A nonparametric estimation of the local Zipf exponent for all US Cities

González-Val, Rafael

Universitat de Barcelona Institut d'Economia de Barcelona (IEB)

13 July 2011

A Nonparametric Estimation of the Local Zipf Exponent for all US Cities

Rafael González-Val

Universitat de Barcelona & Institut d'Economia de Barcelona (IEB)

Abstract: In this paper we apply the methodology proposed by Ioannides and Overman (2003) to estimate a local Zipf exponent using data for the entire twentieth century of the complete distribution of cities (incorporated places) without any size restrictions in the US. First, we run kernel regressions using the Nadaraya–Watson estimator, excluding some atypical observations (5.66% of the sample). The results reject Zipf's Law from a long-term perspective, but the evidence supports Gibrat's Law. In the short term, decade by decade, the evidence in favour of Zipf's Law is stronger. Second, to consider the whole sample we apply the LOcally WEighted Scatter plot Smoothing (LOWESS) algorithm. From a long-term perspective the evidence supporting Zipf's Law increases, but the evidence supporting Gibrat Law's is weaker, as small cities exhibit higher variance than the rest of the cities. Finally, the estimated values by decade are again closer to Zipf's Law.

Keywords: Zipf's Law; Gibrat's Law; urban growth.

JEL: R00; C14.

Address*:* Rafael González-Val,

Departamento de Economía Política y Hacienda Pública, Universitat de Barcelona. Facultat d'Economia i Empresa, Av. Diagonal 690, 08034 Barcelona (Spain). E-mail: r.gonzalez-v@ub.edu

1. Introduction

City size distribution has been the subject of numerous empirical investigations by urban economists, statistical physicists and urban geographers. One of the stylised facts in urban economics is that the city size distribution in many countries can be approximated by a Pareto distribution whose exponent is equal to one. If this is the case, it can be concluded that there is evidence for Zipf's Law[1] (Zipf, 1949) and this means that, ordered from largest to smallest, the size of the second city is half that of the first, the size of the third is a third of the first and so on. Another well-known stylised fact is Gibrat's Law or the Law of Proportionate Growth (Gibrat, 1931), which establishes that the growth rate of a variable is independent of its initial size. Both are considered to be two sides of the same coin. While Gibrat's Law has to do with the population growth process, Zipf's Law refers to its resulting population distribution. They are closely linked; if the city sizes exhibit random growth rates (Gibrat's Law) then the city size distribution will satisfy Zipf's Law (Gabaix, 1999).

These are extensively studied empirical regularities in many countries, especially in the United States (US); see Black and Henderson (2003), Ioannides and Overman (2003), Eeckhout (2004) and González-Val (2010). Ioannides and Overman (2003) propose a nonparametric procedure to estimate Gibrat's Law for city growth processes as a time-varying geometric Brownian motion and to calculate local Zipf exponents from the mean and variance of city growth rates. They use data from metropolitan areas from 1900 to 1990 (112 to 334 metropolitan areas) and arrive at the conclusion that Gibrat's Law holds in the urban growth processes and that Zipf's Law is also fulfilled approximately for a wide range of city sizes. Nevertheless, Black and Henderson (2003) arrive at different conclusions for the same period (probably because

[1] Although Auerbach previously observed in 1913 the Pareto pattern of city size distribution.

they use different metropolitan areas; their data increase from 194 metropolitan areas in 1900 to 282 in 1990). Zipf's Law holds only for cities in the upper third of the distribution, while Gibrat's law would be rejected for any sample size. These results highlight the extreme sensitivity of conclusions to the geographical unit chosen and to the sample size.

Finally, Eeckhout (2004) demonstrates the statistical importance of considering the whole sample, not only the larger cities.[2] The estimated Pareto parameter depends on the truncation point, so when all the cities are considered for the period 1990 to 2000, the empirical city size distribution follows a log-normal rather than a Pareto distribution, and the value of Zipf's parameter is not 1, as earlier works concluded, but is slightly above $1/2$; Gibrat's Law holds for the entire sample. In a recent work, González-Val (2010) generalises this analysis for all of the twentieth century, extracting long-term conclusions: Gibrat's Law holds (weakly; growth is proportionate on average but not in variance, as the smallest cities present a clearly higher variance) and Zipf's Law holds only if the sample is sufficiently restricted to the top, not for a larger sample, because city size distribution follows a log-normal distribution when we consider all cities with no size restriction.

The nonparametric procedure put forward by Ioannides and Overman (2003) is especially relevant because it is based on the statistical explanation of Zipf's Law for cities offered by Gabaix (1999). Gabaix presents a model based on local random amenity shocks, independent and identically distributed, which through migrations between cities generate Zipf's Law. The main contribution of the work is to justify the

[2] In the US, to qualify as a metropolitan area a city needs to have 50,000 or more inhabitants, or the presence of an urbanised area of at least 50,000 inhabitants, and a total metropolitan population of at least 100,000 (75,000 in New England), according to the Office of Management and Budget (OMB) definition. Therefore, data from metropolitan areas impose an implicit truncation point.

fulfilment of Zipf's Law in that cities in the upper tail of the distribution follow similar growth processes, so that the fulfilment of Gibrat's Law involves Zipf's Law.

In this paper, the methodology proposed by Ioannides and Overman (2003) to estimate a local Zipf exponent is applied to a new data set covering the complete distribution of cities in the US (understood as incorporated places) without any size restrictions, for the entire twentieth century. Section 2 presents the data set and summarises the nonparametric procedure and its statistical foundations. Section 3 offers the results and Section 4 concludes.

2. Data and Methodology

Following Eeckhout (2004; 2009), Levy (2009), Giesen et al. (2010) and González-Val (2010), we identify cities as what the US Census Bureau calls places. This generic denomination, since the 2000 census, includes all incorporated and unincorporated places. We use the same data set as González-Val (2010). Table 1 presents the number of cities for each decade and the descriptive statistics. Our base, created from the original documents of the annual census published by the US Census Bureau, consists of the available data of all incorporated places without any size restriction, for each decade of the twentieth century (decennial data from 1900 to 2000). The US Census Bureau uses the generic term incorporated place to refer to the governmental unit incorporated under state law as a city, town (except in the states of New England, New York and Wisconsin), borough (except in Alaska and New York) or village, and which has legally established limits, powers and functions.

Two details should be noted.[3] First, Alaska, Hawaii and Puerto Rico have not been considered due to data limitations. Second, for the same reason, we also exclude all the unincorporated places (concentrations of population that do not form part of any

[3] More details about data sources and definitions are given by González-Val (2010).

incorporated place, but that are locally identified with a name), which began to be taken into account after 1950. However, these settlements did exist earlier, so their inclusion would again present a problem of inconsistency in the sample. Also, their elimination is not quantitatively important; in fact there were 1,430 unincorporated places in 1950, representing 2.36% of the total population of the US, which by 2000 would be 5,366 places and 11.27%.

The empirical strategy commonly used to test Zipf's Law consists in the estimation of log linear regressions of city size (population, P) against rank (R):

$$\log R(P) = \log A - \zeta \log P, \qquad (1)$$

where A and ζ are parameters. Zipf's Law is an empirical regularity, which appears when the Pareto exponent is equal to unity, $\zeta = 1$ (see the surveys of Cheshire, 1999, and Gabaix and Ioannides, 2004, for further explanation). The results are usually presented in double logarithmic graphs of rank compared to population, named Zipf plots, which are used extensively in the specialised literature.

However, this approach has pitfalls, highlighted in the recent literature, and different estimators have been proposed. Gabaix and Ioannides (2004) show that the Hill (maximum likelihood) estimator is more efficient if the underlying stochastic process is really a Pareto distribution, but when the size distribution of cities does not follow a Pareto distribution the Hill estimator may be biased (Soo, 2005). At the same time, the OLS estimate has some problems; see Goldstein et al. (2004) and Nishiyama et al. (2008). Finally, Gabaix and Ibragimov (2007) propose subtracting $\frac{1}{2}$ from the rank to obtain an unbiased estimation of the Pareto exponent using an OLS regression.

In this paper we apply the nonparametric procedure put forward by Ioannides and Overman (2003). This is a completely different empirical strategy, relying on the statistical foundation of Zipf's Law offered by Gabaix (1999). The exposition follows Ioannides and Overman (2003) closely; see also Gabaix (1999) and Gabaix and Ioannides (2004) for more details.[4]

Let S_i denote the normalised size of city i, that is, the population of city i divided by the total urban population. Following Gabaix (1999), city sizes are said to satisfy Zipf's Law if the countercumulative distribution function, $G(S)$, of normalised city sizes, S, tends to

$$G(S) = \frac{a}{S^{\zeta}}, \tag{2}$$

where a is a positive constant and $\zeta = 1$. If Gibrat's Law holds for city growth processes, cities grow randomly, with the same expected growth rate and the same standard deviation; then the limit distribution will converge to $G(S)$, given by Eq. (2).[5]

Gabaix also considers the case where cities grow randomly with expected growth rates and standard deviations that depend on their sizes (a weak Gibrat's Law). That is, the size of city i at time t varies according to:

$$\frac{dS_t}{S_t} = \mu(S_t)dt + \sigma(S_t)dB_t, \tag{3}$$

where $\mu(S)$ and $\sigma^2(S)$ denote, respectively, the instantaneous mean and variance of the growth rate of a size S city, and B_t is a geometric Brownian motion. In this case, the limit distribution of city sizes will converge to a law with a local Zipf exponent,

[4] Eqs. (3) and (4) replicate, respectively, Eq. (11), p. 756, and Eq. (13), p. 757, in Gabaix (1999).
[5] See Gabaix (1999), p. 744.

$$\zeta(S) = -\frac{S}{p(S)} \cdot \frac{dp(S)}{dS},$$

where $p(S)$ denotes the invariant distribution of S. Starting from the forward Kolmogorov equation associated with Eq. (3), the local Zipf exponent, associated with the limit distribution, can be derived and is given by

$$\zeta(S) = 1 - 2 \cdot \frac{\mu(S)}{\sigma^2(S)} + \frac{\partial \sigma^2(S)/\sigma^2(S)}{\partial S/S}, \tag{4}$$

where $\mu(S)$ is relative to the overall mean for all city sizes.[6] Eq. (4) identifies two possible causes of deviations from Zipf's Law: the means and the standard deviations. If $\zeta < 1$ then the distribution has neither finite mean nor finite variance, and if $1 < \zeta < 2$ it has finite mean but not finite variance.

3. Results

3.1 Kernel regressions

First, we use kernel regression techniques that establish a functional form-free relationship between the mean and the variance of city growth rates and city size for the entire distribution. This allows us to test whether Gibrat's Law holds. Second, we use Eq. (4) to estimate the local Zipf exponents directly.

---

[6] Recently Eq. (4) has been strongly criticized by Malevergne et al. (2010) and Malevergne et al. (2011). They claim that some of Gabaix's (1999) assumptions are crucial. First, "Gabaix (1999) considers that firms cannot decline below a minimum size and remain in business at this size until they start growing up again", and second, "all firms (or cities) are supposed to enter at the same time, which is technically equivalent to consider that there is only one firm in the economy". They argue that without these assumptions the distribution arising from Eq. (3) is the log-normal distribution. However, one can still consider for the log-normal distribution some form of "effective" power exponent: $\alpha(x) = \frac{1}{2\sigma^2} \ln\left(\frac{x}{e^{2\mu}}\right)$ (Eq. (5) in Malevergne et al., 2011), where μ and σ^2 are, respectively, the mean and the variance of the log-normal distribution. Malevergne et al. (2010) criticise the assumption that all firms (cities) are born at the same instant; they claim that, once one includes birth and death processes, the formula of Gabaix changes significantly.

In order to analyse the entire twentieth century, all the growth rates are taken between consecutive periods. There are 162,698 population–growth rate pairs in that pool. City size is defined as the normalised size of the city (S), that is, the population of the city divided by the total urban population,[7] and the growth rate $\mu(S)$ is defined as the difference between each city's growth rate and the contemporary average growth rate, as in Ioannides and Overman (2003). Table 2 shows the number of growth rates and descriptive statistics by decade. To calculate the conditional mean and variance on city size, we apply the Nadaraya–Watson method,[8] exactly as it appears in Härdle (1990, Chapter 3). The estimator is very sensitive, both in mean and in variance, to atypical values. Thus, we decided to eliminate the 5% of the smallest distribution observations for each decade because, as Table 2 shows, they are characterised by very high dispersion in mean and in variance, and they distort the results. Therefore, the sample size is reduced to 154,563 observations. Finally, we also eliminate 1,079 observations with a growth rate $\mu(S)$ greater than 2. The reason is that we cannot control for changes in city boundaries; there are more than twenty thousand different cities in the sample, and information on boundaries is only available for the largest cities in some decades. Then, we decide to eliminate the cities with the greatest growth rates to control for the most extreme cases, relying on the huge sample size to make the spurious growth produced by changes in boundaries irrelevant. The final sample size is 153,484 observations (94.34% of the total sample).[9]

Figure 1 shows the nonparametric estimates for the entire twentieth century of the mean growth rate and variance of growth rate conditional on city size, and the local Zipf exponent calculated applying Eq. (4). Figures 1c to 1e also display bootstrapped

---

[7] The US urban population data from 1900 to 1990 come from Table 1 in Overman and Ioannides (2001). The data for the year 2000 is taken from the US Census Bureau (http://www.census.gov).
[8] We use an Epanechnikov kernel and Silverman's kernel bandwidth.
[9] In the next section we will carry out an analysis using all the observations.

95% confidence bands, calculated using 500 random samples with replacement. The results are shown until city sizes of 0.01, the reason is that there is one technical problem with this procedure: the sparsity of data at the upper tail of the distribution, which produces extreme values of the estimations. This means that, as Ioannides and Overman (2003), we must exclude the 77 observations corresponding to the largest cities with shares greater than 0.01 (0.05% of the total sample; see Table 2).

Figures 1a and 1b show the estimates of growth and variance with the experimental points. The estimates seem to be straight lines both for growth and for variance, indicating that growth and variance are independent of the initial city size and supporting Gibrat's Law. Although some of the smallest cities deviate and exhibit higher growth rates and variances, the bulk of the 153,484 observations are close to the estimates. Part of this high variation at the lower tail could be explained by the appearance of new cities that enter with small sizes.[10] To be able to observe some kind of increasing or decreasing pattern in the estimates it is necessary to reduce the x-axis scale; this is shown in Figures 1c and 1d. The results show a very slight increasing behaviour of city growth (observe the very small scale of the growth graph), as well as a slight negative relationship between variance and city size. Thus, small cities exhibit lower growth rates and higher variances than larger cities. However, these differences are not significant for most of the city sizes in growth rates, and for any city size in the case of variance. Therefore, the evidence against Gibrat's Law is weak. Regarding Zipf's Law, one would expect a lower Zipf exponent at the upper-tail distribution for two reasons. First, in Eq. (4) the subtracting term depends on the quotient between the growth rate and the variance ($\frac{\mu(S)}{\sigma^2(S)}$); Figures 1c and 1d show that both $\mu(S)$ and

[10] See González-Val (2010) for an analysis of new entrants.

$\sigma^2(S)$ are positive, ensuring that $-2\cdot\frac{\mu(S)}{\sigma^2(S)}<0$, and as larger cities exhibit higher estimated growth rates and lower variances the expected Zipf exponent must be lower at the upper-tail distribution. Second, the term $\frac{\partial\sigma^2(S)/\sigma^2(S)}{\partial S/S}$ in Eq. (4) will also be negative (as long as the variance is decreasing with size $\frac{\partial\sigma^2(S)}{\sigma^2(S)}<0$; see Figure 1d) and decreasing with city size. Thus, all points to a lower Zipf exponent at the upper-tail distribution. Figure 1e shows that this is the case, as the local estimate of the Zipf exponent is decreasing with city size. The results reject Zipf's Law from this long-term perspective, as the estimated values are close to zero.

Some papers propose theoretical models that give an economic explanation for deviations from Zipf's Law. Rossi-Hansberg and Wright (2007) identify the standard deviation of industrial productivity shocks as the key parameter determining the dispersion in the city size distribution, Eeckhout (2004) presents a model that also relates the migration of individuals between cities with productive shocks, obtaining as a result a log-normal and non-Pareto distribution of cities, and Duranton (2007) offers a model of urban economics with detailed microeconomic foundations for technology shocks, which are the fundamental drivers of the distribution of city sizes in the steady state.

The variation in the estimates is very small, maybe as a consequence of the huge sample size of the pool. Moreover, most of the observations are concentrated at the lower end of the distribution. So, we repeat the exercise for each decade, with lower sample sizes. One advantage is that the influence of new entrant cities is lower from one decade to another than in the whole twentieth century. Also, short-term estimations could reveal interesting behaviours.

Figure 2 shows the estimates[11] of the Zipf exponent for almost all the decades, only excluding the decades 1940–1950 and 1990–2000. The reason is that in these decades the estimated values of the Zipf exponent are negative (in 1940–1950 for all the city sizes and in 1990–2000 for only the largest cities), and therefore we cannot interpret those coefficients. The explanation is that, in 1940–1950, both $\mu(S)$ and $\sigma^2(S)$ are positive and growth is much greater than variance; thus, the sign of $\zeta(S)$ becomes negative (see Eq. (4)).

We can observe in Figure 2 that, except for 1980–1990 when the estimated values are similar to those of the pool for the whole century, decade by decade the estimates of the Zipf exponent are greater than when considering all the twentieth century. We find evidence in favour of Zipf's Law as the estimates by decade are close to one. In fact, value one falls within the confidence bands (not shown for clarity purposes) for most of the distribution in most of the decades. The exception in most of the decades is the upper-tail distribution. We can also see how periods in which the Zipf exponent grows with city size (the decades from 1930 to 1980) are interspersed with others in which the relationship between the exponent and the city shares is negative (the decades from 1900 to 1930 and from 1980 to 2000). As the growth rates and variance show similar patterns for almost all the decades,[12] the differentiated behaviours of local exponents must be a consequence of interactions between the terms $-2 \cdot \frac{\mu(S)}{\sigma^2(S)}$ and $\frac{\partial \sigma^2(S)/\sigma^2(S)}{\partial S/S}$ in Eq. (4). The growth estimates are always negative (except in 1940–1950), which implies that, by decade, $-2 \cdot \frac{\mu(S)}{\sigma^2(S)}$ is positive (and increasing with

[11] Again, we exclude the observations with shares greater than 0.01; the maximum number of these observations is 12 in 1920. See Table 2.

[12] Growth and variance estimates by decade, not shown, are available from the author on request.

size, given the patterns of growth and variance). This helps to explain why decade-by-decade estimates are, in general, higher than when considering the entire twentieth century; the estimated growth for the pool of the whole century is positive, and thus, as noted above, $-2\cdot\frac{\mu(S)}{\sigma^2(S)}<0$ in that case. Regarding Gibrat's Law, in general growth rates by decade show an increasing behaviour with city size in all the decades except 1930–1940 and 1970–1980, while the only exception to the decreasing pattern of variance with city size is 1940–1950. However, the differences in growth rates and variance are not significant, finding evidence supporting Gibrat's Law even in the short term.

However, $\partial\sigma^2(S)/\sigma^2(S)$ is negative as variance is decreasing with size in all the decades (except in 1940–1950 again), so $\frac{\partial\sigma^2(S)/\sigma^2(S)}{\partial S/S}<0$ and decreasing with city size. Therefore, the resulting Zipf exponent by decade is the difference between $1-2\cdot\frac{\mu(S)}{\sigma^2(S)}>0$ and $\frac{\partial\sigma^2(S)/\sigma^2(S)}{\partial S/S}<0$ (see Eq. (4)), and both terms vary with city size. An increasing Zipf exponent means that in that decade the term $1-2\cdot\frac{\mu(S)}{\sigma^2(S)}$ is dominant, while a decreasing coefficient implies that $\frac{\partial\sigma^2(S)/\sigma^2(S)}{\partial S/S}$ is the relevant term.

3.2 A resistant smoothing approach

Kernel estimation of regression functions has been receiving much attention in the recent literature examining Gibrat's Law (Ioannides and Overman, 2003; Eeckhout, 2004; González-Val, 2010; González-Val and Sanso-Navarro, 2010; Giesen and Südekum, 2011) and the most widely used estimator is the Nadaraya–Watson estimator.

Therefore, the results in the previous section can be compared with those of other studies. However, the Nadaraya–Watson estimator is known to be highly sensitive to the presence of outliers in the data, and that is the reason why we exclude some observations. Nevertheless, Eeckhout (2004) demonstrates the importance of considering the whole sample.

In this section we try to reduce this sensitivity by using a resistant smoothing technique, the LOcally WEighted Scatter plot Smoothing (LOWESS) algorithm. This method was proposed by Cleveland (1979), and is based on local polynomial fits; see Härdle (1990, Chapter 6). The advantages of LOWESS are that it is a free-functional form method[13] and that it is robust to atypical values. Therefore, it allows us to obtain robust nonparametric estimates of growth, variance and Zipf exponent, using all the sample including the 5% of the smallest distribution observations and the observations with a growth rate greater than 2, which we exclude in the previous section. However, their inclusion produces an increase in the estimates of both growth and variance, especially at the lower tail of the distribution; this increment is much greater in the case of variance, as the dispersion of these observations is very high (see Table 2). Both growth and variance estimated by LOWESS are decreasing with city size.[14] Although the differences in growth rates by city size are not significant, the variance of the growth rates is clearly greater at the lower tail. Thus, small cities exhibit higher variance than the rest of the cities, indicating that Gibrat's Law does not hold exactly. This possibility has already been considered theoretically; Gabaix (1999) examines the case in which cities grow randomly with expected growth rates and standard deviations that depend on

[13] It does not require the specification of a function to fit a model to all of the data in the sample, LOWESS simply carries out a locally weighted regression of the y variable on the x variable, obtaining a new smoothed variable. We use the lowess command in STATA with the default options: a smoothing parameter equal to 0.8 and a tricube weighting function.

[14] Growth and variance estimates, not shown, are available from the author on request.

their sizes, and Córdoba (2008) introduces a parsimonious generalisation of Gibrat's Law that allows size to affect the variance of the growth process but not its mean.

We will focus on the analysis of the estimates of the Zipf exponent. Figure 3 shows the estimates of the Zipf exponent using LOWESS and considering all the observations. Again, the results are shown until city sizes of 0.01; the problem of the sparsity of data at the upper tail of the distribution still remains, so again we must exclude the observations corresponding to the largest cities (77 observations). Figure 3a displays the results for a pool of 162,698 observations over the whole century. The dotted lines are bootstrapped 95% confidence bands, calculated using 500 random samples with replacement. The exponent is decreasing with city size, as in the previous section (Figure 1e). The difference is that here the estimates are much closer to the value one, finding evidence supporting Zipf's Law except in the upper-tail distribution. The explanation is the high variance; both $\mu(S)$ and $\sigma^2(S)$ are positive, and thus $-2 \cdot \frac{\mu(S)}{\sigma^2(S)} < 0$ (Eq. (4)), but the estimated variance is so high that this term becomes very small, especially at the lower tail of the distribution where the variance is higher. Therefore, the decreasing pattern with city size of the estimated Zipf exponent is robust to the inclusion of all the cities, but the values are closer to Zipf's Law when all the cities are considered.

Figure 3b shows the estimates of the Zipf exponent by decade. All the decades are shown, as all the estimates are positive. The difference from Figure 2 is that now there is only one kind of behaviour: all the estimated values are increasing with city size. Moreover, the estimates are close to Zipf's Law (value one) in most of the decades (the values range from 0.8 to 1.1), with the exception of 1900–1910 and 1980–1990. Again, the cause is the high variance at the lower-tail distribution.

4. Conclusions

In this paper, the methodology proposed by Ioannides and Overman (2003) to estimate a local Zipf exponent is applied to a new data set covering the complete distribution of cities in the US (understood as incorporated places) without any size restrictions for all the twentieth century.

First, we run kernel regressions using the Nadaraya–Watson estimator. As the estimator is very sensitive, both in mean and in variance, to atypical values we must exclude some observations, so the final sample size is 153,484 observations (94.34% of the total sample). The results reject Zipf's Law from a long-term perspective, as the estimated values are close to zero. However, the evidence supports Gibrat's Law. In the short term, decade by decade, we find evidence in favour of Zipf's Law for most of the distribution in most of the decades. We also observe differentiated behaviours: periods in which the Zipf exponent grows with the city size are interspersed with others in which the relationship between the exponent and the city shares is negative.

Second, to consider the whole sample (162,698 observations) we apply the LOcally WEighted Scatter plot Smoothing (LOWESS) algorithm. The evidence supporting Zipf's Law increases, as the estimated values are closer to one, but the evidence supporting Gibrat Law's is weaker, as small cities exhibit higher variance than the rest of cities. Finally, the estimated values by decade are also closer to Zipf's Law.

Acknowledgements

The author acknowledges financial support from the Spanish Ministerio de Educación y Ciencia (ECO2009-09332 and ECO2010-16934 projects), the DGA (ADETRE research group) and FEDER. The comments received from anonymous referees have undoubtedly improved the version originally submitted. All remaining errors are mine.

References

Black D, Henderson V, 2003, "Urban evolution in the USA" *Journal of Economic Geography* **3**(4) 343 – 372

Cheshire P, 1999, "Trends in sizes and structure of urban areas", in *Handbook of Regional and Urban Economics, Vol. 3*, Eds P Cheshire and E S Mills (Elsevier, Amsterdam) Chap. 35, pp 1339 – 1373

Cleveland W S, 1979, "Robust locally weighted regression and smoothing scatterplots" *Journal of the American Statistical Association* **74** 829 – 836

Córdoba J C, 2008, "A generalized Gibrat's law" *International Economic Review* **49**(4) 1463 – 1468

Duranton G, 2007, "Urban evolutions: the fast, the slow, and the still" *American Economic Review* **97**(1) 197 – 221

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

Gabaix X, 1999, "Zipf's Law for cities: an explanation" *Quarterly Journal of Economics* **114**(3) 739 – 767

Gabaix X, Ibragimov R, 2007, "Rank-1/2: a simple way to improve OLS estimation of tail exponents", NBER Technical Working Paper, vol. 342

Gabaix X, Ioannides Y M, 2004, "The evolution of city size distributions" in *Handbook of Urban and Regional Economics, Vol. 4*, Eds J V Henderson, J F Thisse (Elsevier, Amsterdam) pp 2341 – 2378

Gibrat R, 1931 *Les Inégalités Économiques* (Librairie du Recueil Sirey, Paris)

Giesen K, Südekum J, 2011, "Zipf's Law for cities in the regions and the country" *Journal of Economic Geography* **11**(4) 667 – 686

Giesen K, Zimmermann A, Suedekum J, 2010, "The size distribution across all cities – double Pareto lognormal strikes" *Journal of Urban Economics* **68** 129 – 137

Goldstein M L, Morris S A, Yen G G, 2004, "Problems with fitting to the power-law distribution" *The European Physical Journal B – Condensed Matter* **41**(2) 255 – 258

González-Val R, 2010, "The evolution of the US city size distribution from a long-run perspective (1900–2000)" *Journal of Regional Science* **50**(5) 952 – 972

González-Val R, Sanso-Navarro M, 2010, "Gibrat's law for countries" *Journal of Population Economics* **23**(4) 1371 – 1389

Härdle W, 1990 *Applied Nonparametric Regression* (Cambridge Univ. Press, Cambridge)

Ioannides Y M, Overman H G, 2003, "Zipf's Law for cities: an empirical examination" *Regional Science and Urban Economics* **33** 127 – 137

Levy M, 2009, "Gibrat's Law for (all) cities: a comment" *American Economic Review* **99**(4) 1672 – 1675

Malevergne Y, Pisarenko V, Sornette D, 2011, "Gibrat's Law for cities: uniformly most powerful unbiased test of the Pareto against the lognormal" In press in *Physical Review E*. arXiv:0909.1281v1 [physics.data-an]

Malevergne Y, Saichev A, Sornette D, 2010, "Zipf's Law and maximum sustainable growth" arXiv:1012.0199v1 [physics.soc-ph]

Nishiyama Y, Osada S, Sato Y, 2008, "OLS estimation and the t test revisited in rank-size rule regression" *Journal of Regional Science* **48**(4) 691 – 715

Overman H G, Ioannides Y M, 2001, "Cross-sectional evolution of the U.S. city size distribution" *Journal of Urban Economics* **49** 543 – 566

Rossi-Hansberg E, Wright M L J, 2007, "Urban structure and growth" *Review of Economic Studies* **74** 597 – 624

Soo K T, 2005, "Zipf's Law for cities: a cross-country investigation" *Regional Science and Urban Economics* **35**: 239 – 263

Zipf G, 1949 *Human Behaviour and the Principle of Least Effort* (Addison-Wesley, Cambridge, MA)

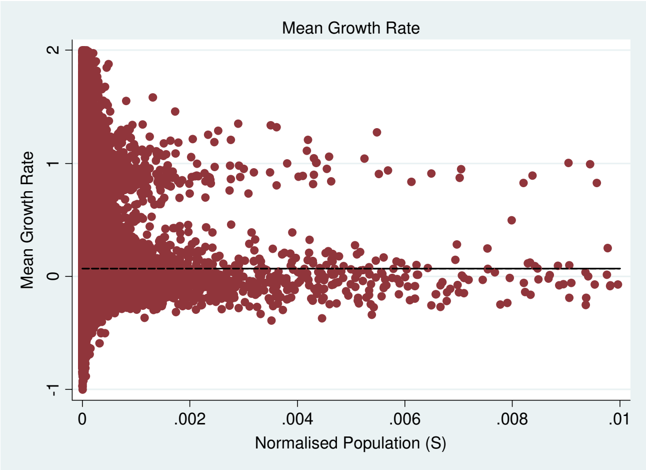
Table 1. Number of cities and descriptive statistics

Year	Cities	Mean	Standard deviation	Minimum	Maximum
1900	10,596	3,376.04	42,323.90	7	3,437,202
1910	14,135	3,560.92	49,351.24	4	4,766,883
1920	15,481	4,014.81	56,781.65	3	5,620,048
1930	16,475	4,642.02	67,853.65	1	6,930,446
1940	16,729	4,975.67	71,299.37	1	7,454,995
1950	17,113	5,613.42	76,064.40	1	7,891,957
1960	18,051	6,408.75	74,737.62	1	7,781,984
1970	18,488	7,094.29	75,319.59	3	7,894,862
1980	18,923	7,395.64	69,167.91	2	7,071,639
1990	19,120	7,977.63	71,873.91	2	7,322,564
2000	19,296	8,968.44	78,014.75	1	8,008,278

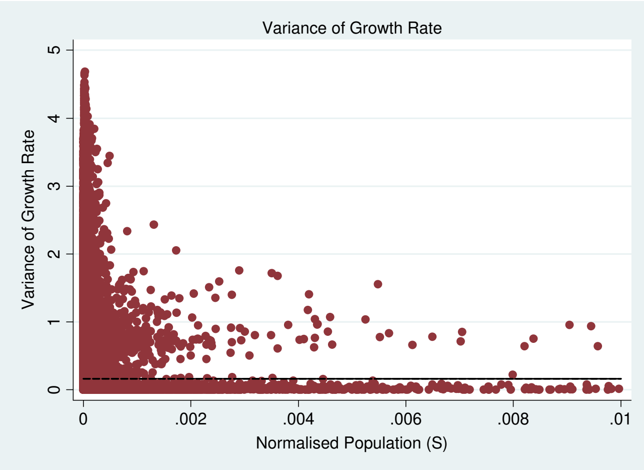
Table 2. Growth rates $\mu(S)$: descriptive statistics

Initial Year	Growth rates	Mean	Standard deviation	Minimum	Maximum
All cities					
1900	10,502	4.30E-11	0.51	-0.94	9.54
1910	13,543	-1.29E-11	0.52	-0.90	31.82
1920	15,085	9.56E-12	0.58	-0.89	22.34
1930	16,199	3.90E-11	0.44	-1.01	31.67
1940	16,416	1.40E-11	1.85	-0.95	231.04
1950	16,943	2.30E-10	1.03	-0.89	92.38
1960	17,831	1.90E-09	36.93	-1.23	4926.66
1970	18,321	-4.93E-11	0.83	-0.90	52.07
1980	18,810	3.70E-11	0.39	-0.84	13.14
1990	19,048	1.42E-10	0.48	-0.96	30.11
Total	162,698	2.51E-10	12.25	-1.23	4926.66
5% smallest cities					
Initial Year	Growth rates	Mean	Standard deviation	Minimum	Maximum
1900	525	0.21	0.85	-0.69	4.98
1910	677	0.16	1.45	-0.81	31.82
1920	754	0.11	1.43	-0.87	22.34
1930	810	0.19	1.56	-0.96	31.67
1940	821	0.27	8.11	-0.88	231.04
1950	847	0.02	1.94	-0.86	34.61
1960	892	5.54	165.08	-1.17	4926.66
1970	916	0.18	1.96	-0.89	26.93
1980	941	0.00	1.01	-0.84	12.41
1990	952	0.12	0.93	-0.94	16.67
Total	8,135	0.68	54.74	-1.17	4926.66
Cities greater than 0.01					
Initial Year	Growth rates	Mean	Standard deviation	Minimum	Maximum
1900	11	-0.05	0.08	-0.15	0.12
1910	11	0.08	0.23	-0.07	0.75
1920	12	0.09	0.25	-0.08	0.78
1930	10	-0.06	0.07	-0.12	0.10
1940	9	-0.05	0.07	-0.10	0.12
1950	8	-0.17	0.09	-0.25	0.04
1960	5	-0.43	0.07	-0.50	-0.32
1970	5	-0.24	0.08	-0.33	-0.11
1980	3	0.02	0.11	-0.09	0.13
1990	3	-0.05	0.02	-0.07	-0.03
Total	77	-0.06	0.20	-0.50	0.78

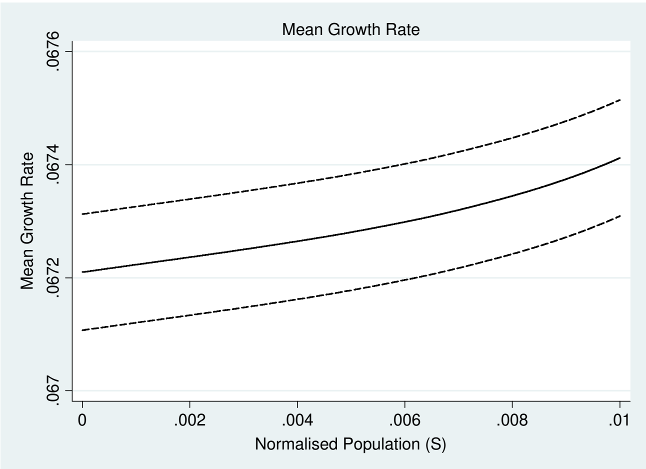
Figure 1. Nonparametric estimates for all the twentieth century (a pool of 153,484 observations, Nadaraya–Watson estimator)



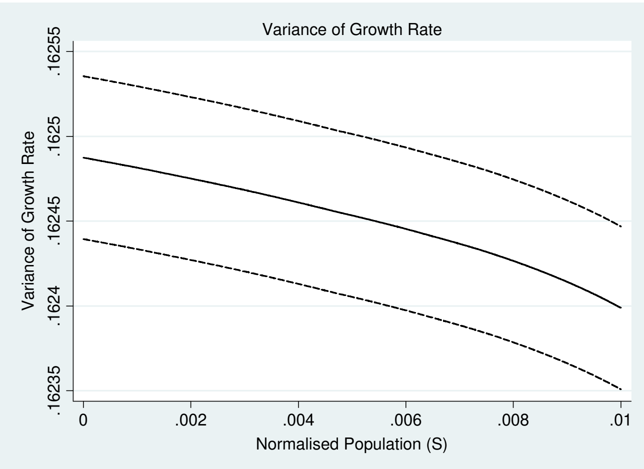
(a) Mean



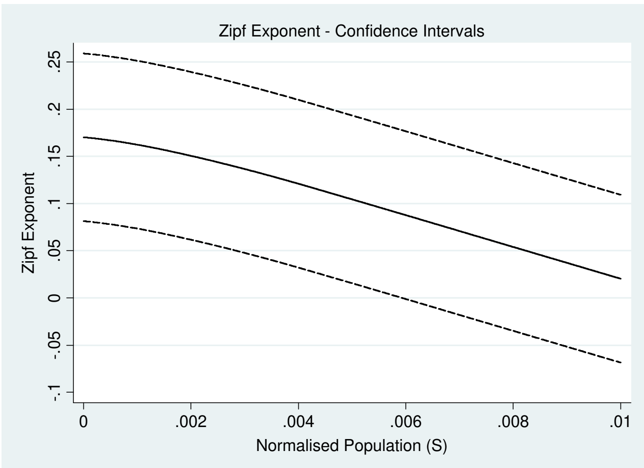
(b) Variance



(c) Mean (reduced scale)

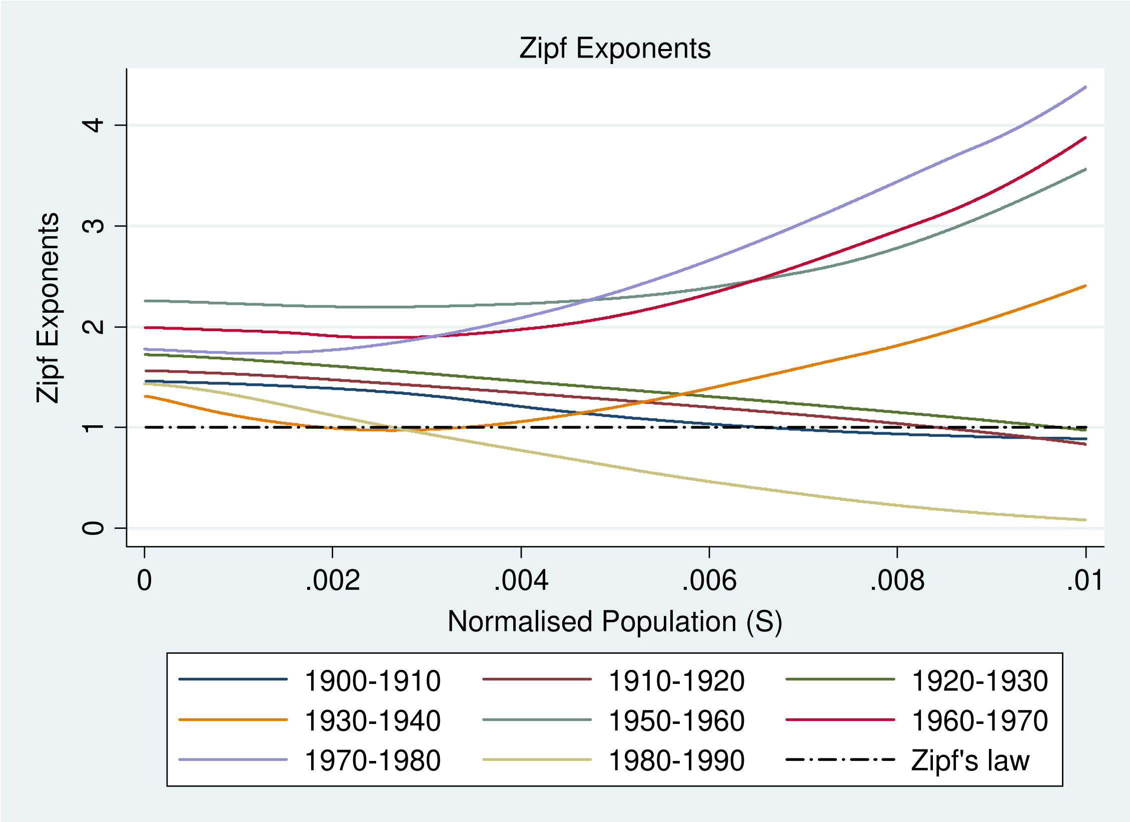


(d) Variance (reduced scale)



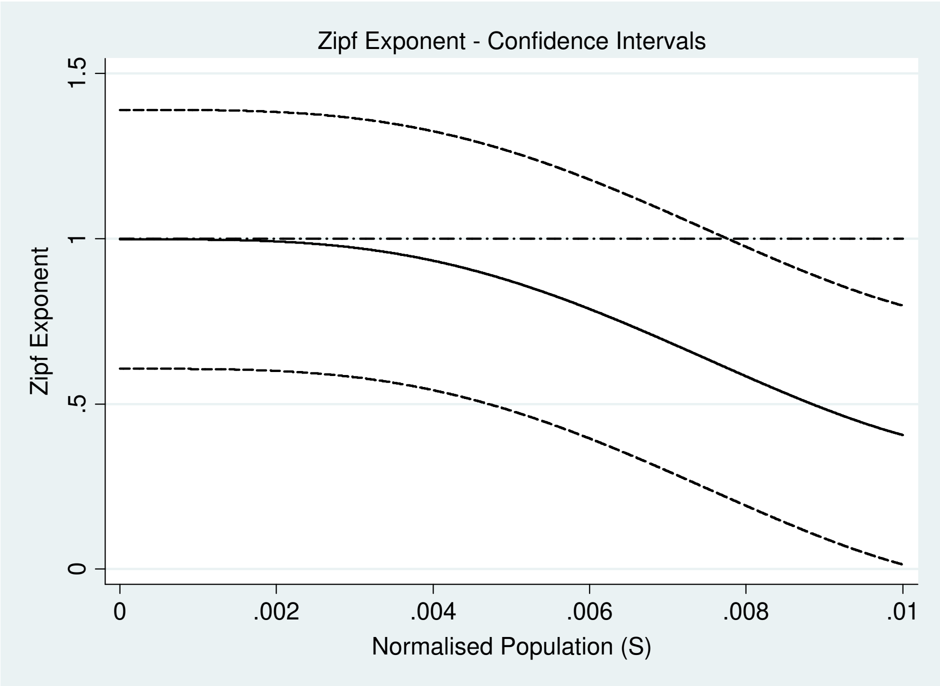
(e) Zipf Exponent

Figure 2. Nonparametric estimates of the Zipf Exponent by decade (Nadaraya–Watson estimator)

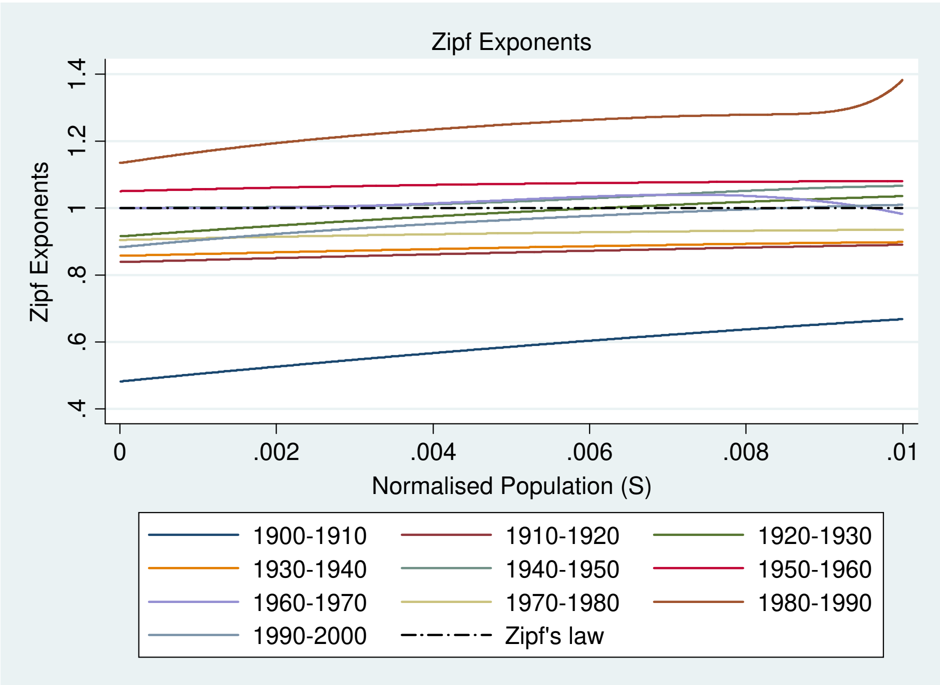


Note: Decades 1940–1950 and 1990–2000 are excluded; see the main text.

Figure 3. Nonparametric estimates for all the twentieth century (a pool of 162,698 observations) using the LOcally WEighted Scatter plot Smoothing (LOWESS) algorithm



(a) Twentieth century



(b) By decade