Letters in Spatial and Resource Sciences manuscript No.
(will be inserted by the editor)

The distribution of urban population and economic activity in the European Union and the United States

Miguel Puente-Ajovín · **Marcos Sanso-Navarro** ·
María Vera-Cabello

Received: date / Accepted: date

Abstract This letter compares the distribution of urban population and economic activity in the European Union and the United States. Economic activity, proxied with nighttime lights, is more unevenly distributed than population, especially in the European Union. This reflects that more population does not necessarily imply a higher level of economic activity. Both Zipf's law and a Pareto distribution are rejected for aggregate nighttime lights within urban extents. Therefore, alternative specifications are required for the distribution of city sizes in terms of economic activity.

Keywords City size distribution · Zipf's law · Population · Nighttime lights

JEL classification O18 · O57 · R12

Miguel Puente-Ajovín
Departamento de Análisis Económico, Universidad de Zaragoza, Zaragoza, Spain
https://orcid.org/0000-0001-9728-668X

Marcos Sanso-Navarro
Departamento de Análisis Económico, Universidad de Zaragoza, Zaragoza, Spain
https://orcid.org/0000-0002-5926-2166

María Vera-Cabello (Corresponding author)
Centro Universitario de la Defensa de Zaragoza, Zaragoza, Spain
https://orcid.org/0000-0001-8973-6804
E-mail: mvera@unizar.es

1 Introduction

As a cornerstone of urban economics, Zipf's law has been widely used to characterise the statistical distribution of city sizes. This empirical regularity – also known as the rank-size rule (Rosen and Resnick, 1980) – quantifies the concept of urban hierarchy, stating that the size of the N-th city is $1/N$ times the size of the largest one. Although the validity of Zipf's law has received a great deal of attention, there is mixed evidence at the country level, even if only the upper tail of the city size distribution is analysed; see Arshad et al. (2018) for a recent review.

Most studies about the distribution of city sizes measure them in population terms. The main exception is the work by Düben and Krause (2021), which aspires to be *"the Zipf paper for the geospatial age"* (p. 190). These authors compare the distribution of population with that of economic activity, measured using nighttime lights (Henderson et al., 2012). They find that while the distribution of urban population adjusts to the rank-size rule in most countries, this is not the case of economic activity. Düben and Krause (2021) use data based on the 'stable night light images' collected by the Defense Meteorological Satellite Program (DMSP) Operational Linescan System. Nevertheless, and among other issues, DMSP nighttime lights data are known to be flawed by blurring, geo-location errors, top-coding, lack of calibration, and coarse resolution. Furthermore, the production of DMSP data ended in 2013; see Gibson (2021), Gibson et al. (2021), and the references therein.

We adopt a supranational approach and analyse the distribution of population and economic activity in the European Union (EU) and the United States (US), see Schmidheiny and Suedekum (2015). In doing so, we consider both consistently defined functional urban areas (FUAs) and urban centres. Moreover, we exploit the more precise nighttime lights data from the Visible Infrared Imaging Radiometer Suite (VIIRS) of instruments onboard the Suomi NPP satellite, available from April 2012. The VIIRS Day/Night Band was designed to measure the radiance of lights on earth in a wide variety of lightning conditions and covers a dynamic range of about seven orders of magnitude (DMSP covers less than two), avoiding saturation problems and top-coding. VIIRS nighttime lights are comparable over time and space, do not have blurring or geo-location errors, and display, at least, 45 times greater spatial resolution than DMSP data (Elvidge et al., 2013). For all these reasons, VIIRS data are superior at attributing lights to the place where they are emitted and, as a consequence, are a better proxy for urban economic activity than DMSP images (Gibson et al., 2021).

2 Data and methods

Instead of administrative units, we use the definition of FUAs proposed by the Organisation for Economic Co-operation and Development (OECD) and the EU, which considers grid cells as spatial building blocks (Dijkstra et al., 2019). This definition of metropolitan areas integrates commuting zones into urban centres, delineated in an harmonised way as high-density areas with more than 50,000 inhabitants. The spatial information of urban centres and FUAs boundaries, as well as their corresponding georreferenced population by one square kilometer grid cells for the year 2015, have been obtained from the Global Human Settlement Layer database (GHSL; Florczyk et al., 2019), provided by the Joint Research Center of the European Commission.

VIIRS nighttime lights have been extracted from the 'vcm-orm-ntl' annual composites[1] for 2015, available at the Earth Observation Group website. This data have been cleaned to exclude background noise, solar and lunar contamination, cloud cover degradation, and features unrelated to electric lighting (Elvidge et al., 2017). At the pixel level, reported radiance values are expressed in nano Watts per square centimeter per steradian, with a resolution of 15 arc seconds (approximately 450 meters at the equator). In the same manner as population, lights data have been aggregated for all pixels included within urban extents to calculate their size.

The rank-size rule implies that the size distribution of cities can be approximated by a Pareto function with power law exponent equal to one. For this reason, cross-sectional empirical analyses of the Zipf's law are generally based on a log-log linear regression between the rank of a city and its size. In order to reduce the bias of the OLS estimator in small samples, Gabaix and Ibragimov (2011) propose the following regression model:

$$log\,(Rank_i - 0.5) = \alpha - \beta \cdot log(Size_i) + \varepsilon_i, \quad i = 1, \ldots, n; \tag{1}$$

where i is a city indicator, and n denotes the sample size. Zipf's law is equivalent to $\beta = 1$. In our context, a coefficient lower (greater) than one reflects that population/economic activity is more unequally (equally) distributed across the urban system than predicted by the rank-size rule.

3 Results

The upper panel of Table 1 shows the estimation results of expression (1) for FUAs. When their size is measured in population terms, the slope coefficient is lower in the US (0.886) than in the EU (1.040). This finding, similar to those in Schmidheiny and Suedekum (2015), does not change if country fixed effects[2] are included in the estimation for EU metropolitan areas. Therefore, it can be claimed that the US displays a more unequal distribution of population across FUAs than the EU. Although this is also the case when their size is calculated using nighttime lights, the estimated coefficients for European urban areas are more similar to the slope parameter for the US. The results included in the lower panel of Table 1 suggest that the same conclusions can be drawn when urban centres are considered as the geographical unit of reference. Contrary to Rosen and Resnick (1980), the Pareto exponent is higher for urban agglomerations than for cities when their size is measured in terms of economic activity.

The results reported in Table 1 also show that, with the exception of population in European metropolitan areas, the null hypothesis that $\beta = 1$ is rejected at the 5%

[1] VIIRS Cloud mask - Outlier removed - Nighttime lights.

[2] Fixed effects at the country level are not jointly statistically significant. These unreported results, available from the authors upon request, may be reflecting the use of a globally harmonised definition of urban entities.

Table 1 OLS estimation of the rank-size regression model in (1), year 2015.

Panel A: Funcional urban areas						
	Population			Nighttime lights		
	EU	EU	US	EU	EU	US
α	18.434***	18.691***	16.006***	14.224***	14.504***	13.331***
	(0.296)	(0.319)	(0.384)	(0.300)	(0.299)	(0.411)
β	1.040***	1.055***	0.886***	0.904***	0.924***	0.805***
	(0.024)	(0.025)	(0.030)	(0.031)	(0.032)	(0.038)
$H_0 : \beta = 1$	2.79*	4.92**	13.97***	9.73***	5.60**	25.89***
Observations	583	583	260	583	583	260
Adjusted R^2	0.96	0.96	0.95	0.91	0.91	0.89
Country FEs	No	Yes	No	No	Yes	No
Panel B: Urban centres						
	Population			Nighttime lights		
	EU	EU	US	EU	EU	US
α	18.184***	18.191***	15.330***	13.524***	13.623***	12.575***
	(0.176)	(0.191)	(0.227)	(0.244)	(0.270)	(0.251)
β	1.069***	1.070***	0.878***	0.893***	0.900***	0.785***
	(0.015)	(0.015)	(0.019)	(0.027)	(0.029)	(0.026)
$H_0 : \beta = 1$	20.95***	20.18***	38.58***	15.45***	12.31***	69.11***
Observations	694	694	323	694	694	323
Adjusted R^2	0.99	0.99	0.99	0.92	0.92	0.95
Country FEs	No	Yes	No	No	Yes	No

Note: Robust standard errors in parentheses. ***p $<$ 0.01, **p $<$ 0.05 and *p $<$ 0.10

significance level. Even so, urban size distributions have been further studied calculating the Kolmogorov-Smirnov (KS) test statistic for the Pareto function and the exact Zipf's law. The corresponding results are included in Table 2. Both distributions are rejected for population and nighttime lights in the FUAs of the two geographical areas analysed. This is also the case of economic activity in urban centres. The KS test statistic provides additional evidence of the fullfilment of Zipf's law in the urban population of the European Union. Our results also suggest that the Pareto function fits the population of urban centres in the upper tail of the size distribution.

Table 2 Kolmogorov-Smirnov test statistics.

Panel A: Functional urban areas				
	Population		Nighttime lights	
	EU	US	EU	US
Pareto distribution	0.422	0.385	0.843	0.556
	[0.000]	[0.000]	[0.000]	[0.000]
Exact Zipf's law	0.408	0.430	0.869	0.631
	[0.000]	[0.000]	[0.000]	[0.000]
Panel B: Urban centres				
	Population		Nighttime lights	
	EU	US	EU	US
Pareto distribution	0.050	0.048	0.832	0.537
	[0.062]	[0.457]	[0.000]	[0.000]
Exact Zipf's law	0.031	0.080	0.861	0.624
	[0.520]	[0.030]	[0.000]	[0.000]

Note: p-values in brackets.

Figure 1 represents urban size distributions. Zipf's law would imply that the observations locate along a decreasing straight line with slope equal to minus one. However, and except population in larger urban centres, the relationship between rank and size is concave. In addition, the estimation results reported in Table 1 show that the slopes are different to minus one. The smaller values for the coefficients of nighttime lights reflect that agglomeration effects are more important when size is measured in economic terms. The rank-size scatter plots in Figure 1 also suggest that the distributions of urban population and economic activity are more similar in the US as compared to the EU.

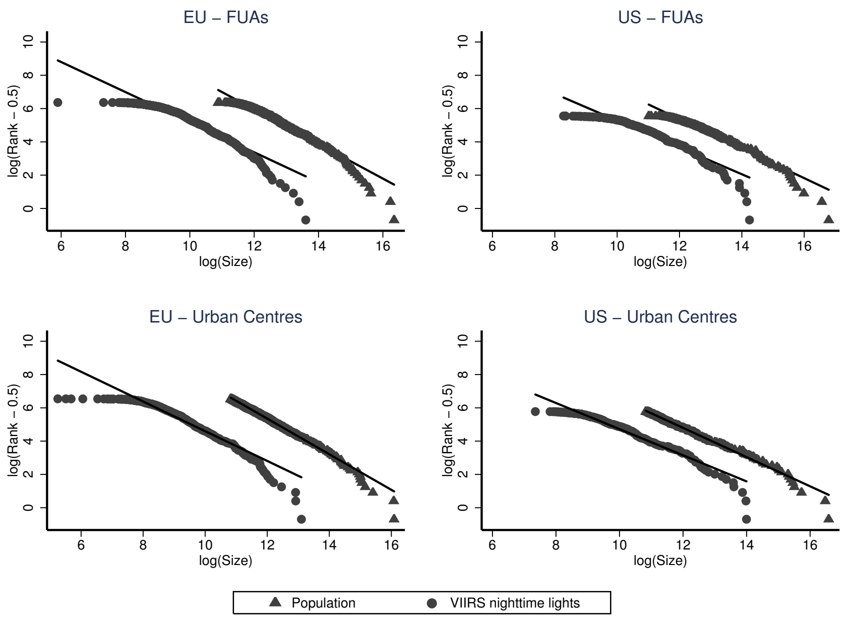

Fig. 1 Size distributions of population (triangles) and VIIRS nighttime lights (circles), year 2015. Solid lines represent the OLS linear regression fit.

4 Conclusion

The distributions of urban population and economic activity in the European Union and the United States have been compared using data for 2015. In doing so, economic activity has been proxied using newer and more accurate nighttime lights information than that commonly used in urban economics to date. Moreover, both metropolitan areas and urban centres have been considered. Agglomeration effects in the European Union and in the United States appear to be similar when urban size is measured in terms of economic activity, which is more unevenly distributed than population. This reflects that more population does not come with a higher level of economic activity, especially in the European Union. The statistical tests that have been implemented reject both Zipf's law and a Pareto distribution for urban aggregate night lights. On the one hand, these results suggest that the distribution of city sizes in terms of economic activity should be further explored with more complex functions. On the other hand, our findings cast doubts on the Pareto assumption established to correct for top-coding in DMSP nighttime lights.

Acknowledgements The authors have benefited from the valuable comments of two anonymous reviewers.

Funding This work has received financial support from Gobierno de Aragón (S39-20R ADETRE Research Group), Ministerio de Economía, Industria y Competitividad (Grant PID2020-112773GB-I00), and Fundación Ibercaja-Universidad de Zaragoza (Grant JIUZ-2020-SOC-11).

Declarations

Conflict of interest The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

References

Arshad, S., Hu, S., Ashraf, B.N.: Zipf's law and city size distribution: a survey of the literature and future research agenda. Physica A: Statistical Mechanics and its Applications **492**(15), 75–92 (2018). https://doi.org/10.1016/j.physa.2017.10.005

Dijkstra, L., Poelman, H., Veneri, P.: The EU-OECD definition of a functional urban area. OECD Regional Development Working Paper No 2019/11 (2019). https://doi.org/10.1787/d58cb34d-en

Düben, C., Krause, M.: Population, light, and the size distribution of cities. Journal of Regional Science **61**(1), 189–211 (2021). https://doi.org/10.1111/jors.12507

Elvidge, C.D., Baugh, K., Zhizhin, M., Hsu, F.C.: Why VIIRS data are superior to DMSP for mapping night time lights. Proceedings of the Asia-Pacific Advanced Network **35**(1), 62-–69 (2013). https://doi.org/10.7125/APAN.35.7

Elvidge, C.D., Baugh, K., Zhizhin, M., Hsu, F.C., Ghosh, T.: VIIRS night-time lights. International Journal of Remote Sensing **38**(21), 5860–5879 (2017). https://doi.org/10.1080/01431161.2017.1342050

Florczyk, A., Corban, C., Ehrlich, D., Carneiro Freire, S., Kemper, T., Maffenini, L., Melchiorri, M., Pesaresi, M., Politis, P., Schiavina, M., Sabo, F., Zanchetta, L.: GHSL Data Package 2019. Publications Office of the European Union, Luxembourg (2019). https://doi.org/10.2760/0726

Gabaix, X., Ibragimov, R.: Rank - 1/2: a simple way to improve the OLS estimation of tail exponents. Journal of Business & Economic Statistics **29**(1), 24–39 (2011). https://doi.org/10.1198/jbes.2009.06157

Gibson, J.: Better night lights data, for longer. Oxford Bulletin of Economics and Statistics **83**(3), 770–791 (2021). https://doi.org/10.1111/obes.12417

Gibson, J., Olivia, S., Boe-Gibson, G., Li, C.: Which night lights data should we use in economics, and where? Journal of Development Economics **149**, 102602 (2021). https://doi.org/10.1016/j.jdeveco.2020.102602

Henderson, J.V., Storeygard, A., Weil, D.N.: Measuring economic growth from outer space. American Economic Review **102**(2), 994–1028 (2012). https://doi.org/10.1257/aer.102.2.994

Rosen, K.T., Resnick, M.: The size distribution of cities: an examination of the Pareto law and primacy. Journal of Urban Economics **8**(2), 165–186 (1980). https://doi.org/10.1016/0094-1190(80)90043-1

Schmidheiny, K., Suedekum, J.: The pan-European population distribution across consistently defined functional urban areas. Economics Letters **133**, 10–13 (2015). https://doi.org/10.1016/j.econlet.2015.05.013