE 2018, and NASMES 2018 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).

1 Introduction

Modern macroeconomic models often feature cross-sectional firm heterogeneity as the firm size distribution affects important outcomes such as economic growth, international trade elasticities, or the sources of aggregate fluctuations. For instance, if an economy is dominated by large firms, idiosyncratic shocks to these firms may be an important source of aggregate fluctuations depending on how heavy the right tail of the firm size distribution is (Gabaix 2011, di Giovanni and Levchenko 2012, Stella 2015). Specifically, Gabaix (2011) shows the idiosyncratic origins of aggregate volatility when the size distribution follows Zipf’s law (that is, Pareto distributed with a shape parameter of 1). More generally, multiple foundational papers in macroeconomics and international economics rely on the Pareto distribution because of its analytical convenience and its seeming empirical regularity (Chaney 2008, Arkolakis et al. 2012, Rossi-Hansberg and Wright 2007, Luttmer 2011, Carvalho and Grassi 2018). In this paper, we revisit the empirical evidence on the nature of the distribution of firm size and establishment size in the United States and provide novel stylized facts along with maximum likelihood estimates of the best-fitting distributions.

Using the Longitudinal Business Database (LBD), 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. establishment and firm employment size distributions. In practice, two classes of distributions are commonly used in the modern macroeconomic literature: the lognormal distribution and the Pareto distribution. We use maximum likelihood estimation (MLE) to estimate the parameters of the lognormal and Pareto distributions, as well as two combinations of lognormal and Pareto distributions. Though both lognormal and Pareto distributions are “heavy-tailed”—that is, their upper tails are “heavier” than an exponential distribution—they are very different in their economic origins and implications.¹

We find the best-fitting Pareto shape parameter to be robustly below 1 for both firms and establishments: a Pareto shape parameter 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 also extend the analysis of Gabaix (2011) to illustrate another striking pitfall and likely unpalatable consequence of the estimated Pareto distributions. We show that, unlike the lognormal case, aggregate volatility does not decrease in the number of firms in the case of Pareto distributions with shape parameter below 1: idiosyncratic shocks generate ‘too much’ aggregate volatility.

Statistically, we clearly find that a lognormal fits the employment size distribution better than a Pareto. This finding holds even when we consider most cuts of the upper tail of the firm size distribution: the Pareto distribution provides a better fit of the right tail for only a narrow range,

¹Head et al. (2014) examine the consequences of a lognormal distribution in international trade, and Rossi-Hansberg and Wright (2007) explores the origins of establishment growth and how scale dependence can vary across industries and generate distributions with thinner tails than Zipf’s law.

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 main reference in the literature.

Moving beyond the simple lognormal and Pareto distributions, we estimate the parameters of a statistical mixture of lognormal and Pareto as well as a convolution of a Pareto random variable multiplied by a lognormal random variable. We find that the mixture provides the best overall fit, but the convolution also beats the fit of the Pareto or lognormal distributions alone. However, we find that the mixture often also has a Pareto shape parameter below 1, while the convolution has a Pareto shape parameter well above 1. If a distribution mixes any Pareto with a shape parameter below 1, some incredibly large firms would be generated in reasonably sized samples, leaving too many employees belonging to the very largest bin of firms. Therefore, when we consider a second criterion of the distribution fit—the fraction of employment accounted for by various bins of establishment or firm size—the convolution provides a markedly better fit. Moreover, because the convolution can arise in a heterogeneous firm model with two sources of firm-level shock, say a demand shock and a productivity shock or two productivity shocks, the convolution may be the more appropriate distribution to use.

Next, we characterize the nature of the firm size distribution across sectors and over time. In the last several decades, the U.S. economy made a substantial transition away from traditional manufacturing to a more service-based economy. This ongoing structural change in the U.S. economy may have implications for economic growth, both at the firm and aggregate level. We explore how such structural transformation affected the employment size distribution across sectors and across time. We find that manufacturing and services are both well-described by a mixture of lognormal and Pareto, but they have notable economic differences in parameter estimates.

Finally, we use estimates of establishment-level total factor productivity (TFP) from Foster et al. (2016) to estimate the best distribution fit for this more-primitive source of establishment size. We find strong evidence that TFP is better described by lognormal than Pareto, and a lognormal distribution often even fits better than a convolution. The mixture still provides the best statistical fit, but the evidence suggests that TFP can be reasonably modeled by a lognormal distribution. This is consistent with an overall size distribution being a convolution, where TFP is lognormal and another firm size determinant, e.g. firm demand, is distributed Pareto.

Overall, our contribution is twofold: first, we adopt the most rigorous statistical techniques to estimate and test the shape of the firm and establishment size distributions, while the previous literature largely relied on demonstrably weaker econometric approaches, such as regression analysis. Second, we use the most comprehensive database for the United States: we have the entire population of firms and establishments in the United States over a time period that spans 30 years, which allows us to not only study the entire population, but also subsamples in the upper tail or sector-year subsamples. This evidence can help develop better and more relevant macroeconomic models of heterogeneous firms and establishments.

The existing literature is mixed about the nature of the best-fitting size distribution. In early work, Simon and Bonini (1958) lays out a simple statistical model of firm growth which can generate both lognormal and Pareto distributions and provides anecdotal evidence that firm sizes might be near Zipf’s law. Using Small Business Administration data and a bin-based regression, Luttmer (2010) finds, like Axtell (2001), that U.S. firm sizes are Pareto with a shape parameter of 1.06. Combes et al. (2012) find evidence that firm-level TFP in France is better described by a lognormal distribution than a Pareto. They use a mixture of Pareto and lognormal and conclude that the mixture parameter is quite close to lognormal and proceed with a straight lognormal distribution for their model.

A few recent papers identify and explore the implications of a firm size distribution that is not Pareto. Fernandes et al. (2015) propose to model the firm productivity distribution with a lognormal distribution to match the empirical evidence on the importance of the intensive margin of trade. Sager and Timoshenko (2017) show that a convolution of lognormal and Pareto fits best using Brazilian export sales data. Nigai (2017) shows that a mixture of lognormal and Pareto fits the firm productivity distribution of French firms the best and that its adoption affects the estimation of the gains from trade. Armenter and Koren (2015) finds that the Pareto shape parameter required to match the exporter size premium (how much larger the average exporter is) would be about 1.65, and thus it is difficult to reconcile models of selection into exporting by firm size with a size distribution generated by realistic Pareto parameterizations.

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 results on the employment size distributions and analyzes the TFP establishment distribution in manufacturing. Finally, 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. It was created using information from a wide array of surveys conducted by the U.S. Census Bureau, such as the Standard Statistical Establishment List, the quinquennial Economic Census, and the annual surveys.² Table 1 shows the number of establishments and firms we use in each year.

²The 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: LBD 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.

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. The LBD firm ID was created using information from the quinquennial Economic Census and the annual Company Organization Survey. The latter is only submitted to large firms and a subset of small firms, so the firm ID is not entirely reliable outside of Census years, which is the years when the quinquennial Census is conducted; for this reason, we only use Census years in our analysis.

In most of the paper, we measure the size of an establishment or firm with its number of employees. Most of the analysis will be conducted on the whole universe, but we also consider two subsamples: manufacturing and services, where the latter excludes retail, wholesale and FIRE.³

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 Total Factor Productivity (TFP) with data from the Annual Survey of Manufactures (ASM) and the quinquennial Census of Manufactures (CM). Since the distribution of productivity shocks represents an important primitive assumption in many theories, we extended our analysis to the distribution of TFP 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 TFP 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 TFP calculation is available for a bit over half of manufacturing establishments.

³We define the manufacturing sector as all establishments with two-digit SIC codes $\in [20, 40)$ for years 1977, 1982, 1987, 1992, and 1997. For 2002, 2007, 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.

Table 2: Sample by sector

Year	Services		Manufacturing		TFP est
	Est.	Firm	Est.	Firm	
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.

3 Parametric Distributions and Estimation Methods

Motivated by the existing literature, we fit four parametric distributions to the data.⁴ 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, much of the endogenous growth literature focuses on generating a Pareto distribution and Pareto is widely used in heterogeneous firm models that assume an exogenous distribution. The CDF of a Pareto with scale parameter x_m and shape parameter α is

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

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. 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 μ and σ is

$$F_L(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{\ln x - \mu}{\sqrt{2}\sigma}\right). \quad (2)$$

For the lognormal distribution, the mean is given by $e^{\mu+\sigma^2/2}$ and the variance $(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$.

⁴See Simon and Bonini (1958) for a simple and seminal statistical model of firm growth which can generate both lognormal and Pareto distributions as special cases.

Besides the two most popular parametric distributions in the literature, we consider two distributions that 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

$$F_M(x) = pF_L(x) + (1 - p)F_P(x), \quad (3)$$

where F_L is the CDF of a lognormal with parameters μ and σ and F_P is the CDF of a Pareto with scale parameter x_m and shape parameter α .

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.⁵ Thus, the CDF of the convolution is

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

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

Most of the previous literature on the firm size distribution evaluated the fit of the Pareto distribution using regression analysis, see Axtell (2001) and 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.⁶ Since we are estimating parametric models with a known and simple likelihood, we use Maximum Likelihood Estimation (MLE) for its excellent statistical properties.⁷ 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 χ_k^2 .

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 and g the model in the denominator. 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.

⁵See Reed (2001) for the stochastic growth processes that can yield such distribution. Sager and Timoshenko (2017) also rationalize this distribution using a model with Pareto productivity shocks and lognormal demand shocks.

⁶Clauset 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 very closely when evaluated with the regression analysis approach used in Axtell (2001) and Gabaix (2009). See also Eeckhout (2009).

⁷We ran Monte Carlo simulations to investigate the accuracy of the MLE in estimating our models and found it to be reliable. Results are in the appendix.

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 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 right by one unit the lognormal, the lognormal component of the mixture, and the convolution.

4 Employment Distribution Results in the U.S.

In this section, we highlight three stylized facts that emerge from the U.S. firm and establishment employment distributions in the years 1982, 1992, 1997, 2002, and 2012.⁸

4.1 Lognormal versus Pareto

Stylized Fact 1 *For both firms and establishments, the best-fitting Pareto distribution has an estimated shape parameter significantly below 1, and differs substantially from the common Zipf’s law benchmark.*

Stylized Fact 2 *A lognormal fits both establishment and firm size employment distributions better than the commonly used Pareto, even far in the upper tail.*

Table 3: Pareto and lognormal estimates using the entire sample

Year	Pareto		lognormal			
	α		μ		σ	
	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

⁸To capture long-run trends we analyze every 10 years starting with the 1982 census year, and 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 1, around 0.6. The estimates are all tightly estimated, standard errors and test statistics are available upon request.

Right Tail Estimates While the lognormal distribution might be a better fit overall, for some economic questions, only the upper tail is the relevant portion of the distribution. Therefore, we present in Table 4 the parameter estimates for a Pareto distribution and a truncated lognormal distribution at various employment thresholds for the 1997 firm size distribution.

In Table 4, the Pareto shape parameter first increases monotonically with the upper tail threshold. 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 non-trivially above one.⁹ Indeed, the lack of economic stability of the shape parameter estimates across cutoffs suggests that the underlying distribution is not Pareto, because a true Pareto distribution would have shape parameter estimates invariant to the cutoff or more stable further in the upper tail.¹⁰

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

Threshold	Pareto	lognormal		N (rounded)
	α	μ	σ	
2	0.76	-0.02	1.93	3,750,000
5	0.97	-5.88	3.11	2,150,000
10	1.05	-14.97	4.26	1,160,000
25	1.11	-14.79	4.22	450,000
50	1.12	-16.44	4.46	210,000
100	1.10	-19.10	4.85	95,000
200	1.05	-22.01	5.28	43,000
300	1.02	-23.80	5.56	28,000
400	1.01	-24.73	5.72	20,000
500	1.01	-21.29	5.40	16,000
1000	1.01	-5.32	3.73	8,100
2500	1.05	1.99	2.68	3,300
5000	1.11	5.70	2.00	1,600
10000	1.23	7.21	1.66	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.

⁹See Table 4 for truncated data sample size.

¹⁰The well-known Hill plot for visually identifying power-law distributions is based on this stability argument.

Does the Pareto Provide a Better Fit? In Table 5, we also use the Vuong statistic to formally test which distribution has a better fit for each truncation threshold. The truncated 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. With 300 employees, neither distribution is preferred. Thus, the lognormal fit typically dominates the Pareto fit even far in the upper tail, except in a relatively narrow truncation window.

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	0.16	None
400	-2.0	0.02	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	0.00	lognormal

Revisiting the Zipf’s Law Evidence How can these results be reconciled with those of Axtell (2001)? Using a very popular methodology, Axtell explores the fit of Pareto by running a regression of the logarithm of the frequency distribution of the data on the logarithm of the firm size. Axtell 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. Moreover, Axtell (2001) computes the frequency distribution using successive bins of increasing size in powers of three. This approach yields thirteen (13) data points and the regression estimation on such binned sample effectively gives more weight to the observations in the right tail.

We first replicated Axtell (2001)’s procedure to ensure that differences in our results were not due to the underlying sample used.¹¹ Axtell (2001) also uses the LBD, but includes non-employee

¹¹In Appendix A.1, we analyze the ability of the Axtell (2001) regression approach to recover the true parameters of

firms. 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 2.057 with a standard deviation of 0.039, which implies that the differences in results are not due to the underlying data, but to the methodologies used.

Visual inspection of Figure 1 (p. 1819) from Axtell (2001) 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 1. 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 right tail of the firm size distribution. However, the best-fitting lognormal may look quite similar to the best-fitting Pareto. Does this mean that, despite the better statistical fit of a lognormal, these different distributions are interchangeable in economic models? Simon (1955) and Simon and Bonini (1958) provide an early and insightful discussion of different ingredients generating various heavy-tailed steady-state size distributions. Moreover, if for a given model the two distributions yield different results, then it must be that the nature of the left tail of the distribution or its right tail is important. In the next section we sharply contrast the implications of these seemingly interchangeable distributions in one such context: Gabaix (2011)'s study of aggregate fluctuations arising from idiosyncratic shocks to very large firms in economies with a finite set of firms, or “granular” economies.

4.2 Lognormal versus Pareto: Some Theoretical Implications for Granularity

Having established that a lognormal distribution is a better fit than a Pareto for the U.S. firm and establishment employment distributions, we now highlight some theoretical implications for aggregate volatility that sharply distinguish the two distributions.

Gabaix (2011) provides a useful framework, emphasizing the potential of large firms to generate sizable aggregate shocks in a “granular” economy. Gabaix (2011) focuses his discussion on a Pareto distribution as it approaches Zipf's law. Gabaix (2011) shows that idiosyncratic shocks can be a more important source of aggregate volatility under Zipf's law, compared to a lognormal. Here, we show the implications of going *past* Zipf's law, with an empirically-plausible shape parameter below one, and contrast it to the implications of a lognormal distribution. In particular, we show that Pareto shape parameter estimates below one have unappealing economic implications in granular economies, further suggesting that it may not be an appropriate modeling assumption in

a Pareto distribution and other distributions.

heterogeneous-firms models. We state below the main results and their intuition. See Appendix C for detailed derivations.

Proposition 1 *Consider N firms facing i.i.d. multiplicative growth shocks with finite variance ζ^2 to firm employment. If the size distribution of firms has finite mean and variance, then the variance of aggregate employment, $\sigma_{N,t}^2$, is decreasing at a rate proportional to $\frac{1}{N}$.*

For a formal proof, see Appendix C. The proof is an application of the law of large numbers. Intuitively, with finite mean and variance, the tail of the distribution is not too fat and uncorrelated shocks to firms will cancel each other out, and aggregate volatility decreases as the number of firms in an economy increases. This holds true of the lognormal distribution with any parameters, but for Pareto, it depends on the shape parameter. Building on Gabaix (2011), we fully characterize the Pareto case with Proposition 2.

Proposition 2 *Consider N firms facing i.i.d. multiplicative growth shocks with finite variance ζ^2 to firm employment. If the size distribution of firms is Pareto with shape parameter α , then the variance of aggregate employment, $\sigma_{N,t}^2$, is proportional to:*

$$\left\{ \begin{array}{ll} \zeta^2 \times 1/N & \text{when } \alpha > 2 \\ \zeta^2 \times 1 / \left(\frac{N}{\ln N} \right) & \text{when } \alpha = 2 \\ \zeta^2 \times 1 / \left(N^{1-\frac{1}{\alpha}} \right)^2 & \text{when } \alpha \in (1, 2) \\ \zeta^2 \times 1 / (\ln N)^2 & \text{when } \alpha = 1 \\ \zeta^2 & \text{when } \alpha \in (0, 1) \end{array} \right. \quad (5)$$

See Appendix C for a formal proof. 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 .¹²

Discussion Zipf's law is considered a good approximation of the size distribution of U.S. firms. Having replicated the main analysis underlying that benchmark, we use detailed Census data to show that the lognormal provides a better fit while the best-fitting Pareto is not stable in the upper

¹²We also show that, quantitatively, the granular origins of aggregate fluctuations would be unrealistically large in the case of the best-fitting Pareto. See Appendix C for the calibrated simulation results in the spirit of Gabaix (2011). See also Stella (2015), Roszbach (2017) and Yeh (2018) for other quantitative investigations of the contribution to aggregate volatility coming from idiosyncratic shocks in granular economies.

tail and yields a shape parameter below one for the entire sample. Since Simon and Bonini (1958), the growth literature has also explored economic and stochastic properties that rationalize a Pareto distribution versus a lognormal. Beyond the origins of the size distribution, we show that seemingly innocuous statistical differences can also lead to strikingly different economic implications. The granularity properties characterized above illustrate this point and the importance of accurately characterizing the size distribution.

The existing literature motivated using the lognormal and the Pareto as natural starting points in our analysis of the size distributions of US. firms and establishments. However, neither distribution seems sufficiently flexible to fit equally well the right and left tails of the distribution. We therefore extend our analysis to consider two more-flexible alternatives: a statistical mixture of lognormal and Pareto, and a convolution, or product of a lognormal random variable and a Pareto random variable.

4.3 Mixture and Convolution Distributions

Stylized Fact 3 *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 fit, but economically, the convolution's fit may be more suitable.*

Mixture Estimates Table 6 provides the parameter estimates for the statistical mixture distribution of a lognormal and a Pareto. The parameters μ , σ , and α 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 very stable across both firms and establishments.

The mixture also has the unique 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.

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

Year	μ		σ		p		x_m		α	
	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 μ and σ parameters are systematically lower than their mixture counterparts, while most notably the Pareto shape parameter α is higher, and always well above 1, a point to which we will return shortly.

Table 7: Parameter Estimates - Convolution

Year	μ		σ		α	
	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 α for firms, relative to establishments, means that the right tail firm size will be thicker than the establishment size right tail.¹³

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 slightly less heavy tail and it is becoming less heavy over time, while the tail of the Pareto component for the firm size distribution 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

¹³This pattern is also apparent for the mixture. We chose not to emphasize the relative thickness of the firm size right tail because the Pareto shape parameters estimated in those cases are below 1. Interestingly, the Pareto estimated on the entire sample does not have the same property, suggesting the importance of using more flexible distributions.

(1989) described earlier. As an alternative, we also computed the AIC for all the distributions and find identical rankings.

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.¹⁴ Again, the AIC produces identical rankings.

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 Business Dynamic Statistics (BDS) and data simulated using the parameter estimates in Tables 3 to 7. The BDS are public and are drawn from the same underlying data that we use for our analysis.¹⁵

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) provides a very poor fit of the U.S. firm size distribution. Even the Pareto with our estimate of the shape parameter, the fourth column, does not fit the data well, putting too much weight on the left and right tails. Lognormal, mixture, and convolution all provide a better fit.

Table 8: Tabulation of 1997 BDS data and simulated data

	BDS	Axtell	Pareto	lognormal	Mixture	Convolution
1 to 4	54.8	81.9	62.6	54.9	55.0	55.8
5 to 9	21.2	9.4	13.0	17.3	20.4	19.4
10 to 19	12.2	4.5	8.6	12.3	12.5	12.3
20 to 49	7.5	2.6	6.9	9.5	7.9	8.1
50 to 99	2.3	0.8	3.2	3.4	2.3	2.6
100 to 249	1.2	0.5	2.6	1.8	1.2	1.3
250 to 499	0.4	0.1	1.2	0.4	0.3	0.4
500 to 999	0.2	0.1	0.8	0.1	0.2	0.2
1,000 to 2,499	0.1	0.0	0.6	0.0	0.1	0.1
2,500 to 4,999	0.0	0.0	0.3	0.0	0.0	0.0
5,000 to 9,999	0.0	0.0	0.2	0.0	0.0	0.0
$\geq 10,000$	0.0	0.0	0.0	0.0	0.0	0.0

Figure 1 provides a graphical representation of the fit of the parametric distributions with 1997 BDS data. It depicts the Complementary Cumulative Distribution Function in log space: for each logarithm of s number of employees, it shows how many firms are larger than s . The black line

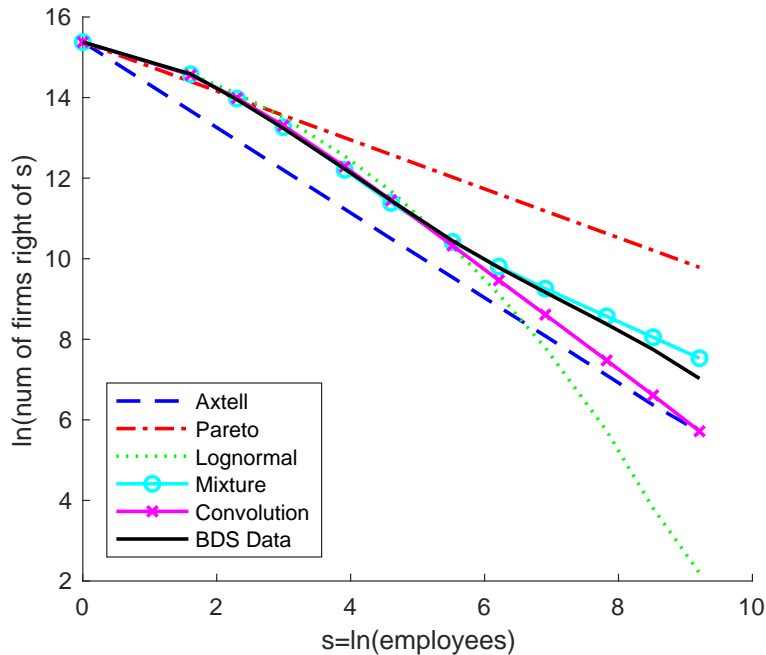
¹⁴Test statistics available upon request.

¹⁵The U.S. Census Bureau does not allow researchers to disclose histograms or other tabulations of the data, so we use data made publicly available by the U.S. Census Bureau.

represents the data available in the BDS described in Table 8. Distributions with fits above the black solid line represent too many larger firms than in the data, while distributions with fits below the black line have too few larger firms relative to the data.

Figure 1 confirms that the mixture and convolution appear to provide the best fit with this binning procedure. By contrast, 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 lined) with shape parameter 1.06 remains below the data, but, overall, it matches the slope of the black line excluding the left tail, which is essentially by design of the regression-based estimation procedure. The lognormal fit seems decent through about $e^5 \approx 150$ employees but produces too few very large firms. The convolution also has a slightly too-thin tail, while the mixture has a slightly too-thick tail. In Appendix B, we evaluate the stability of the employment size distribution in the BDS and the corresponding mixture distribution’s fit across time.

Figure 1: CCDF



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.¹⁶ 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 parametric distributions under study. The second column shows, for example, that firms with

¹⁶See Appendix B for confidence bands and a description of how this table was constructed.

20 to 49 employees account for 10.6 percent of overall employment in the economy. Here, Axtell’s estimate 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).

But this table also makes clear the odd results of a Pareto distribution or Mixture containing a Pareto with a shape parameter below one. Simulating this distribution generates some very, very large firms, whose employment completely dominates the economy. The lognormal, as expected from Figure 1, 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.

Table 9: Fraction of 1997 firm employment

	BDS	Axtell	Pareto	lognormal	Mixture	Convolution
1 to 4	5.65	11.56	0.00	7.04	0.46	6.57
5 to 9	6.53	5.44	0.00	7.52	0.53	7.15
10 to 19	7.73	5.41	0.00	11.06	0.67	9.27
20 to 49	10.62	6.95	0.00	19.07	0.94	13.57
50 to 99	7.52	5.06	0.00	15.56	0.64	9.81
100 to 249	8.72	6.40	0.01	17.99	0.68	11.23
250 to 499	5.54	4.63	0.01	9.76	0.44	7.06
500 to 999	5.09	4.45	0.01	6.18	0.44	5.95
1,000 to 2,499	7.07	5.61	0.02	4.00	0.66	6.44
2,500 to 4,999	5.4	4.04	0.02	1.18	0.61	3.98
5,000 to 9,999	5.46	3.88	0.03	0.44	0.72	3.35
$\geq 10,000$	24.68	36.58	99.90	0.19	93.19	15.64

4.4 Manufacturing versus Services

We have so far discussed results obtained using the entire U.S. population of firms and establishments. In this section, we estimate the parametric distributions on two subsamples: manufacturing and services, where the latter excludes retail, wholesale and FIRE.

Stylized Fact 4 *Manufacturing and services sectors have notably different distribution estimates, the distribution fit ranking is unchanged in these sector-year subsamples.*

Pareto and Lognormal Estimates Table 10 presents the estimates for the lognormal distribution by sector. Focusing on the average line, manufacturing establishments have both a larger μ and σ than services establishments, implying both a greater fitted mean and variance. A similar pattern holds for manufacturing firms relative to services firms. Table 11 presents the Pareto shape parameter α by sector. Notably, manufacturing establishments and firms have an α 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 Pareto on the overall 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 so much in the manufacturing sector.

Table 10: Lognormal by sector

Year	Manufacturing		Services	
	μ	σ	μ	σ
<i>Establishments</i>				
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
<i>Firms</i>				
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 11: Pareto α by sector

Year	<i>Establishments</i>		<i>Firms</i>	
	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 12 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.

Starting with manufacturing establishments, the Pareto scale parameter, x_m , is estimated to take effect around 4 employees and the Pareto shape parameter, α , is estimated to be a reasonable 1.24, but the mixing parameter $p = 0.96$ suggests that 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 1. 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 overtime. The estimated distributions appear to be getting more lognormal over time, especially in the services sector. Over time, the estimated Pareto shape parameters also appear to be declining, but not for establishments.

Table 12: Mixture by sector

Year	Manufacturing					Services				
	μ	σ	p	x_m	α	μ	σ	p	x_m	α
<i>Establishments</i>										
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
<i>Firms</i>										
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.

Convolution Estimates Finally, Table 13 shows the sectoral parameter estimates for the convolution distribution. As with the previous distributions, manufacturing establishments and firms have a much higher estimates of mean μ and standard deviation σ than their services counterparts. The Pareto shape parameter is well above 1 for all subsamples. These results further suggest 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 α 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 overall 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 13: Convolution by sector

	Manufacturing			Services		
	μ	σ	α	μ	σ	α
	<i>Establishments</i>					
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
	<i>Firms</i>					
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

Finally, we formally test which distribution fits the data best by sector. For establishments and firms in both manufacturing and services across 1982, 1992, 1997, 2002, and 2012, the 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.¹⁷ The rankings are statistically significant at least at a 5 percent level and typically at a much tighter level.

¹⁷Detailed results available upon request.

4.5 Manufacturing TFP

Stylized Fact 5 *The distribution of establishment-level total factor productivity is 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 distribution of productivity shocks self-reflects into the firm size distribution, but given selection, demand functions, and potentially many sources of shocks, this is not always the case.¹⁸

Therefore, in this section, we focus directly on estimates of the TFP distribution available for the U.S. manufacturing sector. We use the TFPR measures of TFP as estimated by for establishments in the Census of Manufactures. Our distribution estimates are only for 1982, 1992, 1997, and 2002 because these TFPR estimated by Foster et al. (2016) are not available for 2012.

Table 14 shows the results of the maximum likelihood estimation of the four parametric distributions considered in this paper. Starting with the Pareto shape parameter fit in the top sub-panel, the estimate averages a paltry 0.13, inheriting all of the economic issues inherent with an α below 1. The second sub-panel provides the lognormal fit, which remains remarkably consistent across time. The mixture estimates again show a very high proportion of lognormal, 0.93 on average. The shape parameter is well above 1 in all years, while the lognormal parameters μ and σ are very similar to the estimates from the lognormal alone. This is in contrast to the Pareto shape parameter of the mixture for the employment size distributions, which were below 1. Those lognormal parameters remain fairly similar in the convolution, which includes very high Pareto shape α averaging 3.4.

Using the likelihood-based statistical tests, we find that generally, the mixture distribution outperforms lognormal, but, in 2002, the convolution outperforms both mixture and lognormal, bumping the mixture fit to second place. 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 data. The data strongly suggest that Pareto provides a poor fit, and that lognormal is a reasonable distribution for TFP.

¹⁸ Mrazova et al. (2017) 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.

Table 14: Manufacturing establishment TFP

Year	α	p	μ	σ	
<i>Pareto</i>					
1982	0.11				
1992	0.14				
1997	0.12				
2002	0.15				
Average	0.13				
<i>Lognormal</i>					
1982			1.86	0.57	
1992			1.90	0.54	
1997			1.93	0.54	
2002			1.97	0.57	
Average			1.92	0.56	
Year	x_m	α	p	μ	σ
<i>Mixture</i>					
1982	3.68	1.59	0.87	1.85	0.56
1992	4.75	1.52	0.88	1.86	0.51
1997	19.65	1.83	0.98	1.90	0.50
2002	19.49	1.71	0.97	1.92	0.49
Average	11.89	1.67	0.93	1.88	0.52
<i>Convolution</i>					
1982		4.00		1.61	0.51
1992		3.44		1.61	0.45
1997		3.57		1.65	0.46
2002		2.75		1.61	0.43
Average		3.44		1.62	0.46

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. We establish five stylized facts about these distributions and provide guidance for researchers in parameterizing models that include firm heterogeneity.

The commonly used Pareto distribution is a particularly poor fit for U.S. establishments and firms. The lognormal distribution, a commonly used alternative, provides a better fit in most circumstances and for most truncations of the right tail. Economically, the lognormal fit has markedly different properties, as our Pareto shape parameter is robustly below one, implying neither a well-defined mean or variance. Our estimate is *past* the often discussed Zipf's Law result studied by Gabaix (2011) and others. So for both statistical and economic reasons, the use of Pareto is generally unwise.

Both a mixture and a convolution of a lognormal and Pareto distribution provide a better fit. While the mixture has the best statistical fit, its Pareto shape parameter below 1 implies that it inherits the unappealing characteristics of the Pareto distribution in that range. In addition, 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 might provide a more suitable choice economically for use in other applications. It can also be generated in a very reasonable way as the product of a Pareto distribution random variable and a lognormally distributed one; thus, models incorporating both productivity shocks and taste shocks could easily generate such a distribution.

We estimate our four distributions on manufacturing and services sub-samples. Here too we show that Pareto is strictly dominated by the other three distributions, with the same overall ranking of mixture first, then convolution, then lognormal, and Pareto fitting least well. While the services sector tends to have estimates close to the overall distribution (unsurprising given that much of the overall distribution consists of our services sub-sample), the manufacturing estimates are characterized by even lower Pareto shape parameters. In a mixture, the Pareto shape parameter is less consistent over time for manufacturing establishments, but remains well below one for manufacturing firms. Both manufacturing μ and σ tend to be higher than their services counterparts. And as with the results for all establishments and firms, the convolution yields sensible parameter estimates that can be included tractably in models.

Finally, we make use of manufacturing TFP estimates to consider the distribution of this deeper source of firm heterogeneity. We find that if anything, lognormal provides an even better fit than Pareto, often statistically beating out the convolution. This lends further support to our suggestion that given only one source of heterogeneity, lognormal TFP draws are reasonable. Adding a second source of heterogeneity with Pareto distributed draws would further help fit the overall employment size distribution. Aside from providing guidance about calibrating models with exogenously defined firm heterogeneity, our results also highlight how future models of endogenous growth should not generate straight Pareto size distributions.

References

- Arkolakis, Costas, 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.
- Carvalho, Vasco M and Basile Grassi**, “Large Firm Dynamics and the Business Cycle,” 2018.
- Chaney, Thomas**, “Distorted Gravity: The Intensive and Extensive Margins of International Trade,” *American Economic Review*, August 2008, *98* (4), 1707–1721.
- 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.
- 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 Rodríguez-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.
- Head, Keith, Thierry Mayer, and Mathias Thoenig**, “Welfare and Trade without Pareto,” *American Economic Review*, May 2014, 104 (5), 310–316.
- Luttmer, Erzo G. J.**, “On the Mechanics of Firm Growth,” *The Review of Economic Studies*, July 2011, 78 (3), 1042–1068.
- Luttmer, Erzo G.J.**, “Models of Growth and Firm Heterogeneity,” *Annual Review of Economics*, 2010, 2 (1), 547–576.
- Mrazova, Monika, J Peter Neary, and Mathieu Parenti**, “Sales and Markup Dispersion: Theory and Empirics,” September 2017.
- Nigai, Sergey**, “A tale of two tails: Productivity distribution and the gains from trade,” *Journal of International Economics*, January 2017, 104, 44–62.
- Reed, William J.**, “The Pareto, Zipf and other power laws,” *Economics Letters*, December 2001, 74 (1), 15–19.
- Rossbach, Jack**, “International Competition and Granular Fluctuations,” December 2017, p. 51.
- 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 Aggregate Trade Elasticities,” 2017.
- Simon, Herbert A.**, “On A Class Of Skew Distribution Functions,” *Biometrika*, December 1955, 42 (3-4), 425–440.
- and **Charles P. Bonini**, “The Size Distribution of Business Firms,” *The American Economic Review*, 1958, 48 (4), 607–617.
- 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, p. 41.

A Monte Carlo experiments

A.1 Axtell's regression analysis

To investigate the performance of Axtell's methodology to determine the fit of a Pareto distribution, we generate 250,000 synthetic datasets for each of the following parametric distributions: Pareto with Axtell's estimated parameter (1.06), and Pareto, lognormal, Mixture and Convolution with our parameter estimates for 1997 Census data. We implemented Axtell's methodology to estimate the Pareto shape parameter.¹⁹

Table 15: Axtell's regression analysis on synthetic data

	Axtell's Pareto	Pareto	lognormal	Mixture	Convolution
True α	1.06	0.61		0.74	1.25
$\hat{\alpha}$	1.05	0.61	1.37	0.81	1.12
c.i.	(1.00, 1.10)	(0.58,0.63)	(1.24,1.45)	(0.77,0.85)	(1.08,1.16)
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: α 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 15 shows that Axtell's methodology is able to correctly uncover the Pareto shape parameter when the data is drawn from a Pareto. However, Axtell's methodology produces a shape parameter close but above one even when data is drawn from a Convolution with our estimated parameters.

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.²⁰ 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 16 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

¹⁹Axtell (2001) explains in "References and Notes" 30 how he implements the regression in Figure 1.

²⁰We used a million observations because using the number of observations in the 1997 LBD was not computationally feasible for this exercise.

for the most part tiny. Table 17 presents the percentage of times that the likelihood-based ratio tests with 95% confidence were able to pick the correct distribution. 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, but they struggle when the true distributions Pareto or lognormal are compared to the more flexible Mixture and Convolution.

Table 16: RMSEs

	Pareto	lognormal	Mixture	Convolution
μ		1.17	1.00	0.49
RMSEs		0.00	0.34	0.00
σ		1.74	1.49	1.29
RMSEs		0.00	0.14	0.00
p			0.86	
RMSEs			0.08	
α	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.

Table 17: Likelihood-based ratio tests

<i>True:</i>	Pareto	lognormal	Mixture	Convolution
<i>Alternative:</i>				
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.

B Additional Results

In Table 18, 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, 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 1 and the mixture, also incorporating a Pareto estimate with a shape parameter below 1, consistently draws massive firms which account for nearly all of employment. 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 18: Fraction of 1997 firm employment: 95% confidence intervals

	BDS	Axtell	Pareto	lognormal	Mixture	Convolution
1 to 4	5.65	(5.22,13.95)	(0.00,0.01)	(7.01,7.07)	(0.01,1.10)	(5.43,6.97)
5 to 9	6.53	(2.46,6.56)	(0.00,0.01)	(7.49,7.56)	(0.02,1.27)	(5.91,7.58)
10 to 19	7.73	(2.44,6.52)	(0.00,0.01)	(11.01,11.11)	(0.02,1.59)	(7.66,9.83)
20 to 49	10.62	(3.14,8.39)	(0.00,0.01)	(18.98,19.15)	(0.03,2.24)	(11.22,14.39)
50 to 99	7.52	(2.28,6.11)	(0.00,0.02)	(15.48,15.65)	(0.02,1.52)	(8.11,10.40)
100 to 249	8.72	(2.89,7.73)	(0.00,0.03)	(17.87,18.12)	(0.02,1.63)	(9.28,11.91)
250 to 499	5.54	(2.09,5.60)	(0.00,0.03)	(9.63,9.89)	(0.01,1.05)	(5.83,7.50)
500 to 999	5.09	(2.00,5.39)	(0.00,0.04)	(6.04,6.32)	(0.01,1.05)	(4.91,6.34)
1,000 to 2,499	7.07	(2.53,6.82)	(0.00,0.07)	(3.83,4.17)	(0.02,1.58)	(5.31,6.90)
2,500 to 4,999	5.40	(1.82,4.99)	(0.00,0.07)	(1.03,1.32)	(0.02,1.44)	(3.27,4.36)
5,000 to 9,999	5.46	(1.74,4.88)	(0.00,0.09)	(0.32,0.57)	(0.02,1.73)	(2.73,3.76)
$\geq 10,000$	24.68	(23.48,71.39)	(99.62,100.00)	(0.07,0.35)	(83.80,99.79)	(10.61,30.27)

C Theoretical Appendix

Let ℓ_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.²¹

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 ζ_i :

$$\Delta \ell_{i,t+1} \equiv \ell_{i,t} \varepsilon_{i,t+1}. \quad (6)$$

²¹For the rest of the theoretical exposition, we use firm to denote the individual economic entity.

Then the aggregate employment growth rate is simply:

$$\frac{\Delta L_{N,t+1}}{L_{N,t}} = \sum_{i=1}^N \frac{\Delta \ell_{i,t+1}}{L_{N,t}} \quad (7)$$

and aggregate volatility, the variance of aggregate growth is:

$$\sigma_{N,t}^2 \equiv \text{var}_t \left[\sum_{i=1}^N \frac{\Delta \ell_{i,t+1}}{L_{N,t}} \right] = \sum_{i=1}^N \left(\frac{\ell_{i,t}}{L_{N,t}} \right)^2 \zeta_i^2. \quad (8)$$

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

$$\sigma_{N,t}^2 = \sigma^2 \sum_{i=1}^N \left(\frac{\ell_{i,t}}{L_{N,t}} \right)^2 \equiv \sigma^2 h_{N,t}^2. \quad (9)$$

The Herfindahl, in turn, can be rewritten as:

$$h_{N,t}^2 = \frac{\sum_{i=1}^N \ell_{i,t}^2}{\left(\sum_{i=1}^N \ell_{i,t} \right)^2} = \frac{(N^{-2} \sum_{i=1}^N \ell_{i,t}^2)}{(N^{-1} \sum_{i=1}^N \ell_{i,t})^2} = N^{-1} \frac{(N^{-1} \sum_{i=1}^N \ell_{i,t}^2)}{(N^{-1} \sum_{i=1}^N \ell_{i,t})^2}. \quad (10)$$

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

$$h_{N,t}^2 \times N \xrightarrow{a.s.} \frac{\mathbb{E}[\ell^2]}{(\mathbb{E}[\ell])^2}. \quad (11)$$

Definition 1 Consider a random variable Y , a sequence of random variables ζ_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 $N \rightarrow \infty$ is also denoted $\zeta_N \sim a_N Y$ and ζ_N is said to scale like a_N .

Using equation 10 and the scaling definition above, we can characterize the scaling properties of granular shocks when the moments of the size distribution are finite.

Proposition 3 If the size distribution has finite mean and variance, then the size Herfindahl index is such that:

$$h_N^2 \sim \frac{1}{N} \quad (12)$$

and thus the macro variance σ_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 3 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 α is equal or lower than 2, the distribution has undefined variance, and it has both undefined mean and variance when α is equal 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 4 *If firm size is distributed Pareto with shape parameter α , then the size Herfindahl index is characterized by:*

$$h_N^2(\alpha) \propto \begin{cases} 1/N & \text{when } \alpha > 2 \\ 1/\left(\frac{N}{\ln N}\right) & \text{when } \alpha = 2 \\ 1/\left(N^{1-\frac{1}{\alpha}}\right)^2 & \text{when } \alpha \in (1, 2) \\ 1/(\ln N)^2 & \text{when } \alpha = 1 \\ 1 & \text{when } \alpha \in (0, 1) \end{cases} \quad (13)$$

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

Proof: We will prove Proposition 4 using Theorem 3.8.2 in Durrett (2017, p. 167-168), which is as follows:

Theorem 1 *Suppose that X_1, X_2, \dots are i.i.d. with a distribution that satisfies (i) $\lim_{x \rightarrow \infty} P(X_1 > x)/P(|X_1| > x) = \theta \in [0, 1]$ 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 $n \rightarrow \infty$, $(s_N - b_N)/a_N$ converges in distribution to a nondegenerate random variable Y . When $\alpha < 1$, $b_n = 0$.*

Using equations (9) and (10), we can write

$$\sigma_N = \sigma \frac{(\sum_{i=1}^N \ell_i^2)^{1/2}}{(\sum_{i=1}^N \ell_i)} \quad (14)$$

When $\alpha > 2$, a Pareto random variable has finite mean and variance, and so we simply apply Proposition 3. 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:

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

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 14. Since ℓ_i is Pareto distributed with scale parameter equal to 1 and shape parameter equal to α , 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:

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

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,

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

Calibration of Granularity and Aggregate Volatility Simulations Gabaix (2011) provides a calibration of aggregate fluctuations based on the simple model used in this section. He shows that, using equation (9), with a firm volatility of $\sigma = 12\%$ and a Zipf distribution of firm sizes, GDP volatility, σh , is 1.4%.

We replicated this simulation using our 1997 estimates; we simulated 100,000 samples taking 10^6 draws from several distributions, including Zipf, Axtell’s Pareto, and our parametric distributions. We computed the GDP volatility for each simulation and we show the means in Table 19.

Gabaix compares his calibration with a U.S. aggregate volatility of 1.7% and makes the point 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 picture; as we proved earlier, with a Pareto shape parameter below one idiosyncratic shocks do not cancel out in the aggregate producing too much aggregate volatility. If firms are lognormally distributed, the law of large numbers essentially kicks in and idiosyncratic shocks cancel out in the aggregate. Finally, the convolution, our preferred distribution, allows for idiosyncratic shocks to matter in the aggregate, but with a much more diminished role compared to Zipf’s law.

The values in Table 19 should be taken as upper bounds, as the calibration assumes that all firms have the same volatility, whereas larger firms have lower volatilities thus potentially significantly diminishing the aggregate impact of idiosyncratic shocks.

Table 19: Calibration of Aggregate Volatility under Granularity

Zipf’s law	Axtell’s Pareto	Pareto	Lognormal	Mixture	Convolution
1.43%	0.99%	6.63%	0.05%	4.59%	0.38%

Note: We took 100,000 times 10^6 draws from each distribution, computed the aggregate volatility using equation (9), and took the mean. Zipf’s law is a Pareto with shape parameter equal to 1. Axtell’s Pareto is a Pareto with shape parameter equal to 1.057. Pareto, lognormal, mixture and convolution are parametrized using our 1997 estimates for the firm size distribution from Tables 3, 6, and 7.