Economics Letters 120 (2013) 232–237

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

The size distribution of US cities: Not Pareto, even in the tail

Marco Bee [a], Massimo Riccaboni [b,c,*], Stefano Schiavo [d,a,e]

[a] Department of Economics and Management, University of Trento, via Inama 5, 38122 Trento, Italy
[b] LIME, IMT Institute for Advanced Studies, Piazza San Ponziano 6, 55100 Lucca, Italy
[c] Department of Managerial Economics, Strategy and Innovation (MSI), K.U. Leuven, Naamsestraat 69, 3000 Leuven, Belgium
[d] School of International Studies, University of Trento, Italy
[e] OFCE-DRIC, France

HIGHLIGHTS

- The entire distribution of US city size is neither a Pareto one nor a lognormal one.
- Based on multiple tests, we find that the largest US cities are not Pareto distributed.
- Tests on real data and samples draws from a lognormal distribution yield similar Pareto tails.
- Bootstrap exercises show that the length of the Pareto tail shrinks by increasing sample size.

ARTICLE INFO

Article history:
Received 3 March 2013
Received in revised form
17 April 2013
Accepted 19 April 2013
Available online 24 April 2013

JEL classification:
C14
D30
R12
R23

Keywords:
City size distribution
Zipf distribution
Power-law
Lognormal distribution
Maximum entropy

ABSTRACT

We question the claim that the largest US cities are Pareto distributed. We show that results of multiple tests on real data are similar to those obtained when the true distribution is lognormal, and largely depend on sample sizes.

1. Introduction

Recently, a lively debate has emerged on whether city size data are better approximated by a Pareto distribution or by a lognormal one (Eeckhout, 2004; Levy, 2009; Eeckhout, 2009; Malevergne et al., 2011; Rozenfeld et al., 2011; Ioannides and Skouras, 2013).

Beside the 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) predicts a Zipf's law, Eeckhout (2004) proposes an equilibrium theory to explain the lognormal distribution of cities. This debate is hampered by the difficulty to distinguish lognormal versus Pareto tails (Embrechts et al., 1997; Bee et al., 2011). Moreover, the contention is partly based on the difficulty of properly defining what a city is and, empirically, what is the correct measure to use.[1]

* Corresponding author at: Department of Managerial Economics, Strategy and Innovation (MSI), K.U. Leuven, Naamsestraat 69, 3000 Leuven, Belgium.
E-mail addresses: marco.bee@unitn.it (M. Bee), massimo.riccaboni@kuleuven.be (M. Riccaboni), stefano.schiavo@unitn.it (S. Schiavo).

[1] This point is made 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

0165-1765/$ – see front matter
http://dx.doi.org/10.1016/j.econlet.2013.04.035

While early studies focus on the largest US Metropolitan Statistical Areas (MSAs) only, recent contributions use data for all the populated places of the US and other countries. By so doing, Eeckhout (2004) shows that the size distribution of US cities is a lognormal one, not a power-law one as previously thought (at least since Zipf, 1949). A few years later, Levy (2009) acknowledged that the body of the city size distribution is well approximated by a lognormal distribution, but claimed that there are significant departures in the upper tail. In particular, the top 0.6% of the distribution, i.e., the MSAs, appear to fit better a power-law distribution. Eeckhout (2009) replied to these new findings by highlighting potential problems associated with the procedures used by Levy (2009) to identify the power-law tail.[2] Recently Malevergne et al. (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). Last, Ioannides and Skouras (2013) applied a switching model and found that the distribution is lognormal in the body, but robustly Pareto in the upper tail (top 5%).

We contribute to this debate by providing new evidence based on a through analysis of the tail behavior of the distribution and a number of counterfactual exercises. We conclude that the power-law behavior of the upper tail is less robust than previously claimed, due to the limited power of the available statistical tests (Perline, 2005).

2. Data and methodology

2.1. Data

We analyze the distribution of US city size: information 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 6 127 259 census 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.[3] Since it has been argued 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, we perform our analysis using both the administrative definition of cities and the clusters identified by Rozenfeld et al. (2011).[4]

2.2. Testing for a power-law tail

Discriminating between power-law (Pareto) and lognormal tail behavior is a difficult task. Although asymptotically the two distributions are mathematically different, the convergence of the lognormal to the asymptotic distribution is extremely slow (Perline, 2005), so the difference may be very small, to the extent that they are often practically indistinguishable for any finite sample size.

Given these difficulties, several tests have been proposed: the uniformly most powerful unbiased (UMPU) test developed by del Castillo and Puig (1999) and used by Malevergne et al. (2011); the maximum entropy (ME) test by Bee et al. (2011); and the test proposed by Gabaix and Ibragimov (GI henceforth; see Gabaix and Ibragimov, 2011).

The UMPU test is based on the fact that the logarithm of a truncated lognormal distribution is truncated normal, and the logarithm of a Pareto distribution is exponential. del Castillo and Puig (1999) have shown 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 $\bar{c} = \min\{1, \hat{\sigma}/\hat{\mu}\}$ of the logarithms of the observations, where μ and σ are the parameters of the truncated normal. The UMPU test only compares the null of a power-law distribution against the alternative of a lognormal distribution, and rejects the null hypothesis for small values of the coefficient of variation c. However, the coefficient of variation does not uniquely identify distributions with power-law tails. This implies that the UMPU test 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 distributions and $c < 1$ below the threshold (Bee et al., 2011). On the other hand, if the distribution below the threshold is not a power-law one but nonetheless has $c \geq 1$, as happens, for example, for the Weibull distribution with shape parameter equal to 1, the UMPU test is completely unreliable. A case that illustrates this point is the aggregate city size distribution studied below (see Section 3).

The ME approach entails maximizing the Shannon 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 ith theoretical and sample moments, n is the number of observations, and T is the function defining the characterizing moment.[5] 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 lognormal (i.e., the normal) distribution is an ME with $k = 2$. A log-likelihood ratio (llr) test of the null hypothesis $k = k^*$ against $k = k^* + 1$ is given by

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

From standard limiting theory, the llr test is asymptotically χ^2_1 and is optimal (Cox and Hinkley, 1974; Wu, 2003).[6]

The ME test is a by-product of a more general non-parametric approach to density estimation. It can indeed be shown that, 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 llr test (Bee, 2013). Referring the interested reader to Wu (2003) for details, the main advantages of the technique are that (i) it delivers the best (according to the ME criterion) approximating density, and allows one to assess whether it belongs to certain parametric families; (ii) if the true distribution is a Pareto one, it provides an estimate of the shape parameter; (iii) it does not consider a single alternative model, so pitfalls such as the one discussed for the UMPU test in Section 3 below are avoided.

distribution of 1947 US cities with more than 12,000 inhabitants (about 1000 cities with more than 5000 inhabitants for the UK).

2 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.

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

4 Data on clusters are available at http://lev.ccny.cuny.edu/~hmakse/soft_data.html.

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

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

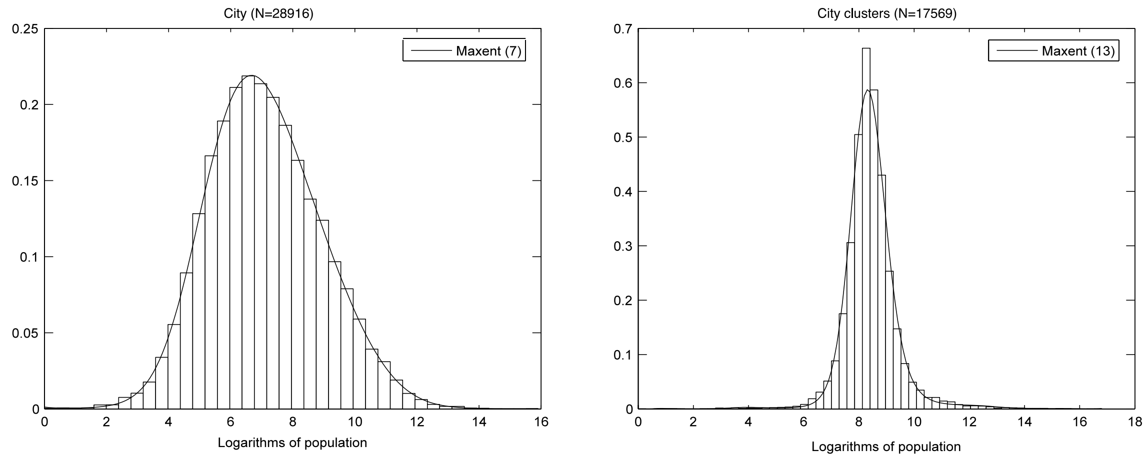

Fig. 1. Maximum entropy estimates of empirical distributions: cities (left) and city clusters (right). The sample sizes reported above the panels stands for the number of observations used to fit the ME distribution and generate the plots (k value in parentheses).

Finally, the GI test is based on the following intuition. Estimate by ordinary least squares (OLS) the regression

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

where ξ is the Pareto shape parameter, q is the quadratic deviation from a Pareto distribution, r is the rank, x_r is the rth order statistic, and $\gamma \overset{\text{def}}{=} \text{cov}((\log(x_r))^2, \log(x_r))/2\text{var}(\log(x_r))$ is a recentering term needed for guaranteeing that ξ is the same whether the quadratic term is included or excluded. Asymptotically, for the Pareto distribution, $q = 0$, so a large value of $|q|$ points towards rejection of the null hypothesis of a power-law distribution. Gabaix and Ibragimov (2011) show that, under the null hypothesis of a Pareto distribution, 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.

To further assess the accuracy of the tests, in this paper we conduct an experiment based on Monte Carlo simulation: we randomly draw from a lognormal distribution with sample size and parameters derived from the data and estimate the probability of obtaining a power-law tail at least as long as the one found by the tests. Hence, for each test, we run N Monte Carlo replications (with N equal to 500) and obtain the same number of independent realizations $t_1, \ldots, t_N$ of the discrete random variable B taking value r, where r is the rank such that the test is below the 95% critical value for ranks $1, \ldots, r-1$ and above this value for ranks $r, r+1, \ldots, n$. We then compute the quantity

$$\hat{p} = \frac{\#\{t_i > t^*\}}{N},$$

where t^* is the rank obtained applying the test to the observed data. A small value of $\hat{p}$ implies that the data display a tail which is significantly longer than the lognormal one used in the simulation.[7]

As a second robustness check, we perform a parametric bootstrap exercise. We simulate the ME distribution that best fits the populated places data and apply the tests to the resulting samples.[8]

Table 1
Test results.

	Cities ($n = 28\,916$)		Clusters ($n = 17\,569$)	
	5%	1%	5%	1%
ME	870	890	1180	1300
	(3.01)	(3.08)	(6.72)	(7.40)
UMPU	760	870	**13 110**	**13 310**
	(2.63)	(3.01)	(74.62)	(75.76)
GI	**962**	**1181**	1219	1566
	(3.33)	(4.08)	(6.94)	(8.91)

Rank (percentile) after which the power-law hypothesis is rejected. Boldface indicates a tail significantly longer than the one obtained simulating from a lognormal distribution with parameters estimated from the data.

3. Empirical analysis

In this section, we apply multiple tests to cities and city clusters to investigate the tail behavior of the distributions. First, we fit the ME density to the whole distribution. Next we run the statistical tests described above. Finally, to control of the presence of a Pareto tail, we perform a set of robustness checks.

3.1. Test results

We start the analysis by fitting the ME density to the empirical distributions of both cities and clusters. Results are displayed in Fig. 1. 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 a Pareto distribution ($k = 1$) and a lognormal distribution ($k = 2$).

Table 1 reports results of three statistical tests (UMPU, ME, GI) on cities and city clusters. The table shows 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.[9] Furthermore, boldface indicates that the length of the tail found in the data is significantly longer than the one found by applying the tests to a sample drawn from a lognormal distribution with parameters estimated from the

[7] In the following, we apply a 5% significance level ($\hat{p} < 0.05$).

[8] The ME distribution is simulated by means of the inverse distribution function method, where the ME distribution function is obtained by numerical integration of the ME density.

[9] This means that the figures 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 crosses the significance level and never go back to the acceptance region, so we disregard instances where a test goes in the critical region but bounces back 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.

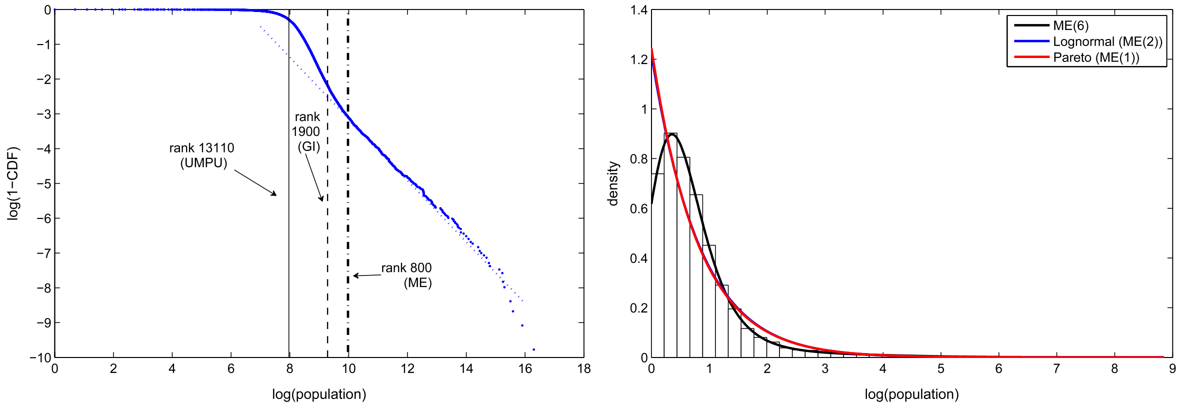

Fig. 2. Left panel: complementary cumulative distribution of cluster size. The vertical lines mark the power-law cutoffs identified by the tests. The dotted reference line has slope equal to the shape parameter as estimated by the ME procedure ($\alpha = 0.884$). Right panel: histogram of the cluster data with superimposed fitted densities.

data (see Section 2.2). Since the big picture is unaffected by the significance level, in what follows we concentrate on 5%, and only discuss other results when they convey specific information.

When we look at populated places, the power-law tail appears to be limited to the top 760 (UMPU), 870 (ME), or 960 (GI) observations. All tests show good agreement on this. The power-law tail starts at a population of about 51 500, 45 200, and 41 100 inhabitants, and it covers between 2.63% and 3.33% of the whole sample.[10] These results are in line with previous findings by Malevergne et al. (2011), Rozenfeld et al. (2011), and Ioannides and Skouras (2013). Note, however, that only the GI test finds a power-law tail significantly longer than the one we would observe if the true data-generating process were a lognormal one.

When we move to the 17 569 city clusters (Rozenfeld et al., 2011), we note that the UMPU test displays a rather odd behavior, as it identifies a power-law tail spanning 75% of the sample (13 110 observations). Results for ME and GI in contrast are in line with those obtained using populated places: according to them, the power-law tail starts at ranks 1180 and 1219 (about 15 400 and 15 000 inhabitants), respectively. The share of population that lives in the clusters comprised in the power-law tail is around 71–73%, significantly larger than the corresponding percentage for populated places (52–60%).

The difference between the results obtained with the UMPU test and the other two tests is macroscopic. To further investigate this behavior, in Fig. 2 (left panel) we plot the complementary cumulative distribution function (CDF) in double log scale, along with the thresholds identified by different tests. As a power-law relationship 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 the UMPU test. The right panel of Fig. 2 illustrates this finding by showing the histogram of the logs of city cluster sizes corresponding to the power-law tail found by the UMPU test (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 based only 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. 2, this means that the UMPU test rejects the null of a power-law distribution for departures from linearity below the straight line, but not for those above. This example is a clear illustration of the limitations of the UMPU test discussed in Section 2.2.

Table 2
Estimates of the shape parameter.

	Cities		Clusters	
	5%	1%	5%	1%
ME	1.30	1.27	0.87	0.89
UMPU[a]	1.32	1.30	1.24	1.23
GI	0.92	0.94	0.95	0.93

[a] These results are computed at the threshold found by the UMPU test, which is unreliable.

An important part in the debate on the size distribution of cities has been played by the value of the shape parameter associated with the power-law tail. This value has important implications, 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 the shape parameter changes significantly at different cutoffs, inferring from this that the distribution cannot be truly a power-law one. Finally, both Malevergne et al. (2011) and Ioannides and Skouras (2013) report a coefficient significantly larger than 1.

Table 2 shows the estimates of the shape parameter obtained using the methodologies associated with the three tests, performed at the cutoff identified by each of them. In particular, the estimate of the shape parameter is a byproduct of both the ME and GI testing procedures, whereas in the case of the UMPU test we rely on the Hill estimator, as done by Malevergne et al. (2011). Different tests yield different estimates, with the ME and UMPU values being larger than 1 and close to each other, whereas the GI result is more in line with the findings in Rozenfeld et al. (2011). When we shift the analysis to clusters, we need to take into consideration that the estimate corresponding to the UMPU test is unreliable, as it is computed at the threshold found by the test, which identifies a power-law tail spanning 75% of the data. Apart from this caveat, the most notable change occurs for the estimate obtained with the ME procedure, which falls significantly from 1.3 to 0.89, closer to the value implied by Zipf's law.

10 As noted elsewhere (e.g. Ioannides and Skouras, 2013), the skewness of the distribution implies that cities in the power-law tail account for a large chunk of total population, in this case 52–60% of inhabitants.

11 Recall from Section 2.2 that the UMPU test rejects the null of a power-law distribution for values of $\hat{c}$ smaller than 1.

So far, our results closely match what has already been found in the literature. However, we show that the UMPU test results leads to the misleading conclusion that almost the entire distribution of the city clusters is approximately of Pareto type. Moreover, in five out of six cases, the tests do not find a significant power-law tail when the results are compared to those obtained from a lognormal distribution.

To further investigate the role of sample size, we run now the tests on a random sample of census blocks data of the same size of cities. Since there is no theoretical prior that blocks should be power-law distributed, and indeed the same tests find almost none, this exercise allows us to single out the effect of sample size on the length of the observed power-law tail.[12]

Results are displayed in Table 3: the sampled data show – at least according to the UMPU and ME tests – a much longer Pareto tail than the one found for actual observations. Indeed, these two tests identify a power-law tail spanning roughly 3000 observations, i.e., more than 10% 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 the data. Indeed, the Pareto tail detected by both the ME and UMPU tests is longer than the one observed when simulating from a lognormal distribution. Hence, although a reduction of the sample size yields an increase in the length of the tail even when data come from a pure lognormal distribution, this effect is magnified in our city data. Such a conclusion is partially tempered by the results of the GI test, which finds a power-law tail limited to the top 655 observations in the sampled dataset, which is longer than in the case of census blocks, but still shorter than the one found for city sizes.

Table 3 suggests that, in the case of cities, the impact of sample size could be substantial, and could 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 et al., 2011), and a test on a larger sample of world cities should be performed.

Similar results are found in the case of city clusters: the beginning of the Pareto tail is set at ranks 810 (ME), 13 130 (UMPU), and 2002 (GI). Once again, this evidence suggests that sample size plays a relevant role in determining the power-law tail 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.

3.2. Simulation experiments

To provide further backing to our results, we run another sampling exercise. This time, we start from the ME distributions that best fit the city data, and sample by means of a parametric bootstrap exercise; we then apply the tests to the resulting samples with 50 000, 100 000, and 200 000 observations. Table 4 reports results obtained with 100 replications.[13]

We first consider a sample of equal size as the actual data ($n = 28\,916$) in order to check that the estimated ME distribution approximates them well and does not generate any bias. Indeed, the outcomes with the bootstrapped sample are analogous to those we get with the actual data (see Table 1): the power-law tail is limited to the largest 500–1000 observations (corresponding to the top 2.0–3.7% of the distribution) depending on the test and the significance level.[14] More interestingly, the table shows that, as expected, the length of the power-law tail decreases substantially by increasing the sample size.[15] Already for $n = 50\,000$ we observe a decline in the share of observations found to follow a power-law distribution, and the fall increases with the rise in the sample size. When we apply the tests to a sample of 200 000 observations, the relative length of the power-law tail drops by roughly one half in the case of the ME and UMPU tests, and by two thirds for the GI test. Correspondingly, the share of the population living in the cities belonging to the power-law tail according to the GI test decreases from around 55% to 37%. Similar results are found using the ME and UMPU tests.

Table 3
Test results on synthetic datasets obtained by random sampling from census blocks the same number of observations as for cities.

	Cities ($n = 28\,916$)		City clusters ($n = 17\,569$)	
	5%	1%	5%	1%
ME	**3000**	**3200**	810	970
	(10.37)	(11.07)	(4.61)	(5.52)
UMPU	**3000**	**3200**	**13 130**	**13 230**
	(10.37)	(11.07)	(74.73)	(75.30)
GI	655	885	2002	2322
	(2.27)	(3.06)	(11.40)	(13.22)

Rank (percentile) after which the power-law hypothesis is rejected. Boldface indicates a tail significantly longer than the one obtained simulating from a lognormal distribution with parameters estimated from the data.

Table 4
Test results obtained by parametric bootstrap from the fitted ME density on city data, for different sample sizes (n).

	$n = 28\,916$		$n = 50\,000$		$n = 100\,000$		$n = 200\,000$	
	5%	1%	5%	1%	5%	1%	5%	1%
ME	816.5	930	1107.5	1235	1617	1725.5	3059.5	3279.5
	(2.82)	(3.22)	(2.22)	(2.47)	(1.62)	(1.73)	(1.53)	(1.64)
UMPU	581.5	753.5	784.5	1006.5	1148.5	1419.5	1909.5	2405
	(2.01)	(2.61)	(1.57)	(2.01)	(1.15)	(1.42)	(0.95)	(1.20)
GI	837.45	1088.6	1138.04	1425.94	1470.31	1794.13	1961.8	2619.51
	(2.90)	(3.76)	(2.28)	(2.85)	(1.47)	(1.79)	(0.98)	(1.31)

4. Discussion and conclusion

Our results cast new doubts on the existence of a genuine Pareto distribution of city sizes. Irrespective of the empirical definition of a city, either as a populated place or as a city cluster, and using different statistical tests, we confirm that the power-law behavior is limited to a very small fraction of the largest cities. Hence, a first strong result is that the whole city size distribution is surely neither a Pareto one nor a Zipf one.

All in all, models predicting either a lognormal or a Pareto distribution for city size do not perform well in fitting the whole distribution. Sometimes, the latter hypothesis is preferred in the literature on the ground that the largest cities belonging to the power-law tail account for the majority of total population. Here, however, we find that even the power-law behavior of the upper tail is questionable, since it may well be due to sample size issues and the associated weaknesses of the testing procedures.

[12] When applied to census blocks, our tests find a power-law tail limited to the top 0.06% of the distribution. Additional details about these results are available upon request.

[13] Results using city clusters are qualitatively similar, although less clear-cut. This is due to the fact that the ME distribution does not provide a good fit for the whole distribution of city clusters (see Fig. 1). Results are available upon requests.

[14] To measure the goodness of fit, we count how many times (out of the 100 replications) the rank obtained by each test on the simulated data is larger (smaller) than the rank obtained by the same test on the actual data. In all cases the numbers range between 16% and 62%, merely reflecting statistical noise, not a significant bias in the procedure.

[15] To check the significance of the reduction in the length of the tail, we compute 90% confidence intervals for the mean quantile. As the confidence intervals do not overlap, we can conclude that there is a significant difference between the quantiles corresponding to increasing sample sizes.

The argument developed in the paper is threefold. First, test results on the actual data are rarely different from those obtained from a lognormal distribution. Second, starting from a distribution that displays almost no power-law behavior (that of census blocks) and reducing the sample size to match actual city data, the available tests find a power-law tail longer than in the real data themselves. Finally, when increasing the sample size by means of a parametric bootstrap exercise, the length of the power-law tail decreases, and the majority of the population is no more living in the Pareto upper tail of the distribution. This implies that deeper investigation on larger samples of world cities should further confine the range of validity of the Pareto behavior.

Overall, based on the best available data and statistical methods, we conclude that the Pareto distribution should not be considered a good first-cut approximation of the shape of the city size distribution. More in general, since Pareto behavior and lognormal behavior are universal properties of size distributions (Gabaix, 2009), and the discriminatory power of statistical tests is limited, theoretical model predictions should be evaluated on an extended set of empirical regularities specific to city growth.

Acknowledgments

The authors wish to thank Xavier Gabaix, Marco Modica, Hernan Rozenfeld, and Didier Sornette for useful comments on previous versions of the paper. M.R. acknowledges support from Progetto di Interesse Nazionale CrisisLab.

References

Bee, M., 2013. A maximum entropy approach to loss distribution analysis. Entropy 15 (3), 1100–1117.
Bee, M., Riccaboni, M., Schiavo, S., 2011. Pareto versus lognormal: a maximum entropy test. Physical Review E 84, 026104.
Cox, D.R., Hinkley, D.V., 1974. Theoretical Statistics. Chapman and Hall.
del Castillo, J., Puig, P., 1999. The best test of exponentiality against singly truncated normal alternatives. Journal of the American Statistical Association 94, 529–532.
Eeckhout, J., 2004. Gibrat's law for (all) cities. American Economic Review 94 (5), 1429–1451.
Eeckhout, J., 2009. Gibrat's law for (all) cities: reply. American Economic Review 99 (4), 1676–1683.
Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer.
Gabaix, X., 1999. Zipf's law for cities: an explanation. Quarterly Journal of Economics 114 (3), 739–767.
Gabaix, X., 2009. Power laws in economics and finance. Annual Review of Economics 1, 255–293.
Gabaix, X., Ibragimov, R., 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.
Ioannides, Y., Skouras, S., 2013. US city size distribution: robustly Pareto, but only in the tail. Journal of Urban Economics 73 (1), 18–29.
Levy, M., 2009. Gibrat's law for (all) cities: comment. American Economic Review 99 (4), 1672–1675.
Malevergne, Y., Pisarenko, V., Sornette, D., 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.
Perline, R., 2005. Weak and false inverse power laws. Statistical Science 20, 68–88.
Rozenfeld, H., Rybski, D., Gabaix, X., Makse, H., 2011. The area and population of cities: new insights from a different perspective on cities. American Economic Review 101 (5), 2205–2225.
Wu, X., 2003. Calculation of maximum entropy densities with application to income distribution. Journal of Econometrics 115, 347–354.
Zipf, G.K., 1949. Human Behavior and the Principle of Last Effort. Addison-Wesley, Cambridge: MA.