DEM Discussion Papers

Department of Economics and Management

2014/03

Where Gibrat meets Zipf: Scale and Scope of French Firms

Marco Bee, Massimo Riccaboni and Stefano Schiavo

ISSN 2282-2801 DEM discussion papers [online]

UNIVERSITÀ DEGLI STUDI DI TRENTO

DEM DISCUSSION PAPER

ISSN 2282-2801 DEM Discussion Papers [online]
Università degli Studi di Trento

Guidelines for authors

Papers may be written in Italian or in English. Faculty members of the Department must submit to one of the editors in pdf format. Management papers should be submitted to R. Gabriele. Economics Papers should be submitted to L. Andreozzi. External members should indicate an internal faculty member that acts as a referee of the paper.

Typesetting rules:

1. papers must contain a first page with title, authors, abstract, keywords and codes. Page numbering starts from the following page;
2. a template is available upon request from the managing editors.

Where Gibrat meets Zipf: Scale and Scope of French Firms

Marco Bee* Massimo Riccaboni† Stefano Schiavo‡

April 1, 2014

Abstract

The proper characterization of the size distribution of business firms represents an important issue in economic literature, with the most common reference distributions being the lognormal and the Pareto varieties. This analysis is related to some methodological issues that are rarely properly addressed in applied work, and may significantly affect the results: the major difficulties arise from low power of the tests caused by limited sample size and the common practice of binning the data. In this paper we contribute to this body of literature by analyzing the size distribution of *all* French companies, strongly rejecting the hypothesis that it is a Pareto distribution. Moreover, we argue that the lognormal distribution is a more reasonable first-cut benchmark for the entire population of firms. This is especially true for single-product firms, while we show the emergence of a Zipf tail for the class of multi-product companies. Our findings are in strong agreement with some recent theoretical contributions, which predict that the size distribution depends on a set of industry specific determinants.

JEL Codes: C46, L11, L25

Keywords: Firm size distribution; multi-product firms; Pareto; Zipf's law; lognormal

*Department of Economics and Management, University of Trento. Via Inama, 5 - 38122 Trento (Italy); e-mail: marco.bee@unitn.it

†IMT Institute for Advanced Studies Lucca & DMSI, KU Leuven. Piazza S. Ponziano, 6 - 55100 Lucca (Italy); e-mail: massimo.riccaboni@imtlucca.it

‡School of International Studies and Department of Economics and Management, University of Trento & OFCE-DRIC. Via T. Gar, 14 - 38122 Trento (Italy); e-mail: stefano.schiavo@unitn.it

1 Introduction

At least since the work of Gibrat (1931), a great deal of attention has been devoted to the investigation of the size distribution of business firms and the mechanics through which it is determined (Luttmer, 2011). In his seminal work, Gibrat (1931) found that French establishments were lognormally distributed, and postulated a process of proportional growth capable of replicating that shape. A few decades later, Simon and Bonini (1958) showed that a Pareto distribution provides a better fit to the upper tail of the largest firms. Since then, the lognormal and the Pareto (or power-law) distributions have been considered to be the two foremost candidate distributions for firm size (see Sutton, 1997; Mitzenmacher, 2004; de Wit, 2005, for broad reviews of the topic). More recently, a particular type of power-law distribution, namely Zipf's law, has come to be considered the best first-cut benchmark for firm size and many other empirical phenomena (Simon, 1955; Axtell, 2001; Luttmer, 2007; Gabaix, 2009).

Ijiri and Simon (1974) noticed that the data feature systematic departures from power-law behavior, as most firm size distributions depict a concave shape. Due to the lack of data on the whole distribution of firms, early works by Simon and co-authors analyzed only the largest (listed) companies (i.e. Fortune 500). Axtell (2001), who represents the main reference among recent works, manages to overcome the limitation in the representativeness of the data by using information taken from the US Census, thus looking at the entire distribution of firms. He concludes that Zipf's law provides the best representation of the data "all the way down to the smallest sizes" (Axtell, 2001, p. 1819), even though (as noted by Axtell himself) a closer inspection reveals the presence of departures similar to those highlighted by Ijiri and Simon (1974), compatible with a lognormal distribution. The concave shape of the size distribution has been further confirmed by Rossi-Hansberg and Wright (2007), who use the same data analyzed by Axtell (2001). This is part of a stream of literature that claims there is no universal functional form for the size distribution of firms, as it depends rather on specific industry characteristics such as production function, innovation, financial constraints, and firm selection (Jovanovic, 1982; Klette and Kortum, 2004; Cabral and Mata, 2003; Rossi-Hansberg and Wright, 2007; Luttmer, 2007).

The typical strategy adopted in this literature considers one candidate distribution (Zipf, Pareto or, less frequently, lognormal) and performs goodness of fit analyses. Replicating existing results is often difficult because of the limitations in accessing official data, and only a few studies investigate the universe of firms in the US (Axtell, 2001; Rossi-Hansberg and Wright, 2007; Luttmer, 2011) or other countries (see Cabral and Mata, 2003 for Portugal, and Eaton, Kortum, and Kramarz, 2011 or Garicano, LeLarge, and Van Reenen, 2013 for France).

This paper studies the behavior of all French firms using multiple statistical tests, while letting the data speak as much as possible. This is particularly important as discriminating between a lognormal and a Pareto upper tail is extremely difficult from a statistical point of view (Malevergne, Pisarenko, and Sornette, 2009; Bee, Riccaboni, and Schiavo, 2011). Our results provide new evidence in support of the recent theoretical models on the determinants of the size distribution of business firms.

In fact, we find that the whole data distribution is neither a Zipf nor a Pareto model, and the lognormal works better as an approximation of the entire size distribution of French firms. Moreover, when we discriminate across firms based on their scope (e.g. number of products/markets served) we see that the size distribution changes and its upper tail converges to a Pareto distribution with a shape parameter approaching 1 from above (Zipf's law) for highly diversified companies.

The rest of the paper is organized as follows: the next section analyzes the most important methodological issues relative to the identification of a power-law behavior in the data and the practice of binning the observations, which may significantly alter the results. Section 3 presents the main empirical outcomes and investigates the size distribution of different groups of firms. Finally, Section 4 discusses some concluding remarks.

2 Methodological issues

The empirical literature dealing with firm size distribution is characterized by a number of often overlooked methodological issues that may have a significant impact on the results. The most important ones concern the power of the tests used to distinguish among different candidate distributions, especially when the sample size is small, the widespread practice of binning the data, and whether the power-law behavior refers to the upper tail or to the whole distribution.

A substantial body of literature addresses the problem of discriminating between power-law (Pareto) and lognormal tail behavior. The two distributions are mathematically different, but only in the limit (Perline, 2005), so that for finite sample size the tests often have low power. Given these premises, several tests have been developed. Here we follow Bee, Riccaboni, and Schiavo (2013) and show the results obtained with the Uniformly Most Powerful Unbiased (UMPU) test developed by Del Castillo and Puig (1999) and used by Malevergne, Pisarenko, and Sornette (2011); the Maximum Entropy (ME) test by Bee, Riccaboni, and Schiavo (2011); the test proposed by Gabaix and Ibragimov (2011, GI henceforth).

The UMPU test exploits the fact that the logarithm of a truncated lognormal is truncated normal, and the logarithm of a Pareto is exponential. Del Castillo and Puig (1999) have established that the likelihood ratio test for the null hypothesis of exponentiality against the alternative of truncated normality is given by the clipped sample coefficient of variation $\bar{c} = \min\{1, \hat{\sigma}/\hat{\mu}\}$, where μ and σ are the parameters of the truncated normal. for the Weibull with shape parameter equal to 1, UMPU is completely unreliable. A case that illustrates this point is the aggregate city size distribution studied below (see Sect. 3.1).

The ME approach starts from the assertion that the Pareto and the lognormal distributions are two special instances of the Maximum Entropy (ME) density. The latter is obtained by maximizing Shannon's information entropy under k moment constraints $\mu^i = \hat{\mu}^i$ $(i = 1, \ldots, k)$, where $\mu^i = \mathrm{E}[T(x)^i]$ and $\hat{\mu}^i = \frac{1}{n}\sum_j T(x_j)^i$ are the i-th theoretical and sample moments, n is the number of observations, and T is the function defining the characterizing moment.[1] The solution (that is, the ME density) takes the form $f(x) = e^{-\sum_{i=0}^{k} \lambda_i T(x)^i}$. If $T(x) = x$, the logarithm of the Pareto (i.e., the exponential) distribution is an ME density with $k = 1$, whereas the logarithm of the truncated lognormal (i.e., the truncated normal) is ME with $k = 2$. It can be shown that a log-likelihood ratio (llr) test of the null hypothesis $k = k^*$ against $k = k^* + 1$ is given by

$$\mathrm{llr} = -2n \left(\sum_{i=0}^{k^*+1} \hat{\lambda}_i \hat{\mu}^i - \sum_{i=0}^{k^*} \hat{\lambda}_i \hat{\mu}^i \right).$$

The llr test is asymptotically χ^2_1 and is optimal (Cox and Hinkley, 1974; Wu, 2003).[2]

More generally, the ME method is a powerful non-parametric approach to density estimation: when the whole distribution is of interest, it can be used for fitting the best approximating density, with the optimal k found by the llr test (Bee, 2013).

Finally, the GI test is based on the following procedure. Estimate by OLS the regression

$$\log\left(r - \frac{1}{2}\right) = \text{constant} - \alpha \log(x_r) + q[\log(x_r) - \gamma]^2, \tag{1}$$

where r is the rank of the observation, α is the Pareto shape parameter, q is the quadratic deviation from a Pareto distribution, x_r is the r-th order statistic and

$$\gamma \stackrel{\text{def}}{=} \mathrm{cov}(\log(x_r)^2, \log(x_r))/2\mathrm{var}(\log(x_r))$$

is a recentering term needed for guaranteeing that α is the same whether the quadratic term is included or not. Asymptotically, for the Pareto distribution, $q = 0$, so that a large value of $|q|$ points towards rejection of the null hypothesis of power-law. Gabaix and Ibragimov (2011) show that, under the null of a Pareto distribution, the statistic

[1] The two most common cases are $T(x) = x$ and $T(x) = \log(x)$, corresponding respectively to arithmetic and logarithmic moments.

[2] The routines for implementing the UMPU and ME tests are available at `https://sites.google.com/site/ sschiavo7788/home/software`.

$\sqrt{2n}q_n/\alpha^2$ converges to a standard normal distribution, which can therefore be used to find the critical points of the test.

The size distribution of firms has often been tested using binned data (see for instance Axtell, 2001; Di Giovanni, Levchenko, and Rancière, 2011). Axtell (2001) uses US firm sizes measured by receipts in dollars (US Census Bureau data for 1997, consisting in 5 541 918 observations). Data are tabulated in successive bins of increasing size in powers of three, so that bins are equally spaced in logarithmic scale. Using the 7 bins obtained in this way, an OLS regression in doubly logarithmic scale suggests an approximate Zipf distribution ($\alpha = 0.994$). Di Giovanni, Levchenko, and Rancière (2011) perform a similar analysis considering data based on the mandatory reporting of firms' income statements to tax authorities in France. The dataset contains 2 182 571 firms, but the authors only consider firms whose annual sales are above 750 000 euros. They employ both binned and raw data: even though the authors say that "in practice the three estimators deliver remarkably similar results" (Di Giovanni, Levchenko, and Rancière, 2011, p. 46), the estimates of the Pareto shape parameter given by the first two methods, which are based on binned data, are significantly different from the third method, which uses the original observations.[3]

In general, binning the observations in a sample implies a loss of information with respect to the original sample. Intuitively, the reason is that, after binning, we only know how many observations are included in a certain interval, but not their exact location. One can think of the original "raw" observations as being masked by the binning process. In applied statistics, binned data are indeed used only when the original observations are not available, either because they were not recorded or because a measuring instrument produces quantized data, for example when it has coarse resolution compared to the variance of the measured values.

More formally, and focusing on the problem at hand, the (adjusted) frequency used by Axtell (2001) and Di Giovanni, Levchenko, and Rancière (2011), located at the geometric mean of the bin endpoints, is not a sufficient statistic for the Pareto shape parameter, and this results in a loss of information. Obviously, as the number of bins gets smaller, the loss of information increases, because all intervals become wider. Hence, any statistical inference procedure based on binned data produces less reliable results than the same procedure based on the actual values of the observations.

Given that the loss of information is negatively related to the number of bins, the inferential results are sensitive to the choice of this number. In general, the largest possible number of bins should be employed. Thus, when the original observations are available,

[3] The three estimates reported by Di Giovanni, Levchenko, and Rancière (2011, Table 1) are 1.017 (0.032), 1.019 (0.031), 0.825 (0.004), with standard errors in parentheses. Given the asymptotic normality of the Gabaix (2009) estimator, the hypothesis that the point estimate obtained with the third method is equal to 1 is rejected at any significance level.

one should use them instead of the binned data, however obtained. As mentioned above, the impact of the binning procedure is clear from the difference in the results obtained by Di Giovanni, Levchenko, and Rancière (2011, Table 1) with binned and raw observations. Finally, a technical problem arises when applying formal tests aimed at finding the starting point of the alleged Pareto tail. If the data are binned, the null distribution is different from the null distribution hypothesized by the tests. In particular, the former is discrete and the latter is continuous. Thus, the tests cannot be used with binned data.

Recently, a detailed analysis of the statistical issues related to tests for power-law behavior with binned data has been performed by Virkar and Clauset (2014). They propose to identify the power-law threshold via a test based on the Kolmogorov-Smirnov goodness-of-fit statistic and to assess the plausibility of a power-law for the tail of the distribution by means of a bootstrap-based test (see Virkar and Clauset, 2014, pp. 8-12, for details). Their conclusion is in line with the remarks above: "the common practice of identifying and quantifying power-law distributions by the approximately straight-line behavior of a histogram on a doubly logarithmic plot should not be trusted: such straight-line behavior is a necessary but not sufficient condition for true power-law behavior. Furthermore, binned data present special problems because conventional methods for testing the power-law hypothesis could only be applied to continuous or integer-valued observations." (Virkar and Clauset, 2014, p. 26).

3 Empirical results

In this section we investigate the properties of the distribution of firm size for the universe of French firms using an approach with methodological foundations that are more solid than many of the existing empirical studies. First, we look at the entire distribution of firms and compare the performance of a lognormal and a power-law against the data. Second, we show that the practice of binning observations does introduce a bias in the results and significantly affects the conclusions derived from the data in the specific case under scrutiny. Third, we concentrate on the upper tail of the distribution and investigate the presence of power-law behavior using the three different methodologies presented in Section 2 above. Last, we look at the relationship between the scope of firms and the properties of the size distribution, providing support for a number of recent models that link the emergence of Zipf's law in the upper tail to firm- and industry-specific characteristics.

The analysis exploits a comprehensive dataset covering the universe of French firms: the data are analogous to those used in Eaton, Kortum, and Kramarz (2011) and have been used elsewhere as well (e.g. Garicano, LeLarge, and Van Reenen, 2013). The data on total revenues that we use to measure firm size are taken from the FICUS (*Fichier complet de*

Système Unifié de Statistique d'Entreprises) database maintained by the French National Statistical Office (INSEE). We focus on the year 2003 (although the choice of year is actually irrelevant in terms of the results), and have information on more than 2 million firms, excluding the very few cases in which a firm reports total revenues equal to zero from the analysis. To investigate the relationship between firm scope and size distribution (Section 3.2 below), we exploit the information collected by the French Customs, which reports values, destinations and product classes of export flows involving French firms. Our definition of a product is therefore a 6-digit code within the Harmonized System classification. We use the number of different products exported and/or the number of foreign destinations that served as a proxy for firm scope. No comparable information is available for domestic transactions. Firms exporting less than 1 000 euros outside the EU, or less than 100 000 euros within the EU are not required to report their transactions; other than that, the dataset is comprehensive. Furthermore, we can link the two sources of information by means of a unique firm identification number.

3.1 The size distribution of French firms

Figure 1 represents the histogram of the logarithms of the data along with the normal (logarithm of the lognormal), the exponential (logarithm of Pareto) and the best fitting ME distributions. We use observations larger than 14 000 euros, as below this threshold the distribution is very irregular. Only the smallest 3.7% of the observations are discarded in this way. The optimal ME distribution has $k = 11$; hence, neither the Pareto ($k = 1$) nor the lognormal ($k = 2$) distributions are good approximations, although it is clear from the graph that the lognormal one is "closer" to the true distribution than the Pareto distribution.

The upper panel of Figure 2 shows the distribution of US and French firm sizes. The graph is the log-log plot of the counter-cumulative distribution function of bin frequencies. For comparison purposes, we use the same number of bins (7) used by Axtell (2001), and the bins are equally spaced on a logarithmic scale. The results look qualitatively similar. Nevertheless, the slope of the regression line is considerably smaller in absolute value for French firms. The residuals from the linear regressions are shown in the lower panel of the same figure: they look far from random, showing instead a clear cyclical pattern. These doubts are confirmed by performing the bootstrap-based test of power-law behavior with binned data developed by Virkar and Clauset (2014): at the 5% level, the null hypothesis of power-law is rejected for both the US data (p-value = 0.02) and the French data (p-value smaller than 0.001).

Two remarks should reinforce this conclusion. First, we ran the corresponding test (see Clauset, Shalizi, and Newman, 2009) with the raw observations of the French database. The test seems to be "less conservative", in the sense that it always results in a longer

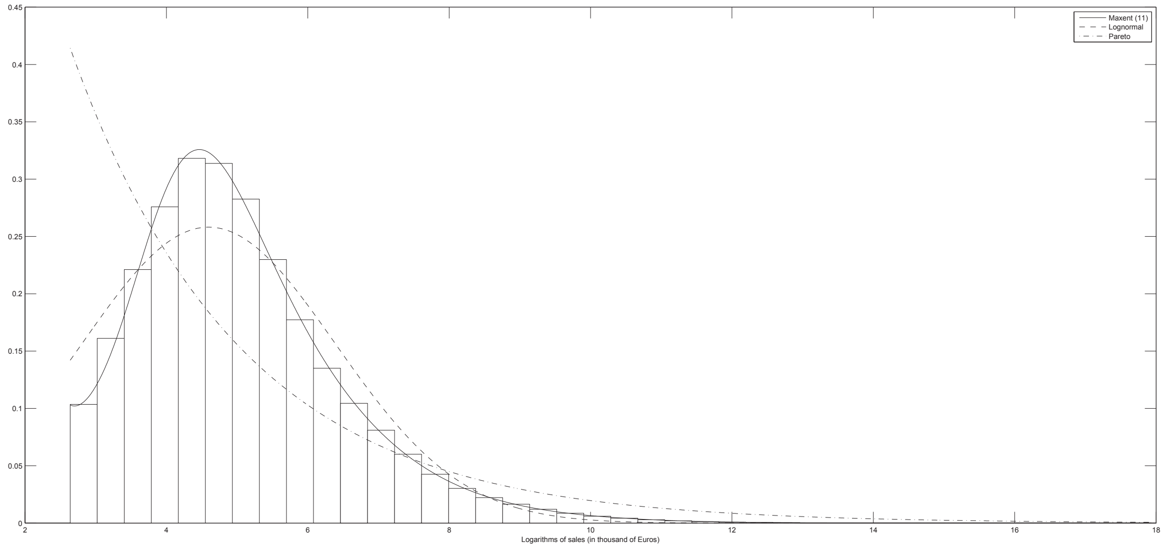

Figure 1: The size distribution of French firms, along with the Pareto, lognormal and best-fitting ME densities. Sales are in thousand of euros.

power-law tail than the other three tests (see Table 2 below). Yet, with binned data it does not find a power-law tail. Second, Virkar and Clauset (2014) note that "for small values of n, or for a small number of bins[...], the empirical distribution may closely follow a power-law shape, yielding a large p[-value], even if the underlying distribution is not a power law". Thus, whereas the risk of a false positive (the test finds a power-law when the true distribution is not power-law) is high, a false negative does not seem to be a major concern.

According to the discussion in Sect. 2, we should trust the tests that use raw data rather than binned data more. Thus, we carried out the UMPU, ME and GI tests using the 2 247 547 observations on the receipts of French firms. The results are reported in Table 1.

Table 1: Test results for French firms ($n = 2\,247\,547$): threshold rank, percentile, share and shape parameter

	UMPU		ME		GI	
	5% level	1% level	5% level	1% level	5% level	1% level
rank	1600	1650	1750	2150	2400	3480
percentile	< 0.1	< 0.1	< 0.1	0.1	0.1	0.2
% total revenues	41.89	42.19	42.75	44.70	45.75	49.35
shape	1.157	1.147	1.119	1.111	1.181	1.151

The values in the table should be interpreted as follows: *rank* gives the number of observations larger than the power-law threshold, *percentile* is the percentile of the empirical distribution corresponding to the rank, *% total revenues* is the percentage of total revenues corresponding to the firms in the power-law tail and *shape* is the estimate of the Pareto shape parameter.

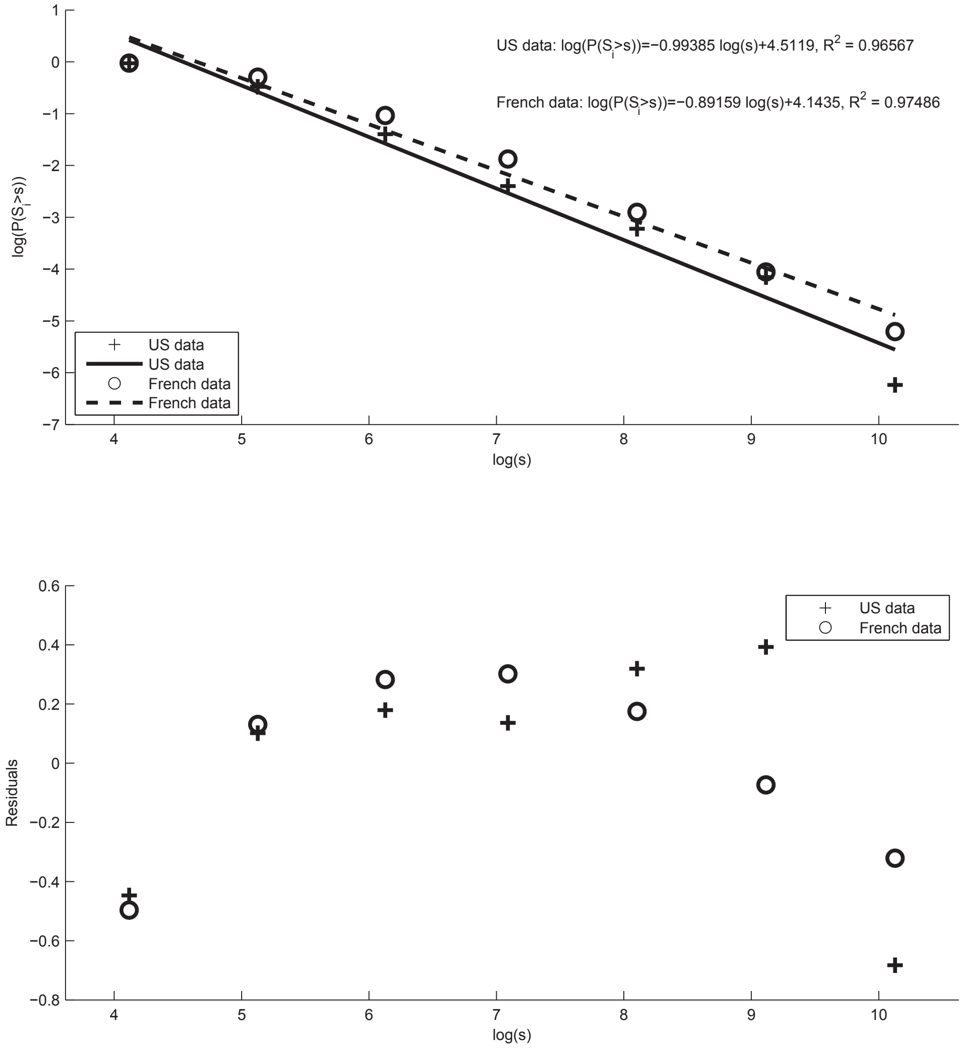

Figure 2: The size distribution of US and French firms: the upper panel shows the counter-cumulative distribution function of bin frequencies in double log scale; the lower panel displays the residuals of the linear regressions on binned data.

Although the GI test finds a slightly longer tail than UMPU and ME, the overall message is clear: the Pareto tail (if any) corresponds to an extremely small fraction of large firms. Furthermore, although these very few observations contain a non-negligible share of the total receipts, a simulation-based analysis (not reported but available upon request) finds that the Pareto tail is not significantly longer than the one identified when applying the tests to a lognormal distribution with parameters matching the empirical data (for details on the methodology adopted see Bee, Riccaboni, and Schiavo, 2013).[4]

3.2 Size and scope of French firms

The economic literature suggests that firms grow according to two possible mechanisms: the first refers to the growth process of existing activities, the so-called intensive margin, and is the one described by Gibrat's law that leads to a lognormal distribution. The second consists in the acquisition of new business opportunities (e.g. development of new products, penetration into new markets, mergers and acquisitions) and is therefore associated with an expansion of firm scope (e.g. Bernard, Redding, and Schott, 2010). This second process of growth —associated with the seminal work by Simon (1955) and, more recently, to the model developed by Chatterjee and Rossi-Hansberg (2012)— leads to Zipf's law.

Another relevant piece of empirical evidence has to do with the assertion that, upon aggregation, the size distribution of firms moves closer to a power-law or, at least, develops a more pronounced Pareto upper tail. Rossi-Hansberg and Wright (2007) find evidence that while US *establishments* seem not to follow a power-law, the distribution of *firm* size more closely resembles a Pareto distribution. Differences across sectors are linked to heterogeneity in terms of human-capital intensity. Growiec, Pammolli, Riccaboni, and Stanley (2008) find a similar behavior for pharmaceutical firms, and explain it by postulating the existence of a stretching factor triggered by the aggregation of different products sold by the same firm.

As establishments mainly tend to grow according to the first process, i.e. by increasing scale, they are more likely to be lognornally distributed. On the other hand, when firms aggregate multiple business units (products) within the same firm (increasing firm scope), the distribution is better approximated by a power-law or a Zipf distribution.

In this section we investigate the extent to which the scope of enterprises influences firm size distribution. This is done by discriminating among different subsets of firms, based on the number of products they sell, the number of foreign destinations they serve,

[4]The ME methodology lends itself in a natural way to testing the hypothesis $H_0 : k = 2$ against $H_1 : k = 3$. As the ME(2) distribution is lognormal, this test allows to find the length of a (possible) lognormal tail. At the 5% level, the test finds a lognormal tail containing the largest 47 000 firms, approximately corresponding to the 0.021 quantile and to the 75.59% of the total revenues.

and the number of product-destination pairs.

Table 2 reports the test results for various subsets of firms. The first row of panel *a* refers to the whole dataset and thus simply replicates the results reported in Table 1, the second ($K > 0$) refers to exporting firms only, while the next three lines concern multi-product firms. The sample size shrinks significantly when we move from the universe of firms in the dataset to exporting firms only (92 000 observations, roughly 4% of the total), whereas only 16 000 firms export more than 10 different products.

All three tests find a monotonic increase in the share of firms belonging to the Pareto tail of the distribution, whose length increases from virtually nil (0.07–0.15% depending on the test) to a value ranging between 5 and 8.5% of the distribution. Similarly, the estimated shape parameter of the power-law decreases monotonically toward 1 once we progressively restrict the analysis to firms exporting a larger number of products.[5] What remains approximately stable is the share of total revenues corresponding to the firms belonging to the power-law tail.

Similar conclusions hold when we look at "very international" firms, i.e. companies shipping their goods to many foreign destinations (see panel *b* of Table 2). Indeed, when compared to the total population or the universe of exporters, the size distribution of firms exporting to more than 10 destinations (which make up less than 1% of all firms) displays a power-law tail spanning between 6 and 13% of the population. Furthermore, the estimated shape parameter moves downward becoming closer to 1.

The change in the behavior of the distribution is all the more apparent when we classify firms on the basis of the number of their product-destination pairs, thus distinguishing between, say, apples shipped to country A and to country B (see panel *c*).

It should be noted that the overall distribution is a weighted average of the distributions of different subgroups, where weights are given by the relative importance of the various classes of firms in the total population of firms. As a result, one could figure out that the size distributions of business firms may differ depending on the weights of the firm types. Therefore, a larger share of firms either exporting to many destination markets, or producing a large number of different products, would lead to a more pronounced power-law behavior and a closer resemblance to Zipf's law (at least in the upper tail). In fact, despite the methodological caveats we have presented in Section 2 above, one way to reconcile the evidence put forward by Axtell about US firms with our own, is to consider that US firms are generally more diversified and/or more internationalized than French ones. This feature would then result in the corresponding size distribution showing a longer Pareto tail and a shape parameter closer to 1.

The tendency toward Zipf's law in the upper tails of the distributions appears clearly

[5] For details on how the shape parameter is estimated with the different methodologies see Bee, Riccaboni, and Schiavo (2013).

Table 2: Test results for firms with different levels of diversification (number of products, K) and internationalization (number of foreign markets, N, and product-market pairs, NK)

	UMPU				ME				GI				
						panel a: number of products							
K	rank	perc.	share	shape	rank	perc.	share	shape	rank	perc.	share	shape	# firms
≥ 0	1600	< 0.1	41.98	1.158	1750	< 0.1	42.83	1.137	2400	0.1	45.84	1.181	2247547
> 0	1350	1.5	42.36	1.136	1450	1.6	43.10	1.112	2410	2.6	48.42	1.135	92057
> 1	1300	2.1	42.62	1.123	1350	2.2	43.03	1.112	2230	3.7	48.50	1.125	60670
> 5	1050	3.7	42.13	1.089	1150	4.1	43.18	1.057	1830	6.5	48.33	1.086	28150
> 10	850	5.3	41.23	1.073	950	5.9	42.51	1.049	1400	8.7	46.98	1.072	16084
						panel b: number of destinations							
N	rank	perc.	share	shape	rank	perc.	share	shape	rank	perc.	share	shape	# firms
> 1	1250	2.4	42.06	1.1241	1350	2.6	42.84	1.1059	2235	4.2	48.65	1.1184	52935
> 5	1050	4.8	41.51	1.1129	1100	4.6	41.96	1.0434	1810	7.6	49.07	1.0700	23651
> 10	850	6.4	39.98	1.1095	900	6.8	40.57	1.0180	1460	11.0	48.92	1.0404	13249
						panel c: number of product-destination pairs							
NK	rank	perc.	share	shape	rank	perc.	share	shape	rank	perc.	share	shape	# firms
> 1	1300	2.0	42.46	1.1326	1350	2.1	42.85	1.1225	2280	3.5	48.54	1.1266	65072
> 5	1150	3.1	41.25	1.1361	1200	3.2	42.91	1.0811	2095	5.7	49.09	1.0982	36974
> 10	1050	4.1	42.16	1.1188	1150	4.5	43.25	1.0825	1800	7.0	47.24	1.0879	25712
> 50	780	10.4	42.20	1.0824	820	11.0	42.87	0.9946	1195	16.0	47.74	1.0311	7463
> 100	520	14.7	40.79	1.1214	600	17.0	42.64	0.9409	780	22.1	46.32	1.0022	3527

The values in the table should be interpreted as follows: *rank* gives the number of observations larger than the power-law threshold, *perc.* is the percentile of the empirical distribution corresponding to the rank, *share* is the percentage of total revenues corresponding to the firms in the power-law tail and *shape* is the estimate of the Pareto shape parameter.

in Figure 3 as well, where we portray the counter-cumulative distribution functions pertaining to the different groups of firms, along with a reference line with a slope equal to -1.[6] Interestingly, the top panel shows that the distribution of all firms is similar to the one found by Rossi-Hansberg and Wright (2007) on US firms, with the central part approximately resembling a Pareto distribution and a more pronounced concave shape in the tails. The plots confirm the results presented in Table 2: the Zipf behavior is more apparent in the case of multiple products (top panel) than in the case of firms serving multiple markets.

4 Discussion and conclusions

This paper investigates the size distribution of the universe of French firms, adopting a more sound methodology than is normally found in the literature. When it comes to discriminating between lognormal and power-law distributions as the preferred approximation of the empirical distribution, we show that binning the data severely biases the results in favor of a power-law.

We strongly reject the hypothesis that the size distribution is Pareto. We argue

[6]The various distributions have been right-shifted to improve readability.

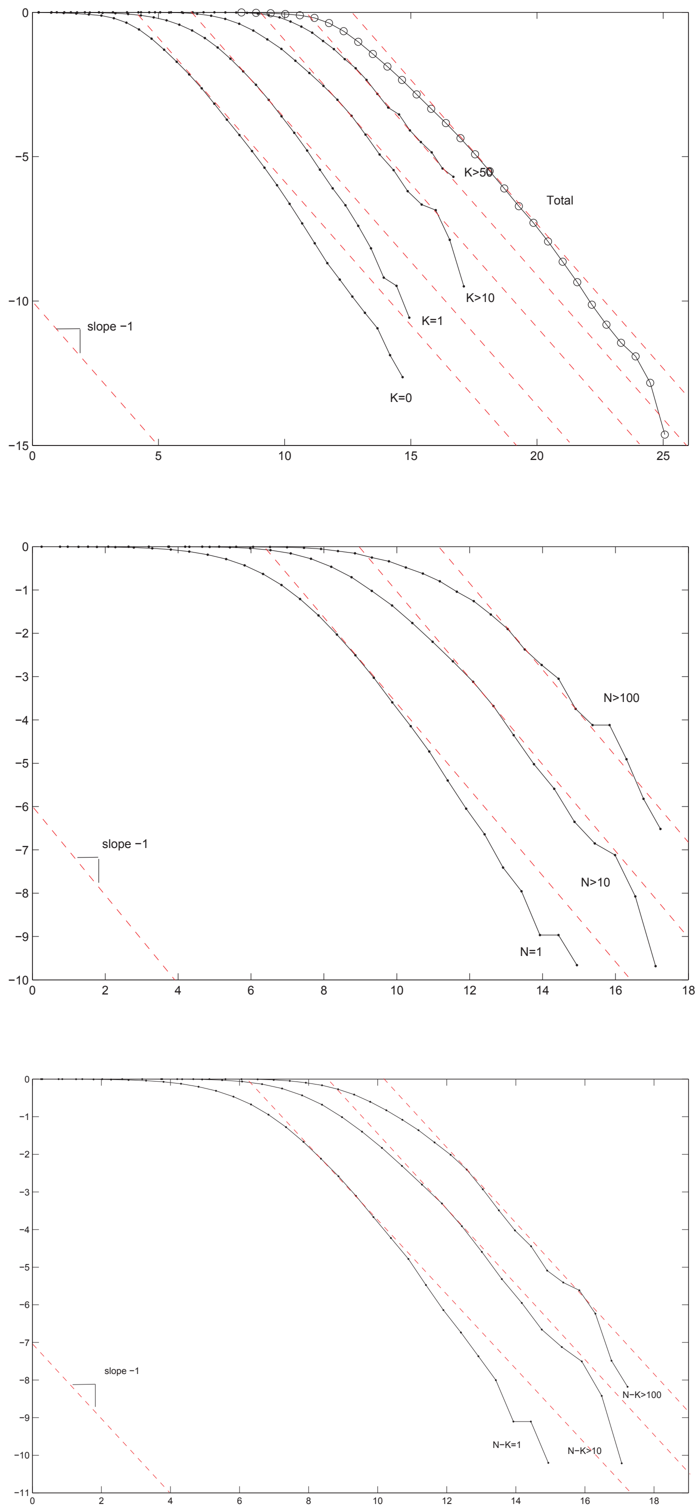

Figure 3: The size distribution of US and French firms by number of products, N, destination markets, K and product-destination NK.

that the lognormal distribution should be favored as a first-cut benchmark especially for single-product firms. Even if we consider Pareto as a tail property, the shape parameter is significantly larger than 1, and only approximately 0.1% of the firms belong to the Pareto tail, corresponding to less than 50% of combined total revenues. On the other hand, the lognormal behavior accommodates the top 2% of firms corresponding to about 75% of total revenues. We also show the emergence of a Zipf tail for large multi-product firms.

Our empirical results confirm this intuition, lending credit to a recent stream of literature that explains departures from power-law distribution and differences across sectors or firm types. In fact, Chatterjee and Rossi-Hansberg (2012) finds that, so long as new ideas are captured mostly by established firms (versus startups), the size distribution of firms converges to Zipf's law. Since few firms are highly diversified from the very beginning, and few new exporters ship many goods to many destinations the first time they enter foreign markets, when looking at firms with large scope, we are focusing on established ventures. Hence, our evidence for highly diversified firms supports the notion that when opportunities are mainly captured by incumbents, the size distribution of firms moves closer to a Zipf distribution. Similarly, Rossi-Hansberg and Wright (2007) argue that higher human capital intensity leads to an approximate power-law size distribution. Since the empirical literature on firm behavior in international markets has found a link between the degree of international involvement and human capital at the firm level (Munch and Skaksen, 2008), our findings are consistent with this notion. Furthermore, as firms with wider scope tend to be older on average, the emergence of a heavier upper tail in the distribution of such firms is also consistent with Cabral and Mata's (2003) claim that financial constraints (more likely to be binding for young firms) affect the size distribution of firms by limiting the ability of startups to operate at optimal scale. With respect to the work by Di Giovanni, Levchenko, and Rancière (2011), while we confirm that the distribution of exporting firms displays a heavier upper tail, only for very international firms ($K >> 0$ or $N >> 0$) do we find a convergence to a Zipf distribution (from above).

From the analysis we derive a number of general lessons. First, from a theoretical point of view, it makes little sense to aim necessarily at reproducing Zipf's law, only to be forced later to find arguments to justify lognormal departures in the body of the distribution. At that point, it is probably wiser to stick to a lognormal for the whole distribution. It is important to note that although the power-law may be analytically more tractable and thus simplify the formal modeling of economic dynamics, predictions derived from assuming a power-law build on shaky foundations and should not be used to inform policy prescriptions (e.g. welfare implications, policy evaluations and the like).[7] Moreover, the assumption of lognormality may be useful in specific contexts. For instance,

[7]A notable example is the benefits from trade liberalization, whose magnitude seems to depend crucially on the shape of the distribution of firm size (see Di Giovanni, Levchenko, and Rancière, 2011).

it allows for the easy derivation of expected concentration indexes and therefore for the quantification of departures from a clear-cut theoretical benchmark (on this subject see Hart, 1975; Davies, 1980).

From a practical point of view we consider the lognormal distribution to be a good approximation of the size distribution of French firms; departures from this benchmark in the lower and upper tail are informative of economic forces and frictions, and deserve further scrutiny. Overall, it makes much more sense to devote future research to the analysis of such departures across countries rather than to a comparison of distributions in search of the best fit.

References

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

Bee, M. (2013): "A Maximum Entropy Approach to Loss Distribution Analysis," *Entropy*, 15, 1100–1117.

Bee, M., M. Riccaboni, and S. Schiavo (2011): "Pareto versus lognormal: A maximum entropy test," *Physical Review E*, 84, 026104.

——— (2013): "The size distribution of US cities: Not Pareto, even in the tail," *Economics Letters*, 120(2), 232–237.

Bernard, A. B., S. J. Redding, and P. K. Schott (2010): "Multiple-Product Firms and Product Switching," *American Economic Review*, 100(1), 70–97.

Cabral, L. M. B., and J. Mata (2003): "On the Evolution of the Firm Size Distribution: Facts and Theory," *American Economic Review*, 93(4), 1075–1090.

Chatterjee, S., and E. Rossi-Hansberg (2012): "Spinoffs And The Market For Ideas," *International Economic Review*, 53(1), 53–93.

Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009): "Power-law distributions in empirical data," *SIAM Review*, 51, 661–673.

Cox, D. R., and D. V. Hinkley (1974): *Theoretical Statistics*. Chapman and Hall.

Davies, S. (1980): "Measuring Industrial Concentration: An Alternative Approach," *Review of Economics and Statistics*, 62(2), 306–309.

de Wit, G. (2005): "Firm size distributions: An overview of steady-state distributions resulting from firm dynamics models," *International Journal of Industrial Organization*, 23(5-6), 423–450.

Del Castillo, J., and P. Puig (1999): "The Best Test of Exponentiality against singly Truncated normal alternatives," *Journal of the American Statistical Association*, 94, 529–532.

Di Giovanni, J., A. 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.

Eaton, J., S. Kortum, and F. Kramarz (2011): "An Anatomy of International Trade: Evidence From French Firms," *Econometrica*, 79(5), 1453–1498.

Gabaix, X. (2009): "Power Laws in Economics and Finance," *Annual Review of Economics*, 1, 255–93.

Gabaix, X., and R. Ibragimov (2011): "Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents," *Journal of Business and Economic Statistics*, 29(1), 24–39.

Garicano, L., C. LeLarge, and J. Van Reenen (2013): "Firm Size Distortions and the Productivity Distribution: Evidence from France," Working Paper 18841, NBER.

Gibrat, R. (1931): *Les Inegalites Economiques*. Sirey, Paris.

Growiec, J., F. Pammolli, M. Riccaboni, and H. E. Stanley (2008): "On the size distribution of business firms," *Economics Letters*, 98(2), 207–212.

Hart, P. E. (1975): "Moment Distributions in Economics: An Exposition," *Journal of the Royal Statistical Society. Series A (General)*, 138(3), 423–434.

Ijiri, Y., and H. A. Simon (1974): "Interpretations of Departures from the Pareto Curve Firm-Size Distributions," *Journal of Political Economy*, 82(2), 315–331.

Jovanovic, B. (1982): "Selection and the Evolution of Industry," *Econometrica*, 50(3), 649–70.

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

Luttmer, E. G. (2007): "Selection, growth, and the size distribution of firms," *Quarterly Journal of Economics*, 122(3), 1103–1144.

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

Malevergne, Y., V. Pisarenko, and D. Sornette (2009): "Gibrat's law for cities: uniformly most powerful unbiased test of the Pareto against the lognormal," *Swiss Finance Institute Research Paper Series*, pp. 09–40.

——— (2011): "Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities," *Physical Review E*, 83(3), 036111.

Mitzenmacher, M. (2004): "A brief history of generative models for power law and lognormal distributions," *Internet mathematics*, 1(2), 226–251.

Munch, J. R., and J. R. Skaksen (2008): "Human capital and wages in exporting firms," *Journal of International Economics*, 75(2), 363–372.

Perline, R. (2005): "Weak and False inverse power laws," *Statistical Science*, 20, 68–88.

Rossi-Hansberg, E., and M. L. Wright (2007): "Establishment size dynamics in the aggregate economy," *American Economic Review*, 97(5), 1639–1666.

Simon, H. A. (1955): "On a class of skew distribution functions," *Biometrika*, 42(3-4), 425.

Simon, H. A., and C. P. Bonini (1958): "The Size Distribution of Business Firms," *American Economic Review*, 48(4), 607–617.

Sutton, J. (1997): "Gibrat's Legacy," *Journal of Economic Literature*, 35(1), 40–59.

Virkar, Y., and A. Clauset (2014): "Power-law distributions in binned empirical data," *Annals of Applied Statistics*, in press.

Wu, X. (2003): "Calculation of Maximum Entropy Densities with Application to Income Distribution," *Journal of Econometrics*, 115, 347–354.

List of recent papers published un the same series

2013/01 A regular multidistance among fuzzy numbers, Franco Molinari

2013/02 Employer moral hazard and wage rigidity. The case of worker-owned and investor-owned firms, Marina Albanese, Cecilia Navarra and Ermanno Tortia

2013/03 Origins and prospects of the Euro existential crisis, Luigi Bonatti and Andrea Fracasso

2013/04 Uses And Motivations For Credit Derivatives: An Empirical Investigation Into Italian Banks, Eleonora Broccardo, Maria Mazzuca and Elmas Yaldiz

2013/05 At the core of the international financial system, Valentina Feroldi, Edoardo Gaffeo

2013/06 Reassessing the spatial determinants of the growth of Italian SMEs, Roberto Gabriele, Diego Giuliani, Marco Corsino, Giuseppe Espa

2013/07 A spatial and sectoral analysis of firm demography in Italy, Giuseppe Espa, Danila Filipponi, Diego Giuliani, Davide Piacentino

2013/08 Using information markets in grantmaking. An assessment of the issues involved and an application to Italian banking foundations, Edoardo Gaffeo

2013/09 Interbank contagion and resolution procedures: inspecting the mechanism, Edoardo Gaffeo e Massimo Molinari

2013/10 Transatlantic contagion 2010-…, Roberto Tamborini

2013/11 Do middle managers matter?, Elena Feltrinelli, Roberto Gabriele, Sandro Trento

2013/12 Approximate Maximum Likelihood Estimation of the Autologistic Model, Marco Bee, Diego Giuliani, Giuseppe Espa

2013/13 Scope and Flaws of the New Neoclassical Synthesis, Ronny Mazzocchi

2013/14 Investment-Saving Imbalances with Endogenous Capital Stock, Ronny Mazzocchi

2013/15 Intertemporal Coordination Failure and Monetary Policy, Ronny Mazzocchi

2013/16 Monetary Policy when the NAIRI is unknown: The Fed and the Great Deviation, Ronny Mazzocchi

2013/17 Resilience and specialization in volatile environments: evidence from the Italian Air Force Tornado crews learning practices, M. Laura Frigotto, Marco Zamarian

2013/18 Bio-Economics of Allocatable Pollination Services: Sequential Choices and Jointness in Sites, L. Pilati and V. Boatto

2014/01 Macroprudential Consolidation Policy in Interbank Networks, E. Gaffeo and M. Molinari

2014/02 Competition in the banking sector and economic growth: panel-based international evidence, E. Gaffeo and R. Mazzocchi