The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

Working Papers

www.cesifo.org/wp

The Pan-European Population Distribution Across Consistently Defined Functional Urban Areas

Kurt Schmidheiny
Jens Suedekum

CESIFO WORKING PAPER NO. 5335
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
APRIL 2015

ISSN 2364-1428

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5335*

The Pan-European Population Distribution Across Consistently Defined Functional Urban Areas

Abstract

We analyze the first data set on consistently defined functional urban areas in Europe and compare the European to the US urban system. City sizes in Europe do not follow a power law: the largest cities are "too small" to follow Zipf's law.

JEL-Code: R110, R120.

Keywords: city size distributions, Zipf's law, functional urban areas, urban systems.

Kurt Schmidheiny
University of Basel
Department of Economics
Peter-Merian-Weg 6
Switzerland – 4002 Basel
kurt.schmidheiny@unibas.ch

Jens Suedekum
Heinrich-Heine-University
Duesseldorf Institute for Competition Economics (DICE)
Universitätsstrasse 1
Germany – 40225 Duesseldorf
suedekum@dice.hhu.de

We thank Lewis Dijkstra for providing us with the EC-OECD city data for Europe on the basis of a noncommercial access agreement for scientific use. This research was supported by the German National Science Foundation (DFG) under grant SU 413/2-1, by the Swiss National Science Foundation under grants Sinergia/130648 and 147668 and by the Spanish Ministerio de Economía y Competitividad under project MINECO-ECO2012-36200.

1) Introduction

Beginning with the seminal contributions by Auerbach (1913) and Zipf (1949), there is a long literature on the distribution of population across cities. Virtually all of this research is concerned with cities of the same country. Gabaix (1999) focuses on the United States (US) and shows that population sizes across metropolitan statistical areas (MSAs) closely follow a Pareto distribution with shape parameter close to minus one, also known as Zipf's law. Further studies on the US city size distribution and the underlying urban growth process include Eeckhout (2004), Overman and Ioannides (2003), and Black and Henderson (1999). Focusing on other countries, Eaton and Eckstein (1999), and Giesen and Suedekum (2011) obtain evidence for France, Japan, and Germany, respectively, that is consistent with the US experience.

Much less is known about city sizes in a wider context than the nation state, however, even though national borders are steadily losing significance in the ongoing process of economic globalization. The reason is that "cities" are usually not consistently classified; instead, each country adopts its own methods of defining urban areas and delineating their boundaries according to administrative or economic criteria. This is even true in Europe, where official approaches and city definitions differ widely across countries. For this research, we use novel and unexplored data, which allow for a harmonized approach to defining urban areas in 31 European countries and, with the same methodology, in the US. Our goal is to address the pan-European distribution of city sizes, and to compare the European to the American urban system.

2) Data

The novel data stem from a collaboration of the European Commission (EC), see Dijkstra and Poelman (2012) and the Organization for Economic Co-operation and Development (OECD),

see Brezzi et al. (2012). The EC-OECD definition of functional urban areas proceeds in three steps: Step 1 partitions the European surface into 1 km^2 grid cells and identifies *high-density cells* with a population density greater than 1,500 inhabitants per km^2 based on categorized satellite images. Step 2 generates clusters of contiguous (sharing at least one border) high-density cells. Low-density cells encircled by high-density cells are added. Clusters with a total population of at least 50,000 inhabitants are identified as *urban centers*. Step 3 uses administrative data to calculate commuting flows from local administrative units (municipalities) into urban centers. Local administrative units with 15% of employed persons working in an urban center are assigned to the urban center. A contiguous set of assigned local administrative units form a *larger urban zone*. Non-contiguous local urban centers with bilateral commuting flows of more than 15% of employed persons are combined into a polycentric larger urban zone.

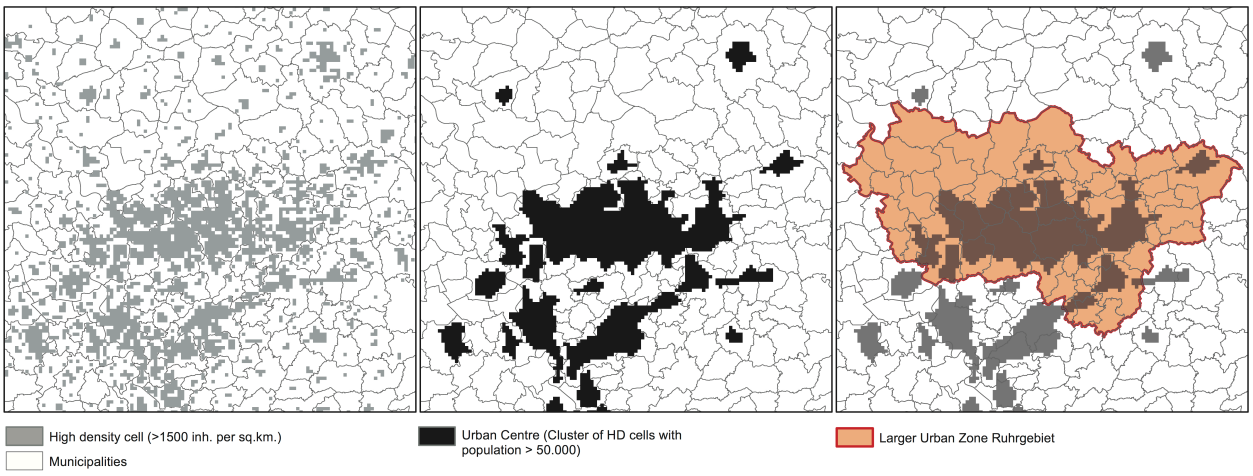

Fig. 1: Construction of the Ruhrgebiet (Germany) functional urban area. The left panel shows the high-density cells with more than 1500 inhabitants per square kilometer and administrative municipal boundaries. The middle panel illustrates the construction of urban centers with a total population of more than 50,000 inhabitants. The right panel shows the construction of the larger urban zone based on bilateral commuting flows. Source: European Commission, Directorate-General Regional and Urban Policy.

Figure 1 provides an example, where the single panels illustrate the three steps for the case of the Ruhr area (*Ruhrgebiet*) in Germany. Table 1 gives an overview of the European urban hierarchy across the resulting 692 functional urban areas in Europe in the year 2006.

Table 1. Population size (number of inhabitants) across 692 urban areas in Europe in 2006.

Rank	Urban area name	Population
1	Paris (FR)	11,370,846
2	London (UK)	11,256,669
3	Madrid (ES)	5,993,683
4	Ruhrgebiet (DE)	5,280,039
5	Berlin (DE)	4,980,394
6	Barcelona (ES)	4,374,747
7	Milano (IT)	4,052,933
8	Athens (GR)	4,045,748
9	Roma (IT)	3,850,688
10	Napoli (IT)	3,545,095
…	…	…
22	Amsterdam (NL)	2,381,265
…	…	…
135	Bydgoszcz (PL)	489,204
…	…	…
282	Algeciras (ES)	263,244
283	Bayreuth (DE)	259,547
…	…	…
570	Targoviste (ROM)	120,141
571	Cáceres (ES)	119,493
…	…	…
690	Acireale (IT)	54,978
691	Santa Lucía de Tirajana (ES)	53,630
692	Mollet del Vallès (ES)	51,648

This EC-OECD definition of urban areas has important advantages over using population data for administratively defined cities. The algorithm, for example, identifies the *Ruhrgebiet* as the largest German city. This larger urban zone comprises the four administrative cities Duisburg, Essen, Bochum, and Dortmund, which form a contiguously populated cluster but are reported as individual cities in traditional data. The algorithm also assigns larger urban zones across national borders, for example Geneva and Basel, which consist of urban centers not only

in Switzerland but extend into France and Germany/France, respectively. Finally, the EC-OECD data is complimentary to other approaches that draw on fine-grained satellite images, such as Rozenfeld et al. (2008, 2011) who build on population clusters, or Henderson et al. (2021) who exploit data on night-lights. While those data also ignore artificial administrative boundaries when defining "cities", they neglect economic linkages across cities such as commuting flows. The EC-OECD data considers such linkages, and thus combines the key advantages of the satellite-based approach and the traditional delineations of functional urban areas. To the best of our knowledge, this paper is the first to analyze this novel data source for city sizes.

3) Analysis for Europe

In Figure 2, we depict the pan-European population distribution across all 692 functional urban areas. The data are arranged as a logarithmic rank-size plot, which is typically used to illustrate city size distributions. We observe that the relationship is straight in the intermediate range of city sizes with populations roughly between $\exp(11.7) \approx 120{,}000$ and $\exp(14.7) \approx 2{,}4000{,}000$ inhabitants. In that range, which spans the ranks 22 to 570 in the urban hierarchy, city sizes can be approximated by a power law distribution. Outside this range, there are three notable deviations.

First, the plot turns concave for smaller cities. This, however, is a typical feature that is observed in many individual countries. See Eeckhout (2004), Rozenfeld et al. (2011), or Giesen and Suedekum (2014), who emphasize that the power law behavior pertains to the upper tail of the distribution only. Second, the plot also turns concave for large cities. On a pan-European scale, the largest cities are thus "too small" relative to a power law distribution. Within single European countries, this often tends to be the opposite. Here we observe that the largest city

within a country is often "too large" for a power law; examples include Vienna in Austria or Budapest in Hungary. Finally, the third key feature of the pan-European urban system is that the two largest cities (Paris and London) roughly have the same population size, which clearly violates the rank-size rule that would pertain under Zipf's law.

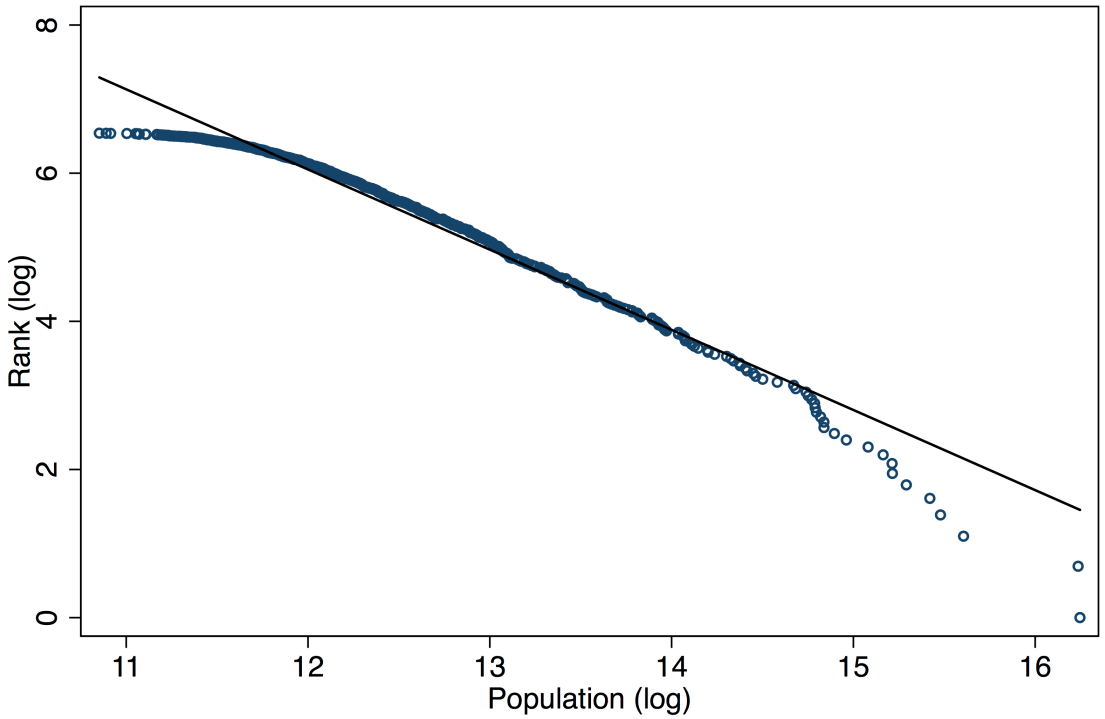

Fig. 2. Population distribution across 692 functional urban areas in Europe, 2006. Horizontal axis depicts the natural logarithm of the city population. Vertical axis depicts the natural logarithm of the city's rank when all cities are ordered according to their population size (see Table 1). The solid line is the OLS regression line (assuming a common intercept).

Next, we investigate the distributional properties of the pan-European city size distribution formally by running a standard rank-size regression:

$$\log\left(rank_i\right) = \alpha - \xi \cdot \log\left(population_i\right) + u_i$$

This regression would deliver a slope coefficient $\zeta \cong 1$ with $R^2 \approx 1$ if Zipf's law held exactly. We use simple ordinary least squares (OLS) estimation with White-robust standard errors, and consider two specifications: one with the assumption of a common intercept term, and one where we allow the intercept α to differ across countries by introducing country-fixed effects. The number of observations is N = 692 in all regressions and results are as follows (robust standard errors in parentheses):

Regression (1): OLS, common intercept

$\hat{\zeta}$ = 1.082 (0.018), adj. R^2 = 0.968

Regression (2): OLS, country-specific intercepts

$\hat{\zeta}$ = 1.090 (0.019), adj. R^2 = 0.970

Two main observations arise: First, allowing for country-specific intercept terms has only a negligible effect on $\hat{\zeta}$, and the country-specific intercepts turn out to be jointly insignificant (p-value = 0.257). This is consistent with the homogeneous city definition across countries in our data set. Second, the null hypothesis $\zeta = 1$ is rejected at 0.1% significance level in both regressions (p-value < 0.001). Hence, Zipf's law cannot describe the city size distribution in Europe. The estimated regression line is steeper than the Zipf hypothesis would predict which reinforces the above finding that the largest European cities are "too small".

Regressions such as (1) and (2) can be problematic, because the ordering of cities by their size may yield a spurious rank-size correlation. Gan el al. (2006) have shown this with Monte Carlo simulations, and they suggest a Kolmogorov-Smirnov nonparametric test as an alternative approach to investigate city size distributions. Following their suggestion, we have performed

that test against the null hypothesis of Pareto distributed city sizes.[1] This hypothesis is strongly rejected statistically (p-value < 0.001). A Kolmogorov-Smirnov test against the null hypothesis of the exact Zipf law (Pareto distribution with $\zeta = 1$) is also strongly rejected (p-value < 0.001), which is in line with our previous result that city sizes in Europe do not follow Zipf's law.[2]

4) Europe versus the United States

For the US urban system, Krugman (1996) and Gabaix (1999) find that Zipf's law holds exactly when considering the 135 largest US MSAs with a minimum population threshold of 280,000 inhabitants. To facilitate a closer comparison, we analyze 2008 US city size data stemming from the same EC-OECD data project and using the analogous definition of functional urban areas in the US as discussed in Section 2 above.

To be consistent with previous studies, we also consider only the largest 135 urban areas, which leads to a minimum population size of 261,952 (Atlantic City, NJ) in our case. Repeating regression (1) with this US data, we obtain $\hat{\xi} = 0.960$ (robust standard error = 0.029) which does not reject a coefficient of one at usual significance levels (p-value = 0.172). A Kolmogorov-Smirnov test also cannot reject the Pareto distribution (p-value = 0.231) or the exact Zipf's law at the 10% significance level (p-value = 0.109). These findings thus corroborate earlier results on the exact validity of Zipf's law in the upper tail of the US city size distribution using the novel data from EC-OECD.

[1] The cumulative distribution function of the Pareto distribution for cities with population (pop) larger than the minimum threshold pop_{min} is $F(pop) = 1 - (pop_{min}/pop)^{\zeta}$. We perform the Kolmogorov-Smirnov test using the OLS estimator for ζ and the population of the smallest city in the estimation sample for pop_{min}.

[2] As another robustness check, we have also considered the regression approach developed by Gabaix and Ibragimov (2011) which addresses the small-sample bias of the OLS estimator but not the spurious correlation problem emphasized by Gan et al. (2006). This approach yields similar R^2 levels and point estimates for ζ as regressions (1) and (2), but considerably larger standard errors.

For Europe, by contrast, we obtain $\hat{\xi}$ = 1.362 (0.035) when using only the largest 135 cities, and $\hat{\xi}$ = 1.291 (0.022) when imposing a minimum threshold of 260,000 (largest 282 cities). Based on the OLS estimates, the Zipf null hypothesis $\xi = 1$ is thus even more decisively rejected after these data truncations.[3]

5) Summary and discussion

Our research reveals that the pan-European urban system to date still differs substantially from the American one. The emergence of Zipf's law in an urban system requires time and evolves in parallel with the general degree of economic integration of that area. See Black and Henderson (1999) and Krugman (1996), who discuss the US urban system in historical perspective.

A candidate explanation for the deviations from Zipf's law in Europe is, therefore, that the area is still much less integrated economically. For example, individuals are less mobile within and particularly across countries, and regulatory regimes are more diverse than within the US. To the extent that the actual degree of economic integration in Europe is likely to increase in the future (which is beyond the scope of this paper to discuss), we may expect a transition towards a clearer Zipf-pattern in the European urban system. This would involve substantial redistribution of the population leading to stronger population concentration in the biggest cities, and potentially the emergence of one primate city (likely Paris or London) at the pan-European level. The current policy debate, for example in the UN development report, acknowledges recent tendencies of increasing urbanization but focuses mainly on trends in Asia, Latin America and Africa. Our research suggests that increasing urbanization might also happen in Europe.

[3] The exact Zipf's law is also rejected at the 10% significance level by the Kolmogorov-Smirnov test using the 260,000 minimum size threshold (p-value = 0.054) but not using the top 135 cities threshold (p-value = 0.225).

Literature

Auerbach, F. (1913), Das Gesetz der Bevölkerungskonzentration, *Petermanns Geographische Mitteilungen* **59**, 74-76.

Black, D., J.V. Henderson (1999), A theory of urban growth, *Journal of Political Economy* **107**, 252-284.

Brezzi, M., M. Piacentini, K. Rosina, D. Sanchez-Serra (2012), Redefining urban areas in OECD countries, in: OECD Publishing, *Redefining "urban": a new way to measure metropolitan areas*. http://dx.doi.org/10.1787/9789264174108-en

Dijkstra, L., H. Poelman (2012), *Cities in Europe: the new OECD-EC definition*, EU Commission: Regional Focus 1/2012.

Eaton, J., Z. Eckstein (1999), Cities and growth: theory and evidence from France and Japan, *Regional Science and Urban Economics* **27**, 443-474.

Eeckhout, J. (2004), Gibrat's law for (all) cities, *American Economic Review* **94**, 1429-1451.

Gabaix, X. (1999), Zipf's law: an explanation, *Quarterly Journal of Economics* **114**, 739-767.

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

Gan, L., D. Li, S. Song (2006), Is the Zipf law spurious in explaining city-size distributions?, Economics Letters **92**, 256-262.

Giesen, K., J. Suedekum (2011), Zipf's law for cities in the regions and the country, *Journal of Economic Geography* **11**, 667-686.

Giesen, K., J. Suedekum (2014), City age and city size, *European Economic Review* **71**, 193-208

Henderson, J.V., A. Storeygard, D.Weil (2012), Measuring economic growth from outer space, *American Economic Review* **102**, 994-1028.

Krugman P. (1996), Confronting the mystery of urban hierarchy, *Journal of the Japanese and International Economies* **10**, 399-418.

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

Rozenfeld, H., D. Rybski, X. Gabaix, H. Makse (2011), The area and population of cities: new insights from a different perspective on cities, *American Economic Review* **101**, 2205-2225.

Rozenfeld, H., D. Rybski, J. Andrade, M. Batty, E. Stanley, H. Makse (2008), Laws of population growth, *Proceedings of the National Academy of Science* **105**, 18702-18707.

Zipf, G.K. (1949), *Human behavior and the principle of least effort*, Cambridge: Addison-Wesley