Gibrat's law for cities, growth regressions and sample size

González-Val, Rafael and Lanaspa, Luis and Sanz, Fernando

November 2012

Gibrat's law for cities, growth regressions and sample size

Rafael González-Val[a]

Luis Lanaspa[b]

Fernando Sanz[b]

[a] Universidad de Zaragoza & Institut d'Economia de Barcelona (IEB)

[b] Universidad de Zaragoza

Abstract: This paper uses un-truncated city population data from six countries (the United States, Spain, Italy, France, England and Japan) to illustrate how parametric growth regressions can lead to biased results when testing for Gibrat's law in city size distributions. The OLS results show non-monotonic behaviours depending on the sample size. Moreover, it is possible to find a critical sample size from which we reject Gibrat's law.

Keywords: Gibrat's law, city size distribution, urban growth

JEL: R00, R11, R12

1. Introduction

One of the stylised facts in urban economics is Gibrat's law, according to which city growth is independent of initial city size. Gibrat's law started as an empirical regularity associated with the distribution of firms, but in the field of urban economics, especially since the 1990s, many empirical studies have tested its validity for city size distributions. The countries considered, statistical and econometric techniques used and sample sizes are heterogeneous, but some of these studies estimate parametric growth regressions (e.g. Eaton and Eckstein, 1997)[1].

Let S_{it} be the size (population) of city i at time t and g be its growth rate, then $S_{it} = S_{it-1}(1+g)$. Taking logarithms and adding that the rate depends on the initial size, we can obtain the following general expression of the growth equation:

$$\ln S_{it} - \ln S_{it-1} = \mu + \beta \ln S_{it-1} + u_{it}, \qquad (1)$$

where $\mu = \ln(1+g)$ and u_{it} is an iid random variable representing the random shocks that the growth rate may suffer, with $E(u_{it}) = 0$ and $Var(u_{it}) = \sigma^2 \; \forall i,t$. If $\beta = 0$, Gibrat's law holds because growth is independent of initial size. Thus, if the estimate of β is significant, we can reject the fulfilment of Gibrat's law. In the case of it being significant and positive, we observe divergent growth, because city growth would depend directly and positively on initial size[2].

In this paper, we argue that parametric growth regressions such as that in Equation (1) can lead to biased results because city size data often violate one of the basic assumptions of the OLS estimator: the error terms are usually not homoskedastic. As González-Val (2010) and González-Val et al. (2012) show, variance of the growth rates usually depends negatively on city size because small cities display higher growth and variance compared with larger cities. Therefore, the results derived from Equation (1) depend on the number and size of the cities included in the sample.

---

[1] Gabaix and Ioannides (2004) describe Eaton and Eckstein's (1997) paper as *"the most noteworthy recent study that focused on the persistence of the city size distribution and one of the most important contributions to the recent urban growth literature"*.

[2] Actually, Equation (1) is an unconditional β-convergence regression, widely known in the economic growth literature.

The remainder of this paper is organised as follows. Section 2 introduces the databases we use. Section 3 analyses the results of estimating Equation (1) with different sample sizes, and Section 4 concludes.

2. Data

There are six countries in our sample: the United States, Spain, Italy, France, England and Japan. For the US, Spain and Italy, we use the same dataset as used by González-Val et al. (2012); this database includes the decennial census for each decade of the 20th century[3]. Data for France, England and Japan are obtained from the national official statistical services, and information only on recent decades is available. These samples include un-truncated city population data from all countries.

The US database is created from the original documents of the annual census published by the US Census Bureau (www.census.gov), and it consists of the available data on all incorporated places without any size restriction. The US Census Bureau uses the generic term *incorporated place* to refer to the governmental unit incorporated under state Law as a city, town, borough or village. Alaska, Hawaii and Puerto Rico are excluded because of data limitations. The number of cities considered by period is: 1900 (10,596 incorporated places), 1910 (14,135), 1920 (15,481), 1930 (16,475), 1940 (16,729), 1950 (17,113), 1960 (18,051), 1970 (18,488), 1980 (18,923), 1990 (19,120) and 2000 (19,296).

For Spain and Italy, the geographical unit of reference is the *municipality*, and the data come from the official statistical information services. In Italy, this is the *Istituto Nazionale di Statistica* (www.istat.it), while for Spain we have taken the census of the *Instituto Nacional de Estadística* (www.ine.es). Because municipalities are the smallest spatial units (local governments), they are the administratively defined "legal" cities. For Italy, the number of cities by period is 7,711 municipalities in 1901 and 1911 and 8,100 municipalities from 1921 to 2001. For Spain, we consider the following years: 1900 (7,800 municipalities), 1910 (7,806), 1920 (7,812), 1930 (7,875), 1940 (7,896), 1950 (7,901), 1960 (7,910), 1970 (7,956), 1981 (8,034), 1991 (8,077) and 2001 (8,077).

[3] More information about the databases and comparisons between these countries can be found in González-Val et al. (2012).

The French dataset is obtained from the national statistical office, the *Institut national de la statistique et des études économiques* (www.insee.fr). Again, we consider the lowest spatial subdivision, the *communes*, which comprise the whole country population. The years considered are 1990 (36,686 *communes*), 1999 (36,685) and 2009 (36,722). For England, we consider data from the 354 local authorities and administrative areas, compiled by the UK Office for National Statistics (www.ons.gov.uk). Only data on 1991 and 2001 are available. Finally, we also include data from the 2,102 Japanese municipalities for 2005 and 2010. The source here is the Statistics Bureau and the Director-General for Policy Planning of Japan (www.stat.go.jp).

In these six countries, we consider administratively defined cities (legal cities); thus, their boundaries may not make economic sense and, in many relevant instances, they may not correspond to a meaningful economic definition of a city. Although metropolitan areas are considered to be more natural economic units, certain factors, such as human capital spillovers, are thought to operate at a local level, and there are statistical reasons to consider an un-truncated city population dataset (Eeckhout, 2004).

3. Results

Eeckhout (2004) demonstrates the importance of choosing sample size in the analysis of city size distribution, as the arbitrary choice of a truncation point can lead to biased results. In the same way, we analyse whether the results for city growth depend on sample size when parametric growth regressions are used. To do this, we estimate Equation (1) for different sample sizes, namely 50, 100, 200, 500, 1000 and so on, adding groups of 500 cities at a time until they are all considered[4]. We estimate starting from the largest cities (from the upper tail)[5]. Figure 1 shows the values estimated by OLS of the beta parameter from Equation (1) for the six studied countries. For the US, Italy and Spain, only selected decades are shown for clarity reasons[6]. We want to point out three fundamental results.

One is that non-monotonic behaviours are observed depending on the sample size, especially in the US. For example, in the 1950–1960 period (although the

[4] For the US and France, where the sample sizes are noticeably larger, we take cities up to 5,000 in batches of 500, and from that number, add them in groups of 1,000.
[5] We also estimate starting from the smallest cities (from the lower tail), but the results do not provide qualitatively new evidence and so are not included.
[6] The results for the decades not shown are available from the authors on request.

behaviour is similar in all decades from 1930–1940 to 1960–1970) estimates begin by accepting Gibrat's law, then a convergent behaviour is obtained (the estimated β is significant and negative), Gibrat's law cannot be rejected for some intermediate sample sizes, and, finally, we have a divergent behaviour (the estimated β is significant and positive). These non-monotonicities strengthen our claim that "everything depends on sample size" when growth regressions are used.

Second, in Italy, and especially in Spain, the predominant behaviour is divergence (very few estimates are significant and negative), which does not occur so intensively in the US. This suggests, as González-Val et al. (2012) indicate, that in these two European countries the biggest cities are those that have grown the most. Third, if we focus on recent decades, similar behaviour can be observed in the six countries because the estimates clearly increase with sample size in all of them (in some cases, after an initial fall with small sample sizes).

We argue that these non-monotonic behaviours are driven by heteroskedastic errors, because the variance of the growth rates usually depends negatively on city size (González-Val, 2010; González-Val et al., 2012). To corroborate this point we calculate White's test for the heteroskedasticity of residuals. The results[7] indicate that, in general, the null hypothesis of homoskedastic residuals cannot be rejected for small sample sizes, but as the number of observations increases the test often points to a rejection of the null. This is especially true for France and the US, the two countries with the highest number of cities.

Figure 2 shows what we have called the *critical sample size*, namely the size from which we reject the null hypothesis $\beta = 0$ in Equation (1) at a significance level of 5% (standard errors are calculated using White's heteroskedasticity-consistent standard errors). The results show that for small initial sizes the null hypothesis $\beta = 0$ is not rejected, finding empirical evidence supporting Gibrat's law[8]. However, as the sample size increases, this conclusion soon changes and the parameter becomes significant. While the critical sample size is variable, for the US, Spain and Italy—in most of the 20th century—it is equal to or less than the 500 biggest cities. In recent decades, critical

[7] The results, not shown for size restrictions, are available from the authors on request.
[8] When Gibrat's law was rejected with the initial sample size (50), we explored the exact critical sample size for these cases (lower than 50). This situation happened six times for Italy and once in Japan (never for the rest of the countries).

sample sizes are similar in the six countries. For any lower sample size, we cannot reject the local fulfilment of Gibrat's law.

Two relevant conclusions can also be derived from Figure 2. First, there is a great deal of interannual variability, even between two consecutive periods. Second, although there are exceptions, critical sample sizes tend to be bigger for the US than for the rest of the countries, which is more evidence in favour of the greater validity of Gibrat's law for the US. Furthermore, over recent decades there has been an increasing trend in US critical sample sizes, meaning that Gibrat's law holds for a growing number of city sizes. The behaviour of French critical sample sizes is similar, although the growth is lower.

The information in Figure 2 should be compared with the sample sizes used in other studies, which tend to be lower (sometimes much lower) than the 500 largest cities. Therefore, if the sample size is low one may be positioning oneself below the critical sample size, and accepting the relative fulfilment of Gibrat's law when the behaviour of the entire distribution may be different.

4. Conclusions

In this paper, we show that parametric growth regressions can lead to biased results using un-truncated city population data from six countries: the US, Spain, Italy, France, England and Japan. We obtain non-monotonic behaviours depending on the sample size. Furthermore, it is possible to find a critical sample size from which we reject Gibrat's law. This evidence suggests that the appropriate tools to test Gibrat's law are nonparametric methods (see Gabaix and Ioannides, 2004; Eeckhout, 2004). It may also suggest that a more precise definition of what a city is may be required.

References

[1] Eaton, J., and Z. Eckstein (1997). Cities and Growth: Theory and Evidence from France and Japan. Regional Science and Urban Economics, 27: 443–474.

[2] Eeckhout, J. (2004). Gibrat's Law for (All) Cities. American Economic Review, 94(5): 1429–1451.

[3] Gabaix, X. and Y. M. Ioannides (2004). The Evolution of City Size Distributions. In Handbook of Urban and Regional Economics, vol. 4, ed. John V. Henderson and Jean. F. Thisse, 2341–2378. Amsterdam: Elsevier Science, North-Holland.

[4] 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.

[5] González-Val, R., L. Lanaspa, and F. Sanz (2012). New Evidence on Gibrat's Law for Cities. IEB Working paper No. 2012/18.

Figure 1. Local estimations of Equation (1)

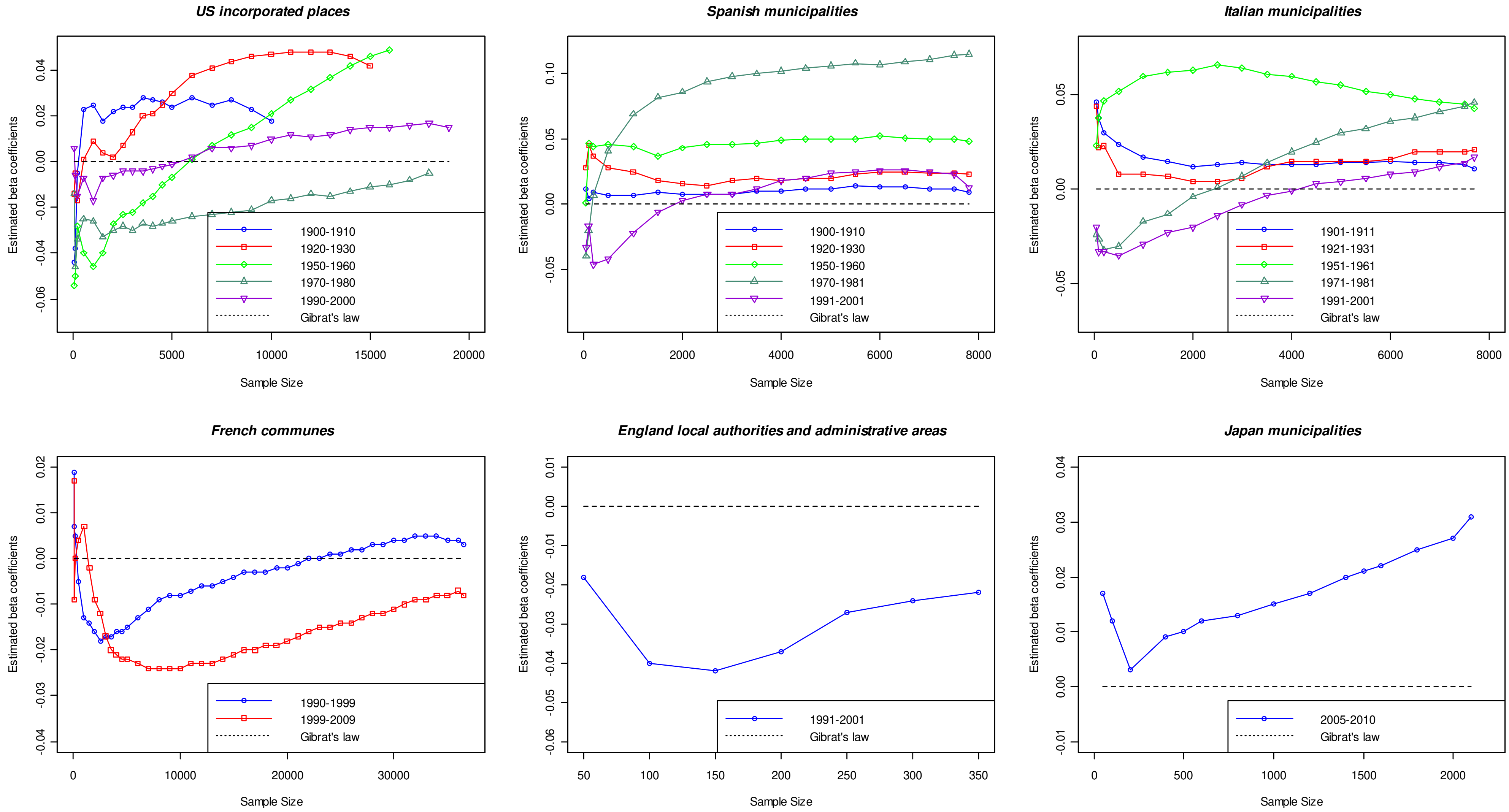

Figure 2. Critical sample sizes from which Gibrat's Law is rejected