

Prieto, Faustino; Sarabia, José María

Conference Paper

Estimation of power laws for city size data: An optimal methodology

50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Prieto, Faustino; Sarabia, José María (2010) : Estimation of power laws for city size data: An optimal methodology, 50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/118918>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Estimation of power laws for city size data: An optimal methodology

Faustino Prieto and José María Sarabia

Department of Economics
University of Cantabria
Av. Los Castros, s/n
39005 Santander, Cantabria, Spain
Phone: (34) 942 201 630; Fax: (942) 201 603
e-mail: faustino.prieto@unican.es; sarabiaj@unican.es

Abstract

Power laws appear widely in many branches of economics, geography, demography and other social sciences. In particular, the upper tail of city size distributions appear to follow power laws, as many researchers have shown for different countries and different periods of times. A crucial point in the estimation of these laws is the correct choice of the truncation point. The aim of this paper is to investigate how to choice this truncation point from an optimal point of view. A new methodology based on the Akaike Information Criterion is proposed. An extensive simulation study is carried out in order to prove the existence of this optimal point, under different assumptions about the underlying population. Several kind of populations are considered, including lognormal and population with heavy tails. Finally, the methodology is used to optimal estimation of power laws in city size data sets for Spain for several years

Key Words: Power law; Pareto distribution; Akaike information criterion; city size.

JEL Classification: C52, O18, R12.

1. Introduction

Power laws appear widely in many branches of economics, finance, physics, computer science, demography and other social sciences. The upper tail of many sets of data: population of cities (Auerbach, 1913; Zipf, 1949; Bosker et al., 2008), personal income (Pareto, 1897; Clementi and Gallegati, 2005), size of firms (Axtell, 2001; Fujiwara et al., 2004; Growiec et al., 2008) and many other examples, all appear to follow a power law behavior, as many researchers have shown for different countries and different periods of times.

Power laws are also known as classical Pareto distributions (Arnold, 2008) or as Rank Size Rules (Stanley et al., 1995; Urzúa, 2000). For example, if we sort the largest cities of any country by population, we ordered them by its *size*, then we associate each city to the number of the sequence obtained, its *rank*, and we plot the *rank* versus the *size* along the y-axis and the x-axis respectively on a log-log scale, we get a straight line with negative slope, confirming empirically that power law behavior.

Many procedures have been proposed to estimate power laws. The two most commonly used (Gabaix and Ioannides, 2004) are Zipf Regression (Gabaix, 1999a; Gabaix, 1999b; Balakrishnan et al., 2008) and Hill Estimator (Hill, 1975). Both of them have several problems (Gabaix and Ioannides, 2004; Bauke, 2007; Clauset et al., 2009). In this paper, we have chosen to follow Zipf Regression method for two reasons: first, because unfortunately empirical research always manage finite samples and the bad non-asymptotic properties of the Hill estimator can be worrisome in finite samples (Embrechts et al., 1997; Gabaix and Ioannides, 2004), and second, because empirical literature has provided a large list of data sets well described with Zipf Regression method (Rosen and Resnick, 1980; Krugman, 1996; Zanette and Manrubia, 1997; Brakman et al., 1999; Davis and Weinstein, 2002; Moura and Ribeiro, 2006; Nitsch, 2005; Soo, 2005).

Power laws are usually valid only in the upper tail. So, a crucial point in the estimation of power laws is the correct choice of the truncation point. With Zipf Regression method, the three most commonly methods used to fix the limits of the upper tail are (Cheshire, 1999): a fixed number of data, a fixed size threshold and a size sample above which that sample accounts for some given proportion of the total. A pitfall of those three methods is that all of them depend on the decision each researcher makes, they are not objective criteria. The aim of this paper is to investigate how to choose this truncation point from an optimal point of view.

The contents of this paper are as follows. In *Section 2*, a new methodology, based on the Akaike Information Criterion, is proposed to estimate power laws optimally. In *Section 3*, an extensive simulation study is carried out to prove the existence of the optimal truncation point, under different assumptions about the underlying population. An empirical application to city size, with Spanish data, is included in *Section 4*. Finally, some conclusions are given in *Section 5*.

2. Choosing the truncation point optimally

Power laws are defined, in terms of cumulative distribution function (cdf), as follows

$$F(x) = \Pr(X \leq x) = 1 - \left(\frac{x}{\sigma}\right)^{-\alpha}, \quad x \geq \sigma \geq 0 \quad (1)$$

and $F(x) = 0$ if $x < \sigma$, where $\alpha > 0$ is a shape parameter and σ is a scale parameter. The α parameter will be called Pareto coefficient. The Zipf's law corresponds to the choice $\alpha = 1$ in Eq. (1) (see, for example, Urzúa, 2000; Fujiwara et al., 2004; Anderson and Ge, 2005). A generalization has been proposed by Sarabia and Prieto (2009). A random variable X with cdf given by Eq. (1) will be represented by $X \square Pa(\alpha, \sigma)$. By taking logarithms of both sides of the Eq. (1), we obtain a linear expression in $\log x$

$$\log[1 - F(x)] = \log(\sigma^\alpha) - \alpha \log x \quad (2)$$

so, a power law with Pareto coefficient α can be seen as a straight line with negative slope α on a log-log plot. Also, in terms of rank, we have

$$\log(rank) = C - \alpha \log(size_{(i)}) \quad (3)$$

where we have considered the ordered sequence of n data values of the random variable X :

$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(i)} \leq \dots \leq x_{(n)}$, and where we have considered $size_{(i)} = x_{(i)}$, $rank = (n+1) - i$ and $C = \log[(n+1)\sigma^\alpha]$.

A new methodology, based on the Akaike Information Criterion (AIC), is proposed, in order to choose the truncation point to estimate a power law from an optimal point of view. Akaike Information Criterion (AIC) is defined by Akaike (1974) as follows

$$AIC = 2 \log l - 2k \quad (4)$$

where k is the number of parameters in the statistical model and $\log l$ is the log-likelihood of the model evaluated at the maximum likelihood estimates. The AIC is a measure of the goodness of fit of an estimated statistical model and a useful tool for model selection. Let $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ be the ordered sequence of n data values of X (a random variable with power law behavior in the upper tail). Let $x_{(p)} \leq x_{(p+1)} \leq \dots \leq x_{(n)}$ the upper tail, defined by a threshold p and by a size $N = (n+1) - p$. Under the assumptions that the model errors are

normally and independently distributed, under Zipf regression method given by $y = X\hat{\beta} + \hat{u}$, AIC can be expressed as follows (Goldberger, 1991; Gujarati, 2005)

$$AIC = e^{2kIN} \frac{\hat{u}^T \hat{u}}{N} = e^{2kIN} \frac{y^T y - \hat{\beta} X^T y}{N} = e^{2kIN} \frac{1}{N} \sum_{i=p}^n \hat{u}_i^2 \quad (5)$$

In the particular case of power laws ($k = 2$),

$$AIC = e^{4IN} \frac{1}{N} \sum_{i=p}^n \hat{u}_i^2 \quad (6)$$

The new methodology proposed to choose optimally the limit of the upper tail to estimate a power law is to find the value p where Akaike Information Criterion (AIC) is minimum.

$$P_{optimal} = \underset{p}{\operatorname{argmin}} AIC \quad (7)$$

where $P_{optimal}$ is the truncation point proposed. Then, the optimal size of the upper tail proposed to estimate a power law, $N_{optimal}$, is

$$N_{optimal} = (n+1) - P_{optimal} \quad (8)$$

3. Montecarlo Simulation Study

In this section, we carry out a simulation experiment to study the performance of the proposed method and to prove the existence of the optimal truncation point, under different assumptions about the underlying population. The experiment has been designed considering two populations, Lognormal and Singh-Maddala, for two reasons: they are the most usual population in the context of this paper (urban economics, etc) and they correspond with two important types of tails (lognormal and heavy tail). Lognormal distribution (Johnson et al., 1994), denoted by $X \square LN(\mu, \sigma^2)$, is defined in terms of cumulative distribution functions (cdf) as follows

$$F(x) = \Pr(X \leq x) = \Phi\left(\frac{\log(x) - \mu}{\sigma}\right), \quad x \geq 0 \quad (9)$$

and $F(x) = 0$ if $x < 0$, where Φ is the standard normal cdf. SinghMaddala distribution, denoted by $X \square SM(a, b, q)$, is defined in terms of cdf by (Singh and Maddala, 1976)

$$F(x) = \Pr(X \leq x) = 1 - \left[1 + \left(\frac{x}{b}\right)^\alpha\right]^{-q}, \quad x \geq 0 \quad (10)$$

First, we generate R times a uniform $[0,1]$ random sample with n values each one, taking $R = 300, 10000$ in order to analyze the influence of the number of samples and taking $n = 500, 1000, 3000$ to analyze the influence of the sample size. Then, we generate R times a random sample of X with n values each one with the inverse of the corresponding cdf (Rios et al., 1997), taking $X \square LN(\mu, \sigma^2)$ and $X \square SM(a, b, q)$ in order to analyze the influence of different assumptions about the underlying population. With respect to the lognormal assumption, *Table 1* shows the different values of the parameters (μ, σ) taken, and the coefficient of variation $CV(X)$, mean $E(X)$, standard $STD(X)$ values corresponding to them, which make 30 different simulation scenarios with R samples of n values each scenario.

Table 1: Design of Lognormal simulation experiment, $CV(X), \mu, \sigma, E(X), STD(X)$ values used

$CV(X)$	μ	σ	$E(X)$	$STD(X)$
0.1003	1.0000	0.1000	2.7319	0.2739
0.5003	1.0000	0.4726	3.0394	1.5205
1.0000	1.0000	0.8326	3.8442	3.8442
1.5001	1.0000	1.0857	4.9007	7.3515
4.9998	1.0000	1.8050	13.8601	69.2978

With respect to the SinghMaddala assumption, *Table 2* shows the different values of the parameters (a, q, b) taken, and the mean $E(X)$, variance $VAR(X)$, coefficient of variation $CV(X)$ values corresponding to them (Kleiber and Kotz, 2003), simulating three types of scenarios, with infinite mean ($q \leq \frac{1}{a}$), finite mean and infinite variance ($\frac{1}{a} < q \leq \frac{2}{a}$), finite mean and finite variance ($q > \frac{2}{a}$), which make 174 different simulation scenarios with R samples of n values each scenario.

In total, 204 different scenarios which were simulated in the ALTAMIRA node of the Spanish Supercomputing Network.

Table 2: Design of SinghMaddala simulation experiment, a , q , b , $E(X)$, $VAR(X)$, $CV(X)$ values used

a	q	b	$E(X)$	$VAR(X)$	$CV(X)$
1/2	3/2	1	Inf.		
1/2	2	1	Inf.		
1/3	2	1	Inf.		
1/3	3	1	Inf.		
1/2	3	1	1	Inf.	Inf.
1/2	4	1	1/3	Inf.	Inf.
1/3	4	1	1	Inf.	Inf.
1/3	5	1	1/4	Inf.	Inf.
1/3	6	1	1/10	Inf.	Inf.
1/2	5	1	1/6	0.9722	5.9161
1/2	6	1	1/10	0.1900	4.3589
1/2	7	1	1/15	0.0622	3.7417
1/2	10	1	1/36	0.0090	3.0472
1/3	7	1	1/20	0.9975	19.9750
1/3	8	1	1/35	0.1420	13.1909
1/3	9	1	1/56	0.0354	10.5356
1/3	12	1	1/165	0.0021	7.6111
2	2	1	$\pi/4$	0.3831	0.7881
2	3	1	$3\pi/16$	0.1530	0.6641
2	4	1	$5\pi/32$	0.0924	0.6192
2	6	1	$63\pi/512$	0.0506	0.5817
5	3/5	1	1.3676	0.8648	1.4892
5	4/5	1	1.1745	0.3111	0.4749
5	1	1	1.0690	0.1787	0.3954
5	2	1	0.8552	0.0615	0.2899
10	2/5	1	1.2831	0.2540	0.3928
10	3/5	1	1.1323	0.0855	0.2582
10	4/5	1	1.0607	0.0493	0.2094
10	1	1	1.0166	0.0354	0.1851

Second, for each scenario, we calculate for each one of the R samples of that scenario, the value of $N_{optimal}$ using the new method proposed in *Section 2*, $Size_{optimal}$ which is the value of X in $P_{optimal}$ and then, the Ordinary Least Square (OLS) estimators, \hat{C} and $\hat{\alpha}$ in Eq. (3). After that, for each scenario, we calculate the mean of the R estimators obtained, $\bar{N}_{optimal}$, $\overline{Size_{optimal}}$, \bar{C} , $\bar{\alpha}$; the standard deviation of those estimators $\sigma_{N_{optimal}}$, $\sigma_{Size_{optimal}}$, $\sigma_{\hat{C}}$, $\sigma_{\hat{\alpha}}$; and we obtain the frequency histogram of those estimators. So, as result of the simulation experiment made, we get 204 values of $\bar{N}_{optimal}$, $\overline{Size_{optimal}}$, \bar{C} , $\bar{\alpha}$; 204 values of $\sigma_{N_{optimal}}$, $\sigma_{Size_{optimal}}$, $\sigma_{\hat{C}}$, $\sigma_{\hat{\alpha}}$; and 204 frequency histograms; which permit us to analyze the behavior of the method proposed in *Section 2* under different assumptions: different number of samples, different size of the sample and different kinds of population, and in order to prove the existence of the optimal truncation point. *Tables 3 and 4* show the results obtained with Lognormal distribution. All the results and histograms obtained are available via email for those people interested in them.

Third, we analyze the results obtained. The most relevant are:

- Existence of the optimal truncation point, obtained by the method proposed in *Section 2*, in all the 204 simulation scenarios.
- Under changes of R : Stability of \bar{C} , $\bar{\alpha}$, $\bar{N}_{optimal}$, $\overline{Size_{optimal}}$ and $\sigma_{\hat{C}}$, $\sigma_{\hat{\alpha}}$, $\sigma_{N_{optimal}}$, $\sigma_{Size_{optimal}}$ (with exception of $\overline{Size_{optimal}}$ and $\sigma_{Size_{optimal}}$ in a few scenarios of SinghMaddala simulation with $E(X)$ infinite).
- Under changes of n : \bar{C} , $\bar{\alpha}$, $\bar{N}_{optimal}$, $\overline{Size_{optimal}}$ and $\sigma_{N_{optimal}}$, $\sigma_{Size_{optimal}}$, increase when n increases; there is no relationship between changes of n and changes of $\sigma_{\hat{C}}$, $\sigma_{\hat{\alpha}}$.
- Under changes of (μ, σ) or (a, b, q) : Stable patterns in the changes of \bar{C} , $\bar{\alpha}$, $\bar{N}_{optimal}$, $\overline{Size_{optimal}}$ and $\sigma_{\hat{C}}$, $\sigma_{\hat{\alpha}}$, $\sigma_{N_{optimal}}$, $\sigma_{Size_{optimal}}$ values in agreement with the (μ, σ) or (a, b, q) values changed. Some examples: under lognormal assumptions, increments of coefficient of variation make decrements of $\bar{C}, \bar{\alpha}, \sigma_{\hat{C}}, \sigma_{\hat{\alpha}}$, make increments of $\overline{Size_{optimal}}$, $\sigma_{Size_{optimal}}$ and have no relation with $\bar{N}_{optimal}, \sigma_{N_{optimal}}$ changes; under SinghMaddala

assumptions, increments of q make decrements of \bar{C} , $\bar{N}_{optimal}$, $\overline{Size}_{optimal}$ and $\sigma_{N_{optimal}}$ and increments of $\bar{\alpha}$, stability of \bar{C} , $\bar{N}_{optimal}$, under changes of a .

In summary, stability of the method proposed under changes in the number of the samples, under changes in the size of the sample and under different assumptions about the population, and existence of the optimal truncation point obtained by the method proposed in all the simulation scenarios.

4. An empirical application to city size

The methodology proposed can be used for optimal estimation of power laws in city size data. In this paper, we have considered Spanish city size data. In short term, migration is the main driver for city size fluctuations and Spain is the country with the biggest change in the international migrant stock between years 2000 and 2005 (see International Migration Wall Charts 2002 and 2006, Department of Economic and Social Affairs of United Nations). Data sets used are composed of information of the population of the Spanish cities for ten years, from 1998 to 2007 and dated on January 1st of each year, published by the Spanish National Statistic Institute (INE).

There is no agreement in the definition of a city in the literature related. We have adopted the definition of a city as a municipality corresponding to local administrative and jurisdictional entity considered by Anderson and Ge (2005).

Some relevant information about data sets used appears in *Table 5*. *Table 6* shows the minimum AIC and $N_{optimal}$ values obtained, providing empirical evidence of the existence of the optimal truncation point. It can be seen that AIC is very close to zero in those points. *Figure 1* shows, for five of the ten year considered, on the left all the values of AIC and on the right the values of AIC next to the optimal truncation point. *Table 7* shows C and α values obtained by Zipf Regression, using the optimal truncation points of *Table 6*. *Figure 2* shows graphically how Spanish city size data, in the ten years considered, are well described in the upper tail by a power law estimated by the method proposed in *Section 2*.

Table 3: Lognormal simulation results, mean.

n / R	$CV(X)$	\bar{C}	$\bar{\alpha}$	$\bar{N}_{optimal}$	$\bar{Size}_{optimal}$
500/300	0.1003	29.1887	22.2639	97.0633	3.0076
	0.5003	11.8245	4.8114	92.2233	4.5723
	1.0000	9.8382	2.7690	91.4000	7.2134
	1.5001	8.9885	2.0661	95.0900	9.3360
	4.0998	8.2585	1.2598	93.2667	26.9000
500/10000	0.1003	29.4069	22.4620	95.2007	3.0116
	0.5003	11.7479	4.7783	93.9560	4.5345
	1.0000	9.6757	2.7106	94.5506	6.9330
	1.5001	8.9961	2.0628	94.8192	9.5824
	4.0998	8.2578	1.2587	92.9501	26.9370
1000/300	0.1003	32.8920	24.8024	126.9833	3.0914
	0.5003	13.1186	5.1332	135.1233	4.9327
	1.0000	10.9346	2.9300	132.1567	8.3607
	1.5001	10.2189	2.2379	134.1000	11.9519
	4.0998	9.3108	1.3355	141.0400	39.3109
1000/10000	0.1003	31.8335	23.9148	138.8054	3.0744
	0.5003	13.0239	5.0818	138.2511	4.9668
	1.0000	10.8169	2.8833	139.4899	8.1089
	1.5001	10.1604	2.2152	136.9196	11.8117
	4.0998	9.3221	1.3405	135.5760	38.1956
3000/300	0.1003	35.2587	25.8373	246.7233	3.1663
	0.5003	14.8216	5.4339	263.9733	5.6067
	1.0000	12.7447	3.1732	237.1500	10.5888
	1.5001	11.9789	2.4336	237.7200	16.1167
	4.0998	11.0054	1.4600	240.3000	59.7772
3000/10000	0.1003	35.8401	26.3118	240.0072	3.1719
	0.5003	15.1222	5.5784	240.9278	5.7303
	1.0000	12.7422	3.1774	239.4232	10.4551
	1.5001	11.9937	2.4340	241.8122	16.2408
	4.0998	10.9940	1.4585	242.4453	61.2042

Table 4: Lognormal simulation results, standard deviation.

n/R	$CV(X)$	$\sigma_{\hat{c}}$	$\sigma_{\hat{\alpha}}$	$\sigma_{N_{optimal}}$	$\sigma_{Size_{optimal}}$
500/300	0.1003	7.8060	6.4851	55.5822	0.1516
	0.5003	3.0794	1.5649	53.6610	1.1327
	1.0000	2.4316	0.9046	56.1974	3.3394
	1.5001	2.2716	0.7254	52.7239	5.5848
	4.0998	1.8412	0.3869	54.7384	29.8910
500/10000	0.1003	8.3662	6.9787	54.7430	0.1489
	0.5003	2.9189	1.4920	54.2249	1.1209
	1.0000	2.3036	0.8609	54.9270	3.1394
	1.5001	1.9693	0.6201	55.7112	5.8816
	4.0998	1.9164	0.4109	54.4663	29.6718
1000/300	0.1003	10.3587	8.5277	80.5126	0.1407
	0.5003	3.3777	1.6181	75.1204	1.0481
	1.0000	2.3418	0.8262	84.1862	3.5902
	1.5001	2.1532	0.6338	77.5331	7.2801
	4.0998	2.1866	0.4314	87.2587	46.3137
1000/10000	0.1003	8.2094	6.7285	83.8760	0.1448
	0.5003	3.0510	1.4938	83.7424	1.1691
	1.0000	2.3599	0.8324	84.2074	3.5759
	1.5001	2.1830	0.6385	83.4262	6.9791
	4.0998	2.0288	0.4021	83.3587	42.4830
3000/300	0.1003	7.1013	5.7131	165.9498	0.1373
	0.5003	2.7734	1.2733	168.3849	1.2879
	1.0000	2.3787	0.7629	155.1701	4.3872
	1.5001	2.1242	0.5700	155.5715	8.7128
	4.0998	1.8196	0.3347	160.0925	62.2124
3000/10000	0.1003	8.1120	6.4902	155.6780	0.1375
	0.5003	3.1347	1.4297	156.6891	1.2519
	1.0000	2.5131	0.8136	155.9896	4.2704
	1.5001	2.4948	0.6619	158.1800	9.2960
	4.0998	2.0455	0.3682	158.4850	67.4772

Table 5: Some relevant information about Spanish city size data sets used

Country	Year	Minimum city size considered	Cities considered		Population considered	
		(People)	(Number)	(% of total)	(Number)	(% of total)
Spain	1998	26	8,027	99.12	39,851,234	99.9964
	1999	26	8,030	99.12	40,200,753	99.9965
	2000	25	8,039	99.20	40,498,574	99.9970
	2001	24	8,049	99.28	41,115,778	99.9974
	2002	23	8,054	99.33	41,836,950	99.9977
	2003	22	8,056	99.35	42,716,171	99.9979
	2004	20	8,070	99.52	43,197,070	99.9986
	2005	19	8,074	99.57	44,108,016	99.9988
	2006	18	8,077	99.59	44,708,495	99.9990
	2007	17	8,087	99.70	45,200,426	99.9993

5. Conclusions

In this work, we have proposed a new methodology to choose the truncation point of the upper tail of a power law from a optimal point of view. The new methodology is based on Zipf regression method and on Akaike Information Criterion (*AIC*). We have carried out a simulation experiment to study the performance of the proposed method and we have obtained two main results: first, stability of the method proposed under changes in the number of samples, under changes in the size of the sample and under different assumptions about the underlying population, and second, existence of the optimal truncation point of a power law obtained by the method proposed in all the different simulation scenarios. Finally, we have used the new methodology proposed, with Spanish city size data for ten years, from 1998 to 2007, and we have obtained the following empirical results in all the data set used: graphically power law is reasonably accuracy to the real data, minimum *AIC* values obtained are very close to zero and existence of the optimal truncation point.

Acknowledgements

The authors thank to Ministerio de Ciencia e Innovación (project SEJ2007-65818) for partial support of this work.

Table 6: Minimum AIC and $N_{optimal}$ values obtained for ten years in Spain, 1998-2007

Country	Year	Minimum AIC	$N_{optimal}$
Spain	1998	0.004005343	729
	1999	0.004072735	736
	2000	0.004183031	730
	2001	0.004323567	718
	2002	0.004438122	675
	2003	0.004554755	655
	2004	0.004639344	652
	2005	0.004871440	631
	2006	0.005103109	632
	2007	0.005165000	631

Table 7: \hat{C} and $\hat{\alpha}$ obtained by the method of minimum AIC proposed, for ten years in Spain, 1998-2007

Country	Year	\hat{C}	$\hat{\alpha}$
Spain	1998	16.637266154	1.105718717
	1999	16.673086891	1.107709287
	2000	16.720934802	1.110892196
	2001	16.780948539	1.114463351
	2002	16.884058381	1.121690818
	2003	16.951746426	1.125299404
	2004	17.011058941	1.129263185
	2005	17.117212268	1.136311063
	2006	17.187940802	1.140825431
	2007	17.271660320	1.147095898



Figure 1: *AIC* values for Spain 1998, 2001, 2003, 2005 and 2007
(on the left all the values of *AIC*, on the right the values of *AIC* next to the optimal truncation point)

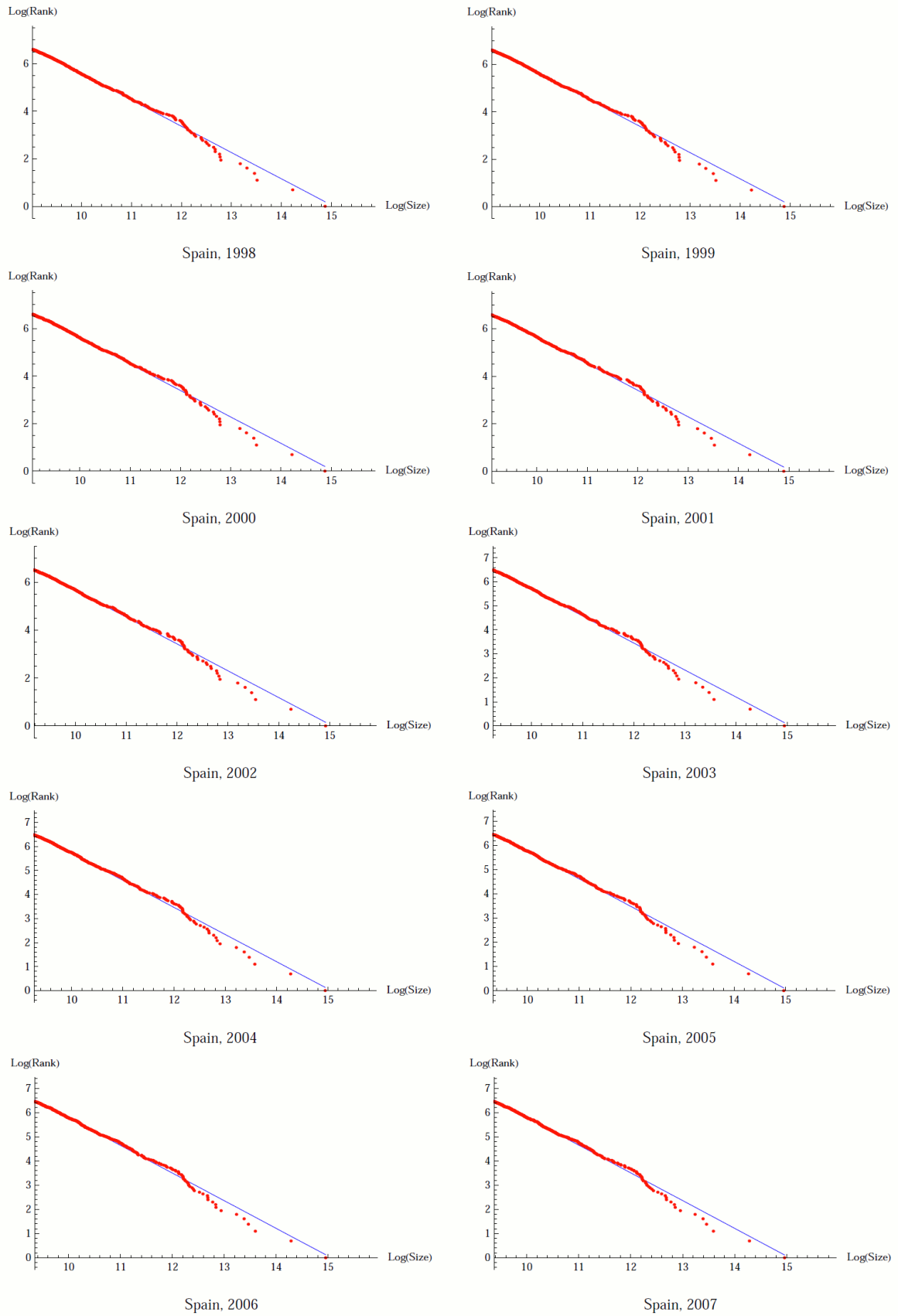


Figure 2: Upper tail of Spanish city size data on log-log scale and the corresponding power laws estimated by the method proposed (in solid lines)

References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19: 716–723
- Anderson G, Ge Y (2005) The size distribution of Chinese cities. *Regional Science and Urban Economics*, 35: 756-776
- Arnold BC (1983) *Pareto distributions*. International Cooperative Publishing House, Fairland, Maryland
- Axtell RL (2001) Zipf Distribution of U.S. Firm Sizes. *Science*, 293: 1818
- Auerbach F (1913) Das Gesetz der Bevölkerungskonzentration. *Petermanns Geographische Mitteilungen*, 59: 74-76
- Balakrishnan PVS, Miller JM, Shankar SG (2008) Power laws and evolutionary trends in stock markets. *Economics Letters*, 98(2): 194-200
- Bauke H (2007) Parameter estimation for power-law distributions by maximum likelihood methods. *The European Physical Journal B*, 58(2): 167-173
- Bosker M, Brakman S, Garretsen H, Schramm M (1949), A century of shocks: the evolution of the German city size distribution, 1925-1999. *Regional Science and Urban Economics*, 38(4): 330-347
- Brakman S, Garretsen H, van Marrewijk C, van de Berg M (1999) The return of Zipf: Towards a further understanding of the rank-size distribution. *Journal of Regional Science*, 39(1): 182-213
- Cheshire P (1999) *Trends in sizes and structures of urban areas*. In: P Cheshire, ES Mills Eds. *Handbook of Regional and Urban Economics* 3: 1339-1372, Elsevier, Amsterdam

Clauset A, Rohilla C, Newman MEJ (2009) Power-law distributions in Empirical Data. *SIAM Review*, 51(4): 661-703

Clementi F, Gallegati M (2005) Power law tails in the Italian personal income distribution. *Physica A: Statistical Mechanics and its applications*, 350(2): 427-438

Davis DR, Weinstein DE (2002) Bones, bombs, and break points: the geography of economic activity. *The American Economic Review*, 92(5): 1269-1289

Embrechts P, Klupperberg C, Mikosch T (1997) *Modelling External Events for Insurance and Finance*. Springer, Berlin

Fujiwara Y, Guilmi CD, Aoyama H, Gallegati M, Souma W (2004) Do Pareto-Zipf and Gibrat laws hold true? An analysis with European firms. *Physica A: Statistical Mechanics and its applications*, 335: 197-216

Gabaix X (1999a) Zipf's law and the growth of cities. *The American Economic Review*, 89: 129-132

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

Gabaix X, Ioannides YM (2004) *The evolution of city size distributions*, In Handbook of Regional and Urban Economics, 4: 2341-2378, JV Henderson, JF Thisse (Eds), Elsevier

Goldberger AS (1991) *A course in Econometrics*. Harvard University Press, Cambridge, MA

Growiec J, Pammolli F, Riccaboni M, Stanley HE (2008) On the size distribution of business firms. *Economics Letters*, 98(2): 207-212

Gujarati DN (2005) *Econometría*. McGraw-Hill, México

Hill BM (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3: 1163-1174

Spanish National Statistic Institute (INE), Spanish Official Municipal Registers 1998-2007, <http://www.ine.es>

Johnson NL, Kotz S, Balakrishnan N (1994) *Continuous univariate distributions*. John Wiley, NY

Kleiber C, Kotz S (2003) *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley and Sons, New York

Krugman P (1996) *The Self-Organizing Economy*. Cambridge, MA: Blackwell Publishing

Moura NJr, Ribeiro MB (2006) Zipf law for Brazilian cities. *Physica A: Statistical Mechanics and its applications*, 367: 441–448

Nitsch V (2005) Zipf zipped. *Journal of Urban Economics*, 57: 123-146

Pareto V (1897) *Cours d'Economie Poloitique*. Rouge et Cie, Paris

Rios D, Rios S, Martin J (1997) *Simulación. Métodos y aplicaciones*. Ra-Ma Editorial, Madrid

Rosen KT, Resnick M (1980) The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics*, 8(2): 165-186

Sarabia JM, Prieto F (2009) The Pareto-positive stable distribution: a new descriptive method for city size data. *Physica A: Statistical Mechanics and its applications*, 388(19): 4179-4191

Singh SK, Maddala GS (1976) A function for the size distribution of income. *Econometrica*, 44: 963-970

Soo KT (2005), Zipfs law for cities: a cross-country investigation. *Regional Science and Urban Economics*, 35(3): 239-263

Stanley MHR, Buldyrev SV, Havlin S, Mantegna RN, Salinger MA, Stanley HE (1995) Zipf plots and the size distribution of firms. *Economics Letters*, 49: 453-457

United Nations, Department of Economic and Social Affairs, International Migration Wall Charts, 2002 & 2006, <http://www.un.org>

Urzúa CM (2000) A simple and efficient test for Zipf's law. *Economics Letters*, 66: 257-260

Zanette DH, Manrubia SC (1997) Role of intermittency in urban development: a model of large-scale city formation. *Physics Review Letters*, 79(3): 523-526

Zipf GK (1949) *Human behavior and the principle of least effort*. Addison-Wesley Press, Cambridge, MA