Economics Bulletin

Volume 41, Issue 4

Zipf's exponent and Zipf's law in the BRICS: a rolling sample regressions approach

Guillermo Peña
Universidad de Zaragoza

Fernando Sanz-Gracia
Universidad de Zaragoza

Abstract

Using urban data from the last available census of the five BRICS countries we have tested, by means of a rolling sample regressions approach, whether, as Eeckhout (2004) proposed, the Pareto exponent in a standard Zipf equation is decreasing as more cities are added to the sample. The results are very conclusive: Eeckhout's hypothesis is satisfied for Brazil, Russia and South Africa, but for India and China there are non-negligible parts of the distribution where it is not fulfilled. We also test the fulfilment of Zipf's law: it holds in the upper tail of the five countries (except South Africa) but for the rest of the distribution the predominant outcome is rejection.

Acknowledgments: We want to express our gratitude for the fund received from the Project ECO2017-82246-P and for the anonymous referees and Editors because of their insightful comments and suggestions that have improved the paper. Declaration of Competing Interest: Declarations of interest: none.
Citation: Guillermo Peña and Fernando Sanz-Gracia, (2021) "Zipf's exponent and Zipf's law in the BRICS: a rolling sample regressions approach", *Economics Bulletin*, Vol. 41 No. 4 pp. 2543-2549.
Contact: Guillermo Peña - gpena@unizar.es, Fernando Sanz-Gracia - fsanz@unizar.es.
Submitted: May 06, 2021. **Published:** December 29, 2021.

1. INTRODUCTION

The study of city size distribution has a long tradition in Urban Economics. One of the more frequently used densities is the Pareto or power law and a particular case of it, the so-called Zipf's law. There exists a certain consensus in the literature that the upper tail of the city size distribution follows a power law (Ioannides and Overman, 2003; Gabaix, 2016), something that can be proved by analysing the Zipf or Pareto exponent. Zipf's law holds when this exponent is statistically one. Many papers have exhaustively analysed the behaviour of the Zipf exponent; see, to cite only two, Rosen and Resnick (1980) and Soo (2005).

In this paper we want to empirically test for the BRICS (Brazil, Russia, India, China and South Africa) a proposition first raised by Eeckhout (2004) and corroborated, among others, by Luckstead and Devadoss (2014a): The Zipf exponent declines systematically as the sample size increases. We also test the fulfilment of Zipf's law. To do so, we adopt a rolling sample regressions approach.

But, why the BRICS[1]? The relative economic and urban importance of the BRICS has steadily increased since the end of WWII and the rise of their cities in the world urban network is hard to deny (Liu et al., 2014). Indeed, in 1900 only three of the twenty largest cities in the world belonged to the BRICS and none of them were in the top ten; in 2010 this number is nine and five of them are in the top ten (see Table 1 in Jedwab and Vollrath, 2015). In terms of absolute population, four of the BRICS rank among the ten most populated countries in the world, being China the first and India the second. Regarding surface, again four of them are among the largest ten, being Russia the first and China the third.

Our main results are two. First, Eeckhout hypothesis is satisfied for Brazil, Russia and South Africa, while for India and China is not fulfilled in some relevant parts of their distributions. Second, Zipf's law, as expected, is accomplished in the upper tail of the five countries, except South Africa, while for the rest of the distribution the predominant behaviour, especially for India and China, is rejection.

The rest of the paper is organized as follows. Section 2 describes the data and Section 3 the methodology. The results are presented in Section 4 and Section 5 is devoted to the conclusions.

2. DATA

We use the official administrative division of each country as urban units, creating a consistent database that acknowledges nearly all the population of each country while maintains a certain structure that allows us to compare them. For doing that, in order to allow for comparability between the five BRICS, we have considered the smallest spatial division available as a basis. We use the last census available.

For the case of Brazil, we have used the census of 2010, taking the 5.565 municipalities as the urban unit. The data has been provided by the Instituto Brazileiro de Geografia e Estatistica, covering 100% of the population of the country. In the case of Russia, we have obtained the data from the Federal State Statistics Service. It covers 21.354 urban units and represents 99.99% of the population in 2010. For India, we collected the data from the

[1] For a review of the literature regarding city size distribution for each of the five BRICS see Puente-Ajovín et al. (2020a).

Census of India of 2011. The administrative units used are towns and villages. The 601,791 urban units cover 93.3% of the Indian population. For China we use the census of 2010 and we have obtained the information from the website http://www.citypopulation.de/China.html. The 37.358 urban units cover 99.4% of the entire population of Mainland China. For South Africa we have obtained the complete dataset from http://blog.adrianfrith.com/, from which we have used the Main Places urban unit. In this case, 13.942 urban units cover all the population of the country in 2011.

3. METHODOLOGY

The standard Zipf equation is as follows, taking into account the correction introduced by Gabaix and Ibragimov (2011):

$$ln(R-0.5) = \alpha - \beta \, ln(S) \qquad ,$$

where R is the rank of the city (1 for the largest, 2 for the second and so on), S is its size or population, α is a parameter and β is the so called Zipf exponent or Pareto exponent which, by construction, is always positive. The asymptotically correct standard error of the estimated β, which is used to test whether the Pareto exponent is statistically equal to one, is $\hat{\beta}\left(2/N\right)^{1/2}$, with N being the number of cities considered in the sample. When β=1 from a statistical point of view, Zipf's law is fulfilled and, in any case, β is interpreted as a measure of the degree of inequality in the city size distribution: when β increases (decreases) the distribution becomes more equal (unequal).

Eeckhout (2004), in Proposition 1, proved that if the underlying distribution is lognormal, then ($d\beta/dN$)<0, beginning with the upper tail. In other words, according to Eeckhout (2004), as we successively add smaller and smaller cities to the sample, the inequality in the distribution will grow. This statement, along with the direct test of Zipf's law, is what we want to test empirically with data of the BRICS. To do so, we employ rolling sample regressions (see Peng, 2010), so we incorporate cities one by one to the sample, beginning with the largest one hundred cities, and each time we calculate the estimated Pareto exponent β. The first sample size for each country is one hundred and it increases one by one until we reach the smallest urban unit. Therefore, the largest rank R considered in each regression is equal to the one by one varying sample size N.

4. RESULTS

Figures 1 to 5 show the OLS estimated[2] Pareto exponent with the rolling sample regressions for the five BRICS. In red the estimated Pareto exponents that do not reject Zipf's law and, at the top of each Figure, the percentage of non-rejections. For Brazil, except for a very small length in the very upper tail, the Zipf exponent is always decreasing and its range of variation (0.84; 1.37) is reasonably close to one. The same behaviour holds for Russia and South Africa, although the range of variation of the Pareto exponent is different in each country: (0.82; 1.40) for Russia, where the estimated β is very close to one for most of the sample sizes, and (0.63; 1.66) for South Africa. For these three countries the hypothesis of Eeckhout is valid.

---

[2] We are conscious that there are other estimation procedures, such as maximum likelihood (Hill estimator). It is out of the scope of this note to elucidate which is best.

Figure 1. Pareto exponent for Brazil (16.65% of non-rejections of Zipf's law)

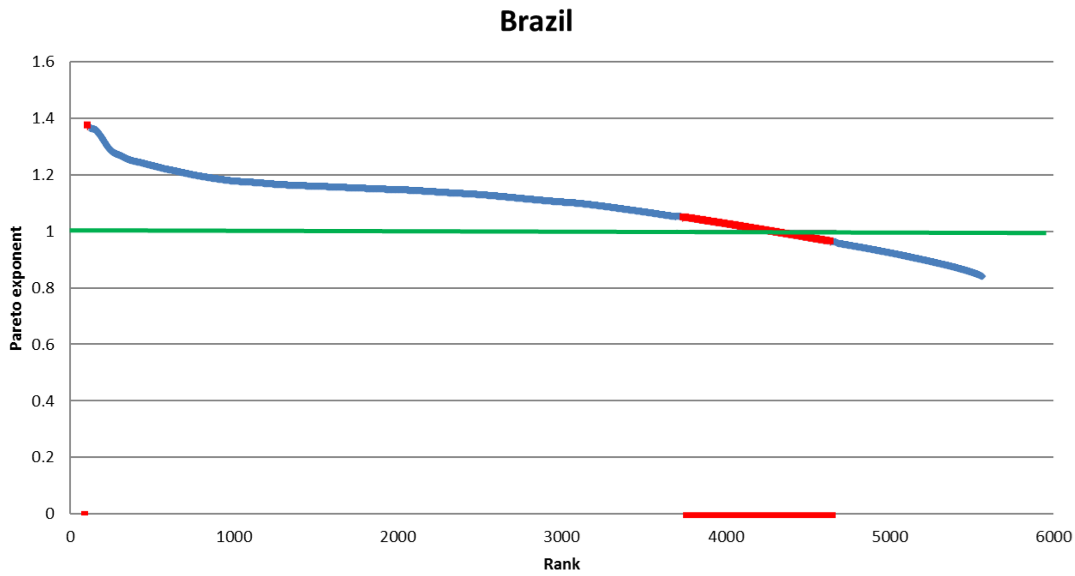
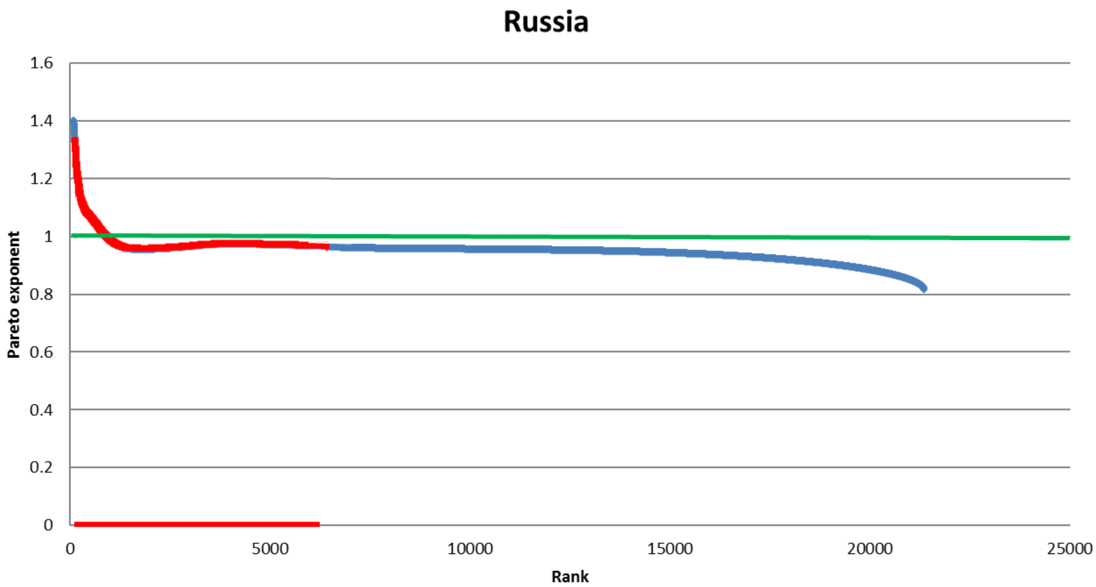


Figure 2. Pareto exponent for Russia (29.13% of non-rejections of Zipf's law)



China and India's evolution of the Pareto exponent is similar[3]. First, a very fast drop, then an increasing segment until reaching a maximum for a sample size of roughly 150,000 for India and nearly 15,000 for China and, finally, a continuous decreasing behaviour until the end of the sample. The range of variation is also different: (0.64; 1.51) for India and (0.68; 1.60) for China. We want to emphasize that for these two countries there is an important part of the distribution where ($d\beta/dN$)>0, in such a way that Proposition 1 of Eeckhout is not satisfied there and, therefore, adding more cities to the sample makes the distribution less unequal.

---

[3] City size distribution of China has been studied in detail in Anderson and Ge (2005) and of China and India in Luckstead and Devadoss (2014b).

Figure 3. Pareto exponent for South Africa (22.60% of non-rejections of Zipf's law)

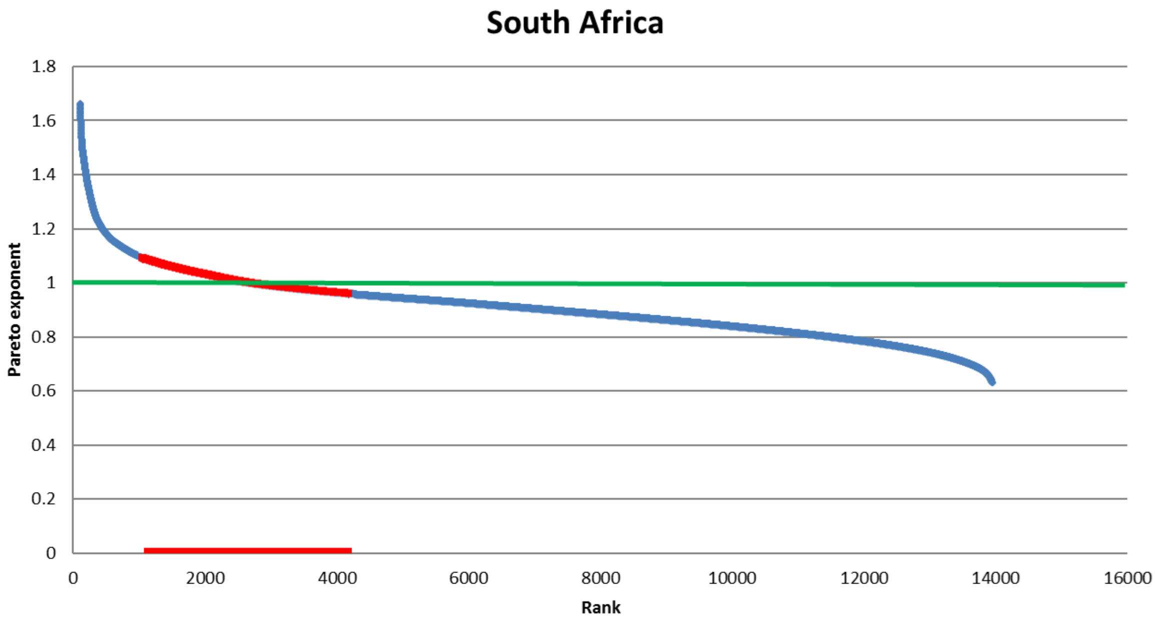
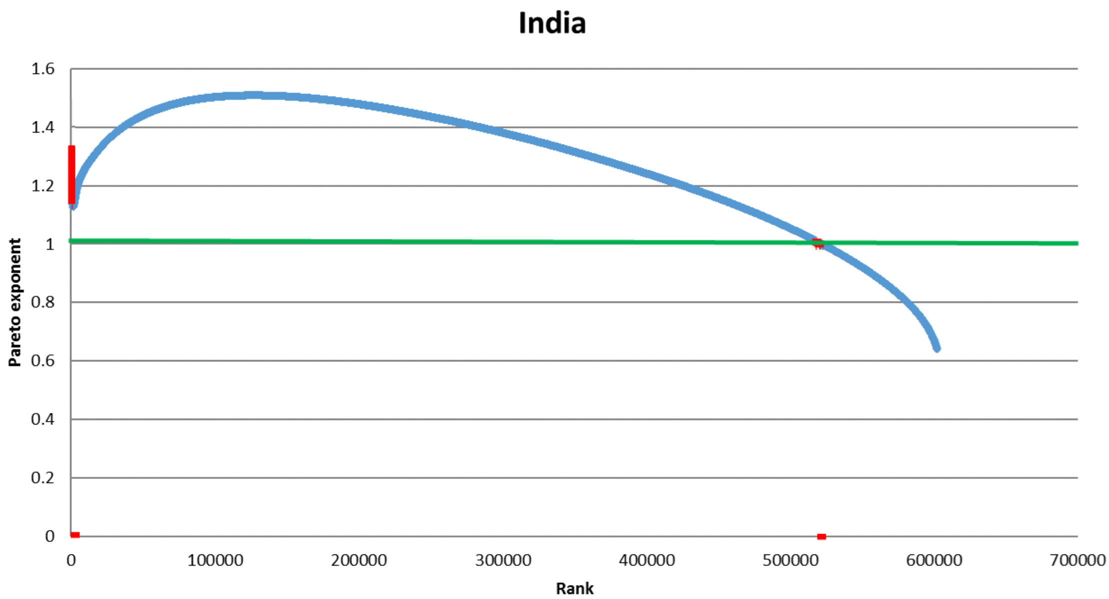
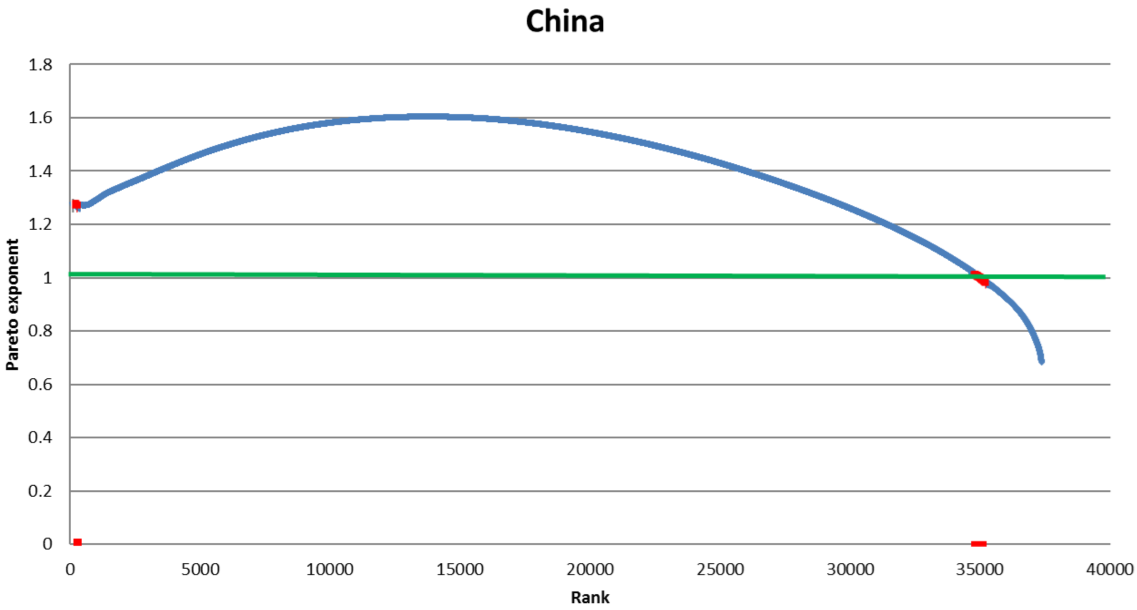


Figure 4. Pareto exponent for India (0.55% of non-rejections of Zipf's law)



Eeckhout (2004) establishes Proposition 1 in a context of a model of free mobility and suggesting a strong tendency towards urban agglomeration, in addition to increased workers' productivity. We have tried to test in a very simple way these hypotheses with our data, generating a dummy variable that takes the value 1 for the countries that fulfil the Proposition 1 of this author (Brazil, Russia and South Africa) and 0 otherwise (India and China). This variable is regressed by applying OLS over variables of mobility, productivity and density, as explanatory variables, obtained from the World Bank for the sample of BRICS for the years 2009-2018. The sample size in each regression is 50.

Figure 5. Pareto exponent for China (1.30% of non-rejections of Zipf's law)



Analysing the correlation with the density (inhabitants per km^2), it is obtained a negative and statistically significant relationship with adjusted R^2 of 0.65. This is the main outcome with respect to agglomeration. Regarding mobility, using the travel services as percentage of commercial service exports, it is found a positive and significant relationship with adjusted R^2 of 0.12. Finally, taking the GDP per worker as measure of labour productivity, again a positive and significant relationship is obtained with an adjusted R^2 of 0.76. So, this can indicate that the Eeckhout hypothesis is not fulfilled for density, but it is for labour productivity and for mobility. It is worth to mention that the previous findings are only the result of mere and simple correlations, without any aim of obtaining causality relationships. Therefore, the main differences for India and China with respect to the rest of the BRICS could be related to the higher density, lower labour productivity and, finally, a reduced mobility in these two countries. This outcome helps to explain why, in terms of Eeckhout's proposition, the behaviour of China and India is different to the behaviour of Brazil, Russia and South Africa.

According to the meta-analysis about the Pareto exponent carried out by Nitsch (2005) the estimated β belongs to the interval (0.49; 1.96). All our estimations are in that range of variation. In this context of interpretation of the magnitude of the Pareto exponent, Nitsch (2005) finds that the average β estimated of all the works he analyses is 1.09; in our case the average estimated β for all the regressions carried out for Brazil is 1.09, 0.95 for Russia, 0.91 for South Africa, 1.28 for India and 1.40 for China. Therefore, on average, the more equal distribution is that of China, followed by India, Brazil, Russia and South Africa.

Regarding Zipf's law, it is satisfied in the upper tail of the five countries[4], except South Africa. This is a standard outcome. The percentage of non-rejections (Zipf's law is fulfilled) for the whole rolling sample regressions ranges from a minimum of 0.55% for India to a maximum of 29.13% for Russia.

[4] For an interesting discussion in this context see Levy (2009) and Eeckhout (2009).

5. CONCLUSIONS

In this note we have carried out a very simple empirical exercise. Using urban data from the last available census of the five BRICS countries we have tested, by means of a rolling sample regressions approach, first, whether, as Eeckhout (2004) proposed, the Pareto exponent in a standard Zipf equation is decreasing as more cities are added to the sample, and second, the validity of Zipf's law. The results are very conclusive: Eeckhout's hypothesis is satisfied for Brazil, Russia and South Africa, but for India and China there are non-negligible parts of the distribution where it is not fulfilled; with regards to Zipf's law, it is satisfied in the upper tail of the five countries, except South Africa, but for the rest of the distribution is mostly rejected, especially in India and China.

The simple exercise carried out in this paper brings out two main reflections that we hope are relevant for Urban Economics. First, the Pareto distribution describes reasonably well the upper tail, while the lognormal (which is behind Eeckhout's, 2004, proposition), might be an optimum distribution for some countries, in our case Brazil, Russia and South Africa[5]. Second, our results for China and India open the door for the possible consideration of other densities than the Pareto alone or the lognormal alone as good descriptions of city size distribution; in effect, a combination of Pareto and lognormal (Giesen et al., 2010; Luckstead et al., 2017) or a mixture of lognormals (Puente-Ajovín et al., 2020b) have been revealed as good alternatives for the analysis of city size distribution.

REFERENCES

Anderson, G. and Y. Ge (2005) "The size distribution of Chinese cities" *Regional Science and Urban Economics* **35**(6), 756-776.

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), 1676-83.

Gabaix, X. (2016) "Power laws in Economics: An introduction" *Journal of Economic Perspectives* **30**(1), 185-206.

Gabaix, X., and R. Ibragimov (2011) "Rank-1/2: A simple way to improve the OLS estimation of tail exponents" *Journal of Business Economics and Statistics* **29**(1), 24-39.

Giesen, K., Zimmermann, A. and J. Suedekum (2010) "The size distribution across all cities: double Pareto lognormal strikes" *Journal of Urban Economics* **68**(2), 129–137.

Ioannides, Y. M., and H. G. Overman (2003) "Zipf's law for cities: An empirical examination" *Regional Science and Urban* Economics **33**(2), 127-137.

[5] Of course, this statement should be corroborated by the implementation of direct and standard lognormality tests.

Jedwab, R., and D. Vollrath (2015) “Urbanization without growth in historical perspective” *Explorations in Economic History* **58**, 1-21.

Levy, M. (2009) “Gibrat's law for (all) cities: Comment” *American Economic Review* **99**(4), 1672-75.

Liu, X., Derudder, B. and P. Taylor (2014) “Mapping the evolution of hierarchical and regional tendencies in the world city network, 2000-2010” *Computers, Environment and Urban Systems* **43**, 51-66.

Luckstead, J. and S. Devadoss (2014a) “Do the world’s largest cities follow Zipf’s and Gibrat’s laws?” *Economics Letters* **125**, 2, 182-186.

Luckstead, J. and S. Devadoss (2014b) “A comparison of city size distributions for China and India from 1950 to 2010” *Economics Letters* **124**(2), 290-295.

Luckstead, J., Devadoss, S., and D. Danforth (2017) “The size distributions of all Indian cities” *Physica A* **474**, 237-249.

Nitsch, V. (2005) “Zipf zipped” *Journal of Urban Economics* **57**(1), 86-100.

Peng, G. (2010) “Zipf’s law for Chinese cities: Rolling sample regressions” *Physica A* **389**, 3804-3813.

Puente-Ajovín, M., Ramos, A., and F. Sanz-Gracia (2020a) “City size distribution in the BRICS” *Mimeo*, Universidad de Zaragoza.

Puente-Ajovín, M., Ramos, A. and F. Sanz-Gracia, (2020b) “Is there a universal parametric city size distribution? Empirical evidence for 70 countries” *The Annals of Regional Science*, **65**, 727-741.

Rosen, K. T. and M. Resnick (1980) “The size distribution of cities: an examination of the Pareto law and primacy” *Journal of Urban Economics* **8**, 165-186.

Soo, K. T. (2005) “Zipf’s law for cities: a cross-country investigation” *Regional Science and Urban Economics* **35**, 239-263.