



UNIVERSITY
OF TRENTO - Italy

Department of Economics

A Trick of the (Pareto) Tail.

Marco Bee, Massimo Riccaboni and Stefano Schiavo

n.
06/2012

A Trick of the (Pareto) Tail

Marco Bee* Massimo Riccaboni† Stefano Schiavo‡

April 13, 2012

Abstract

Several economic phenomena are found to follow an approximate Pareto distribution, at least in the upper tail. The debate is well established for the distribution of wealth and business firms, and has recently been particularly animated with respect to city sizes. In this paper we contribute to this stream of the literature by showing that the power-law tail emerges upon aggregation, and this holds true across three different domains: cities, firms and trade flows. We explore different mechanisms that could give rise to this effect, from mere sample size to correlation among the number of constituent parts of aggregate entities and their size, to the aggregation rule, and discuss their impact on the Pareto tail. Our results suggest that the debate on the shape of the distribution of city size is not yet closed and deserves further scrutiny. Using multiple statistical tests we find that the existence of a Pareto tail for the city size distribution is questionable due to sample size issues. Furthermore, the presence of a positive correlation between the number of elementary units (products) comprised in each aggregate level (firm) and their average size is key to explain why the size distribution of business firms displays a significant power-law tail. Conversely, we do not find any Pareto tail for trade flows. The paper casts new light on the mechanisms through which idiosyncratic shocks do not average out upon aggregation, so that idiosyncratic shocks to individual units are not washed away in economic aggregates (as the central limit theorem would predict), but can even be magnified.

Keywords: Zipf distribution; lognormal distribution; maximum entropy; cities, size distribution; firms, size distribution; international trade.

JEL classification: C14, C51, C52

*Department of Economics - University of Trento, via Inama 5 – 38122 Trento (Italy), marco.bee@unitn.it.

†IMT Institute for Advanced Studies, Piazza San Ponziano 6 – 55100 Lucca (Italy) and Department of Managerial Economics, Strategy and Innovation (MSI), K.U.Leuven, massimo.riccaboni@imtlucca.it.

‡Department of Economics - University of Trento, via Inama 5 – 38122 Trento (Italy), and OFCE-DRIC (France), stefano.schiavo@unitn.it.

1 Introduction

Several phenomena in economics and finance follow a Pareto distribution, at least in the upper tail (Gabaix, 2009). Well-known examples are the distribution of wealth (Pareto, 1896; Champernowne, 1953; Benhabib, Bisin and Zhu, 2011), firm size (Gibrat, 1931; Ijiri and Simon, 1977; Axtell, 2001; Cabral and Mata, 2003; Luttmer, 2007), and city size (Zipf, 1949; Gabaix, 1999). Recently, the last one has been at the center of a lively debate on whether it is better approximated by a Pareto or by a lognormal distribution (Eeckhout, 2004; Levy, 2009; Eeckhout, 2009; Malevergne, Pisarenko and Sornette, 2009; Rozenfeld et al., 2011). This debate is hampered by the difficulty to distinguish lognormal versus Pareto tails (Embrechts, Klüppelberg and Mikosch, 1997; Bee, Riccaboni and Schiavo, 2011). To ascertain the exact shape of the size distribution of economic systems at different levels of aggregation is of crucial importance since the Pareto distribution of firm sizes is a building block of many economic models, such as for instance Melitz (2003) and Helpman, Melitz and Yeaple (2004), whereas other models have been developed to generate a Pareto firm size distribution (Steindl, 1965; Ijiri and Simon, 1977; Luttmer, 2007). Moreover, it has been recently demonstrated that the impact of idiosyncratic shocks on the economic system as a whole depends on the size distribution of economic units (Gabaix, 2011) as well as their relations (Acemoglu et al., 2011).

Here we present new evidence concerning the size distribution of firms and cities —two domains where the issue has been studied at length— as well as bilateral trade flows, at different levels of granularity of the economic system from elementary units (products and census blocks) to aggregate entities (firms, cities and bilateral trade flows). Although the literature on trade flows is not very large, the size distribution of trade flows has been recently analyzed by Easterly, Reshef and Schwenkenberg (2009), who derive important policy implications from its alleged power-law distribution.¹ The issue appears then relevant not only from the statistical standpoint, but also for its broader implications. Easterly, Reshef and Schwenkenberg (2009) analyze bilateral trade flows for 151 countries (at the lowest comparable level of aggregation, i.e. using 6 digits of the Harmonized System commodity classification) and conclude that their distribution is a mixture of a lognormal body with a Pareto upper tail. To account for this shape the paper develops a model incorporating both (Pareto distributed) productivity shocks and demand shifts (which follow a lognormal distribution). Furthermore, building on the existence of a power-law upper tail, the authors

¹Recent papers dealing with the distribution of trade flows are Bhattacharya et al. (2008), Fagiolo, Schiavo and Reyes (2008), Fagiolo, Reyes and Schiavo (2009), Riccaboni and Schiavo (2010), all of which tend to agree with lognormality. When reviewing the economic literature on power-laws, for what concerns international trade Gabaix (2009) quotes works that look at rather different phenomena such as the size of exporting firms (Helpman, Melitz and Yeaple, 2004), the number of markets served by each firm (Eaton, Kortum and Kramarz, 2004), and the Balassa index of comparative advantage (Hinloopen and van Marrewijk, 2012). Chaney (2011) only focuses on the extensive margin of trade, i.e. the number of destinations served by each country and not to trade flows per se, whereas di Giovanni, Levchenko and Rancière (2011) analyze the impact of openness on the size distribution of firms.

argue against active industrial policies since the skewness of the distribution makes it very hard to pick a ‘big hit’, i.e. a product whose export can provide a meaningful contribution to aggregate trade and economic growth.

The second domain we study is the size of business firms, where there exists a long-standing debate on whether empirical observations are better approximated by a lognormal or a power-law distribution Sutton (1997); Gabaix (2009). The issue dates back at least to Gibrat (1931), who postulates a model of firm growth leading to a lognormal distribution. Years later, Hart and Prais (1956) confirm that such a distribution fits the data quite well, using a sample of British manufacturing firms listed on the London Stock Exchange. A similar conclusion is reached more recently by Stanley et al. (1995), who use data on publicly listed US firms taken from *Compustat*.

However, already Simon and Bonini (1958) note that the empirical distribution displays a thicker upper tail than a lognormal would imply, so that a Pareto fits better that part of the distribution.² The presence of a Pareto upper tail (and a lognormal body) is later confirmed by many studies, such as Marsili (2005) or Growiec et al. (2008) among others. The controversy is not yet over, as there are at least two very influential papers claiming that the *whole* distribution of firm size is better approximated by a power-law, not just its upper tail (Axtell, 2001).

Three issues emerge forcefully from this literature. First, the coverage of the samples used in the analysis does matter: while early studies only look at large listed firms (Hart and Prais, 1956; Simon and Bonini, 1958; Stanley et al., 1995), Axtell (2001) uses Census data covering the universe of US firms, and claims this is the main explanation for the novelty of his results.³ Second, the level of aggregation at which phenomena are observed, typically establishments and firms, does matter. Quandt (1966) notes that regularities observed for aggregate data do not hold at a disaggregate level. More recently, Growiec et al. (2008) perform the analysis at different levels and show that even if elementary units are lognormally distributed, a power-law may emerge upon aggregation as the result of a skewed aggregation rule. Third, at a certain level of resolution, results are robust to different definitions of size (employment, sales, market valuation), and over time (Axtell, 2001).

Our last application concerns the size distribution of cities: in recent years a remarkable effort has been devoted to determining its proper shape or, at least, the correct shape of its upper tail (see Gabaix, 1999; Eeckhout, 2004; Levy, 2009; Malevergne, Pisarenko and Sornette, 2011, to name just a few). Scholars are engaged in an ongoing debate on whether such a distribution displays a Pareto tail and whether this conforms to a Zipf’s law (i.e. a power-law distribution with shape parameter equal to one) or not. Beside the

²Simon and Bonini (1958) and later Ijiri and Simon (1977) refer more precisely to the Yule distribution, which is a discrete version of a power-law.

³Comparing Census data with Compustat, Axtell (2001) notes that the latter disregards all the large privately-held firms; furthermore, the former displays increasing frequencies of increasingly smaller firms, a shape that is not consistent with a lognormal distribution.

specific intellectual curiosity the issue may raise, there are broader theoretical reasons for investigating the matter, as competing models yield different implications. Indeed, while the seminal paper by Gabaix (1999) predict a Zipf's law, Eeckhout (2004) proposes an equilibrium theory to explain the lognormal distribution of cities.⁴

The contention is partly based on the difficulty of properly defining what a city is and, empirically, what is the correct measure to use.⁵ Indeed, while early studies focus on the largest US Metropolitan Statistical Areas (MSA) only (Gabaix, 1999), Eeckhout (2004) uses data for all the US populated places identified by the US Census in years 2000 and 2001. By so doing the author shows that the size distribution of cities is lognormal, not power-law as previously thought (at least since Zipf, 1949). A few years later Levy (2009) acknowledges that the body of the city size distribution is well approximated by a lognormal, but claims that there are significant departures in the upper tail. Specifically, the top 0.6 percent of the distribution, i.e. the MSA, appear to fit better a power-law. Eeckhout (2009) replies to these new findings by highlighting potential problems associated with the procedures used by Levy (2009) to identify the power-law tail. Specifically, Eeckhout (2009) suggests that the graphical procedure based on visual inspection of a log-log plot introduces significant biases in the right tail of the distribution. Moreover, he warns against the practice of estimating a truncated subsample of the distribution only, while testing its significance against a complete distribution with specific parameters. Recently Malevergne, Pisarenko and Sornette (2011) have suggested that the debate rests on the small power of the tests employed by both Eeckhout (2004) and Levy (2009). They claim the issue can be definitely settled by adopting a better testing procedure, namely the uniformly most powerful unbiased test of the exponential versus truncated normal distribution in log-scale developed by del Castillo and Puig (1999). Hence, for what concerns city size, methodological issues related to the proper way of identifying a power-law tail seems at least as relevant as sample coverage.

We contribute to this debate by providing new empirical evidence that cuts across different domains, and points to the fact that the tail behavior of the distributions changes upon aggregation, giving rise to a more pronounced (longer) power-law tail. To the best of our knowledge this is the first time that such a regularity is documented across different empirical applications. Furthermore, we investigate three different mechanisms that can generate this phenomenon, namely sample size, correlation between the number of elementary units and their average size, and the aggregation function (i.e. the shape of the distribution of the number of elementary units associated to each aggregate element). We find the last factor

⁴The same holds true in the industrial organization literature, where models to generate different skewed firm size distributions have been contrasted.

⁵This point is particularly stressed in Rozenfeld et al. (2011), who propose a new methodology to define cities based on microdata and a clustering algorithm that identifies a city as the maximal connected cluster of populated sites. By applying this methodology to both US and UK data, the authors find that a Zipf's law approximates well the distribution of 1,947 US cities with more than 12,000 inhabitants (1,007 cities with more than 5,000 inhabitants for the UK).

exerts the largest influence and needs to be carefully scrutinized when proposing models aiming at explaining the shape of the distribution of particular phenomena Rozenfeld et al. (2011). However, in the case of cities, sample size appears to crucially determine the length of the power-law tail found in the data. In this respect then, the debate is not yet closed.

More in general, our work casts new light on the fact that upon aggregation one observes the emergence of a thicker tail, so that extreme events become more likely (Perline, 2005; Growiec et al., 2008). Such a behavior is counterintuitive since the common wisdom based on the Central Limit Theorem assumes idiosyncratic shocks to cancel out upon aggregation, so that aggregate economies might be more stable than constituent entities, and reinforces the claim recently made by Gabaix (2011).

The paper is organized as follows: the next section describes methodology and data used in the analysis to detect the presence and origin of power-law tails; Section 3 presents the results and explores mechanisms through which a power-law could emerge upon aggregation. Finally, Section 4 concludes.

2 Methodology and Data

In this paper we apply several statistical tests to the size distributions of economic systems in different domains. Most papers apply one method to a specific field of investigation. By comparing our findings across domains and methods we aim at identifying robust stylized facts and domain-specific effects. We analyze the distribution of US city sizes, pharmaceutical firms and world trade flows. In all the three cases we decompose aggregate entities (i.e. cities, firms and trade flows) into constituent parts: census blocks, product sales and product-level trade flows. We analyze the size distribution of aggregate entities, $P(S_i)$ and the size of constituent parts $P(s_j)$, with $S_i = \sum_{j=1}^{K_i} s_j$ where K_i is the number of parts of aggregate i and s_j is the size of the building blocks. A traditional “replication” argument states that if an integrated entity of larger size had higher unit costs, then it should be possible to split the business into independent and separately managed units so that any such disadvantage is eliminated. Thus non-diminishing returns at the level of aggregated entities and constituent units imply that at the aggregate level economic organizations are made by an uneven number of units of different sizes.⁶

2.1 Testing for a power-law tail

Discriminating between power-law (Pareto) and lognormal tail behavior is a difficult task. The methodological reference is Extreme Value Theory (EVT), which studies the statistical properties of the distributions of upper order statistics. It is well-known that they belong to

⁶Aggregate economies can thus be represented and as (random) partitions (Sutton, 2002; Aoki and Yoshikawa, 2011) or as sums of a random number of random variables.

the domain of attraction of one of three distributions, namely Fréchet, Gumbel or Weibull (Embrechts, Klüppelberg and Mikosch, 1997). Whereas the distributions in the Fréchet domain of attraction are definitely heavy-tailed and the distributions in the Weibull domain are light-tailed, the Gumbel domain includes both distributions with a relatively light tail (exponentially decreasing, such as the normal) and with a relatively heavy tail (such as the lognormal). Things are further complicated by the fact that there exist several definitions of “heavy-tailed distributions”, corresponding to different degrees of tail heaviness (see Embrechts, Klüppelberg and Mikosch, 1997, pp. 49-50).

For the purposes of testing between Pareto and Lognormal, the main result is that the upper order statistics of the lognormal converge to the Gumbel distribution, whereas the upper order statistics of the Pareto converge to the Fréchet. This implies that the asymptotic tail behaviors of the two distributions are mathematically different. However, the convergence of the lognormal to the asymptotic distribution is extremely slow (Perline, 2005), so that the difference may be very small, at the extent that they are often practically indistinguishable for any finite sample size.

A similar conclusion is reached by recalling that a continuous random variable (r.v.) is in the domain of attraction of the Fréchet if and only if its density is a regularly varying function (Embrechts, Klüppelberg and Mikosch, 1997, pp. 131-132). Although the Pareto density is regularly varying and the lognormal is not, Malevergne, Pisarenko and Sornette (2009) point out that, when the variance is large, the lognormal probability density function (pdf) can be rewritten in a form similar to the Pareto pdf, the only difference being that the exponent of the lognormal, unlike the Pareto one, varies with x . However, the lognormal exponent is almost constant with respect to x , so that in practice, unless the sample size is huge and/or the variance is very small, discriminating between a constant and an “almost constant” exponent is problematic.

Given these difficulties, several tests have been proposed, in an attempt to find the one that guarantees the best performance. We mention here, and employ in the following, the Uniformly Most Powerful Unbiased (UMP) test based on the clipped sample coefficient of variation developed by del Castillo and Puig (1999) and used by Malevergne, Pisarenko and Sornette (2009), the Maximum Entropy (ME) test by Bee, Riccaboni and Schiavo (2011) and a test recently proposed by Gabaix and Ibragimov (Gabaix, 2009; Gabaix and Ibragimov, 2011; Rozenfeld et al., 2011) (GI henceforth).

The UMP test is uniformly most powerful, but only in the class of unbiased tests. A more serious drawback is that it is a test of the null of power-law against the alternative of lognormal, and rejects the null hypothesis for small values of the coefficient of variation c . Implicitly, this implies that it works well (i.e., its power is high) in cases such as the lognormal-Pareto mixture, namely when the data generating process is such that $c \geq 1$ above the threshold that separates the lognormal and the Pareto and $c < 1$ below the threshold

(Bee, Riccaboni and Schiavo, 2011). However, if the distribution below the threshold is not power-law but nonetheless has $c \geq 1$, as happens, for example, 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).

The ME approach entails maximizing the Shannon's information entropy under k moment constraints $\mu^i = \hat{\mu}^i$ ($i = 1, \dots, k$), where $\mu^i = E[T(x)^i]$ and $\hat{\mu}^i = \frac{1}{n} \sum_j T(x_j)^i$ are the i -th theoretical and sample moments and n is the number of observations. This can be solved by introducing $k + 1$ Lagrange multipliers λ_i ($i = 0, \dots, k$), so that the solution (that is, the ME density) takes the form $f(x) = e^{-\sum_{i=0}^k \lambda_i T(x)^i}$. The Pareto distribution is an ME density with $k = 1$, whereas the lognormal is ME with $k = 2$.

A log-likelihood ratio (llr) test of the null hypothesis $k = k^*$ against $k = k^* + 1$ is given by

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

where n is the population size. From standard limiting theory the llr test is asymptotically χ_1^2 and is optimal (Cox and Hinkley, 1974; Wu, 2003).

When the whole distribution is of interest, the method can be used for fitting the best approximating density, with the optimal k found by the log likelihood ratio (llr) criterion. The procedure is based on the following steps: (1) estimate sequentially the ME density with $k = 1, 2, \dots$; (2) perform the test for each value of k ; (3) stop at the first value of k (k_0 , say) such that the hypothesis $k = k_0$ cannot be rejected and conclude that $k^* = k_0$. If the aim consists in testing a power-law against a lognormal tail, we just test $k = 1$ against $k = 2$.⁷

When ME tests for the optimal value of k , it is computed iteratively starting from $k = 1$ and stopping only when the p -value is sufficiently small. When the true distribution might be neither Pareto nor lognormal, the test should be carried out for some values of k larger than 2, even though the p -value for $k = 2$ against $k = 1$ may be relatively small. Typically, a very small p -value will be obtained for the optimal value of k , which is expected to be larger than 2. In other words, in such a case it may be worth to use a rather high level α , such as 10%, in order to avoid accepting the null hypothesis when $k = 2$. It is also recommended to look at the graphs of the ME densities for various values of $k \geq 2$, superimposed on the histogram of the data, in order to ascertain whether the rejection of $k = 2$ was the correct decision.

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

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

⁷The routines for the ME test are available at <https://sites.google.com/site/sschiavo7788/home/software>.

where ξ is the Pareto shape parameter, q is the quadratic deviation from a Pareto, r is the rank, and x_r is the r -th order statistic. 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, the statistic $\sqrt{2n}q_n/\xi^2$ converges to a standard normal distribution, which can therefore be used to find the critical points of the test.

2.2 Data description

Trade data are taken from the COMTRADE database maintained by the United Nations. This collects data on bilateral trade flows among 157 reporting countries (sources) and 230 destinations. The finest disaggregation is the 6-digit level of the Harmonized System classification, which consists of roughly 5 000 products. Data are then aggregated up to total trade for each country-pair. In the analysis we focus on year 2007, which results in 6 002 617 non-null disaggregate bilateral flows, adding up to 20 767 country pairs exchanges. Data are expressed in thousands of US dollars (USD), and display a lower cutoff at 1,000 USD.

The firm size distribution is investigated by means of a unique longitudinal database that records sales figures of 340,560 products commercialized by 5 721 firms in 28 countries from 1994 to 2004, covering the whole size distribution for products and firms, and monitoring the flows of entry and exit at both levels. Data cover the worldwide pharmaceutical industry (Fu et al., 2005; Buldyrev et al., 2007). The pharmaceutical industry offers a unique context for empirical investigation relevant to our model, because it consists of many independent submarkets corresponding to different therapeutic groups within the industry (Sutton, 1997). Information is available both at the disaggregate level of product sales, as well as reaggregated by assigning each product to the firm that sells it. Data are in thousands pounds (GBP) with a lower cutoff at 1,000 GBP.

Information on the population of US cities is derived from the 2010 Census Data collected by the US Census Bureau. The elementary unit of analysis, corresponding to disaggregate data, is the population of each city block: we have data for 6 127 259 blocks. These figures are then aggregated into administrative units that represent populated places. As in Eeckhout (2004) we take populated places as the unit of analysis at the aggregate level.⁸ Rozenfeld et al. (2011) have recently claimed that the way cities are defined (i.e. the way elementary units are aggregated) is not neutral with respect to the shape of the resulting city size distribution. At present we are unable to replicate the aggregation using the clustering algorithm proposed in that paper and therefore rely on the administrative definition of cities. We do, however, perform our analysis on the clusters identified by Rozenfeld et al.

⁸In the rest of the paper the terms city and populated place are used interchangeably.

(2011), which are available on one of the authors' website.⁹

3 Empirical Results

We start the empirical analysis by fitting the maximum entropy density to the empirical distributions of both aggregate $P(S)$ and disaggregate $P(s)$ data. Results are displayed in Figure 1. Small observations are characterized by jumps and discontinuities that make estimation problematic, thus we have to truncate the distributions below a certain threshold. This is particularly true in the case of city size data at the disaggregate level, where we exclude all blocks with a population smaller than 54 (4 in natural logs). The fit with the ME distribution reveals that $k > 2$ for all the distributions, thus the best fit for the whole distributions is significantly different from Pareto $k = 1$ and Lognormal $k = 2$.

3.1 Test results

We first look at disaggregated data, i.e. commodities traded by country pairs, products sold by pharmaceutical companies in the world and block sizes in the US. Table 1 reports results for the three tests (UMPU, ME, GI) at the 5 and 1 percent level. Since the big picture is unaffected, in what follows we concentrate on the 5 percent level and only discuss other results when they convey further information. The table reports the highest numerical values of the rank associated with rejection of the null hypothesis of a power-law tail, as well as the associated percentile in parentheses. This means that the figures in the table represent the length of the Pareto tail in terms of number of observations (and in percentage of the sample size). We report the rank at which the tests start staying in the critical region and never go back to the acceptance region, so that we disregard instances where a test goes in the critical region but then is unable to reject the null hypothesis once we increase the sample size. In so doing we are giving more chances to the null hypothesis, which implies a possible overestimation of the length of the power-law tail.

Table 1: Test results on disaggregate data.

| | Trade ($n = 5\,152\,700$) | | Firms ($n = 536\,577$) | | Cities ($n = 1\,547\,203$) | |
|------|--------------------------------|----------------|-----------------------------|-----------------------------|---------------------------------|----------------|
| | 5 percent | 1 percent | 5 percent | 1 percent | 5 percent | 1 percent |
| ME | 2212 (0.04) | 2354 (0.04) | 8300 [†] (1.55) | 8500 [†] (1.58) | 3600 (0.06) | 3800 (0.06) |
| UMPU | 1276 (0.02) | 1637 (0.03) | 100 (0.02) | 200 (0.04) | 3600 (0.06) | 3800 (0.06) |
| GI | 482 (< 0.01) | 1573 (0.03) | 15 (< 0.01) | 18 (< 0.01) | 1870 (0.03) | 2953 (0.05) |

Rank (percentile) after which the power-law hypothesis is rejected.

[†] Between rank 100 and rank 8300 ME p-value is close to 5 percent.

⁹http://lev.ccny.cuny.edu/~hmake/soft_data.html.

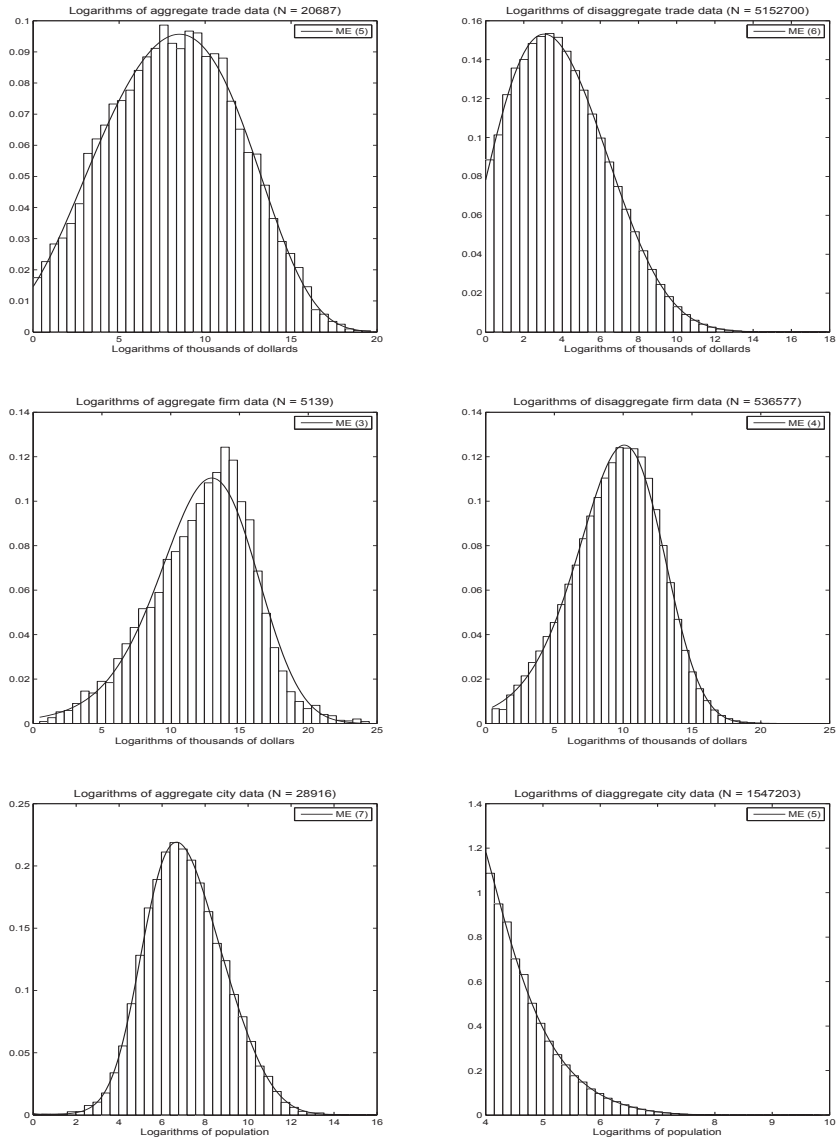


Figure 1: Maximum entropy estimates of empirical distributions for aggregate and disaggregate data. The sample sizes reported above the panels stands for the number of observations used to fit the ME distribution and generate the plots (k value into parentheses).

In all three domains the power-law tail appears to be limited to the very top of the distribution, and all tests show good agreement on this. The main difference can be found in the case of firms, where the UMPU and GI tests find almost no evidence of a power-law tail (the length is limited to the top 100 and top 15 observations respectively), whereas the ME test manages to definitely reject the null hypothesis only at rank 8300. This still represents just 1.5 percent of the sample and, as we show in Figure 2, the p -value of the ME test is often below the 5 percent threshold for ranks ranging between 100 and 8300.

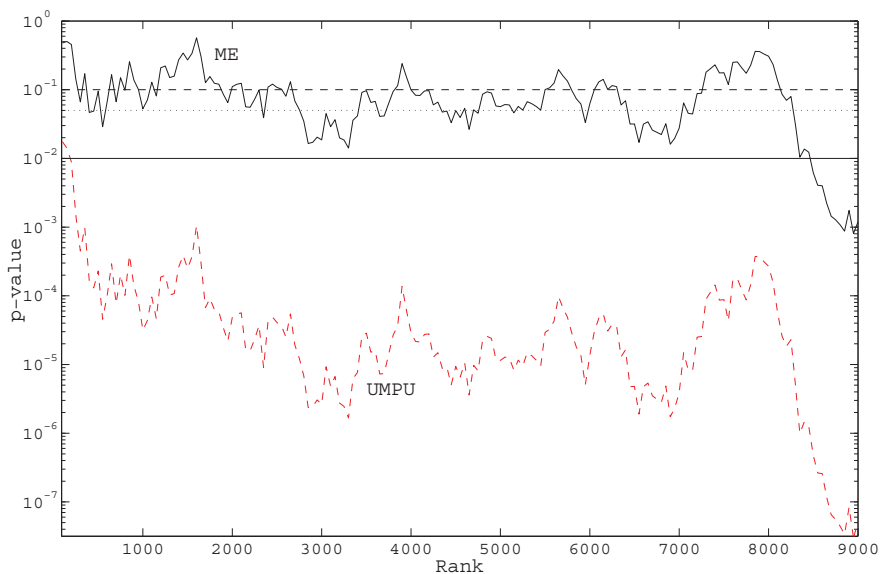


Figure 2: Log of p -value of the UMPU (dashed line) and ME (solid line) tests on disaggregate firm data. The horizontal lines represent the 5 and 1 percent significance thresholds.

Moving to aggregate data at the level of countries, firms or cities, we see that the length of the Pareto tail significantly increases in terms of percentile. Again, there is good agreement among the three tests, with the notable exception of the GI test, which in the case of firms rejects the null of a power-law for ranks larger than 14 at the 1 percent significance level.¹⁰ In any case, the three domains display a different behavior. The length of the power-law tail remains rather small in the case of aggregate trade flows, comprising only the top 107 or 110 observations for the UMPU and GI tests, or the top 676 according to ME, which corresponds to 3.26 percent of the sample. Firm size on the contrary displays a rather long power-law tail: even the most conservative estimate (provided by the UMPU test) suggests it spans 15.73 percent of the sample, reaching 23.6 percent according to ME. Cities lie

¹⁰At the 1 percent level the test identifies a power-law only for ranks between 562 and 1212, rejecting the null both before and after.

somewhat in between: moving from disaggregate to aggregate data does imply a marked difference, but the power-law tail here is limited to the top 1000 (ME and UMPU) or 1700 (GI) cities. These correspond to a population of about 39 500 and 24 300 inhabitants and represent between 3.4 and 6.1 percent of the whole sample. Moreover, they are in line with previous findings by Malevergne, Pisarenko and Sornette (2009) and Rozenfeld et al. (2011).

Table 2: Test results on aggregate data.

| | Trade ($n = 20\,687$) | | Firms ($n = 5139$) | | Cities ($n = 28\,916$) | |
|------|----------------------------|---------------|-------------------------|------------------------------|-----------------------------|----------------|
| | 5 percent | 1 percent | 5 percent | 1 percent | 5 percent | 1 percent |
| ME | 676 (3.26) | 772 (3.72) | 1350 (23.60) | 1400 (24.47) | 1030 (3.56) | 1250 (4.32) |
| UMPU | 110 (0.53) | 419 (2.02) | 900 (15.73) | 1100 (19.23) | 990 (3.42) | 1050 (3.63) |
| GI | 107 (0.51) | 165 (0.80) | 14 (< 0.01) | 1212 [†] (21.19) | 1759 (6.08) | 2159 (7.47) |

Rank (percentile) after which the power-law hypothesis is rejected.

[†] The test rejects the power-law hypothesis for ranks between 17 and 512 as well.

Taken together, Tables 1 and 2 provide evidence that in all the three domains under consideration here, the level of aggregation at which economic phenomena are studied plays an important role in determining the results and, in particular, the findings about the shape of the upper tail of the empirical distributions. In what follows we explore three different mechanisms through which such a behavior may emerge in the data, that cut across different domains and are not specific to the way international trade, firm size, or city growth evolve: one has to do with sample size and the associated power of the tests, the second with the shape of the aggregation function, and the third with the correlation between the size and the number of elementary units that compose aggregate entities.

3.2 Estimates of the shape parameter

Before moving to discuss the mechanisms that drive the change in the tail behavior upon aggregation, we take a look at the estimates of the shape parameter of the power-law portions of the distributions. Its magnitude has played an important part in the debate, especially with respect to cities, since Gabaix’s model implies a shape parameter equal to 1 (Zipf’s law). This prediction finds empirical support both in Gabaix (1999) and more recently in Rozenfeld et al. (2011); on the other hand, Eeckhout (2004) finds that the value of shape parameter changes significantly at different cutoffs, inferring from this that the distribution cannot be truly power-law. Finally, Malevergne, Pisarenko and Sornette (2009) report a coefficient significantly larger than 1.

Table 3 reports the estimates of the shape parameter obtained using the methodologies associated with the three tests. The estimation is performed at the cutoff identified by each of the tests. In particular, the estimate of the shape parameter is a byproduct of both the

Table 3: Estimates of the shape parameter

| | Aggregate data | | | | | |
|------|----------------|-----------|-----------|-----------|-----------|-----------|
| | Trade | | Firms | | Cities | |
| | 5 percent | 1 percent | 5 percent | 1 percent | 5 percent | 1 percent |
| ME | 1.40 | 1.40 | 0.50 | 0.53 | 1.30 | 1.27 |
| UMPU | 1.34 | 1.05 | 0.57 | 0.54 | 1.32 | 1.30 |
| GI | 1.82 | 1.82 | 1.74 | 0.59 | 0.92 | 0.94 |

ME and GI testing procedures, whereas in the case of UMPU we rely on the Hill estimator as done by Malevergne, Pisarenko and Sornette (2009). From the last two columns of the table we can see that the various estimates are in line with results presented by Malevergne, Pisarenko and Sornette (2009): the shape parameter takes values around 1.3. Different tests yield different estimates, with ME and UMPU being rather close to each other, whereas the GI result is more in line with the findings in Rozenfeld et al. (2011).

3.3 Emergence of a power-law upon aggregation

We investigate three candidate mechanisms that could explain the length of the power-law tail upon aggregation: the sample size, the shape of the aggregation function (i.e. the number of elementary units of which aggregate entities are made), and the correlation between the number of elementary units comprised in each aggregate element and their average size.

3.3.1 Sampling

To verify the impact of sample size on the results, for each of the three domains we run the tests on a sample of the same size of the aggregate datasets, obtained by simple random sampling from disaggregate data. So, for instance, in the case of trade we randomly select 20 687 observations among the 5 152 700 ones that constitute our disaggregate sample. Since detecting the difference between a lognormal and a Pareto tail is difficult and the tests have low power, in particular when n is small, the tests might well suffer the smaller sample size associated with more aggregate datasets and therefore have more troubles rejecting the null Pareto hypothesis. Our sampling exercise aims precisely at investigating the impact of such an effect.

In the case of trade, sampling does not modify the results reported in Table 2 above. Indeed, once we reduce the sample size of disaggregate data, the length of the Pareto tail grows as large as for aggregate data, but remains confined to the top 0.5 percent (UMPU and GI) or the top 3 percent (ME) of the distribution.

For what concerns firm size, the length of the power-law tail in sampled data, though longer than what observed in disaggregate data, is much smaller than the one displayed in Table 2. This implies that even if sample size does play a role in determining the length of

Table 4: Test results on synthetic datasets obtained by random sampling from disaggregate data with the size of aggregate ones.

| | Trade ($n = 20687$) | | Firms ($n = 5139$) | | Cities ($n = 28916$) | |
|------|--------------------------|---------------|-------------------------|----------------|---------------------------|-----------------|
| | 5 percent | 1 percent | 5 percent | 1 percent | 5 percent | 1 percent |
| ME | 690 (3.34) | 790 (3.82) | 610 (10.66) | 630 (11.01) | 3000 (10.37) | 3200 (11.07) |
| UMPU | 110 (0.53) | 430 (2.08) | 430 (7.52) | 610 (10.66) | 3000 (10.37) | 3200 (11.07) |
| GI | 105 (0.51) | 165 (0.80) | 690 (12.06) | 855 (14.94) | 655 (2.27) | 885 (3.06) |

Rank (percentile) after which the power-law hypothesis is rejected.

the power-law tail, its effect is unable to fully account for the difference we observe when moving from disaggregate to aggregate data.

The situation is reversed in the case of city size, where sampled data display —at least according to the UMPU and ME tests— a much longer power-law tail than the one found for actual aggregate observations. Indeed, as reported in Table 4, these two tests identify a power-law tail spanning roughly 3000 observations, i.e. more than 10 percent of the sample. This seems to imply that the reduction in the power of the tests associated with smaller sample size accounts for most of the power-law tail observed in city size data. Such a conclusion is partially tempered by results of the GI test, which finds a power-law tail limited to the top 655 observations in the sampled dataset. Table 4 suggests that in the case of cities, the impact of sample size could be substantial and explain a great deal of the length of the power-law tail found in the data. In this respect, the debate appears far from being closed, as claimed elsewhere (Malevergne, Pisarenko and Sornette, 2009) and a test on a larger sample of world cities should be performed.

To further investigate the issue we replicate the sampling exercise using, as reference point for aggregate data, the 17 569 clusters identified by Rozenfeld et al. (2011) as a better definition of cities with respect to the administrative definition of populated places. Table 5 reports test results for both the actual cluster data and a sample of the same size obtained by randomly selecting observations among the disaggregate data on block population. First, we note that the UMPU test displays a rather odd behavior as it identifies a power-law spanning 75 percent of the sample (13 110 observations). Results for ME and GI on the contrary are in line with those obtained using populated places (see Table 2 above): according to them the power-law tail starts at ranks 800 and 1990 (about 21 500 and 10 700 inhabitants) respectively. Second, when we perform the tests on the random samples, we find almost the same results as for actual data: the beginning of the Pareto tail is set at ranks 810 (ME), 13 130 (UMPU), 2002 (GI). Once again, this evidence suggests that sample size plays a relevant role in determining the power-law tail that is found in the distribution of city size. If possible, this conclusion is even stronger than before, as now all three tests point in

the same direction.

Table 5: Test results on data for population clusters, actual and sampled.

| | actual data ($n = 17\,569$) | | sampled data ($n = 17\,569$) | |
|------|----------------------------------|------------------|-----------------------------------|------------------|
| | 5 percent | 1 percent | 5 percent | 1 percent |
| ME | 800 (4.55) | 950 (5.41) | 810 (4.61) | 970 (5.52) |
| UMPU | 13110 (74.62) | 13310 (75.76) | 13130 (74.73) | 13230 (75.30) |
| GI | 1990 (11.33) | 2320 (13.21) | 2002 (11.40) | 2322 (13.22) |

Rank (percentile) after which the power-law hypothesis is rejected.

The difference between the results obtained with UMPU and the other two tests are macroscopic. To further investigate this behavior, in Figure 3 (left panel) we plot the complementary cumulative distribution function in double log scale, along with the thresholds identified by different tests. As a power-law should result in a straight line, we also plot a reference line with slope equal to the shape parameter estimated by the ME method (0.884). The graph shows a marked departure from linearity for population values well above the threshold found by UMPU. To better illustrate this finding, the right panel of Figure 3 shows the histogram of the logs of city cluster (CCA) corresponding to the power-law tail found by UMPU (rank 13 130) together with the optimal ME density ($k = 6$), the exponential (log of Pareto) and the truncated normal (log of lognormal). The last two are almost indistinguishable and fit the data rather poorly, whereas the optimal ME provides a very good fit. The sample coefficient of variation \hat{c} of the largest 13 110 observations is equal to 0.983 (and is even larger for larger thresholds): hence, being only based on \hat{c} , the UMPU test overestimates the length of the power-law tail.¹¹ Graphically, in a complementary CDF log-log plot such as the left panel of Fig. 3, this means that UMPU rejects the null of power-law for departures from linearity below the straight line, but not for those above. This example is a clear illustration of the possible UMPU pitfall described in Sect. 2.1.

3.3.2 Aggregation rule

By noting that aggregates are obtained by summing the size of the elementary units associated with each aggregate element, it is fairly easy to conclude that a very simple mechanism giving rise to a power-law tail upon aggregation is the shape of the aggregating function. Indeed, calling K_i the number of disaggregate elements (say products) comprised in aggregate object i (say firm), then a power-law distribution for K gives rise to a power-law distribution of aggregate sizes, if disaggregate units are (approximately) of the same size. Denoting aggregate size with S_i we have that $S_i = K_i \times \bar{s}_i$ where \bar{s}_i is the average size of the elementary units of aggregate i . If K_i is Pareto, and \bar{s} is independent of K_i for sufficiently

¹¹Recall from Section 2.1 that UMPU rejects the null of power-law for values of \hat{c} smaller than 1.

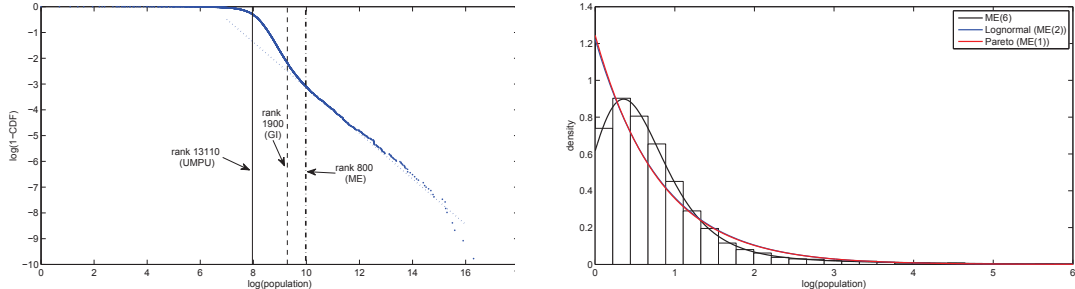


Figure 3: Left panel: Complementary cumulative distribution of cluster size, double logarithmic scale. The vertical lines mark the power-law cutoffs identified by the tests. The dotted reference line has slope equal to the shape parameter as estimated by the ME procedure ($\alpha = 0.884$). Right panel: histogram of the cluster data with superimposed fitted curves for Pareto, lognormal and ME(6) distributions.

large K , S_i will also be Pareto, and this holds true even if the sizes of elementary units are not themselves Pareto distributed (see Growiec et al., 2008, for a detailed explanation).

Figure 4 shows the complementary cumulative distribution (CCDF) of the number of elementary units associated with each aggregate element, in each of the three domains under investigation (in double logarithmic scale). We can see the number of products sold by each pharmaceutical firms and the number of blocks in each city are approximately Pareto in the tail, whereas the distribution of commodities traded by country-pairs is far less skewed.

Table 6 reports the results of the tests applied to the distribution of K . The intuition coming from the visual inspection of the CCDFs is confirmed: trade data display almost no power-law behavior, whereas a Pareto tail is present for both firms and cities (although in the former case the GI test does not fully agree with the others).¹²

Table 6: Test results on the aggregation function, $P(K)$.

| | Trade ($n = 20687$) | | Firms ($n = 5139$) | | Cities ($n = 28916$) | |
|------|--------------------------|---------------|-------------------------|----------------------------|---------------------------|-----------------|
| | 5 percent | 1 percent | 5 percent | 1 percent | 5 percent | 1 percent |
| ME | 330 (1.60) | 370 (1.79) | 1070 (20.82) | 1270 (24.71) | 2350 (8.12) | 2600 (8.99) |
| UMPU | 50 (0.24) | 70 (0.34) | 630 (12.26) | 810 (15.76) | 1300 (4.49) | 1650 (5.70) |
| GI | 41 (0.20) | 55 (0.27) | 26 (0.51) | 339 [†] (6.60) | 1890 (6.53) | 2900 (10.02) |

Rank (percentile) after which the power-law hypothesis is rejected.

[†] The test rejects the power-law hypothesis for ranks between 29 and 158 as well.

¹²This behavior of the GI test could be caused by the presence of a finite-size cutoff in the upper tail of the firm size distribution (Buldyrev et al., 2007). Since this cut-off has been repeatedly observed in many samples (see also Figure 4 for firms), it has been recently argued (Fujimoto and Watanabe, 2011) that both an upper and a lower threshold should be imposed before testing for a power-law.

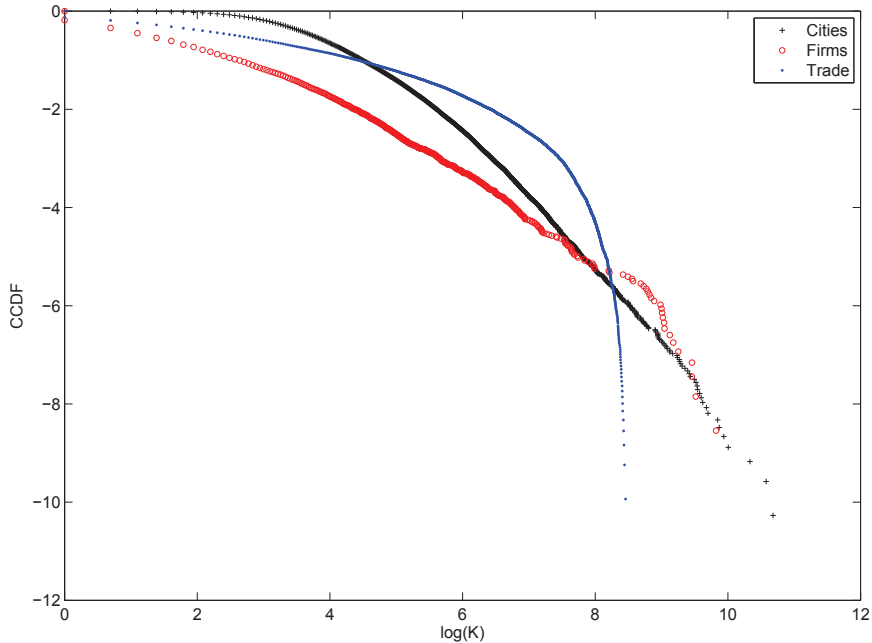


Figure 4: Complementary cumulative distribution of the number of elementary units in each aggregate element (number of commodities traded by country pairs, number of products by firm, number of blocks by city). Double logarithmic scale.

However, \bar{s} is neither constant nor independent of K , so that the presence of a power-law aggregation rule is not sufficient to generate a power-law in aggregate data. When looking at the relationship between the number of elementary units (K) and their average size (\bar{s}), we find they are positively correlated in all the three domains.

A positive correlation between the number of elementary units assigned to each aggregate element and their average size (which makes the ‘identity’ of units relevant) may result from a very skewed distribution for disaggregate data: this makes convergence to the central limit theorem rather slow upon aggregation. This correlation may end up stretching the upper tail of the aggregate distribution as it is more probable that aggregate elements made up of a larger number of units (large K), sum units of larger (average) size.

3.3.3 Correlation

To clean this effect from the data, we run the usual tests on three synthetic datasets obtained by aggregating elementary units according to a random rule. This means that in aggregating, say, firm data, we still allocate 536 577 products to 5139 firms, but instead of assigning the

actual products to each firms, we do it randomly under the condition that each firm receives the correct number of products. In so doing we do not modify the distribution of the number of products, just their sizes.

Table 7: Test results on synthetic datasets obtained by reshuffling data.

| | Trade ($n = 20\,687$) | | Firms ($n = 5139$) | | Cities ($n = 28\,916$) | |
|------|----------------------------|----------------|-------------------------|----------------|-----------------------------|----------------|
| | 5 percent | 1 percent | 5 percent | 1 percent | 5 percent | 1 percent |
| ME | 655 (3.17) | 720 (3.48) | 508 (9.89) | 698 (13.58) | 2231 (7.72) | 2227 (7.70) |
| UMPU | 618 (2.99) | 619 (2.99) | 189 (3.68) | 192 (3.74) | 1361 (4.71) | 1354 (4.68) |
| GI | 1052 (5.09) | 1172 (5.67) | 5 (< 0.01) | 102 (1.98) | 1278 (4.42) | 1278 (4.42) |

Rank (percentile) after which the power-law hypothesis is rejected.

Table 7 shows the average length of the power-law tail detected in 100 synthetic samples obtained by reshuffling the disaggregate data (e.g. products) before assigning them to the aggregate elements (e.g. firms). Comparing these results with those reported in Table 2 we can see that for trade and cities the power-law tends to be longer after reshuffling than in the original aggregate datasets, whereas it is much shorter for firms. More in details, for trade ME finds results in line with the original dataset, while both UMPU and especially GI find much heavier tails; as for cities, ME and UMPU report longer power-law tails in the synthetic data, whereas GI reports a shorter one. For what concerns firms instead, the length of the power-law tail tends to be much shorter according to all tests (bar GI at the 1 percent level, due to the peculiar behavior of the test in the original aggregate dataset, where it rejects the null of power-law for all ranks), so that washing out the correlation among elementary units greatly affects the results.

The overall message we derive from this exercise is that for trade and city data, very large aggregate elements are rarer than one would expected following a purely random assignment of elementary units (given the number of units assigned to each aggregate). So, for instance, if blocks were randomly assigned to cities, we should find more large cities than we actually see. One possible way to rationalize these results is by means of agglomeration diseconomies or congestion effects. More plainly, not all the blocks of a city are made up of densely populated skyscrapers.

On the contrary, a random allocation of products across firms would result in fewer very large firms than we find in the data. This can be explained by means of scale and scope economies, reputation effects, or positive spillovers within the firm (for instance in R&D, which is particularly important in the case of the pharmaceutical firms studied here).

The above results suggest that, first, the sample size plays a relevant role in determining the length of the power-law tail attributed to data on US cities and, second, that the correlation among the size and the number of elementary units explains almost half of the

power-law tail characterizing the size distribution of firms. How is it possible to explain the different aggregate behavior of the two domains given the similar shape of the aggregation rule depicted in Figure 4?

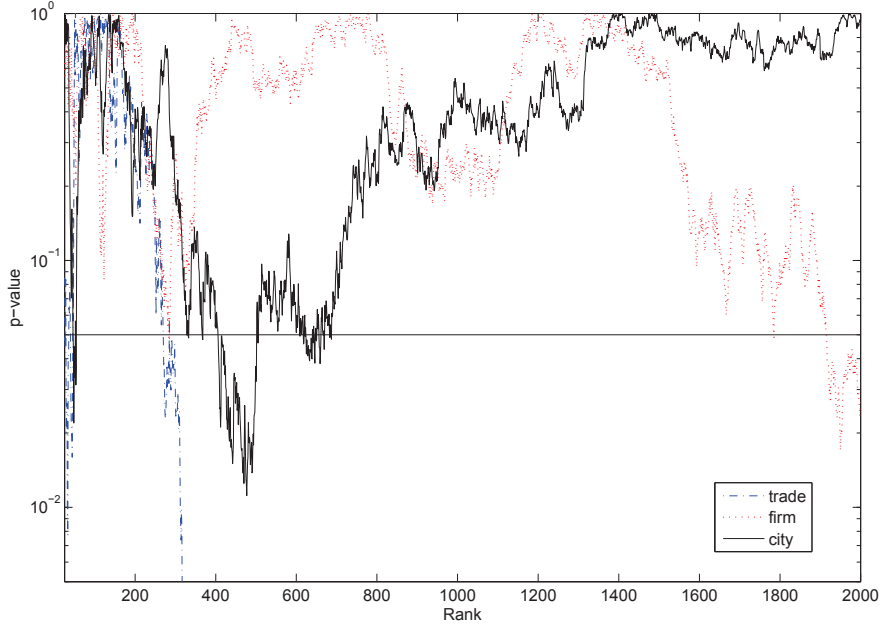


Figure 5: p -values of the quadratic term in the regression of the (log of the) average size of elementary units on the number of elements and its square term (in logs). The p -value for the regression on trade data drops to very small magnitudes at rank 300 and is therefore truncated.

Given the dependence of \bar{s} on K , we conjecture that to find a significant Pareto tail in the aggregate data, we need both a power-law aggregation rule and a power-law relationship between \bar{s} on K .¹³ To investigate the point we have run a series of regression of the type

$$\log(\bar{s}_i) = \beta_0 + \beta_1 \log(K_i) + \beta_2 (\log(K_i))^2$$

on the largest N observations in the three domains, with $N = 100, \dots, 2000$. A significant quadratic term ($\beta_2 > 0$) implies the rejection of the hypothesis of a power-law relationship. Figure 5 reports the p -value of a t -test on β_2 : we see that in the case of trade the null of a linear relationship (in double logs) is quickly rejected, whereas for both firms and cities

¹³A series of simulations confirms this idea: given a power-law distribution for K_i , when \bar{s}_i independent of K_i or when the dependence follows a power-law ($\bar{s}_i = K_i^\gamma$), then aggregate size S_i is Pareto, whereas if the dependence is, say, exponential ($\bar{s}_i = e^{K_i}$) this is no longer the case.

the picture is different. In particular, for what concerns firms, there is a linear, power-law relationship between the two variables, that spans the largest 1800 observations. Data on cities on the contrary reject linearity well before, suggesting that the relationship between \bar{s} and K is not power-law even for the largest observations.¹⁴ All in all, the Pareto tail of the firm size distribution is due to (i) a Pareto tail in the distribution of the number of products by firm, $P(K)$ and (ii) a power-law positive relationship between the number of products sold and their size, at least for large firms.

4 Discussion and Conclusion

The exact shape of the size distribution of economic aggregates is of crucial importance and it has been largely investigated both theoretically and empirically for wealth, income as well as city and firm sizes. However, despite it has been recognized that it is extremely difficult to discriminate between lognormal and Pareto extreme values, most of the literature so far has focused on a single aspect at a time and applied a single testing strategy, typically assumed to be the best one. In this paper we take a broader perspective considering (i) multiple economic distributions; (ii) alternative tests and (iii) different levels of aggregation of economic systems. We analyze the shape of distributions spanning three different domains, namely international trade flows, firm size and city size at different levels of aggregation. We find that the tail behavior of the three distributions changes upon aggregation, with the emergence of a power-law upper tail. However, the extent to which this happens is different: the power-law tail remains almost negligible in the case of trade, becomes longer in the case of cities and much longer in the case of firms.

More in general, despite the Pareto distribution has some nice properties that have made it a natural building block for economic modeling, by cross-checking our findings with multiple and rigorous statistical methods, we discover that the existence of the Pareto tail is limited to some rather peculiar circumstances, satisfied only in a limited number of empirical domains. If one adopted the restrictive criterion that there is a Pareto tail only when all tests agree, and the same tests do not find a Pareto tail for a lognormal distribution of the same size, then there would be no clear-cut evidence supporting the presence of a Pareto tail in any of the three domains under investigation.

Even if we do not follow this route, a number of considerations are still in order. First, it is worth noticing that when testing lognormal versus Pareto we are deliberately excluding possible alternative distributions both for the tail and for the whole distribution. Second, the validity of the results is severely constrained by sample size. This is particularly important whenever the variance of the distribution is large. Based on this consideration, we cast

¹⁴We have also estimated a semi-log version of the regression $\log(\bar{s}_i) = \beta_0 + \beta_1 K_i + \beta_2 K_i^2$: in this case the null of hypothesis of linearity represents an exponential relationship. This null is always rejected for trade and firms, whereas it cannot be rejected using the largest 1000 observations for cities.

new doubts as for the existence of a genuine Pareto tail in the case of the US city size distribution and conclude that further analysis should be performed on a larger sample of world cities. Third, since a Pareto tail emerges upon aggregation in all the three domains we analyze, we argue that the shape of the aggregation function is critically important. This has been recently stressed by Rozenfeld et al. (2011), who use a clustering algorithm to define cities, instead of the standard administrative definition. In particular, when the aggregation function is Pareto and the size of the elementary units is independent from their number in each aggregate entity, the shape of the aggregate size distribution is essentially the same as the shape of the distribution of the number of units. As for trade, since the intensive margin is not Pareto, aggregate trade data do not display a (significant) Pareto tail. Conversely, city and firm sizes are composed by a number of units that is Pareto distributed, at least in the upper tail. In case of no relationship between the number of elementary units and their average size, this would be sufficient to generate a power-law tail in aggregate data. However, we find evidence of a sizable relationship between the number and size of constituent parts of aggregate entities. In such a case, the functional relationship between the two matters.

Even if it is difficult to derive analytical results for the distribution of the sum of dependent heavy-tailed distributed random variables,¹⁵ numerical experiments suggest that when the size of the units is a power-law function of the number of elements of which an aggregate entity is composed and the number of units is Pareto distributed, the size of aggregate entities is also Pareto distributed. This is the case of firm size, for which a power-law relationship exists for the top 30 percent of the distribution; on the contrary, for cities, we reject the null of a power-law relationship already at the top 2 percent.

All in all, Pareto distributions emerge upon aggregation. When the size of the units is approximately independent from their number, the shape of the aggregation function is crucial, as already noticed by Gabaix (1999) and Gabaix and Ibragimov (2011). On the contrary if the size and number of units are interdependent such as in the case of firms, the relationship between the two should also be power-law for the aggregate distribution to display a Pareto tail. The dependence of the size of elementary units on their number deserves further research, as it may signal the presence of non-constant returns to scale which, only under specific conditions, give rise to a power-law distribution.¹⁶ Our empirical analysis suggests that the presence of increasing returns is crucial for the existence of the observed Pareto tail in firm data, and may not hold across all sectors of the economy.

The level at which the various phenomena are investigated has a great influence on results, that is on the length of the power-law tail found in the data. Theoretical models

¹⁵See Asmussen and Rojas-Nandayapa (2008) for some asymptotic results in the special case of lognormal random variables with dependence structure given by the Gaussian copula.

¹⁶For a discussion of the role of increasing and decreasing returns to scale in determining the size of cities, see Fujita, Krugman and Venables (2001, p. 225).

that aim at explaining the shape of the distribution should devote more attention to this aspect. Our results suggest that to adequately explain the emergence of a power-law tail one should focus on the factors that determine the shape of $P(K)$, i.e. on what is known as the extensive margin in international trade, namely the number of products exported, or the number of blocks entering a city, or the number of products sold by each firm, or, in general, the number of elementary units in each aggregate. When the main cause of a Pareto tail at the aggregate level is the skewed shape of the aggregation rule, the usual argument that assumes idiosyncratic shocks to cancel out upon aggregation breaks down. Hence, the aggregate distribution is still skewed. Along the same line, Gabaix (2011) shows that, if idiosyncratic shocks come from an heavy-tailed distribution, the central limit theorem and the familiar “ $1/\sqrt{n}$ ” rule do not hold, with the consequence that single (big) shocks are very important in determining aggregate fluctuations.

References

- Acemoglu, Daron, Vasco M Carvalho, Asuman E. Ozdaglar, and Alireza Tahbaz-Salehi.** 2011. “The Network Origins of Aggregate Fluctuations.” MIT Department of Economics Working Paper 11-23.
- Aoki, M., and H. Yoshikawa.** 2011. *Reconstructing macroeconomics: a perspective from statistical physics and combinatorial stochastic processes*. Cambridge University Press.
- Asmussen, S., and L. Rojas-Nandayapa.** 2008. “Sums of Dependent Lognormal Random Variables: Asymptotics and Simulation.” *Statistics & Probability Letters*, 78: 2709–2714.
- Axtell, Robert L.** 2001. “Zipf Distribution of U.S. Firm Sizes.” *Science*, 293(5536): 1818–1820.
- Bee, Marco, Massimo Riccaboni, and Stefano Schiavo.** 2011. “Pareto versus lognormal: A maximum entropy test.” *Physical Review E*, 84: 026104.
- Benhabib, J., A. Bisin, and S. Zhu.** 2011. “The distribution of wealth and fiscal policy in economies with finitely lived agents.” *Econometrica*, 79(1): 123–157.
- Bhattacharya, Kunal, Gautam Mukherjee, Jari Saramäki, Kimmo Kaski, and Subhrangshu S. Manna.** 2008. “The International Trade Network: weighted network analysis and modelling.” *Journal of Statistical Mechanics: Theory and Experiment*, 2.
- Buldyrev, S.V., J. Growiec, F. Pammolli, M. Riccaboni, and H.E. Stanley.** 2007. “The growth of business firms: Facts and theory.” *Journal of the European Economic Association*, 5(2-3): 574–584.

- Cabral, Luís M.B., and José Mata.** 2003. “On the evolution of the firm size distribution: Facts and theory.” *American Economic Review*, 1075–1090.
- Champernowne, D.G.** 1953. “A model of income distribution.” *The Economic Journal*, 63(250): 318–351.
- Chaney, Thomas.** 2011. “The Network Structure of International Trade.” NBER Working Paper Series 16753.
- Cox, David R., and David V. Hinkley.** 1974. *Theoretical Statistics*. Chapman and Hall.
- del Castillo, Joan, and Pedro Puig.** 1999. “The Best Test of Exponentiality against singly Truncated normal alternatives.” *Journal of the American Statistical Association*, 94: 529–532.
- di Giovanni, Julian, Andrei A. Levchenko, and Romain Rancière.** 2011. “Power laws in firm size and openness to trade: Measurement and implications.” *Journal of International Economics*, 85(1): 42–52.
- Easterly, William, Ariell Reshef, and Julia Schwenkenberg.** 2009. “The power of exports.” The World Bank Policy Research Working Paper Series 5081.
- Eaton, Jonathan, Samuel Kortum, and Francis Kramarz.** 2004. “Dissecting Trade: Firms, Industries, and Export Destinations.” *American Economic Review*, 94(2): 150–154.
- Eeckhout, Jan.** 2004. “Gibrat’s Law for (All) Cities.” *American Economic Review*, 94(5): 1429–51.
- Eeckhout, Jan.** 2009. “Gibrat’s Law for (All) Cities: Reply.” *American Economic Review*, 99(4): 1676–83.
- Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch.** 1997. *Modelling Extremal Events for Insurance and Finance*. Springer.
- Fagiolo, Giorgio, Javier Reyes, and Stefano Schiavo.** 2009. “World-trade web: Topological properties, dynamics, and evolution.” *Physical Review E*, 79(3): 036115.
- Fagiolo, Giorgio, Stefano Schiavo, and Javier Reyes.** 2008. “On the topological properties of the world trade web: A weighted network analysis.” *Physica A*, 387(15): 3868–3873.
- Fu, Dongfeng, Fabio Pammolli, Sergey V. Buldyrev, Massimo Riccaboni, Kaushik Matia, Kazuko Yamasaki, and H. Eugene Stanley.** 2005. “The growth of business firms: Theoretical framework and empirical evidence.” *Proceedings of the National Academy of Sciences of the United States of America*, 102(52): 18801–18806.

- Fujimoto, Shouji, Ishikawa Atushi Mizuno Takayuki, and Tsutomu Watanabe.** 2011. "A New Method for Measuring Tail Exponents of Firm Size Distributions." *Economics: The Open-Access, Open-Assessment E-Journal*, 5(2011-20).
- Fujita, Masahisa, Paul Krugman, and Anthony J. Venables.** 2001. *The Spatial Economy: Cities, Regions, and International Trade*. Vol. 1 of *MIT Press Books*, The MIT Press.
- Gabaix, Xavier.** 1999. "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics*, 114(3): 739–67.
- Gabaix, Xavier.** 2009. "Power Laws in Economics and Finance." *Annual Review of Economics*, 1: 255–93.
- Gabaix, Xavier.** 2011. "The Granular Origins of Aggregate Fluctuations." *Econometrica*, 79: 733–772.
- Gabaix, Xavier, and Rustam 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.
- Gibrat, Robert.** 1931. *Les Inegalites Economiques*. Paris:Sirey.
- Growiec, Jakub, Fabio Pammolli, Massimo Riccaboni, and H. Eugene Stanley.** 2008. "On the size distribution of business firms." *Economics Letters*, 98(2): 207–212.
- Hart, Peter E., and Sig J. Prais.** 1956. "The Analysis of Business Concentration: A Statistical Approach." *Journal of the Royal Statistical Society. Series A (General)*, 119(2): 150–191.
- Helpman, Elhanan, Mark J. Melitz, and Stephen R. Yeaple.** 2004. "Export versus FDI with heterogeneous firms." *American Economic Review*, 94(1): 300–316.
- Hinloopen, Jeroen, and Charles van Marrewijk.** 2012. "Power laws and comparative advantage." *Applied Economics*, 44(12): 1483–1507.
- Ijiri, Yuji, and Herbert A. Simon.** 1977. *Skew Distributions and the Sizes of Business Firms*. Amsterdam:North Holland.
- Levy, Moshe.** 2009. "Gibrat's Law for (All) Cities: Comment." *American Economic Review*, 99(4): 1672–75.
- Luttmer, E.G.J.** 2007. "Selection, growth, and the size distribution of firms." *Quarterly Journal of Economics*, 122(3): 1103.

- Malevergne, Yannick, Vladilen Pisarenko, and Didier Sornette.** 2009. “Gibrat’s law for cities: uniformly most powerful unbiased test of the Pareto against the lognormal.” *Swiss Finance Institute Research Paper Series*, 09–40.
- Malevergne, Yannick, Vladilen Pisarenko, and Didier Sornette.** 2011. “Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities.” *Phys. Rev. E*, 83(3): 036111.
- Marsili, Orietta.** 2005. “Technology and the Size Distribution of Firms: Evidence from Dutch manufacturing.” *Review of Industrial Organization*, 27(4): 303–328.
- Melitz, Mark J.** 2003. “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity.” *Econometrica*, 71(6): 1695–1725.
- Pareto, Vilfredo.** 1896. *Cours d’économie politique professé à l’Université de Lausanne*. Lausanne:F. Rouge.
- Perline, Richard.** 2005. “Weak and False inverse power laws.” *Statistical Science*, 20: 68–88.
- Quandt, Richard.** 1966. “On the size distribution of firms.” *American Economic Review*, 56: 416–432.
- Riccaboni, Massimo, and Stefano Schiavo.** 2010. “The structure and growth of weighted networks.” *New Journal of Physics*, 12(023003): 1–14.
- Rozenfeld, Hernan, Diego Rybski, Xavier Gabaix, and Hernan Makse.** 2011. “The Area and Population of Cities: New Insights from a Different Perspective on Cities.” *American Economic Review*, 101(5): 2205–25.
- Simon, Herbert A., and Charles P. Bonini.** 1958. “The Size Distribution of Business Firms.” *American Economic Review*, 48(4): 607–617.
- Stanley, Michael H. R., Sergey V. Buldyrev, Shlomo Havlin, Rosario N. Mantegna, Michael A. Salinger, and H. Eugene Stanley.** 1995. “Zipf plots and the size distribution of firms.” *Economics Letters*, 49(4): 453–457.
- Steindl, J.** 1965. *Random processes and the growth of firms: A study of the Pareto law*. Griffin London.
- Sutton, J.** 2002. “The variance of firm growth rates: the scaling puzzle.” *Physica a: statistical mechanics and its applications*, 312(3): 577–590.
- Sutton, John.** 1997. “Gibrat’s Legacy.” *Journal of Economic Literature*, 35(1): 40–59.

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

Zipf, George K. 1949. *Human Behavior and the Principle of Last Effort*. Cambridge: MA:Addison-Wesley.

Elenco dei papers del Dipartimento di Economia
pubblicati negli ultimi due anni

- 2011.1 *Leaving home and housing prices. The Experience of Italian youth emancipation*, Francesca Modena e Concetta Rondinelli.
- 2011.2 *Pareto versus lognormal: a maximum entropy test*, Marco Bee, Massimo Riccaboni e Stefano Schiavo.
- 2011.3 *Does a virtuous circle between social capital and CSR exist? A "network of games" model and some empirical evidence*, Giacomo Degli Antoni e Lorenzo Sacconi
- 2011.4 *The new rules of the Stability and Growth Pact. Threats from heterogeneity and interdependence*, Roberto Tamborini
- 2011.5 *Chinese reserves accumulation and US monetary policy: Will China go on buying US financial assets?* Luigi Bonatti e Andrea Fracasso
- 2011.6 *Taking Keller seriously: trade and distance in international R&D spillovers*, Andrea Fracasso e Giuseppe Vittucci
- 2011.7 *Exchange Rate Exposure under Liquidity Constraints*, Sarah Guillou e Stefano Schiavo
- 2011.8 *Global Networks of Trade and Bits*, Massimo Riccaboni, Alessandro Rossi e Stefano Schiavo
- 2011.9 *Satisfaction with Creativity: A Study of Organizational Characteristics and Individual Motivations*, Silvia Sacchetti e Ermanno Tortia
- 2011.10 *Do Monetary Incentives and Chained Questions Affect the Validity of Risk Estimates Elicited via the Exchangeability Method? An Experimental Investigation*, Simone Cerroni, Sandra Notaro e W. Douglass Shaw
- 2011.11 *Measuring (in)security in the event of unemployment: are we forgetting someone?* Gabriella Berloff e Francesca Modena
- 2011.12 *The firm as a common. The case of accumulation and use of common resources in mutual benefit organizations*, Ermanno C. Tortia
- 2012.1 *The implications of the elimination of the multi-fibre arrangement for small remote island economies: A network analysis*, Shamnaaz B. Sufrauj
- 2012.2 *Post Mortem Examination of the International Financia Network*, Matteo Chinazzi, Giorgio Fagiolo, Javier A. Reyes, Stefano Schiavo
- 2012.3 *International R&D spillovers, absorptive capacity and relative backwardness: a panel smooth transition regression model*, Andrea Fracasso and Giuseppe Vittucci Marzetti

2012.4 *The Impact of Financialization on the WTI Market*, Sergio Galli
Lazzeri

2012.5 *Business change in Italian regions. A spatial shift-share approach to
plant-level data*, Giuseppe Espa, Danila Filipponi, Diego Giuliani and
Davide Piacentino

2012.6 *A Trick of the (Pareto) Tail*, Marco Bee, Massimo Riccaboni and
Stefano Schiavo

PUBBLICAZIONE REGISTRATA PRESSO IL TRIBUNALE DI TRENTO