**MISCELLANEA GEOGRAPHICA** – REGIONAL STUDIES ON DEVELOPMENT
Vol. 20 • No. 4 • 2016 • pp. 5-10 • ISSN: 2084-6118 • DOI: 10.1515/mgrsd-2016-0020

# **Does Zipf's law** hold for Polish cities?

**Abstract**
In this paper we study Zipf's law, which postulates that the product of a city's population and its rank (the number of cities with a larger or equal population) is constant for every city in a given region. We show that the empirical literature indicates that the law may not always hold, although its general form, the rank-size rule, could be a good first approximation of city size distribution. We perform our own empirical analysis of the distribution of the population of Polish cities on the largest possible sample to find that Zipf's law is rejected for Poland as the city sizes are less evenly distributed than it predicts.

**Keywords**
Zipf's law • rank-size rule • city size distribution • Poland



**Andrzej Cieślik, Jan Teresiński**

Department of Macroeconomics and International Trade Theory
Faculty of Economic Sciences
University of Warsaw
*e-mail: cieslik@wne.uw.edu.pl*

European University Institute of Florence
*email: Jan.Teresinski@EUI.eu*

Received: 19 February 2016
Accepted: 4 November 2016

## Introduction

As early as Marshall (1920), economists have been attempting to identify the agglomeration forces that lead to the concentration of economic activity in a given place – the city. Contemporary economic theory explains this concentration by referring to localization and urbanization externalities. Although the theory is able to explain why cities are created and why they grow, so far relatively little is known about how human populations are distributed within an urban system and why cities differ in size.

In particular, an astonishing empirical regularity of population distribution can be found, which can be described by the so-called rank-size rule. This rule, and especially its particular form known as Zipf's law, is a striking phenomenon that characterizes the pattern of city size distribution in different countries and regions of the world, as argued by certain authors. For example, according to Gabaix (1999, p. 129) Zipf's law is "the most accurate regularity in economics" that "appears to hold in virtually all countries and dates for which there are data". However, more recent empirical studies have called these results into question.

Many empirical studies on the applicability of Zipf's law have already been carried out for specific countries, as well as cross-sections of countries. However, empirical evidence on Zipf's law for the new members of the European Union still remains relatively scarce. In particular, only a few studies have been devoted to Poland and they have yielded mixed results. Poland is the largest and most populous country of the new EU member states and has an interesting history where path-dependence could play a vital role in shaping the location and distribution by size of its cities.

Therefore, the main goal of this paper is to extend the previous studies and to test whether Zipf's law holds for the distribution of cities in Poland, using the most recent and largest dataset, which in the extended version accounts for 908 observations. The baseline hypothesis of our empirical analysis postulates that Zipf's law is a good description of the Polish city rank distribution. We validate this hypothesis using the extensive dataset on the population of Polish cities based on the last population census from 2011 (derived from the Polish Central Statistical Office (CSO)) using the simple ordinary least squares (OLS) method, as well as the more recent approach proposed by Gabaix and Ibragimov (2011), which allows correcting for small sample bias. Our empirical results show that, irrespective of the employed empirical approach, Zipf's law is generally rejected for Poland, as city sizes are less evenly distributed than it predicts.

The remainder of the paper is organized as follows. First, we provide a literature review on the previous empirical tests of Zipf's law. Next, we describe our dataset, the research methodology and the main hypotheses of the empirical investigation for Polish cities. Subsequently, we report our estimation results. Finally, we summarize and provide directions for future research.

## Literature review

Zipf's law is a special case of rank size distribution. City size is usually measured by its population. When all the cities in a given region are taken into account, they can be ranked according to size, with a ranking of one given to the largest city. The rank-size rule of city size distribution states that the population of a city is

Vol. 20 • No. 4 • 2016 • pp. 5-10 • ISSN: 2084-6118 • DOI: 10.1515/mgrsd-2016-0020

given by the quotient of a constant term and the rank of the city taken to a power close to 1, according to the following formula:

$$N_j = \frac{c}{R_j^{a}} \qquad (1)$$

where $N_j$ is the population of the city $j$, $R_j$ is the rank of the city $j$, $\alpha$ is an exponent close to 1, and $\alpha$ is a constant term.

Zipf's law is a kind of Pareto distribution applied to city sizes. However, it is stricter than the general Pareto distribution, since it imposes restrictions on the exponent and the constant of the aforementioned formula.[1] In its canonical form, Zipf's law holds if and only if $\alpha = 1$ . It is also claimed that the constant $\alpha$ is equal to the population of the largest city, which means that on average the largest city is *j* times larger than the *j*-th largest city.

For estimation purposes, the rank-size rule is often expressed in logarithmic terms as:

$$\log N_j = \log c - \alpha \log R_j \qquad (2)$$

The test of the law's existence for city size distribution in a given region is based on estimating equation (2). This equation allows the verification that the estimated value of exponent $\alpha$ does not differ significantly from unity, while the estimated constant $\alpha$ does not differ significantly from the size of the largest city.[2]

If the obtained estimate of $\alpha$ is close to 1 and does not differ significantly from this value, it means that, on average, Zipf's law holds. If $\alpha$ is estimated to be above 1, then the distribution of city sizes is more uneven than predicted by Zipf's rule (on average, the largest city is more than *j* times larger than the *j*-th largest city). If the estimate of $\alpha$ is smaller than 1, it means that the distribution of city sizes is even more than that described by Zipf's law (on average, the largest city is less than *j* times larger than the *j*-th largest city). In an extreme case when $\alpha = 0$, all cities have the same population.

The first widely cited paper that provided extensive empirical evidence of Zipf's rule for the majority of countries was a study by Rosen and Resnick (1980). Using data from a sample of 44 countries, they found that, for the majority of the national metropolitan size distributions, the rank-size rule was a good description and exponent $\alpha$ was not too far from unity. The mean value of the estimated exponent $\alpha$ for the whole sample was 0.88, while the estimates ranged from 0.509 (for Australia) to 1.236 (for Morocco). The estimate obtained for Poland in 1970 was 0.887. However, it should be emphasized that, for the majority of countries, the obtained estimates of the Zipf's law exponent were below 1. This result suggested that the populations of the cities were more evenly distributed than Zipf's law would predict.[3]

It must be remembered, however, that Rosen and Resnick (1980) results may seem somewhat problematic nowadays, since the authors used data that was relatively old (samples from 1970). In addition, their study suffered from estimation bias, as later suggested by Gabaix and Ioannides (2004). These facts motivated a large number of follow-up studies.[4] In particular, Soo (2005) verified the Zipf's rule validity using data for 73 countries, including Poland. He employed two standard approaches: the ordinary least squares (OLS) method and the Hill (1975) estimator, a maximum likelihood estimator robust to the small-sample bias of OLS.

The OLS regression revealed that, on average, the Zipf's law exponent was 0.9 and the law was rejected for the majority of countries (53), including Poland, as the value of the estimated exponent differed significantly from unity. According to Soo's (2005) OLS estimates for 180 cities, the parameter obtained for Poland in 1998 was 0.845. On the other hand, when the Hill estimator was used, Zipf's rule was rejected only for the minority (29) of countries and the mean of $\alpha$ was 0.860. In the case of the Hill estimator, the parameter obtained for Poland was 0.917. Hence, the outcome of the estimation depended heavily on the estimation method used.

Another extensive empirical study on Zipf's law for a large number of countries, including Poland, can be found in the popular textbook on geographical economics by Brakman, Garretsen and Marrewijk (2009, pp. 306-309). Using the simple OLS technique, they found that, depending on the definition of the city, the mean value of $\alpha$ was 0.78 for a proper city (in administrative boundaries) and 0.91 for an urban agglomeration. The values of $\alpha$ ranged from 0.48 to 1.19 (city proper) and from 0.59 to 1.22 (urban agglomeration). The estimated value for Poland obtained for the 42 biggest cities was 0.674.

In the context of Central and Eastern European countries, Necula et al. (2010) analysed the city size distributions of transition states. In particular, they also used data for Poland. Their sample included the largest Polish cities of over 100,000 inhabitants for the period 1970–2007. Depending on the year, the number of cities ranged from 23 to 43. They estimated the value of $\alpha$ for each year using the simple OLS method, as well as the maximum likelihood estimator (MLE). For OLS, estimates for the value of $\alpha$ ranged from 0.687 to 0.746, while in the case of MLE, estimates ranged from 0.712 to 0.855. These results are generally in line with the results obtained by Rosen and Resnick (1980), Soo (2005), and Brakman, Garretsen and Marrewijk (2009), who found that Polish cities were more evenly distributed than Zipf's law predicts.

To the best of our knowledge, to date there has been only one study devoted to researching the applicability of Zipf's rule exclusively for Poland, by Deichmann and Henderson (2000). They estimated the value of parameter $\alpha$ for six years: 1950, 1960, 1970, 1980, 1990 and 1995. Their sample in each year consisted of the five urban agglomerations, plus 26 municipalities. They found that the value of $\alpha$ ranged from 1.351 in 1950 to 1.111 in 1995. This means that the distribution of Polish cities was more uneven than that predicted by Zipf's law. Their results differ from the results obtained by Rosen and Resnick (1980), Soo (2005), Brakman, Garretsen and Marrewijk (2009), and Necula et al. (2010). These conflicting results clearly indicate the need for a new and more extensive study for Poland.

---

[1] The law is named after George Zipf (1949), who popularized the rule. However, the original discoverer of the law seems to be Felix Auerbach (1913).

[2] Care should be taken in analysing the results of previous studies since many authors prefer to estimate a version of equation (2) in which the rank and population of a city are switched with one another. In order to avoid confusion, we converted all the estimates of exponent $\alpha$ reported in this paper so that they are consistent with equation (2). The specification in which the logarithm of the city population is a dependent variable is sometimes referred to as Lotka's (1925) formulation of the rank-size relationship.

[3] In addition, the results also indicated that the exponent got closer to unity when the metropolitan areas were more carefully defined.

[4] There are many interesting studies on the applicability of Zipf's law, but summarizing their results goes beyond the scope of this paper. Instead, due to the space constraints, in this literature review we limit our attention to the studies related to Poland. An extensive review of the empirical literature on Zipf's law was provided by Nitsch (2005), who analyzed 29 studies that accounted for a total of 515 estimates of exponent $\alpha$. Those estimates ranged from 0.51 to 2.04. Having performed the statistical meta-analysis, he found that the coefficient was significantly smaller than unity and was not constant over time. Thus, on average, cities in different countries around the world were more evenly distributed than Zipf's law predicted. He also showed that the estimates of $\alpha$ got closer to 1 as agglomerations rather than administrative cities were considered.

Vol. 20 • No. 4 • 2016 • pp. 5-10 • ISSN: 2084-6118 • DOI: 10.1515/mgrsd-2016-0020

In summary, it can be argued that there is no clear empirical support in the literature for the validity of Zipf's law in general. Furthermore, the specific results from the previous empirical studies for Poland are mixed. In addition, these results were based only on a very limited number of Polish cities, while our study covers a bigger and more recent sample. Therefore, the main value added by this paper lies in the empirical verification of Zipf's law for Poland using a more recent and more extensive dataset compared to the previous literature.

## Statistical data and research methodology

In this section we provide details concerning the dataset and the research methodology employed to empirically verify whether Zipf's law holds for the distribution of cities in Poland. For the purpose of this verification, we use population data from the results of the National Census conducted in 2011 by the Polish Central Statistical Office (CSO, 2011).[5] The CSO database shows the population of different administrative units, from voivodeships to the smallest entities called *gmina* (communities).[6]

The *gmina* is the principal unit of administrative division of Poland. There are three types of *gmina* – urban, which consists of one city or town; urban-rural, which consists of a town and the surrounding villages; and rural, which consists only of villages and countryside. The size of *gminas* can vary. The largest urban *gmina* is Warsaw, the capital of Poland with almost 2 million inhabitants, while the smallest urban *gmina* is Krynica Morska with as few as just over a thousand inhabitants.

In our benchmark empirical test, we use only urban communities, giving us 306 observations. However, it has often been argued in the literature that estimation results might be affected by the choice of the unit of analysis (Berry and Okulicz-Kozaryn, 2011; Modica, 2014; and Veneri, 2016). Therefore, for the sensitivity analysis we extend the baseline sample by adding the municipal part of the urban-rural communities.[7]

The resulting extended sample has 908 observations and is almost three times larger than the baseline for urban communities alone. The extended sample is, to our best knowledge, the largest dataset used to date in validating Zipf's law for Poland. Figure 1 provides an initial insight into the distribution of the population of Polish cities.[8]

As indicated above, we estimate the Zipf's regression for the two samples – baseline and extended – in order to use all available information on Polish cities. However, there is an intense debate in the literature on the threshold that should be used in verifying Zipf's law (see Nitsch 2005). For example, as noted by Eeckhout (2004), the estimation of the Zipf's coefficient decreases with the truncation point. The way to define the right truncation point was discussed in recent studies by Bee et al. (2013), Ioannides and Skouras (2013), and Fazio and Modica (2015), who proposed different possible criteria for choosing where to draw the cut-off line for the sample of cities that should be included in the empirical analysis.

In order to take these issues into account in our empirical study, we also apply a certain reasonable threshold and estimate the regression in order to be as close as possible to the strict version of Zipf's law. However, in contrast to the aforementioned studies, in our empirical approach we adjust the cut-off somewhat

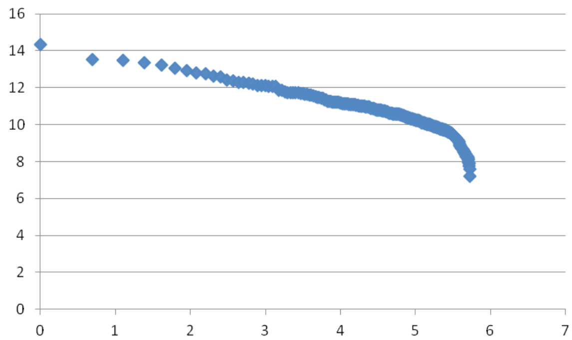


*Figure 1. City size distribution in Poland (baseline sample)*
*Source: Own calculation on the basis of CSO (2011) data.*

to the result we would like to obtain, in order to see the threshold under which the desired outcome can be reached.

In addition, in some of our regressions for the baseline and extended samples, we also exclude Warsaw from the dataset, since some authors claim that Zipf's law may not hold due to the fact that the largest city in many countries is much larger than would be predicted by the Zipf's distribution.

When it comes to the econometric methodology, in order to maintain comparability with the previous studies, we employ the standard ordinary least squares method to obtain our baseline estimation results.[9] Therefore, for the purpose of our empirical verification of Zipf's law, we first estimate the specification described by equation (2) via the standard OLS estimation method.

However, Gabaix and Ioannides (2004) demonstrated using Monte Carlo simulations that, for the small samples usually used in the city size distributions literature, the OLS estimates of the Zipf's law exponent were biased upwards, while standard errors were underestimated. The latter resulted in too many rejections of Zipf's law. As an alternative approach to estimating the rank-size rule without bias, they suggested using the Hill (1975) estimator, which was used for example by Soo (2005), as mentioned in the literature review section.

Subsequently, Gabaix and Ibragimov (2011) showed that the bias of the OLS could be overcome using an even simpler method, namely by shifting the rank by one half. Once the OLS method was applied over the shifted rank, the obtained estimate was unbiased. Therefore, due to the aforementioned bias of the OLS estimates in small samples, as indicated by Gabaix and Ioannides (2004), we also employ the approach proposed by Gabaix and Ibragimov (2011), shifting the rank by one half to overcome this problem.[10] To overcome the potential bias problem, we estimate an equation with the shifted rank that takes the following form:

$$\log N_j = \log c - \alpha \log(R_j - \frac{1}{2}) \qquad (3)$$

Our benchmark research hypothesis states that Zipf's law is a good description of the population distribution of Polish cities. To verify this, the estimated value of $\alpha$ should be close to 1. If it is above one, as in Deichmann and Henderson (2000), it would mean that the distribution of Polish cities is more uneven than that predicted

[5]http://stat.gov.pl/obszary-tematyczne/ludnosc/ludnosc/ludnosc-w-gminach-wedlug-stanu-w-dniu-31-12-2011-r-bilans-opracowany-w-oparciu-o-wyniki-nsp-2011,2,1.html

[6]http://isap.sejm.gov.pl/DetailsServlet?id=WDU20160000446

[7]Extending the size of the sample by including small or very small communities imposes the significant problem of the definition of what a city actually is, since the populations of some rural communities in Poland (*gminas*) are larger than those of urban communities.

[8]Summary statistics for both the baseline and extended samples are reported in Table A1 in the Appendix.

[9]We also considered the use of panel data methods, but we had to reject them due to data constraints as we do not have a sufficiently large sample for years other than 2011. In addition, the panel data analysis is somewhat problematic since some cities change their rankings over time, which complicates the empirical investigation.

[10]However, it should be mentioned that, compared to previous studies for Poland, our sample is relatively large.

Vol. 20 • No. 4 • 2016 • pp. 5-10 • ISSN: 2084-6118 • DOI: 10.1515/mgrsd-2016-0020

*Table 1. Estimation results*
*(robust standard errors in parentheses)*

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 1.114*** (0.037) | 1.186*** (0.022) | 1.135*** (0.036) | 1.196*** (0.021) | 1.077*** (0.024) | 1.167*** (0.027) | 1.008*** (0.026) | 1.000*** (0.015) |
| Constant | 15.538*** (0.175) | 16.072*** (0.132) | 15.644*** (0.170) | 16.143*** (0.123) | 15.350*** (0.118) | 15.957*** (0.158) | 15.142*** (0.121) | 15.204*** (0.081) |
| $R^2$ | 0.874 | 0.939 | 0.873 | 0.940 | 0.865 | 0.934 | 0.940 | 0.975 |
| Sample | Baseline with Warsaw | Extended with Warsaw | Baseline without Warsaw | Extended without Warsaw | Baseline | Extended | Baseline | Extended |
| number of observations | 306 | 908 | 305 | 907 | 306 | 908 | 280 | 600 |
| estimation method | OLS | OLS | OLS | OLS | ½ | ½ | OLS | OLS |
| t-test for $a$ = 1 | 3.034 | 8.384 | 3.707 | 9.576 | 3.133 | 6.271 | 0.321 | 0.010 |
| p-value | (0.003) | (0.000) | (0.000) | (0.000) | (0.002) | (0.000) | (0.749) | (0.992) |
| t-test for constant = log of largest city population | 6.771 | 13.043 | 12.371 | 21.163 | 8.468 | 10.138 | 6.542 | 10.485 |
| p-value | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |

*Notes: *** - statistically significant at the 1% level; ** - statistically significant at the 5% level, * - statistically significant at the 10% level.*
*Source: Own calculations in STATA.*

by Zipf's law. If it is below unity, as in Rosen and Resnick (1980), Soo (2005), Brakman, Garretsen and Marrewijk (2009), and Necula et al. (2010), then Polish cities are more evenly distributed than Zipf's law predicts. In addition, we also verify the hypothesis that the estimated constant $\alpha$ does not differ significantly from the size of the largest city.

## Estimation Results

In this section we report and discuss the results of our empirical investigation. Four sets of our estimation results are shown in Table 1. First, in columns (1)-(2), we report the OLS estimates for the baseline and extended samples respectively, including the capital city, Warsaw.[11] Next, in columns (3)-(4), we report the OLS estimates for the baseline and extended samples without the capital city. Subsequently, in columns (5)-(6), we report the estimates for the baseline and extended samples, including the capital city, obtained using the approach proposed by Gabaix and Ibragimov (2011). Finally, in columns (7)-(8), we report the OLS results for the truncated baseline and extended samples, including the capital city, in the way that allows us to obtain the estimation results that are as close as possible to the strict version of Zipf's law.

It can be noted that, when the standard OLS procedure is employed for both the baseline and extended samples, the estimated value of $\alpha$, reported in columns (1) and (2), is higher than one (1.114 and 1.186, respectively). The t-statistics tests confirm that, in both cases, the estimated values of $\alpha$ are significantly different from unity. Moreover, in both cases, the constant terms are estimated to be significantly different from the logarithm of the population of the largest city (Warsaw). Hence, the simple OLS estimates reject Zipf's law, for both the baseline and extended samples, since the Polish cities are less evenly distributed that the law predicts.

Since the largest city is much bigger than the other cities in Poland, which may contribute to the rejection of Zipf's law, we also estimated the OLS regressions on smaller samples from which the capital city of Warsaw was excluded. The estimation results for the baseline and extended samples without Warsaw are reported in columns (3) and (4), respectively. Despite slightly smaller values of $\alpha$, again, Zipf's law is rejected by the t-test. In addition, the constant is estimated to differ significantly from the log of the population of the current largest city (Kraków).

In our subsequent estimations for the baseline and extended samples, reported in columns (5) and (6) respectively, we employed the method proposed by Gabaix and Ibragimov (2011) to overcome the potential bias of the OLS estimation. It turns out that, indeed, the estimated values of $\alpha$ for both baseline and extended samples are lower (1.077 and 1.167, respectively) compared to the estimates obtained from the simple OLS method reported in columns (1) and (2). This may indicate the upward bias in the simple OLS regressions, as suggested by Gabaix and Ioannides (2004). Nevertheless, in qualitative terms, these results do not differ greatly from the previous results, since Zipf's law is still rejected by the t-test.

Finally, in columns (7) and (8), we used a certain threshold over the baseline and extended samples to see where to make the cut-off to obtain the desired outcome – Zipf's law. In the case of the baseline sample we used 280 observations, while for the extended sample we estimated the specification with 600 observations. It should be noted, however, that these thresholds were chosen arbitrarily for the fixed number of cities. In the extended sample, which includes the urban parts of rural-urban communities along with the urban communities, the previously existing ranks of the baseline urban communities change. In this case, some of the additional observations take the positions of others in the baseline sample.

For instance, in the baseline case, our threshold city of Radymno is in position 280, while in the extended sample its rank becomes 570. Not surprisingly, the inclusion of additional

[11] In order to avoid the potential problem of heteroskedasticity we performed estimations with robust standard errors.

Vol. 20 • No. 4 • 2016 • pp. 5-10 • ISSN: 2084-6118 • DOI: 10.1515/mgrsd-2016-0020

observations in between the baseline group (the group necessary for Zipf's law to hold) induces a downwards shift in the cut-off point in the case of the extended sample. The analysed relationship holds (if at all) not for a fixed number of observations, but for the group of municipalities of a size above a certain threshold. This size is similar in both cases (5,533 inhabitants in the baseline and 4,785 inhabitants in the extended sample). Therefore, the differences in threshold rankings between columns (7) and (8) are not a cause for concern.

The estimated values of $\alpha$ for both baseline and extended samples are much lower and the t-tests do not reject their equality to one. It is quite clear that the smaller the sample, the larger the average city size in the sample, and the lower the estimated value of $\alpha$. It appears that the 'adjusted' thresholds are the only case for which Zipf's law ($\alpha = 1$) holds for Poland. Therefore, by appropriately adjusting the cut-off thresholds, it is possible to show that Zipf's law works in the upper tail of city distribution. However, these thresholds were chosen arbitrarily and do not have any theoretical justification. Moreover, the estimated constants still differ significantly from the log of the largest city size.

Summing up the results of our empirical investigation, it can be stated that Zipf's law is rejected for Poland in every case, unless a special threshold is used. Thus, it is necessary to reject the hypothesis that Zipf's law is a good description of the population distribution of Polish cities. The estimated exponent is above unity, which means that the distribution of Polish city sizes is more uneven than Zipf's law predicts, while the estimated constant differs significantly from the log of the largest city in every case.

## Conclusions

In this paper we conducted an empirical study of Zipf's law, which states that the product of a city's population and the number of cities with a larger or equal population is constant for every city in a given region. According to the empirical literature, this law may not always hold, although its more general form, the rank-size rule, could be regarded as a good first approximation of city size distribution. We performed our own empirical analysis of the distribution of Polish city sizes based on population data from the Central Statistical Office for the year 2011 over the largest possible sample. We found that Zipf's law is generally rejected for Poland and city sizes are less evenly distributed than it predicts. Additional sensitivity tests have confirmed this finding.

Therefore, one might wonder whether the rank-size distribution is the best approximation of city sizes for Poland. It is obvious that their distribution is right-skewed, but this does not mean that the rank-size rule is the only distribution that can describe the cities' population concentration. A variety of other distributions could potentially serve to describe the distribution of city sizes. This is why the empirical literature includes papers that estimate different distribution rules and obtain plausible results for other countries.

For instance, Cameron (1990) showed that the Weibull distribution fitted the city size distribution better than the rank-size rule, while Alperovich and Deutsch (1995) obtained similar results for the generalized Box-Cox transformation function. More recently, the double Pareto log-normal distribution for cities proposed by Giesen et al. (2010) seems to be gaining popularity in the literature. For example, Gonzalez-Val et al. (2015) have recently demonstrated that, among various distributions, the double Pareto log-normal distribution fitted the city size distribution best. Therefore, in future empirical work on Polish cities, it would be useful to also consider distributions other than the rank-size rule.

In addition, in future work it would be useful, using the approach proposed by Bosker et al. (2008), to compare the coherence of the current distribution of cities for entire Poland to the situation in the three former partitions of Poland – Prussian, Russian and Austro-Hungarian in order to verify the claim that the distribution could still be affected by historical divisions. It would thus be possible to identify the potential traits of path-dependence in the size-rank distribution of the Polish cities.

## Appendix

*Table A1. Summary statistics for baseline and extended samples*

| Variable | Observations | Mean | Standard deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| Number of inhabitants | 306 | 61,447.35 | 133,926.30 | 1,353 | 1,708,491 |
| Rank | 306 | 153.50 | 88.48 | 1 | 306 |
| Number of inhabitants | 908 | 25,755.32 | 81,915.46 | 914 | 1,708,491 |
| Rank | 908 | 454.50 | 262.2613 | 1 | 908 |

Vol. 20 • No. 4 • 2016 • pp. 5-10 • ISSN: 2084-6118 • DOI: 10.1515/mgrsd-2016-0020

*References*


Alperovich, G & Deutsch, J 1995, 'The size distribution of urban areas: Testing for the appropriateness of the Pareto distribution using a generalized Box-Cox transformation function', *Journal of Regional Science*, vol. 35, pp. 267–276.

Auerbach, F 1913, 'Das Gesetz der Bevölkerungskonzentration', *Petermann's Geographische Mitteilungen*, vol. 59, pp. 74–76.

Bee, M, Riccaboni, M & Schiavo, S 2013, 'The size distribution of US cities: Not Pareto, even in the tail', *Economics Letters*, vol. 120, no. 2, pp. 232–237.

Berry, BJL & Okulicz-Kozaryn, A 2011, 'The city size distribution debate: Resolution for US urban regions and megalopolitan areas', *Cities*, vol. 29, Supplement 1, pp. 17–23.

Black, D & Henderson, V 2003, 'Urban evolution in the USA', *Journal of Economic Geography*, vol. 3, pp. 343–372.

Bosker, EM, Brakman, S, Garretsen, H & Schramm, M 2008, 'A century of shocks: the evolution of the German city size distribution 1925 – 1999', *Regional Science and Urban Economics, vol.* 38, pp. 330–347.

Brakman, S, Garretsen, H & van Marrewijk, C 2009, *The New Introduction to Geographical Economics,* Cambridge University Press. Cambridge.

Cameron, TA 1990. 'One-stage structural models to explain city size'. *Journal of Urban Economics*, vol. 27, pp. 294–307.

CSO 2011, *Ludność w gminach. Stan w dniu 31 marca 2011 r. - wyniki spisu ludności i mieszkań 2011 r.*, Warsaw.

Deichmann, U & Henderson, V 2000, 'Urban and regional dynamics in Poland', *World Bank Policy Research Working Paper No. 2457*.

Eeckhout, J 2004, 'Gibrat's law for (all) cities', *American Economic Review*, vol. 94, no.5, pp. 1429–1451.

Fazio, G & Modica, M 2015, 'Pareto or log-normal? Best fit and truncation in the distribution of all cities', *Journal of Regional Science*, vol. 55, no. 5, pp. 736–756.

Gabaix, X 1999, 'Zipf's law and the growth of the cities', *American Economic Review*, vol. 89, no. 2, pp. 129–132.

Gabaix, X & Ioannides, YM 2004, 'The evolution of city size distributions' in *The Handbook of Urban and Regional Economics*, vol. IV, eds. JV Henderson and J-F Thisse, North-Holland, Amsterdam, pp. 2341–2380.

Gabaix, X & Ibragimov, R 2011, 'Rank-1/2: A simple way to improve the OLS estimation of tail exponents', *Journal of Business & Economic Statistics*, vol. 29, no.1, pp. 24–39.

Giesen, K, Zimmermann, A & Suedekum, J 2010, 'The size distribution across all cities-double Pareto lognormal strikes', *Journal of Urban Economics*, vol. 68, no. 2, pp. 129–137.

Gonzalez-Val, R, Ramos, A, Sanz-Gracia, F & Vera-Cabello, M 2015, 'Size distributions for all cities: Which one is best?', *Papers in Regional Science*, vol. 94(1), pp. 177–196.

Hill, BM 1975, 'A simple general approach to inference about the tail of a distribution', *Annals of Statistics*, vol. 3(5), pp. 1163–1174.

Ioannides, Y & Skouras, S 2013, 'US city size distribution: Robustly Pareto, but only in the tail', *Journal of Urban Economics*, vol. 73, no. 1, pp. 18–29.

Lotka, AJ 1925, *Elements of Physical Biology*, Williams & Wilkins, Baltimore.

Marshall, A 1920, *Principles of Economics*, Macmillan. London.

Modica, M 2014, 'Does the EU have homogeneous urban structure area? The role of agglomeration and the impact of shocks on urban structure' in *ERSA conference papers,* no. 14, p. 229, European Regional Science Association.

Necula, C, Ibragimov, M, Valetka, U, Bobeica, G, Radu, AN, Mukhamedkhanova, K & Radyna, A 2010, 'City size distribution dynamics in transition economies. A cross-country investigation'. Available from: <https://www.cerge.cuni.cz/pdf/gdn/irp/IRP-cities-growth-rates-necula-ibragimov-valetka.pdf>. [10 February 2016].

Nitsch, V 2005, 'Zipf zipped', *Journal of Urban Economics*, vol. 57, pp. 86–100.

Rosen, KT & Resnick, M 1980, 'The size distribution of cities: an examination of the Pareto law and primacy', *Journal of Urban Economics*, vol. 8, pp. 165–86.

Soo, KT 2005, 'Zipf's law for cities: a cross-country investigation', *Regional Science and Urban Economics*, vol. 35, pp. 239–63.

Veneri, P 2016, 'City size distribution across the OECD: Does the definition of cities matter?', *Computers, Environment and Urban Systems*, vol. 59, pp. 86–94.

Zipf, GK 1949, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology,* Addison-Wesley, New York.