Looking for power-laws in all the wrong places: estimating firm size distribution tails across countries and datasets

Jesse Tweedle

March 27, 2018

Abstract

This paper applies recent developments in power law estimation ("Power-Law distributions in empirical data", Clauset, A., C. Shalizi, and M. Newman (2009), SIAM Review 51(4), 661–703) to reject or fail to reject the null hypothesis that the firm size distribution is best fit with a power law. I use data from Compustat and OSIRIS on several countries, and confidential microdata on Canadian establishments and firms. I fit power-law, log-normal and power-law with exponential cutoff distributions to each dataset, and test the following two hypotheses: does the data reject the power-law fit? If not, does the data reject a power law fit in favour of an alternative distribution (specifically, one with thinner tails). I find that a power law distribution fits the U.S. firm size distribution for most years, but France and Germany confidently reject the null hypothesis that their firm size distributions are best fit with power laws. Canadian firms, both public and private, reject the power-law null hypothesis, but Canadian establishments do not. As an application, I use the estimated power laws to calculate firm size herfindahls to estimate the implied contribution of idiosyncratic shocks to aggregate volatility in different countries. That the power law distributions produce herfindahl results that run strongly counter to the data is a consequence of a seemingly well-fit distribution failing exactly where it matters in economics—in the top 10 or 20 firms.

1 Introduction

We are always looking for simple laws to explain economic behaviour. The power law is a perfect example; the upper tail of distributions of several economic phenomena seem to obey a scale-free law relating the size and rank of individuals (Gabaix, 2009). The firm size distribution is an especially important application of this law; a scale-free firm size distribution has dramatic consequences for several fields of economics, including studies on gains from trade (Di Giovanni et al., 2011; Di Giovanni and Levchenko, 2012, 2013; Nigai, 2017), as well as idiosyncratic volatility (Gabaix, 2011).

However, the empirical evidence supporting power law distributions is weak. In many applications, the methodology is: (1) eyeball the upper tail of the distribution on a log-log plot of rank vs. size and guess where it starts to look linear; (2) estimate via OLS the slope of the line (on the data above the eyeballed cutoff); (3) claim a high R^2 means a power

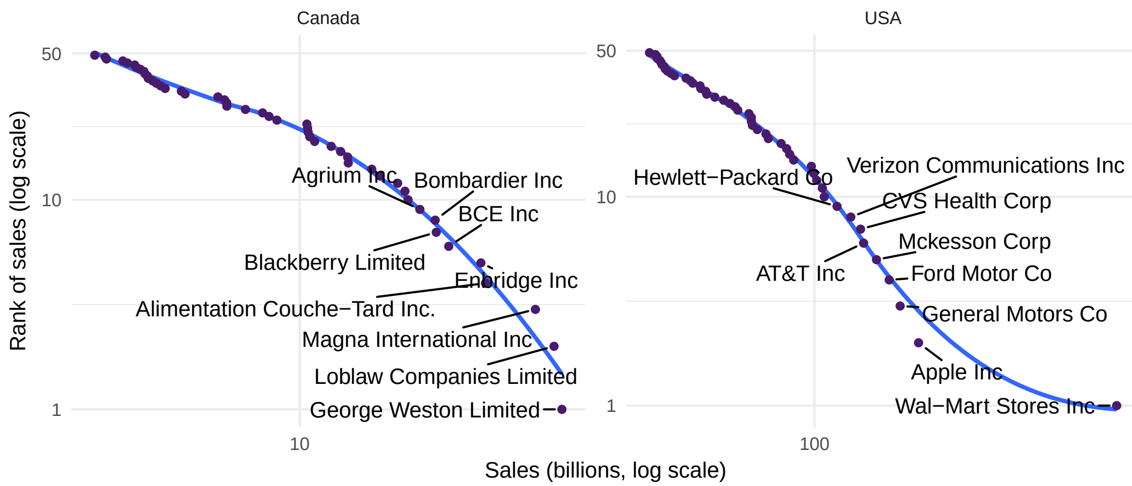

Figure 1: Difference in curvature on rank-size plot for the US and Canada.

Notes: Canadian firm data from OSIRIS, US firm data from Compustat. Petroleum and Finanial firms have been removed. Note also the trouble with public company sales data: George Weston Ltd. has a controlling interest in Loblaw Companies Ltd. and reports Loblaw Co. Ltd.'s sales as its own (see, e.g., page 9 of the George Weston Ltd. Annual report (`http://www.weston.ca/en/pdf_en/gwl_2016ar_en.pdf`), which means Loblaw's $40B sales are reported twice. This double counting skews the relationship a bit; to correct this, I also look at confidential survey and administrative data that does not suffer from this issue.

law fits the data well. The estimated slope is used as the scale exponent for the power-law. For instance, in Figure 1, the shapes of the distributions are clearly very different, but an OLS fit to each plot will give a reasonable scale exponent and a high R^2. One should not conclude that both of these distributions are best fit with a power-law.

Despite the obvious drawbacks noted (even in the papers that use it), this methodology is used to justify many empirical power laws. The OLS methodology is so easy to use, and power laws are so simple and enticing to use in theory, that we don't apply the same statistical rigour that instrumental variables, for instance, would attract. My goal here is to apply recent developments in power law estimation (Clauset et al., 2009; Broido and Clauset, 2018) to reject or fail to reject the null hypothesis that the firm size distribution is best fit with a power law.

I use several sources of firm microdata: Compustat, OSIRIS and confidential microdata sources on Canadian establishments and firms. In each dataset, and each year, I estimate the upper tail cutoff x_{min}, the power law scale exponent α, as well as alternative distributions with thinner tails (the log-normal, and the power law with exponential cutoff). I calculate p-values for each important hypothesis: does the data reject the power law fit? If not, does the data reject a power law fit in favour of an alternative distribution (specifically, one with thinner tails). The conclusions matter for our understanding of firm heterogeneity—how skewed are our firm distributions?

This work contributes to two main strands of literature. First, the estimation of fat tailed distributions. Several methods have been proposed and refined, although all revolve around a rule-of-thumb for finding the cutoff of the upper tail. For distributions that aren't truly power laws, the estimated scale exponent changes non-trivially with the cutoff, which means a researcher can easily draw the conclusions they'd like by varying the cutoff, and justify their choice with the resulting high R^2 given by the OLS estimate. Here, I adopt the much stricter methodology proposed by Clauset et al. (2009); Broido and Clauset (2018) that are used to estimate power laws in scale-free networks. A few estimation studies have changed track from estimating the upper tail to estimating the whole distribution. However, the conclusions of these papers still rely on a power law fitting the best on the upper tail, and then arguing for or against other distributions in the middle and lower tail of the distributions. If the upper tail is not truly a power law, these methods also have room to improve.

Secondly, this work is important for the application of power laws to economics. For an overall review of power laws in economics, see Gabaix (2009). Some of the proposed power laws are more robust than others. The city size distribution, for instance, seems robust to different specifications, as long as one adopts a sprawling definition of city: e.g., the US Census Bureau created 'combined statistical areas' (CSAs) to better represent the size of cities that sprawl over several municipal areas, and power law estimations of this type of city size distributions are more robust to estimation methods (see, e.g., the evolving research on Zipf's law for city sizes, including Gabaix (1999); Eeckhout (2004); Rozenfeld et al. (2011)).

However, research on power laws for firm sizes hasn't received the same attention. Theory suggests that trade increases the skew of the power law tail as size increases, because the most productive firms get access to more and more markets, which further increases their sizes. The data do not seem to support this. Since gains from trade in some models depends on the shape of the firm size distribution (Nigai, 2017; Head et al., 2014; Feenstra, 2018), correctly estimating the upper tail of the distribution is very important. An even stronger motivation comes from research on the microfoundations of aggregate volatility. The argument is that if the firm size distribution is skewed enough, the biggest firms are so big that idiosyncratic shocks to them aren't washed out by random shocks to other firms in the economy (Gabaix, 2011; Acemoglu et al., 2012; Di Giovanni and Levchenko, 2012). Without a power law tail, the argument for idiosyncratic shocks causing aggregate volatility falls apart. Here, the estimation of the power law passes from important to strictly necessary.

I should clarify that even if the data reject a power law for a specific application, it does not mean that the power law is not a useful tool; for instance, it considerably reduces complexity in firm heterogeneity models like Melitz (2003), or models of preferential attachment in social networks (Jackson, 2010). However, the existence and importance and results of models of firm heterogeneity, for example, do not depend on the specific shape of the upper tail of the distribution; the power law just makes the algebra easier.

I find results that are in line with expectations (that a power law distribution fits the U.S. firm size distribution for many years), and some new results. Depending on the specification (e.g., whether petroleum companies like BP are included or not), France and Germany confidently reject the null hypothesis that their firm size distributions are best fit with power laws. Canadian datasets for firms reject power law distributions across the board, but when the unit of observation is an establishment, Canadian size distributions do not reject the power-law hypothesis.

As an application, I use the estimated power laws to estimate the implied contribution of idiosyncratic shocks to aggregate volatility in different countries. The crucial element is the herfindahl of the size distribution; I use the power law to calculate the implied herfindahl and the resulting aggregate volatility, and compare it to the empirical herfindahl and the herfindahl of the alternative tail distributions. As the tail of a dataset deviates farther from a power law, the power law herfindahl deviates much farther from the empirical herfindahl, vastly overstating the contribution to idiosyncratic shocks to aggregate volatility. That the power law distributions produce results that run strongly counter to the data is a consequence of a seemingly well-fit distribution failing exactly where it matters in economics—in the top 10 or 20 firms.

The paper proceeds as follows: Section 2 describes the methodology, Section 3 outlines the datasets used, and Section 4 gives results. Section 5 gives an application of how the estimates can affect economic phenomena, and Section 6 concludes. More estimation and dataset details can be found in Appendices 7 and 8.

2 Methodology

The methodology consists of two parts: (1) for each dataset, estimate each alternative distribution; (2) perform statistical tests to differentiate between possible hypotheses regarding the existence of power-laws. There are three distributions I consider: the power-law, log-normal, and power-law with exponential cutoff. The log-normal is a commonly proposed alternative to the power-law, and the power-law with exponential cutoff is an alternative proposed for degree sequences in social networks. Both have thinner tails compared to power-laws.

2.1 Upper-tail distribution estimation

Given a dataset of firm sizes $X = \{x_i : i \in 1, \dots, N\}$, we need to estimate where the upper-tail $x_{\min}$ begins, and the shape of the distribution above it. We start with estimating the cutoff and scale exponent of the power-law, and then apply that same cutoff to the other distributions to ensure a fair and accurate comparison.

2.1.1 Power-law

A power-law distribution in the upper-tail follows the following density function

$$f(x) = Cx^{-\alpha}, \alpha > 1, x \geq x_{\min} > 0, \tag{1}$$

where α is the scale exponent and C is a constant, and $x_{\min}$ is the value that defines the upper tail. For details on the sources of power-laws in the world of economics, see Gabaix (2009). For our purposes, a power-law means a linear relationship between the size of an individual and the empirical counter-cdf (one minus the empirical cdf) on a log-log plot, *everywhere* in the upper tail. If the top 5 or 10 firms in the data diverge from the linear relationship, then the firm size distribution is not truly a power-law, in the sense that the important implications of power-law distributions don't hold (e.g., the granular hypothesis of aggregate volatility no longer applies).

Given $x_{\min}$, I use the MLE $\hat{\alpha}(x_{\min})$ as the estimate of the scale exponent. Then the estimate $\hat{x}_{\min}$ is the value of $x_{\min}$ that minimizes the Kolmogorov-Smirnov statistic D, the maximum distance between the cdf of the power-law fit and the ecdf, $E(x)$.

$$D = \max_{x \geq x_{\min}} |E(x) - F(x|\hat{\alpha})| \tag{2}$$

We use $\hat{x}_{\min} = \min_{x_{\min}} D$ as the cutoff for all distributions, and $\hat{\alpha} = \hat{\alpha}(\hat{x}_{\min})$ as the estimated power-law scale exponent. For more details, see Appendix 8.

2.1.2 Log-normal

The log-normal distribution is a common alternative to the power-law due to its skewness and association with Gibrat's law. The log-normal density is defined as:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}, \quad x > 0 \tag{3}$$

To compare it directly to the power-law distribution, I truncate the distribution at $x_{\min}$. Write the truncated distribution as:

$$h(x) = \frac{f(x)}{1 - F(x_{\min})} \tag{4}$$

so that the log-normal distribution is only defined in the upper tail and sums to 1 on the interval $[x_{\min}, \infty)$. MLE estimation of the parameters, after using $\hat{x}_{\min}$ from the power-law estimation to compare it fairly to the estimated power-law distribution.

2.1.3 Power-law with exponential cutoff

The power-law with exponential cutoff has a power-law-like tail up to a point, then has a exponential-like tail after that. The power-law is a special case of the power-law with exponential cutoff, which means it, by definition, can't fit worse than a power-law. Nevertheless, except in that special case, it is not scale-free, and thus does not display the same economic properties as a true power-law. Its density is

$$f(x) = Cx^{-\alpha}e^{-\lambda x}, \tag{5}$$

where the constant $C = [e^{-x_{\min}\lambda}\Phi(e^{-\lambda}, \alpha, x_{\min})]$, and $\Phi(z, s, a) = \sum_{i=0}^{\infty} \frac{z^i}{a+i}^{s}$ is the Lerch Phi function. Again, it takes $\hat{x}_{\min}$ as given from the power-law KS statistic minimization.

When a distribution looks like a power-law everywhere but curves down at the end, one usually sees the justification "the distribution is a power-law but for finite-size effects" or "the power-law may hold only over a bounded range", meaning the very tip of the distribution *doesn't* extend to the very largest firms. These distributions are more likely to be power-law with exponential cutoffs than true power-laws, and since we care *most* about the very biggest firms, we would like to be able to statistically differentiate between the two types of distributions.

2.2 Tests

I start with the null hypothesis that the upper tail of the firm size distribution is a power-law, consistent with the literature. If the tests reject that hypothesis for a certain dataset, then that informs our understanding of the firm size distributions across countries or dataset types. There are two relevant tests: first, does the data directly reject the power-law? Second, does the data reject the power-law in favour of an alternative distribution?

2.2.1 Null hypothesis H_0^d (direct): the upper tail is a power-law

Clauset et al. (2009) proposes a semi-parametric bootstrap test to generate p-values for the power-law null hypothesis H_0^d. The idea is to simulate the data as if it were really a power-law, run the estimation procedure again. Simulating this 1000 times gives a null distribution of KS-statistics $Pr(D)$. If D^* is the KS-statistic for the best fitting power-law distribution, then the p-value for this model is defined as the probability of observing, under the null distribution, a KS-statistic at least as extreme as D^*. So $p = Pr(D \geq D^*)$ is the fraction of simulated datasets with KS statistics larger than that of the empirical dataset.

2.2.2 Null hypothesis H_0^a (alternative): the upper tail is fit equally well by a power-law and the alternative distribution

A set of likelihood-ratio tests can distinguish between the power-law and alternative distributions. Given the log-likelihoods of the power-law ($\mathcal{L}_{pl}$) and an alternative distribution ($\mathcal{L}_{alt}$), the test likelihood-ratio test statistic is

$$\mathcal{R} = \mathcal{L}_{pl} - \mathcal{L}_{alt}, \tag{6}$$

where the sign of R, if deemed significantly different from 0, gives evidence for or against the null hypothesis H_0^a, that the data are fit equally well by a power-law and the alternative distribution.

3 Data

I use data on firm sizes from several sources. First, Bureau van Dijk's OSIRIS database contains sales and identifying information on globally listed public companies, including 34,000 listed, 3,500 unlisted and 7800 delisted companies between 1900–2100. Second, the Fundamentals Annual section of the Compustat North America database. Both datasets come from Wharton Research Data Services (WRDS, 2018). I also investigate datasets with public and private firms in Canada. These are the Annual Survey of Manufactures, and the T2-LEAP administrative tax and employment dataset. For each, I use gross sales as a measure of size to be consistent with the other datasets.

Table 1 shows the summary statistics for each country. The US Compustat data is labelled "USA (Compustat)", while every other country is OSIRIS. OSIRIS data typically covers the period 1984–2016, with some exceptions (China's data starts in 1992 and Taiwan in 1995), while the Compustat data starts in 1961. Each country dataset is normalized by the median firm within each year, so the median for each dataset is 1. The data are clearly

Table 1: Summary statistics of firm sizes by country and dataset

Country	Period	N obs.	Mean	3rd Qu.	Max
Australia	2000–2016	510	35.5	19.9	1541.0
Bermuda	2001–2016	313	8.0	7.3	130.9
Canada	1996–2016	610	42.3	26.4	1043.9
Canada (ASM)	1973–1999	32622	9.0	4.0	11728.8
Canada (ASM)	2000–2011	52937	10.0	3.5	16518.5
Canada (ASM, firms)	1973–1999	28057	14.8	3.5	24122.6
Canada (ASM, firms)	2000–2011	47566	12.8	3.1	26131.9
Canada (T2, firms)	2001–2009	1412787	10.8	3.3	14035.5
Cayman Islands	2006–2016	452	7.6	7.5	152.4
China	2000–2016	1351	13.8	7.1	2861.8
France	1998–2016	363	58.4	25.4	2699.7
Germany	1998–2016	340	49.7	21.1	1674.5
Great Britain	1987–2016	685	44.7	18.2	6270.1
India	2000–2016	1288	29.9	14.7	3876.3
Japan	1996–2016	1159	10.6	7.0	567.2
Korea	2008–2016	288	19.5	5.6	2899.7
Malaysia	1996–2016	445	9.9	6.1	1047.3
Taiwan	2007–2016	297	16.0	7.7	762.3
USA	1984–2017	2544	26.3	18.2	3096.1
USA (Compustat)	1961–2014	2974	20.2	13.8	1675.8

Notes: the statistics are averages of all available years. E.g., 'N obs.' is the average number of observations per year. Each dataset is normalized within years by the median firm. Sources of datasets other than OSIRIS are indicated in parentheses. Non-OSIRIS Canadian datasets are confidential microdata. ASM is the Annual Survey of Manufactures (all manufacturing establishments with more than $30,000 in sales. The samples are divided into 1973–1999 and 2000–2011 because of a survey changes. If labelled 'firms', the establishment data are aggregated up to the ultimate parent (firm) level, to better compare with the public firm data sources. T2 is data from all firms in Canada derived from administrative tax records. The sample is much larger and not restricted to manufacturing.

skewed right, with means up to 20 times higher than medians for some countries, and even 4 times higher than the 75th percentiles of the distributions.

Later, I test robustness of the results to removing financial and petroleum firms from the datasets.

4 Results

In this section, I present the results of the power-law estimations, the alternative distribution estimations, and the tests that distinguish between them.

4.1 Power-law estimations

For a power-law to exhibit scale-free behaviour, the estimated scale exponent must satisfy $2 \leq \hat{\alpha} < 3$. A scale parameter less than 2 isn't consistent with a stable distribution, but variations in the data can result in estimating a scale exponent less than 2. Therefore, for each country, one must consider all possible estimates across years to analyze the behaviour of a single country; Broido and Clauset (2018) use a similar approach to analyze separate components of a single network dataset.

Figure 4 plot the densities of scale exponents for each country. In this figure, I only plot estimated exponents that later aren't rejected by the statistical tests. The Zipf law's exponent of 2 is denoted with a vertical red line in each plot. In the left panel, all industries are included in the dataset, and in the right panel, finance, insurance, real estate and petroleum-related industries are removed. The results are striking—across countries, for all

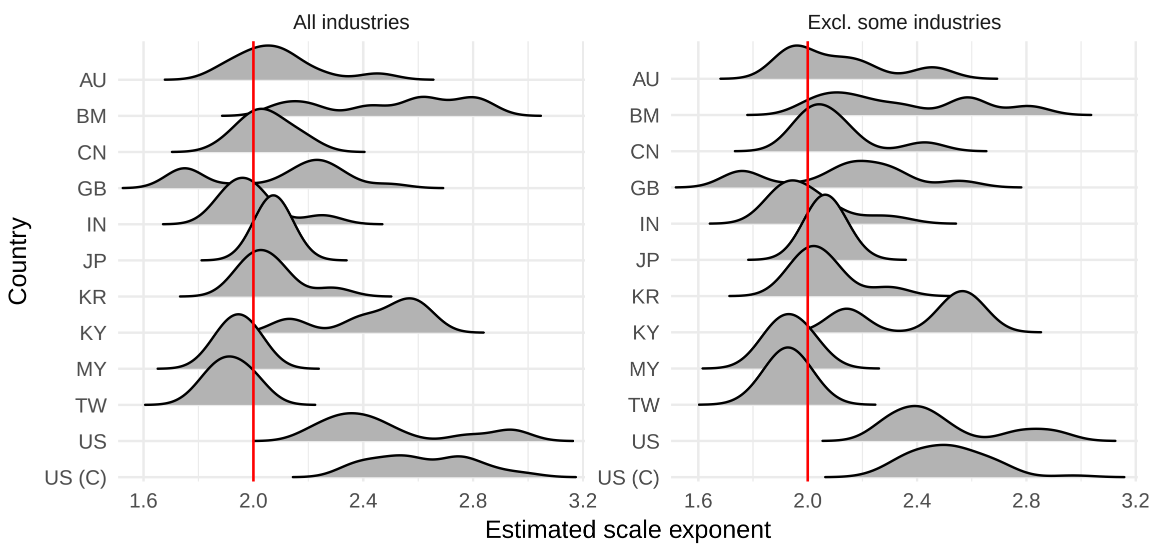

Figure 2: Exponent distributions

Notes: the left panel includes all non-bank firms in all industries. The right panel drops all firms in Finance, Insurance, Real estate and petroleum (NAICS 211, 52 and 53). The 'Zipf' power law exponent of 2 is denoted with a red vertical line in each plot. Estimated scale exponents are only shown for distributions that aren't rejected in favour of an alternative distribution.

distributions that aren't rejected in favour of alternative distributions, the mean estimated scale exponent is around 2, with the U.S. coming in slightly higher at a little less than 2.5. For power-law estimations that aren't rejected by the data, the results bolster theories that require scale-free firm size distributions.

However, there seems to be significant variation in the estimated exponents, and a closer look suggests there is a relationship between the estimated exponent and the number of observations in the tail of the distribution. In Figure 3, I plot the estimated scale exponents vs. the number of observations in the tail of the distribution (which depend on the estimated $\hat{x}_{\min}$). For smaller countries (or a strict $x_{\min}$, the scale exponent tends to be

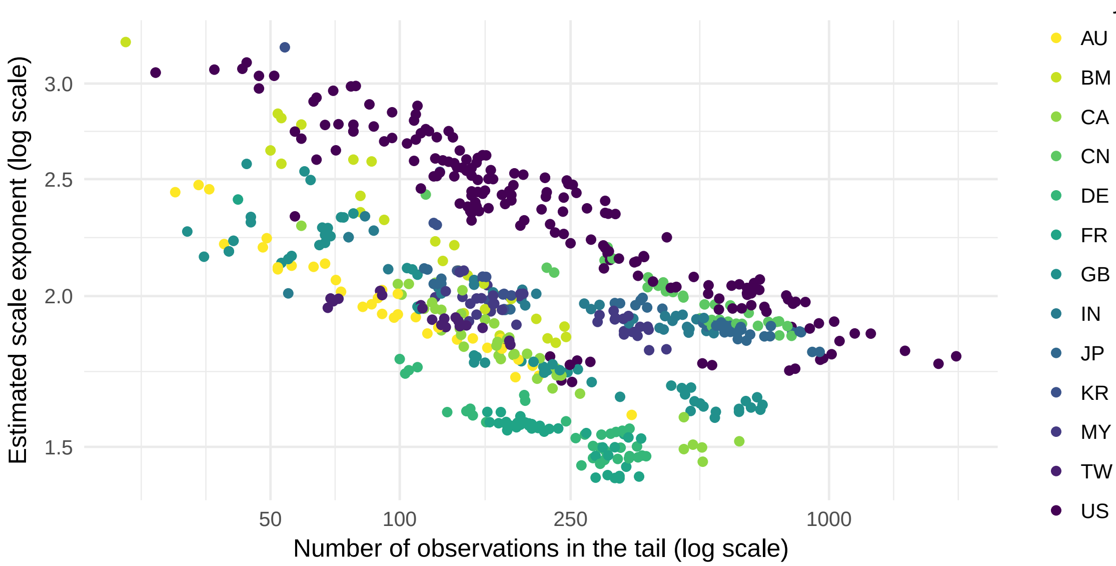

Figure 3: Exponents vs. number of observations

Notes: estimated scale exponents are shown for all datasets, whether or not they are rejected in favour of alternative distributions. There is a clear negative relationship between the number of observations in the tail (which depends on the estimated $\hat{x}_{\min}$ and the estimated scale exponent $\hat{\alpha}$.

a bit larger, while larger countries (and countries with a smaller, less restrictive $x_{\min}$) tend to be closer to satisfying Zipf's law. This contradicts some ideas about the effect of trade on the firm size distribution—theory predicts that smaller countries would have as large or larger power-law exponents for their firm size distributions, because the most productive firms get access to markets much larger than other domestic firms, which skews the firm size distribution more than it otherwise would (Di Giovanni et al., 2011; Di Giovanni and Levchenko, 2012, 2013). It's difficult to get a precise answer to that question here, since the lack of data on private firms may bias our results, but for the datasets listed here, countries that satisfy Zipf's law tend to be larger than others; and that's conditional on the data not rejecting the power-law fit, let alone the smaller countries that we will see that reject the power-law outright. Furthermore, in one case in which the power-law is rejected, I re-estimate the model fits with survey and administrative datasets, confirming the rejected power-law.

4.2 What do the estimated distributions look like?

To get a feel for the shape of each possible distribution, here are visual representations of the differences between the alternatives (including the empirical cdf). For a few examples, I plot the estimated power-law, log-normal and power-law with exponential cutoff distributions,

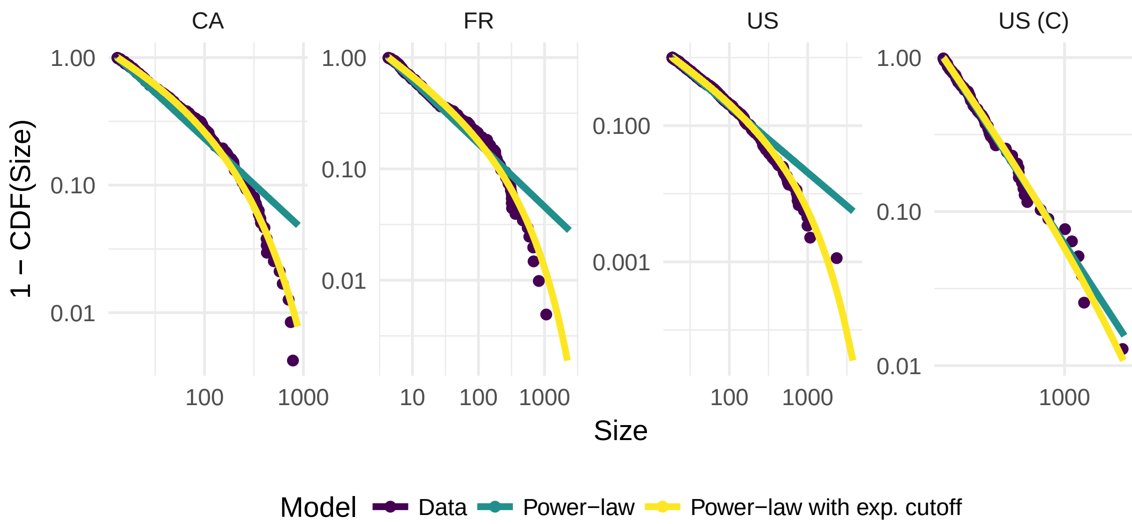

Figure 4: Estimated and empirical distributions

Notes: these figures for the year 2005. CA is Canada, FR is France, US is USA, US (C) is USA data from Compustat. The data are plotted as points, and each estimated distribution is plotted as a line.

along with the empirical cdf. Firm size (log-scale) is on the x-axis, and the counter cdf is on the y-axis $(1 - F(\text{Size}))$. The power-law with exponential cutoff typically fits the data very well. For the US, however, the dataset with more observations (OSIRIS, with 881 observations in the tail) is a power-law with exponential cutoff, while the other dataset (Compustat, with 78 observations in the tail) looks like a power-law. If one restricted the OSIRIS US dataset to a similarly high $x_{\min}$, one might succeed in fitting a power-law equally well as the power-law with exponential cutoff (or log-normal). However, the $x_{\min}$ was chosen to fit a power-law as well as possible; the only problem is there's a lot more data in the lower part of the upper-tail (i.e., closer to $x_{\min}$ than to the maximum size x), and *that* part of the distribution acts more like a power-law (a linear relationship in Figure 4) than the upper part of the upper tail. Keep in mind that $\hat{x}_{\min}$ is chosen to match the empirical cdf $E(x)$, *not* the relationship between $E(x)$ and x.

To assess the fits statistically (instead of visually), I now move to the hypothesis tests.

4.3 Direct and alternative hypothesis tests

The results of the tests, shown in Table 2, seem to strongly separate countries into those with power-law firm size distributions and those that reject power-law firm size distributions. Looking at the direct tests first, the null hypothesis H_0^d is rejected by the median p-value in Canada, Germany, France, Japan, and the USA (OSIRIS dataset). Canadian microdata (as opposed to data on public firms only) give a slightly different answer: when the unit of observation is a firm, the results are consistent with using data on public firms only. However, when the unit of observation is an establishment, the size distributions do not reject the power-law directly or in favour of any alternative distribution I tested. For a more detailed look at Canadian microdata results, see Figure 5.

All other countries do not reject the null hypothesis that the upper tail of the data is generated by a power-law. However, this doesn't answer whether or not there's an alternative distribution that fits the data *better*. For that, we turn to the p-values for $H_0^{\text{log-n}}$ and H_0^{pexp}. The alternative distribution tests show similar results: the same countries reject power-laws for alternative distributions. A few more countries come closer to rejecting power-law distributions for alternatives, but p-values less than 0.1 (e.g., Australia, Bermuda, Great Britain, Cayman Islands, and the USA (Compustat). The other countries still do not reject the null hypothesis that the power-law fits the data as well as the alternative distributions. To make an overall conclusion, I take into account the proportion of years a country rejects each null hypothesis. If 70% or more of the tests reject the null hypothesis, I classify the country as rejecting the power-law distribution. These countries are: Canada, Germany, France, and Japan. Most other countries reject the alternative distribution hypothesis less than 40% of the time, with less rejecting the direct null hypothesis H_0^D. The USA is the only one more difficult to classify. The Compustat dataset rejects the direct hypothesis only 26% of the time, while the OSIRIS dataset rejects it 58% of the time. The proportion rejecting in favour of the power-law with exponential cutoff distribution are closer at 55 and 61%; however, this isn't conclusive evidence, especially with the existing literature supporting the power-law firm size distribution of the USA. One wonders whether a test incorporating the panel nature of the data could help differentiate the alternatives: for instance, can a power-law distribution suffer a few negative shocks at the very tip of the distribution to temporarily thin out the tail.

5 Application

The contribution of idiosyncratic shocks to aggregate volatility depend on the herfindahl of firm sizes in the economy (Gabaix, 2011; Di Giovanni and Levchenko, 2013; Di Giovanni et al., 2014),

$$h = \sqrt{\sum_i^N w_i^2}, \tag{7}$$

where w_i is the weight of firm i in the economy. The bigger the herfindahl, the bigger the contribution that idiosyncratic shocks make to aggregate volatility. The bigger the biggest firms are, the bigger the herfindahl, and finally, a power-law exponent between 2

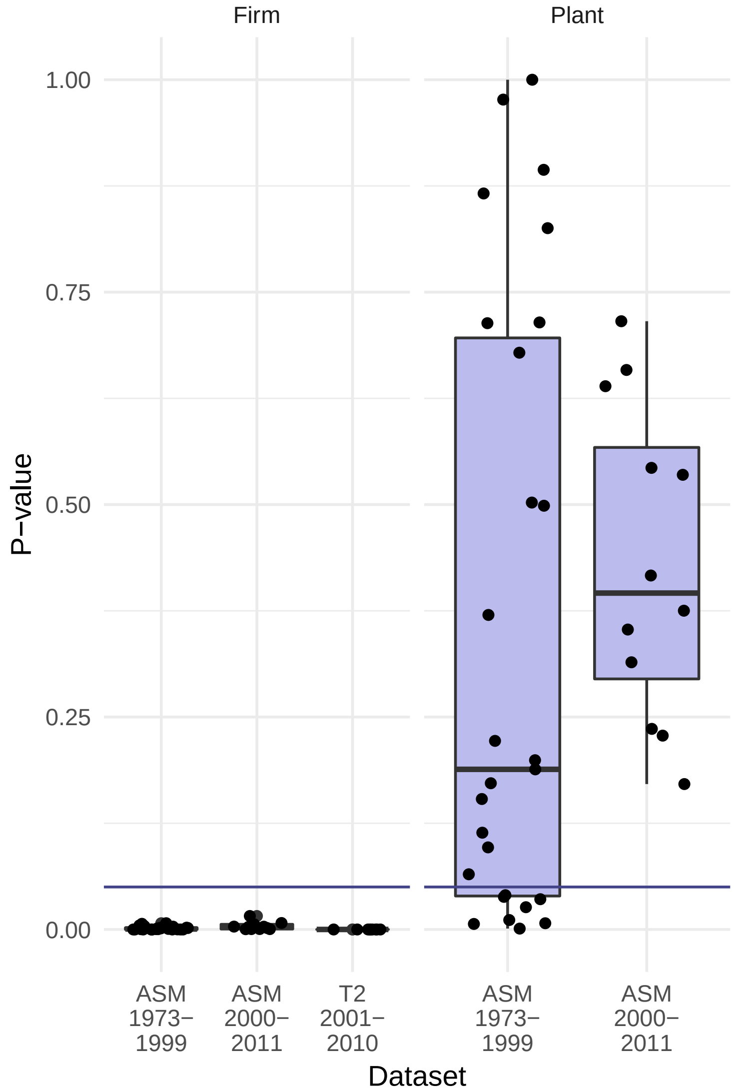

Figure 5: Boxplot of p-values of the test of null hypothesis H_0^{pexp} for each Canadian micro-dataset. Each dot represents a p-value for one year of a dataset.

Notes: These are boxplots of the p-values from the test of power-law versus the alternative hypothesis that the distribution is better fit by a power-law with exponential cutoff (H_a^{pexp}). The three datasets on the left use firms as the units of observation, and definitively reject the null hypothesis for all years. The two datasets on the right use establishments as the units of observation, and do not reject the null hypothesis that the data is fit well by a power-law.

Table 2: Summary of p-values for different tests, by country and dataset.

	Median p-value for:			% rejecting:		
Country	H_a^D	$H_a^{\text{log-n}}$	H_a^{pexp}	H_a^D	$H_a^{\text{log-n}}$	H_a^{pexp}
Australia	0.28	0.16	0.08	33	13	40
Bermuda	0.28	0.17	0.06	14	7	43
Canada	0.00	0.01	0.00	100	84	100
Canada (ASM 73–99)		0.46	0.19		0	26
Canada (ASM 00–11)		0.86	0.40		0	0
Canada (ASM 73–99, firms)		0.18	0.00		0	100
Canada (ASM 00–11, firms)		0.19	0.00		0	100
Canada (T2, firms)		0.00	0.00		100	100
China	0.14	0.20	0.19	33	0	40
Germany	0.00	0.02	0.00	94	94	100
France	0.01	0.01	0.00	100	94	100
Great Britain	0.20	0.16	0.07	46	39	43
India	0.54	0.38	0.29	20	0	13
Japan	0.03	0.02	0.00	74	63	84
Korea	0.34	0.70	1.00	14	0	0
Cayman Islands	0.13	0.18	0.07	44	33	44
Malaysia	0.18	0.18	0.14	21	5	21
Taiwan	0.51	0.41	0.46	0	0	0
USA	0.01	0.01	0.00	58	55	61
USA (Compustat)	0.38	0.37	0.42	29	29	39

Notes: H_a^D is the alternative hypothesis that the data are not well fit by a power-law; $H_a^{\text{log-n}}$ is the alternative hypothesis that the data are better fit with a log-normal; H_a^{pexp} is the alternative hypothesis that the data are better fit with a power-law with exponential cutoff. Sources of datasets other than OSIRIS are indicated in parentheses. Non-OSIRIS Canadian datasets are confidential microdata. ASM is the Annual Survey of Manufactures (all manufacturing establishments with more than $30,000 in sales. The samples are divided into 1973–1999 and 2000–2011 because of a survey changes. If labelled 'firms', the establishment data are aggregated up to the ultimate parent (firm) level, to better compare with the public firm data sources. T2 is data from all firms in Canada derived from administrative tax records. The sample is much larger and not restricted to manufacturing.

and 3 is required for the biggest firms to be big in an economy with a lot of firms. I focus on the last point. If we claim the firm size distribution is best fit by a power-law, the estimated power-law distribution should imply a herfindahl close to the data. If not, the power-law distribution doesn't agree with the main statistic that governs the existence of the microfoundations of aggregate fluctuations.

Therefore, I use each estimated power-law and power-law with exponential cutoff distributions to simulate 1000 datasets for each country and year, and calculate the implied herfindahls. Then I take the mean across all the simulated datasets. The results are in Table 3. The empirical herfindahls are listed in the first column; the simulated herfindahls of the fitted power-law distributions are in the second column (labelled 'Power'), and the simulated herfindahls of the fitted power-law with exponential cutoff distributions are in the third column 'P-exp'. The p-exp distributions are much closer to the empirical herfind-

Table 3: Herfindahls for each predicted distribution and their differences from the data, by country

Country	Herfindahls Data	Herfindahls Power	Herfindahls P-exp	% difference from data Power	% difference from data P-exp
Australia	0.196	0.399	0.200	103.1	2.0
Bermuda	0.162	0.276	0.162	70.4	0.1
Canada	0.124	0.464	0.132	275.5	6.9
Canada (ASM 73–99)	0.100	0.156	0.092	55.9	-8.2
Canada (ASM 00–11)	0.087	0.167	0.100	92.2	15.8
Canada (ASM 73–99, firms)	0.102	0.381	0.094	274.8	-7.9
Canada (ASM 00–11, firms)	0.089	0.343	0.093	284.1	4.3
Canada (T2, firms)	0.019	0.176	0.023	820.6	20.1
Cayman Islands	0.138	0.288	0.137	107.9	−1.1
China	0.210	0.309	0.164	47.5	−21.7
France	0.176	0.643	0.182	265.4	3.6
Germany	0.181	0.635	0.199	250.4	9.7
Great Britain	0.271	0.434	0.210	60.3	−22.4
India	0.198	0.335	0.205	69.0	3.4
Japan	0.110	0.363	0.117	231.3	6.4
Korea	0.549	0.321	0.321	−41.5	−41.4
Malaysia	0.242	0.344	0.183	42.0	−24.3
Taiwan	0.274	0.391	0.285	42.7	3.7
USA	0.120	0.282	0.112	134.4	−7.3
USA (Compustat)	0.112	0.216	0.116	92.5	3.5

Notes: The first three columns are the herfindahls given by the distributions: data, the estimated power-law distribution, and the estimated power-law with exponential cutoff. The last two columns are the percentage differences between the estimated distribution's herfindahl and the data's herfindahl.

ahls, with most being less than 10% different than the data, whereas the implied power-law herfindahls can be almost 3 times higher than the data's herfindahls. The Power and P-exp herfindahls are equally bad in some countries that did not reject the previous hypothesis tests, like Great Britain and China.

In Canadian microdatasets, the power-law with exponential cutoff distributions produce herfindahls much closer to the data than a power-law alone, even though the tests do not reject the fact that the data is represented by a power-law. For example, in Canadian establishment data from 2000–2011, of which all tests do not reject the hypothesis that a power-law generated the data, power-law herfindahls are 92.2% higher than the data on average, and the power-law with exponential cutoff generates herfindahls only 15.8% higher than the data. The results from Canadian microdatasets when the unit of observation is a firm are consistent with other public firm data, with power-law distributions generated herfindahls two to three times higher than the data on average, versus herfindahls of less than 10% in absolute value from power-law with exponential cutoff.

This highlights the importance of estimating distributions rigourously if one is to use the estimated distribution in crucial model calculations later. One should not rely solely on an OLS estimate of a power-law exponent as a justification for and basis of statistics that generate counterfactuals for economic effects and policies.

6 Conclusion

My goal here is to apply recent developments in power law estimation (Clauset et al., 2009; Broido and Clauset, 2018) to reject or fail to reject the null hypothesis that the firm size distribution is best fit with a power law. I use several sources of firm microdata: Compustat, OSIRIS and confidential microdata sources on Canadian establishments and firms. In each dataset, and each year, I estimate the upper tail cutoff x_{min}, the power law scale exponent α, as well as alternative distributions with thinner tails (the log-normal, and the power law with exponential cutoff). I calculate p-values for each important hypothesis: does the data reject the power law fit? If not, does the data reject a power law fit in favour of an alternative distribution (specifically, one with thinner tails).

I find that a power law distribution fits the U.S. firm size distribution for most years, but France, Germany, and Canada confidently reject the null hypothesis that their firm size distributions are best fit with power laws. Results from confidential Canadian microdata support these results: when the unit of observation is a firm, Canadian size distributions consistently reject the hypothesis that the data are generated by a power-law distribution; on the other hand, when the unit of observation is an establishment, the Canadian size distributions cannot reject the power-law null hypothesis.

As an application, I use the estimated power laws to estimate the implied contribution of idiosyncratic shocks to aggregate volatility in different countries. The crucial element is the herfindahl of the size distribution; I use the power law to calculate the implied herfindahl and the resulting aggregate volatility, and compare it to the empirical herfindahl and the herfindahl of the alternative tail distributions. As the tail of a dataset deviates farther from a power law, the power law herfindahl deviates much farther from the empirical herfindahl, vastly overstating the contribution to idiosyncratic shocks to aggregate volatility. That the power law distributions produce results that run strongly counter to the data is a consequence of a seemingly well-fit distribution failing exactly where it matters in economics—in the top 10 or 20 firms.

References

Aban, I. B., M. M. Meerschaert, and A. K. Panorska (2006). Parameter estimation for the truncated Pareto distribution. *Journal of the American Statistical Association 101*(473), 270–277.

Acemoglu, D., V. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The Network Origins of Aggregate Fluctuations. *Econometrica 80*(5), 1977–2016.

Broido, A. D. and A. Clauset (2018). Scale-free networks are rare. pp. 26–28.

Clauset, A., C. Shalizi, and M. Newman (2009). Power-Law distributions in empirical data. *SIAM Review 51*(4), 661–703.

Di Giovanni, J. and A. Levchenko (2012). Country size, international trade, and aggregate fluctuations in granular economies. *Journal of Political Economy 120*(6), 1083–1132.

Di Giovanni, J. and A. Levchenko (2013). Firm Entry, Trade, and Welfare in Zipf's World. *Journal of International Economics 89*(2), 283–296.

Di Giovanni, J., A. Levchenko, and I. Méjean (2014). Firms, Destinations, and Aggregate Fluctuations. *Econometrica 82*(4), 1303–1340.

Di Giovanni, J., A. Levchenko, and R. Rancière (2011). Power Laws in Firm Size and Openness to Trade: Measurement and Implications. *Journal of International Economics 85*(1), 42–52.

Eeckhout, J. (2004). Gibrat's Law for (All) Cities. *American Economic Review 94*(5), 1429–1451.

Feenstra, R. C. (2018). Restoring the product variety and pro-competitive gains from trade with heterogeneous firms and bounded productivity. *Journal of International Economics 110*, 16–27.

Gabaix, X. (1999). Zipf' s Law for Cities: An Explanation. *The Quarterly Journal of Economics 114*(3), 739–767.

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

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

Head, K., T. Mayer, and M. Thoenig (2014). Welfare and Trade without Pareto. *American Economic Review: Papers and Proceedings 104*(5), 310–316.

Jackson, M. O. (2010). *Social and Economic Networks.* Princeton University Press.

Melitz, M. J. (2003). The impact of trade on intraindustry reallocations and aggregate industry productivity. *Econometrica 71*(6), 1695–1725.

Nigai, S. K. (2017). A Tale of Two Tails: Productivity Distribution and the Gains from Trade. *Journal of International Economics 104*, 44–62.

Rozenfeld, H. D., D. Rybski, X. Gabaix, and H. A. Makse (2011). The Area and Population of Cities: New Insights from a Different Perspective on Cities. *The American Economic Review 101*(5), 2205–2225.

WRDS (2018). Wharton Research Data Services.

7 Appendix–Robustness

To test robustness of the results to different specifications, I estimate and test all distributions again after removing petroleum and FIRE industries (fire, insurance and real estate). The results are very consistent with the results in the text, suggesting the shape of the firm size distribution doesn't depend on excluding certain types of firms.

7.1 The truncated Pareto distribution

The truncated Pareto distribution is a common alternative proposed to limit the upper tail of the size distribution.

$$f(x) = Cx^{-\alpha}, \alpha > 1, \infty \geq x_{\max} \geq x \geq x_{\min} > 0, \tag{8}$$

However, I did not find this distribution to fit all that well. Figure 6 reproduces Figure 4 with an estimated truncated Pareto distribution. The truncated Pareto distribution does poorly compared to the power-law with exponential cutoff, especially when the data are closer to a power-law (e.g., in the US).

8 Appendix—Theory

This appendix describes the empirical methodology in more detail. Much of this is also described in detail for a different application in Clauset et al. (2009); Broido and Clauset (2018).

Table 4: Summary of p-values for different tests, by country and dataset.

	Median p-value for:			% rejecting:		
Country	H_a^D	$H_a^{\text{log-n}}$	H_a^{pexp}	H_a^D	$H_a^{\text{log-n}}$	H_a^{pexp}
Australia	0.33	0.25	0.19	20	7	33
Bermuda	0.37	0.21	0.08	0	0	27
Canada	0.00	0.02	0.00	94	78	100
China	0.10	0.13	0.09	47	7	47
Germany	0.00	0.02	0.00	94	94	100
France	0.00	0.01	0.00	94	94	94
Great Britain	0.29	0.30	0.48	39	25	29
India	0.36	0.34	0.27	13	0	7
Japan	0.04	0.03	0.00	68	63	84
Korea	0.30	0.64	1.00	0	0	0
Cayman Islands	0.07	0.08	0.02	50	38	75
Malaysia	0.23	0.19	0.17	19	0	6
Taiwan	0.58	0.45	0.53	0	0	0
USA	0.01	0.02	0.00	65	55	68
USA (Compustat)	0.24	0.39	0.47	32	29	39

Notes: H_a^D is the alternative hypothesis that the data are not well fit by a power-law; $H_a^{\text{log-n}}$ is the alternative hypothesis that the data are better fit with a log-normal; H_a^{pexp} is the alternative hypothesis that the data are better fit with a power-law with exponential cutoff. Sources of datasets other than OSIRIS are indicated in parentheses.

8.1 Distributions

8.1.1 Power-law distribution

A power-law of firm sizes above $x_{\min}$ follows this distribution:

$$f(x) = Cx^{-\alpha}, \alpha > 1, x \geq x_{\min} > 0, \tag{9}$$

where α is the scale exponent and C is a constant, and $x_{\min}$ is the value that defines the upper tail. On a log-log scale, this has the form

$$\log f(x) = \log C - \alpha \log x \tag{10}$$

which leads one to suggest OLS as an appropriate method to estimate α, after guessing an appropriate $x_{\min}$.

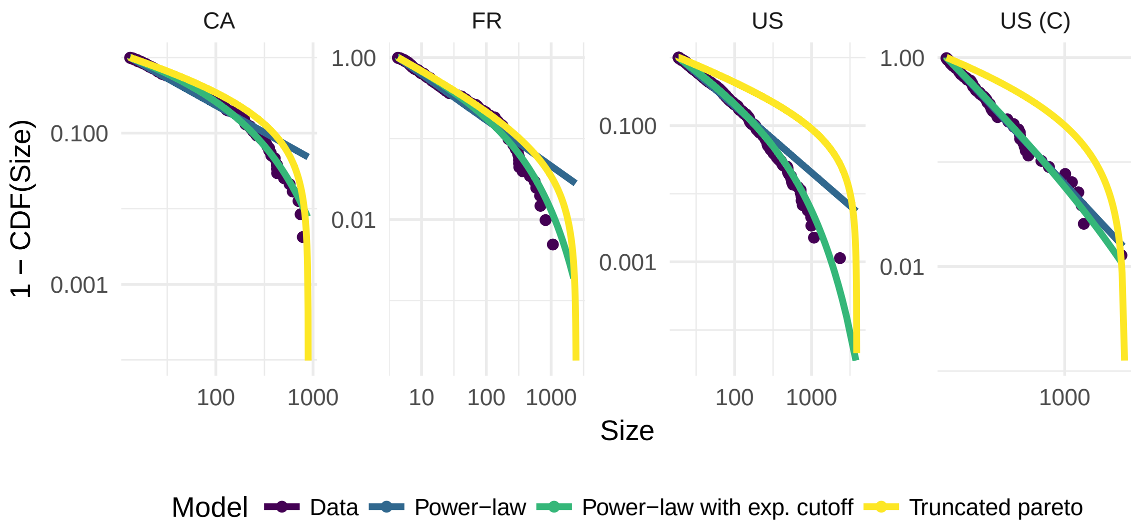

Figure 6: Estimated and empirical distributions, with truncated Pareto

Notes: these figures for the year 2005. CA is Canada, FR is France, US is USA, US (C) is USA data from Compustat. The data are plotted as points, and each estimated distribution is plotted as a line. The truncated Pareto distribution is estimated MLE as described in Aban et al. (2006).

8.1.2 Log-normal distribution

The log-normal distribution is another alternative distribution that can have heavy tails that also happens to be consistent with Gibrat's law:

$$f(k) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}, x > 0 \tag{11}$$

Then write the distribution truncated below at $x_{\min}$ as

$$h(x) = \frac{f(x)}{1 - F(x_{\min})} \tag{12}$$

so that the log-normal distribution is only defined in the upper tail and sums to 1 on the interval $[x_{\min}, \infty)$. MLE estimation of the parameters, after using $x_{\min}$ estimated as if it were a power-law distribution, to compare it correctly.

8.1.3 Power-law with exponential cutoff

$$f(x) = [e^{-x_{\min}\lambda}\Phi(e^{-\lambda}, \alpha, x_{\min})]x^{-\alpha}e^{-\lambda x}, \tag{13}$$

where $\Phi(z, s, a) = \sum_{i=0}^{\infty} \frac{z^i}{a+i}^s$ is the Lerch Phi function.

8.2 Fitting the model, estimating $x_{\min}$ and α

Given $x_{\min}$, one can estimate the scale exponent α via maximum likelihood. A typical method to pick $x_{\min}$ is to plot rank-size on a log-log plot, and eyeball where the upper tail of the distribution "starts to look linear". In contrast, we use the Kolmogorov-Smirnov (KS) minimization method described in Clauset et al. (2009) and Broido and Clauset (2018).

The KS method selects the $x_{\min}$ that minimizes the maximum difference in absolute value between the empirical cumulative distribution (ecdf) $E(x)$ on the observed tail $x \geq x_{min}$ and the cdf of the best fitting power-law $F(x|\hat{\alpha})$ on those same observations. The $\hat{\alpha}$ is estimated via MLE as described above, given the $x_{\min}$ of the current step of the KS method. The KS statistic is defined as

$$D = \max_{x \geq x_{\min}} |E(x) - F(x|\hat{\alpha})| \tag{14}$$

$\hat{x}_{\min}$ is the value that minimizes D, and $\hat{\alpha}$ is the corresponding MLE estimate given $\hat{x}_{\min}$.

8.2.1 Testing goodness-of-fit

The power-law-fitting method will estimate $(\hat{x}_{\min}, \hat{\alpha})$ for any distribution, whether or not the data is from a power-law. To assess the fit, I estimate the p-value with a standard semi-parametric bootstrap approach (Clauset et al., 2009; Broido and Clauset, 2018).

Given firm size data, of which n_{tail} are $x \geq \hat{x}_{\min}$, with MLE $\hat{\alpha}$, I generate a synthetic dataset as follows. For each of n synthetic values, with probability n_{tail}/n I draw a random number from the fitted power-law model, with parameters $\hat{x}_{\min}$ and $\hat{\alpha}$. Otherwise, I choose a value uniformly at random from the empirical distribution below the upper tail, $x < \hat{x}_{\min}$. After n draws, this produces a synthetic dataset that closely follows the empirical distribution below $\hat{x}_{\min}$ and follows the fitted power-law model at and above $\hat{x}_{\min}$.

Then applying the previously defined power-law fitting procedure yields the null distribution of KS-statistics $Pr(D)$. Let D^* denote the value of the KS-statistic for the best fitting power-law model for the empirical distribution. The p-value for this model is defined as the probability of observing, under the null (power-law) distribution, a KS-statistic at least as extreme as D^*. Hence, $p = Pr(D \geq D^*)$ is the fraction of synthetic datasets with KS statistic larger than that of the empirical dataset. Following standard practice, if $p < 0.1$, I reject the power-law as a plausible model of the distribution, and if $p \geq 0.1$, then I fail to reject the model. Failing to reject does not imply the model is correct, only that it is a plausible data generating process.

8.3 Likelihood-ratio tests

Given two candidate distributions that fit the data, I use likelihood-ratio tests to differentiate between them. Let $\mathcal{L}_F$ be the log-likelihood of the fit of distribution F, where F could be pl (power-law), log-n (log-normal), or p-exp (power-law with exponential cutoff). The likelihood-ratio statistic (LRT) is given by the difference between the log-likelihood of the power-law and the log-likelihood of the alternative distribution, $\mathcal{R} = \mathcal{L}_{\text{pl}} - \mathcal{L}_{\text{alt}}$.

When $\mathcal{R} > 0$, the power-law is a better fit to the data, and when $\mathcal{R} < 0$, the alternative is a better fit to the data. When $\mathcal{R} = 0$, the data cannot distinguish between the two models. From here, I calculate a p-value against the null model of $\mathcal{R} = 0$, and reject the null hypothesis if $p < 0.05$ and interpret the sign of $\mathcal{R}$ as evidence for one distribution over another. A p-value of 0.05 is slightly more strict than the p-value of 0.1 used in Broido and Clauset (2018), but is more consistent with existing economics research on power-laws, and more conservative at the same time.