



# Reconsidering Zipf's law for regional development: The case of settlements and cities in Croatia

## Abstract

Zipf's law is a striking regularity in the field of urban economics that states that the sizes of cities should follow the rank-size distribution. Rank-size distribution, or the rank-size rule, is a commonly observed statistical relationship between the population size and population rank of a nation's cities. The goal of this paper is to test Zipf's law as applied to data for settlements and cities in Croatia using the Census of Population Survey for the year 2011. The results of the analysis have shown that Zipf's law for settlements in Croatia holds true for the majority of the settlement sizes. However, the rank-size distribution does not hold true for extremely small and extremely large settlement sizes. When city proper and urban agglomeration of 127 Croatian cities were examined, Zipf's law was found to hold true only for urban agglomerations. The results of the study are discussed in terms of regional development.

## Keywords

Zipf's law • urban agglomeration • settlements • the city proper • Croatia

© University of Warsaw – Faculty of Geography and Regional Studies

Hrvoje Josić<sup>1</sup>, Maja Bašić<sup>2</sup>

Faculty of Economics and Business  
University of Zagreb, Croatia

<sup>1</sup>e-mail: hjosic@efzg.hr

<sup>2</sup>e-mail: mbasic1@net.efzg.hr

Received: 20 October 2017

Accepted: 12 February 2018

## Introduction

Power law is an observed regularity that occurs when an event's value or quantity is inversely proportional to the power of that event's value (Newman 2005). Our attempts to understand natural occurrences form a rationale for the power law's existence and are not founded in any valid theory (Krugman 1996a). Examples of mathematical regularities that are attributed to power laws include Pareto's law and Zipf's law.

The aim of our study is to check whether Zipf's law holds for Croatian settlements, cities and agglomerations, and what implications Zipf's law has for the country's regional development. Our study looks at the differences between a city proper and an urban agglomeration by examining Zipf's law as applied to Croatian cities and settlements. It explains the Republic of Croatia's regional development differences by analysing cities and settlements, and through the comparison of the city proper and the urban agglomeration adherence to a rank-size rule.

There are several reasons for our study. First, the migration that occurred during and after the Croatian War of Independence caused differences in the sizes of cities and, consequently, differences in regional development (Dimou & Schaffar 2009). Second, the changing nature of cities, and their mutual competition and interaction, demand a continuous analysis that will contribute to conclusions about the country's regional development (Jiang, Yin & Liu 2015). This notion is especially relevant within the context of the European Union's regional development policies, where this study builds on Cieślík and Teresiński's (2016) study to provide an empirical examination of Zipf's law for new members of the European Union. Third, studies such as Šolak and Dobrić's (2010), and Dimou and Schaffar's (2009), which examined the regularity of

Zipf's law in the Balkans, did not include cities and settlements in their analysis, nor did they compare the city proper with urban agglomeration. There exists a large number of unanalysed cities that could exhibit power law regularities and explain differences in the regional development. Therefore, this study is a necessary prerequisite to our understanding of Croatia's regional economic development.

## Zipf's law and its application in urban economics

Zipf's law is a statistical, mathematical regularity that shows that the probability or the frequency of an observed event is inversely proportional to its rank (Zipf 1949). There are numerous events in which cumulative distribution patterns follow Zipf's law (Gan, Li & Song 2006): word frequency (Zipf 1949), the citation of scientific papers, web hits, copies of books sold, telephone calls, the magnitude of earthquakes, the diameter of moon craters, the intensity of solar flares, the intensity of wars, and the frequency of family names (Newman 2005). In economics, Zipf's law has been observed in world income distributions (Sinclair 2001), the number of employees in firms (Knudsen 2001), and the gravity model of international trade (Jošić & Nikić 2013). Nonetheless, its most prominent use is in the urban economic studies of the rank-size distributions of cities (Auerbach 1913; Gabaix 1999b; Jiang & Jia 2011).

Zipf's law states that the probability that a city has a size greater than  $S$  decreases with the size of the city, that is,  $1/S$  (Gabaix 1999b). "The size  $S(i)$  of a city of rank  $i$  will follow a power law's rank-size rule which states that the size of the city of rank  $i$  varies as  $1/i$ , and the ratio of the second largest city to the largest city should be  $1/2$ ," etc. (Gabaix 1999b, p 752–753). Zipf's

law aims to answer two essential questions: (1) Does the law hold for a particular country or a region? and (2) Why has this law emerged? (Jiang & Jia 2011).

#### Does Zipf's law hold true for a particular country or a region?

When analysing Zipf's law for the Republic of Croatia, we reconsidered three aspects of the analysis in line with Chesire (1999), and Jiang and Jia (2011): the number of cities, the size of the cities, and their conglomeration.

Zipf law regularities used in urban economics demand a proper definition of their unit of measurement. A study of the size distributions of cities in 44 countries found that Zipf's law is sensitive to how cities are defined and the choice of city sample size (Rosen & Resnick 1980). Subsequent sections explain both the definition of a city and the choice of sample size consisting of the city proper and the urban agglomeration.

#### Definition of the city

City sizes in China during the mid-nineteenth century, in India during the early twentieth century, in the early and modern United States (Gabaix 1999a), and in national and regional Germany, follow a rank-size distribution (Giesen & Suedekum 2009). In each of the studies a cities' size is defined differently based on the respective country's definition.

Different statistical and socio-cultural definitions of cities and cities' boundaries in urban economics disable effective intercountry comparisons (Soo 2005; Holmes & Lee 2009; Dimou & Schaffar 2009; Jiang & Jia 2011; Jiang, Yin & Liu 2015). City size is defined either by urban population census data, or as a *natural city*. An urban population census counts the inhabitants in a particular geographical area based on a country's political or socioeconomic definition of a city. A natural city is measured either through clustered street nodes or 6x6 mile geographical grids (Rozenfeld et al. 2011; Jiang & Jia 2011). Studies involving natural cities often draw data from an urban population census. Veneri (2016) made a distinction between cities defined by administrative boundaries, and functional urban areas that coincide with the economic definition of a city. Functional urban areas are composed of cores and commuting zones; this definition is consistent across OECD countries and enables comparisons. Although there are various definitions, this study uses a definition of a city that is homogeneously defined and based on a definition used by the Croatian urban population census, which defines a city as a place that has a local self-government or any other place consisting of at least 10,000 inhabitants. Our study's definition of a city differs between the urban agglomeration and the city proper, thereby encompassing both the economic and administrative definitions of a city.

#### City sample size

Statistical distributions used to portray Zipf's law have been debated in the literature. Histograms of the distribution of city rank-size are right-skewed, enabling power laws to hold true only for cities above a certain threshold of population size (Gabaix 1999a; Gabaix 1999b; Cieřlik & Teresiński 2016). In the majority of studies the tails of a distribution exhibit different distributional characteristics from the body of a distribution due to large and rare events that occur in the tails of distributions (Clauset, Shalizi & Newman 2009). Eeckhout (2004) argues that the entire distribution of the city size is lognormal. In contrast, Bee, Riccaboni and Schiavo (2011), and Ioannides and Skouras (2013) state that a lognormal distribution can change to a Pareto distribution in the upper tails for the majority of city size distributions. In addition, Giesen, Zimmermann and Suedekum's (2010) study of the distribution of the city-rank size of cities in eight countries, and Luckstead and Devadoss' (2017) study of all US cities, have found the existence of a double Pareto

lognormal distribution, that is, lower and upper tail Pareto and lognormal body distributions. Clauset, Shalizi and Newman (2009) use the maximum likelihood method, which has also been proposed by Akinsete, Famoye and Lee (2008).

Therefore, when examining the city size distribution in Balkan countries, Šolak and Dobrić (2010), and Dimou and Schaffar (2009) truncated the analysis to exclude smaller cities from the analysis. Šolak and Dobrić (2010) excluded cities with less than 7,000 inhabitants in their analysis, while Dimou and Schaffar (2009) found that the city-size distribution for the Balkans holds true only for cities with a population greater than 70,000 inhabitants.

Although 80% of the world's population live in metropolitan areas (Holmes & Lee 2010; Jiang, Yin & Liu 2015), the majority of cities tend to be small (Jiang & Jia 2011). Small cities are influenced by neighbouring or distant countries or regions (Jiang, Yin & Liu 2015), which is explained by a theory of a new economic geography (Krugman 1991). Hence, studies have attempted to exemplify the differences between small and large cities' rank-size distributions. Cieřlik and Teresiński's (2016) study of the rank-size distribution of Polish cities consisted of two parts: (a) examining Zipf's law for urban communities with more than 1,000 inhabitants (306 observations), and (b) adding the municipal part of the urban-rural communities into the study (908 observations) in order to account for the differences between smaller and larger urban areas. In line with Cieřlik and Teresiński's (2016) study we decided to analyse all cities and settlements in the Republic of Croatia (6,606 observations), subsequently truncating the analysis and omitting smaller sized settlements in order to test and illustrate the validity of Zipf's law in Croatia. The rationale is that by not including all city sizes, we would neglect a significant part of the country's regions that make up for regional differences. Moreover, our study gives an insight into regional differences by including cities and settlements of all sizes and comparing the city proper with the urban agglomeration of all cities and settlements in the Republic of Croatia.

#### How is the city rank-size rule connected to regional development?

The rank-size rule has its dynamic application in regional development. Gibrat's law is a dynamic form of the static Zipf's law (Modica, Reggiani & Nijkamp 2015). Gibrat's law analyses a city's rank-size growth rate rule, while Zipf's law analyses a city's rank-size distribution. An examination of the dynamic form of Zipf's law illustrated that city sizes converge according to Zipf's law (Gabaix 1999a; Gabaix 1999b; Ioannides & Overman 2000).

Zipf's law holds true because the majority of economic shocks decline with size (Gabaix 1999b, p. 760). The inertia to create jobs causes the number of new jobs in the cities to be proportional to the number of existing jobs, thereby allowing wages to adjust to variations in amenities and causing externalities in line with productivity increases, thereby influencing city growth (Gabaix 1999b). A study of Malaysian cities found that smaller cities and state capitals grow faster than other cities (Soo 2007). Additionally, young cities may grow faster than older cities, which might affect regional development (Giesen & Suedekum 2013).

Cities grow from the bottom up and their territories result from competition for space (Batty 2008). As a result of social forces relating to industrial capitalism (Harvey 1973), city size distribution is meaningful as it shows different production patterns applicable to industry specialisation (Tabuchi & Thisse 2006); namely, city location and size portray the employment in a particular area, which enables greater technological progress and intra-industry specialisation (Henderson 1986). Greater specialisation induces wage differentials (Muth 1961; Roback 1982). On the other hand, urban costs are higher in larger cities and affect consumers' incomes and their quality of life (Tabuchi & Thisse 2006), which can lead to social segregation and interaction problems (Meen

& Meen 2003) based on the majority of goods being produced in a particular area (Tabuchi & Thisse 2006). Herein, the role of institutional policy is important as it creates the possibility for a top down urbanisation processes (Batty 2008).

Specifically, higher trade openness, the mobility of people and capital, and the introduction of a single currency might influence the changes in the European Union's urban structures, resulting in unexplored patterns of urbanisation that would need to be anticipated from a transport policy, land planning, and urban design perspective (Modica 2017). Therefore, although not always (Soo 2005), uneven regional development can be explained by economic geography, and Zipf's law might be used to account for its differences and to anticipate changes.

#### Why is Zipf's law important for Croatian regional development?

Significant and persistent economic regional disparities exist between Croatian regions (Borožan 2017). As the pace of the convergence between the economic growth of Croatia's regions and the European Union's average is stronger on a national than on a regional level (Mikulić, Lovrinčević & Nagyszombaty 2013), this study tests Zipf's law on settlements and cities of all sizes and differentiates between the city proper and urban agglomeration in order to make recommendations for a more even Croatian regional development.

A deviation from the linearity of Zipf's law can be attributed to a growing urbanisation process in the biggest cities (Rosen & Resnick 1980). As deviations from Zipf's law minimise in subsequent stages of urbanisation, the institutional and demographic changes, and political crises do not influence the rank-size distribution of cities (Rosen & Resnick 1980; Dimou & Schaffar 2009). On the other hand, Modica (2017) 79 has shown that the cities in the European Union Member States are different due to historical, geographical and economic research, and that the introduction of the euro positively impacted the city size distribution leading to a greater dispersion of cities, while the entry of European Union Member States into the Schengen area resulted in a higher concentration of population in larger cities. Our analysis classifies Croatian cities according to the Modica (2017) hierarchical or polycentric structure, and based on a concentration and dispersion processes identified by the new economic geography (Krugman 1991; Modica 2017).

Agglomeration effects within regional economic development of cities exist in central and south-eastern European countries (Romisch 2015). In Croatia, emigration from smaller to larger cities, and abroad during and after the independence war (Dimou & Schaffar 2009), and again after the economic crisis in 2008 might have affected its regional development. Poor industry structure and the quality of human capital affect emigration and poor regional development (Mikulić, Lovrinčević & Nagyszombaty 2013). The rank-size distribution of settlements and cities, and the analysis of the differences between the city proper and the urban agglomeration aim to provide an insight into these regional development differences.

#### Methodology and data

F. Auerbach (1913) was the first to suggest the formula for the size distribution of cities based on the Pareto distribution:

$$R = CS^{-\alpha} \quad (1)$$

where  $R$  is the rank of the city (cities are ranked from largest to smallest, rank=1 is the city with the highest population),  $C$  is a constant,  $S$  is the population of the city, and  $\alpha$  is the Pareto exponent or the Zipf value. Another method of writing this equation is to use natural logarithms due to the right-skewed city size distribution, which provides a better fit for samples encompassing small cities (Gabaix 1999b; Jiang, Yin & Liu 2015):

$$\ln(R) = \ln(C) - \alpha \ln(S) \quad (2)$$

According to Zipf (1949), the Pareto exponent,  $\alpha$ , is employed as a measure of population concentration among cities of different sizes. The rank-size rule is validated when  $\alpha=1$  because the values are then centred around an average value (Jiang & Jia 2011). Zipf's law is a specific example of rank size distribution. Smaller values of the Pareto coefficient imply that the urban system is highly concentrated. On the other hand, higher values of the Pareto exponent imply more equality between cities and less hierarchy. The city rank-size relationship appears linear in the logarithms, yielding a very high  $R^2$  (close to 1) (Gan, Li & Song 2006). Gabaix and Ibragimov (2011) showed that the previous standard approach for estimating the Pareto exponent provided biased estimations for small samples. The authors derived a simple but effective solution to correct this bias by using the variable  $R - 1/2$  instead of  $R$  (Deliktas, Önder & Karadag 2013). This correction is optimal since it minimizes the small-sample bias in the OLS estimator. The corrected formula is presented in equation (3):

$$\ln(R - 1/2) = \ln(C) - \alpha' \ln(S) \quad (3)$$

In order to test Zipf's law on settlements and cities in Croatia, two hypothesis were tested:

*H1: The acceptance of Zipf's law depends on the settlement size in Croatia.*

*H2: Zipf's law for cities in Croatia is valid for the city proper and the urban agglomeration.*

For the purpose of testing the aforementioned hypothesis, the sizes of settlements and cities were measured based on the administrative definition of the city, which encompassed urban, historical, natural, economic, and social definitions. Data were taken and aggregated from the Croatian Bureau of Statistics Census of Population, Households and Dwellings 2011. Croatian Law on local and regional self-government, Article 5, states that "The town is a unit of local self-government where the seat of the county is located, as well as any other place with more than 10,000 inhabitants." The population of the city proper for Croatian cities is defined as cities containing a local self-government unit, while the urban agglomeration population adds the population of the surrounding settlements to the population of the city proper. Both the city proper and the urban agglomerations were ranked according to their size, that is, by population (Jiang and Jia 2011).

The hypotheses were tested using methods of inferential statistics. Zipf's law was tested using a simple regression model based on Equation 2. The sample size comprised of 6,756 settlements with 127 towns and the city of Zagreb. There was no issue with biased estimation. Rank is a sequential variable, and the size of the city in terms of population is a quantitative variable. The dependent variable in the analysis is the variable RANK, which takes the form of a natural logarithm. Regressions were made with respect to the different sizes of the independent variable SIZE and took the form of a natural logarithm.

#### Results and discussion

Statistics that are available after 1991 show that the number of inhabitants in Croatia has been in constant decline (Croatian Bureau of Statistics: Census of Population, Households and Dwellings 2011). This decline has been caused by the Croatian War of Independence, the low rate of both natural and economic growth, emigration, an aging population, etc. (Croatian Bureau of Statistics: Census of Population, Households and Dwellings 2011).

According to the 2011 census there were 6,756 settlements with 127 cities, as well as the capital city of Zagreb, which holds special urban status in the Republic of Croatia. Table 1 presents

Table 1. Croatian settlement structure and population

Settlement size	Number of settlements	Population	Share of	
			Settlement	Population
No population	150	0	2.22%	0%
Up to 100 people	2,653	113,914	39.27%	2.66%
101–200	1,318	192,193	19.51%	4.49%
201–500	1,448	461,114	21.43%	10.76%
501–1 000	658	462,788	9.74%	10.80%
1,001–1,500	195	240,133	2.89%	5.60%
1,501–2,000	113	194,258	1.67%	4.53%
2,001–5,000	143	434,201	2.12%	10.13%
5,001–10,000	39	264,060	0.58%	6.16%
10,001–20,000	20	274,938	0.30%	6.42%
20,001–50,000	11	342,971	0.16%	8.00%
50,001–100,000	5	320,651	0.07%	7.48%
100,001–200,000	2	295,505	0.03%	6.90%
200,001 or more	1	688,163	0.01%	16.06%

Source: Croatian Population Survey, 2011.

the structure and population size of Croatia's settlements. Settlements are divided up according to population size. There is a total of 6,756 settlements in Croatia. Of these 150 settlements have no inhabitants; hence, there are 6,606 settlements with at least one inhabitant in Croatia. The hypotheses were tested on these 6,606 settlements with at least one inhabitant. The majority of the settlements were populated by up to 1,000 inhabitants (92.17%); however, the proportion of the population in the settlements with up to 1,000 inhabitants out of the total population was only 28.71%. In comparison, the capital city of Zagreb comprises 16.06% of the total Croatian population. Additionally, population distribution of Croatia's largest cities does not follow the usual Zipf's law, which states that the second largest city in a country will be approximately 1/2 the population of the first city, which is the rationale for testing Zipf's law on the sample of Croatian settlements.

Figure 1 is a graphical representation of Zipf's law for settlements in the Republic of Croatia based on a sample of 6,606 settlements with at least one inhabitant. Settlements without populations were not included in the analysis.

Figure 1 shows the concave shape of the curve, especially with respect to the smaller sized settlement. Table 2 depicts the results of OLS regression testing on Zipf's law for settlements in Croatia.

Table 2 depicts R squared ( $R^2=0.7887$ ), which indicates that a good proportion of the variance in the dependent variable is explained by the independent variable. The Pareto coefficient is -0.5561 and is statistically significant at a 1 percent significance level. The value of the Pareto coefficient indicates a dispersion of population size relative to the settlements' size. Therefore, Zipf's law does not hold true for all settlements in the Republic of Croatia.

As the analysis showed that Zipf's law did not hold true for settlements of all sizes in the Republic of Croatia, we decided to test this by consecutively omitting groups of settlements with small population sizes from the analysis. Figure 2 shows the results of the regression analyses in which settlements with less than 100, 500, 1000, and 2000 inhabitants were omitted consecutively.

Figure 2 displays Zipf's law for settlements in the Republic of Croatia that contain populations larger than (a) 100, (b) 500,

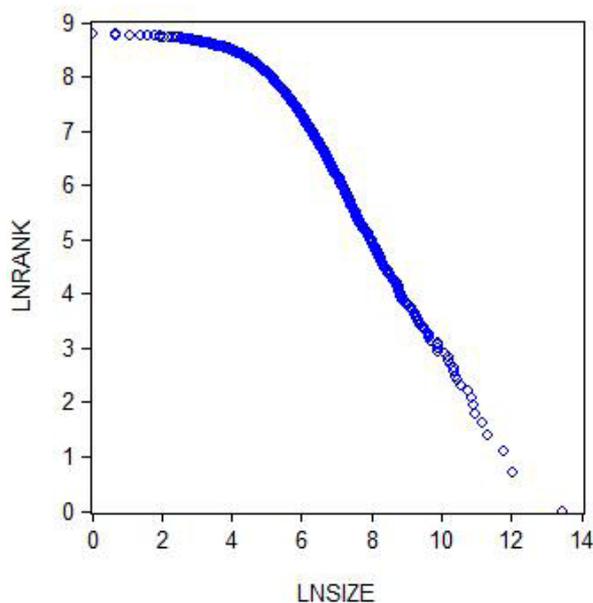


Figure 1. Zipf's law for settlements in Croatia  
Source: Authors' calculations

(c) 1000, and (d) 2000 inhabitants based on samples of (a) 3,953, (b) 1,187, (c) 529, and (d) 221 observations, respectively. If the settlement's size was greater than 100 inhabitants, Zipf's law holds perfectly. The Pareto coefficient is -0.9941 and is statistically significant at a level of 1 percent, indicating a perfect match with Zipf's law, or rank-size distribution. R squared is 0.9843, which indicates that the independent variable model explains the model well.

In cases in which settlements have populations larger than (b) 500, (c) 1000, or (d) 2000 inhabitants ( $\ln(\text{Size} > 500)$ ,  $\ln(\text{Size} > 1000)$ ,  $\ln(\text{Size} > 2000)$ ), Zipf's law also holds true. The

Table 2. OLS for testing Zipf's law as applied to settlements in Croatia

Dependent variable Ln(Rank)	Independent variable				
	Ln (Size>=1)	Ln (Size >100)	Ln (Size >500)	Ln (Size >1000)	Ln (Size >2000)
Method: Least Squares					
Constant	10.5069 (0.0181)	13.1298 (0.0118)	14.1597 (0.0148)	13.8333 (0.0299)	13.1519 (0.0411)
Pareto coefficient	-0.5561*** (0.0035)	-0.9941*** (0.0019)	-1.1401*** (0.0021)	-1.1013*** (0.0038)	-1.0263*** (0.0048)
R-squared	0.7887	0.9843	0.9961	0.9936	0.9952
S.E. of regression	0.4567	0.1243	0.0619	0.0778	0.0664
Prob. (F-statistic)	0.0000	0.0000	0.0000	0.0000	0.0000
Mean dependent variable	7.7931	7.2833	6.0829	5.2786	4.4145
S.D. dependent variable	0.9935	0.9947	0.9872	0.976896	0.9574
Akaike info criterion	1.2707	-1.3315	-2.7252	-2.2640	-2.5774
Durbin Watson stat	0.0003	0.0109	0.2293	0.2906	0.7605
Observations	6606	3953	1187	529	221

Note: OLS estimates show White heteroskedasticity – consistent standard errors and covariances; t-statistics in parentheses; significant at a level of 1 percent: \*\*\*, at a level of 5 percent: \*\*, at a level of 10 percent: \*.

Source: Authors' calculations

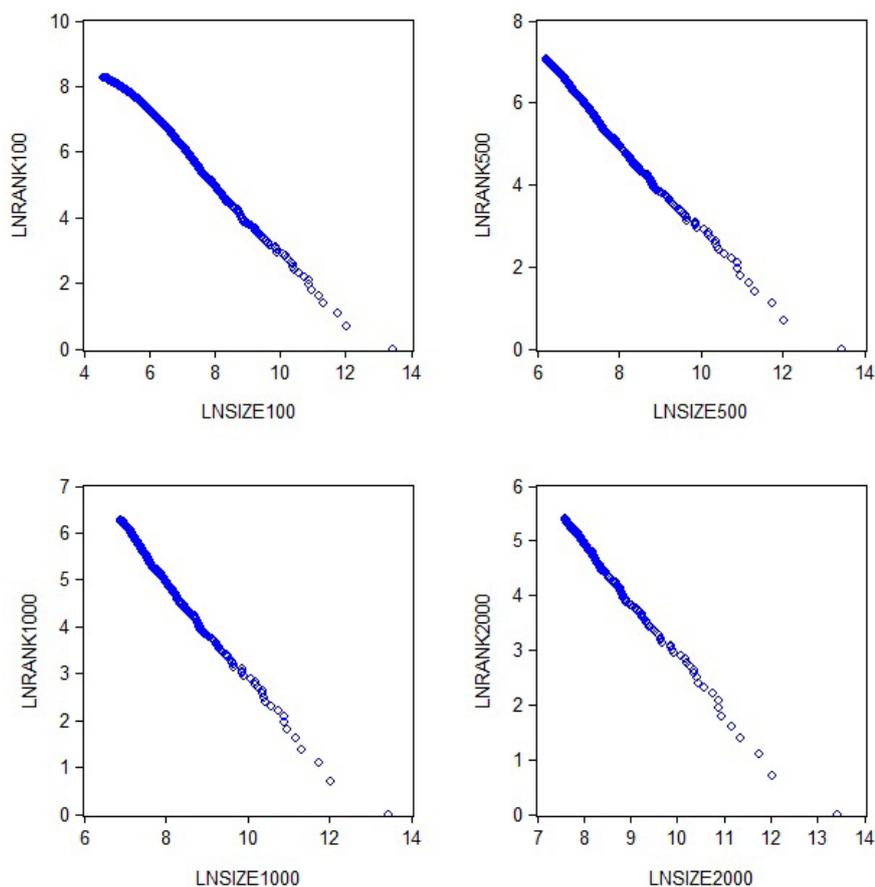


Figure 2. Zipf's law for settlements in Croatia with populations over 100, 500, 1000, and 2000 inhabitants

Source: Authors' calculations

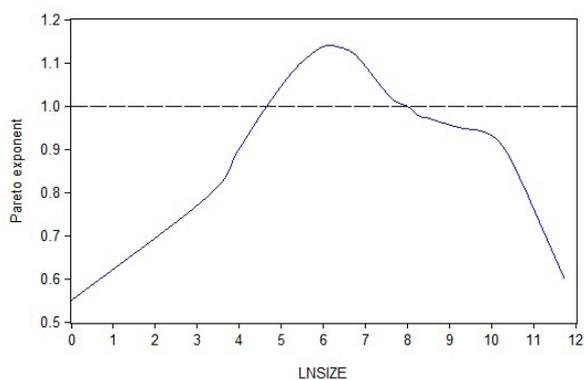


Figure 3. Pareto coefficient and settlement size in Croatia  
Source: Authors' calculations

values of the Pareto coefficients are (b) -1.1401, (c) -1.1013 and (d) -1.0263 respectively. The independent variable is significant for all cases at a significance level of 1 percent, while the value of R squared is very high (above 0.99). The values of the Pareto coefficient are slightly higher than 1, which indicates that the concentration of settlements is high and uniformly distributed.

Hypothesis 1 was tested by calculating the Pareto coefficient in relation to different sizes of settlement based on 20 truncation points (Figure 3). Figure 3 shows that the values of the Pareto coefficient equal 1 (absolute term) for settlements whose size is greater than 100 inhabitants and 2000 inhabitants. For the area in between these two sizes the value of the Pareto coefficient is higher than 1 with a maximum value for settlement size above 500 inhabitants. Therefore, Hypothesis 1 has been confirmed.

Regarding settlements in Croatia, Zipf's law holds true for the majority of settlement sizes in Croatia. Rank-size distribution does not hold true for the lower and upper-tail of the settlement size distribution.

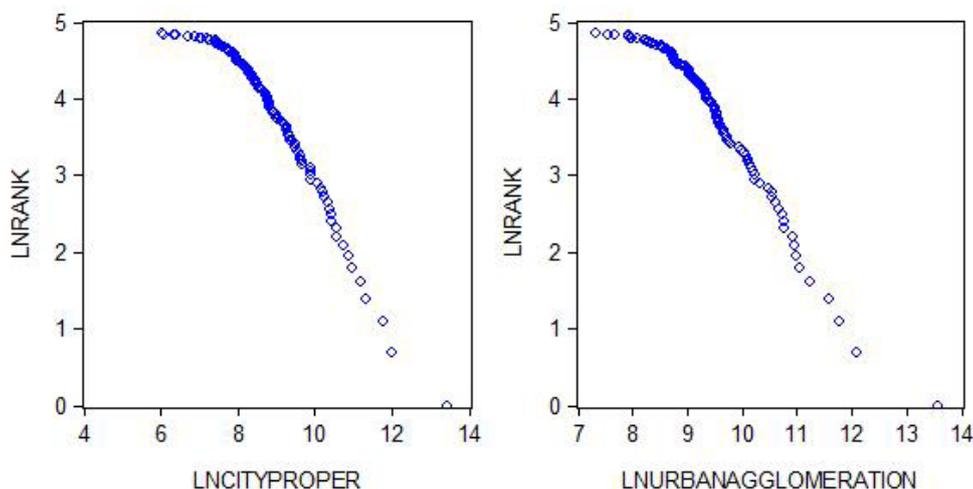


Figure 4. Zipf's law for urban agglomerations and the city proper in Croatia  
Source: Authors' calculations

Table 3. OLS for testing Zipf's law as applied to the city proper and urban agglomeration in Croatia

Dependent variable Ln (Rank)	Independent variable	
	Ln (City proper)	Ln (Urban agglomeration)
<b>Method: Least Squares</b>		
Constant	10.3162 (0.1572)	12.7571 (0.1876)
Pareto coefficient	-0.7397*** (0.0178)	-0.9501*** (0.0199)
Adjusted R-squared	0.9314	0.9473
S.E. of regression	0.2456	0.2153
Prob. (F-statistic)	0.0000	0.0000
Mean dependent variable	3.8705	3.8705
S.D. dependent variable	0.9383	0.9383
Akaike info criterion	0.0460	-0.2174
Durbin Watson stat	0.0525	0.1347
Observations	127	127

Note: OLS estimates show White heteroskedasticity – consistent standard errors and covariances; t-statistics in parentheses; significant at a level of 1 percent: \*\*\*, at a level of 5 percent: \*\*, at a level of 10 percent: \*.

Source: Authors' calculations

The second goal of the analysis was to test Zipf's law for the city proper and urban agglomerations in Croatia – presented in Table 3 and Figure 4.

OLS regression for the definition of the city proper in the Republic of Croatia showed that Zipf's law cannot be confirmed. The Pareto coefficient value is -0.7397, which depicts a highly concentrated urban system or divergent city sizes. Out of the 127 Croatian cities and urban agglomerations in Croatia there are six cities (city proper populations) with less than 1000 inhabitants (Obrovac, Vrljka, Skradin, Klanjec, Rab and Čabar); this points to the previously described exceptions in the definition of city size that affected the hypothesis' confirmation. On the other hand, Zipf's law for urban agglomeration population in Croatia is valid. The value of the Pareto coefficient was -0.9501 and is statistically significant. It can be concluded that Zipf's law is valid for urban agglomerations in Croatia but cannot be confirmed for the city proper sample.

### Conclusion

The aim of this paper was to test whether the Zipf's law holds true for Croatian settlements, cities and agglomerations, and what implications this has for the country's regional development. It tested Zipf's law on data for settlements and cities in Croatia using the 2011 Census of Population Survey. The results of the analysis have shown that Zipf's law when applied to settlements in Croatia holds true for the majority of settlement sizes. Although universal, it is limited to the scope of observation (Jiang, Yin & Liu 2015).

The rank-size distribution does not hold true for extremely small or extremely large sizes of settlement, which is in accordance with previous studies (Gabaix 1999b; Newman 2005). Although Eeckhout (2004) argues that the entire distribution of city sizes is lognormal, studies by authors that affirm that the distribution of city sizes have a Pareto tail and a lognormal body (Bee, Riccaboni & Schiavo 2011, 2013; Clauset, Shalizi & Newman 2009; Ioannides and Skouras, 2013; Luckstead and Devadoss, 2017; Malavergne, Pisarenko & Sornette 2011) are consistent with our results. Our study follows the study of Fazio and Modica (2015), which examined alternative city size distributions depending on differing truncation points and proved the validity of Eeckhout's (2009) study, stating that the choice of Pareto or lognormal distribution depends on the truncation point, wherein the upper tail is longer than assumed.

Moreover, our study is consistent with Veneri's (2016) findings in which the actual size of the city is more accurately identified through its economic functions, which in our study coincide with the definition of urban agglomeration. Urban agglomeration comprises the population of the city proper and includes economic functional areas that are comprised of the commuting flows of urban populations. Our analysis shows that the Croatian system of cities, especially cities based on economic functions, comprise a hierarchical structure. This finding is consistent with Modica's (2017) observation of the hierarchical structure of systems of cities in the European Union's Member States who are experiencing a transition from socialism to capitalism.

Krugman (1996b) found that the size and growth of the few first ranked cities can be slowed down through the opening up of trade and reduced government intervention. However, international trade is not to be assumed to be the key determinant of the regional convergence process in Croatia due to work by Modica (2017) who found the differences in the concentration and dispersion processes of cities in the European Union Member States; namely, the effect of the Schengen treaty induced a concentration of population in larger cities, while the introduction of a single currency has enabled the dispersion processes.

Regional growth in Croatia is primarily determined using the quality of human capital, investments in fixed assets, and structural features of individual Croatian counties (Mikulić & Nagyszombaty 2015). Hence, recommendations concerning regional convergence and regional development include policies that concern education (Deliktas, Önder & Karadag 2013), as demographic changes have not been found to influence rank-size distribution in the long run (Rosen & Resnick 1980; Dimou & Schaffar 2009). The stress should be put on smaller cities that have a tendency to grow faster; therefore, there is room for policy implementation in both very large and very small cities.

Cieślak and Teresiński's (2016) study of rank-size distribution of Polish cities used a smaller sample (306 and 908 observations) than our study and did not find evidence of a rank-size distribution for samples of different sizes. Our study uses a larger sample (6,606 settlements), that is, all cities and settlements in the Republic of Croatia, and tests illustrate the validity of Zipf's law based on samples of settlements and cities of different sizes (6,606; 953; 1,187; 529; and 221 observations). Moreover, by examining the differences between the city proper and urban agglomerations, our study found that when the city proper and the urban agglomerations for 127 Croatian cities were examined, the Zipf's law holds true only in cases of urban agglomerations. Policywise, this implies that populations have started to move out of cities towards secondary urban centres (Dimou & Schaffar 2009). Possible consequences for regional development based on our study include a lack of institutional and demographic policy effects in accordance with the Yule distribution (Rosen & Resnick 1980; Dimou & Schaffar 2009; Newman 2005). The rich-get-richer mechanism could lead to limitations in urban agglomeration development and policies should target secondary urban centres to achieve a more even regional development. As mentioned above, policies aimed at raising regional productivities of secondary urban centres through education, structural features of each county, and investment in fixed assets, should be the primary focus for a more even urban development.

### Limitations and future studies

Our study has several drawbacks to do with individual country analysis and limited census data. Using census data for one period of time is limited to a static view of city-size distribution. The differences in policy recommendations arising from measures by conventional statistical data or big data (Jiang, Yin & Liu 2015) will require new approaches to solving regional development problems and enable us to gain a better insights into urban settlements, sizes, and regional development processes.

Arshad, Hu and Ashraf's (2017) review of the literature concerning Zipf's law applicability to city size distribution has shown that even for the upper tails of the distribution, Zipf's law is not universal. Hence, additional power laws should be considered in future papers, especially ones concerning the cities in the lower tail of distribution as they comprise the majority of cities.

Finally, it will be interesting to observe whether city size distribution has changed over time depending on the inclusion of Croatia in the Schengen area and the introduction of a single currency. Applicable structural and cohesion funds from the European Union, as well as its Smart Specialisation policy should enable a more even development that would mitigate the possible contingencies connected with the uneven concentration and agglomeration of cities.

## References

- Akinsete, A, Famoye, F & Lee, C 2008, 'The beta-Pareto distribution', *Statistics*, vol. 42, no. 6, pp. 547-563.
- Arshad, S, Hu, S & Ashraf, BN 2017, 'Zipf's law and city size distribution: A survey of the literature and future research agenda', *Physica A: Statistical Mechanics and its Applications*, vol. 492, pp. 75-92.
- Auerbach, F 1913, 'Das gesetz der bevölkerungskonzentration', *Petermanns Geographische Mitteilungen*, vol. 59, pp. 74-76.
- Batty, M 2008, 'The size, scale, and shape of cities', *Science*, vol. 319, no. 5684, pp. 769-771.
- Bee, M, Riccaboni M & Schiavo, S 2011, 'Pareto versus lognormal: a Maximum entropy test', *Physical Review E*, vol. 84, no. 2, DOI: 10.1103/PhysRevE.84.026104.
- Bee, M, Riccaboni M & Schiavo, S 2013, 'The size distribution of U.S. cities: Not Pareto, even in the tail', *Economics Letters*, vol. 120, pp. 233-237.
- Borozan, D 2017, 'Internal migration, regional economic convergence, and growth in Croatia', *International Regional Science Review*, vol. 40, no. 2, pp. 141-163.
- Cheshire, P 1999, 'Trends in sizes and structure of urban areas', in *Handbook of Regional and Urban Economics*, vol. 3, eds P Cheshire & ES Mills, Elsevier, Amsterdam, pp. 1339-1373.
- Cieślak, A & Teresiński, J 2016, 'Does Zipf's law hold for Polish cities?', *Miscellanea Geographica – Regional Studies on Development*, vol. 20, no. 4, pp. 5-10.
- Clauset, A, Shalizi, CR & Newman, MEJ 2009, 'Power-law distributions in empirical data', *SIAM Review*, vol. 51, no. 4, pp. 661-703.
- Croatian Bureau of Statistics: *Census of population, households and dwellings 2011*. Available from: <[http://www.dzs.hr/Hrv\\_Eng/publication/2011/SI-1441.pdf](http://www.dzs.hr/Hrv_Eng/publication/2011/SI-1441.pdf)>. [10 November 2016]
- Croatian Law on Local and Regional self-government 2001. Available from: <<http://www.parliament.am/library/Tim/XORVATIA.pdf>>. [10 November 2016]
- Deliktas, E, Önder, AÖ & Karadag, M 2013, 'The size distribution of cities and determinants of city growth in Turkey', *European Planning Studies*, vol. 21, no. 2, pp. 251-263.
- Dimou, M & Schaffar, A 2009, 'Urban hierarchies and city growth in the Balkans', *Urban Studies*, vol. 46, no. 1, pp. 2891-2906.
- Eeckhout, J 2004, 'Gibrat's law for (all) cities', *The American Economic Review*, vol. 94, no. 5, pp. 1429-1451.
- Eeckhout, J 2009, 'Gibrat's law for (all) cities: Reply', *The American Economic Review*, vol. 99, no. 4, pp. 1676-1683.
- Fazio, G & Modica, M 2015, 'Pareto or lognormal? Best fit and truncation in the distribution of all cities', *Journal of Regional Science*, vol. 55, no. 5, pp. 736-756.
- 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.
- Gabaix, X 1999a, 'Zipf's law and the growth of cities', *American Economic Review AEA papers and proceedings*, pp. 129-132.
- Gabaix, X 1999b, 'Zipf's law for cities: An explanation', *Quarterly Journal of Economics*, vol. 114, no. 3, pp. 739-767.
- Gan, L, Li, D & Song, S 2006, 'Is the Zipf law spurious in explaining city-size distributions?', *Economic Letters*, vol. 92, pp. 256-262.
- Giesen, K & Suedekum, J 2009, 'Zipf's law for cities in the regions and the country', Forschungs Institute zur Zukunft der Arbeit-Institute for the Study of Labor, IZA DP No. 3928, Bonn, Germany.
- Giesen, K & Suedekum, J 2013, 'City age and city size', *DICE Discussion paper No. 120*. Düsseldorf University Press, ISSN 2190-9938, Düsseldorf, Germany.
- 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.
- Henderson, JV 1986, 'Efficiency of resource usage and city size', *Journal of Urban Economics*, vol. 19, pp. 47-70.
- Holmes, TJ & Lee, S 2009, 'Cities as six-by-six-mile squares: Zipf's law?' in *The economics of agglomeration*, ed EL Glaeser, University of Chicago Press, Chicago, pp. 105-131.
- Ioannides, YM & Overman HG 2000, 'Zipf's law for cities: An empirical examination', *Regional Science and Urban Economics*, vol. 33, no. 2, pp. 127-137.
- Ioannides, YM & Skouras, S 2013 'U.S. city size distribution: Robustly Pareto but only in the tail', *Journal of Urban Economics*, vol. 73, no. 1, pp. 18-29.
- Jiang, B & Jia, T 2011, 'Zipf's law for all the natural cities in the United States: A geospatial perspective', *International Journal of Geographical Information Science*, vol. 25, no. 8, pp. 1269-1281.
- Jiang, B, Yin, J & Liu, Q 2015, 'Zipf's law for all the natural cities around the world', *International Journal of Geographical Information Science*, vol. 29, no. 3, pp. 498-522.
- Jošić, M & Nikić, M 2013, 'Gravity model and Zipf's law: An in-depth study into the nature of international trade', *Academic Journal of Interdisciplinary Studies MCSER Publishing-Rome, Italy*, vol. 2, no. 9, pp. 583-588.
- Knudsen, T 2001, 'Zipf's law for cities and beyond: The case of Denmark', *American Journal of Economics and Sociology*, vol. 60, no. 1, pp. 123-146.
- Krugman, P 1991, 'Increasing returns and economic geography', *Journal of Political Economy*, vol. 99, no. 3, pp. 483-499.
- Krugman, P 1996a, 'Confronting the mystery of urban hierarchy', *Journal of the Japanese and International Economies*, vol. 10, no. 4, pp. 399-418.
- Krugman, P 1996b, 'Urban Concentration: The role of increasing returns and transport costs', *International Regional Science Review*, vol. 19, no. 1/2, pp. 5-30.
- Luckstead, J & Devadoss, S 2017, 'Pareto tails and lognormal body of US cities size distribution', *Physica A: Statistical Mechanics and its Applications*, vol. 465, pp. 573-578.
- Malevergne, Y, Pisarenko V & Sornette, D 2011, 'Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities', *Physical Review E*, vol. 83, no. 3, DOI: org/10.1103/PhysRevE.83.036111.
- Meen, D & Meen G 2003, 'Social behaviour as a basis for modelling the urban housing market: A review', *Urban Studies*, vol. 40, no. 5-6, pp. 917-935.
- Mikulić, D & Nagyszombat, AG 2015, 'Does international trade cause regional growth differentials in Croatia?', *Zbornik Radova Ekonomskog Fakulteta u Rijeci*, vol. 33, no. 1, pp. 81-102.
- Mikulić, D, Lovrinčević, AZ & Nagyszombat, AG 2013, 'Regional convergence in the European Union, new member states and Croatia', *South East European Journal of Economics and Business*, vol. 8, no. 1, pp. 9-21.
- Modica, M 2017, 'The impact of the European Union integration on the city size distribution of the Member States', *Habitat International*, vol. 70, pp. 103-113.
- Modica, M, Reggiani, A & Nijkamp, P 2015, 'A comparative analysis of Gibrat's and Zipf's law on urban population', *Quaderni - Working Paper DSE N°1008*, Università di Bologna, Bologna, Italy.
- Muth, RF 1961, 'The spatial structure of the housing market', *Regional Science*, vol. 7, no. 1, pp. 207-220.

- Newman, MEJ 2005, 'Powel laws, Pareto distributions and Zipf's law', *Contemporary Physics*, vol. 45, no. 5, pp. 323-351.
- Roback, J 1982, 'Wages, rents, and the quality of life', *Journal of Political Economy*, vol. 90, no. 6, pp. 1257-1278.
- Romisch, R 2015, 'Estimating agglomeration in the EU and the Western Balkan regions', *The wiiw Balkan Observatory. Global Development Network Southeast Europe. Working paper 117*, The Vienna Institute for International Economic Studies, Vienna, Austria.
- Rosen, KT & Resnick, M 1980, 'The size distribution of cities: An examination of the Pareto law and primacy', *Journal of Urban Economics*, vol. 8, no. 2, pp. 165-186.
- Rozenfeld, HD, Rybski, D, Gabaix, X & Makse, HA 2011, 'The area and population of cities: New insights from a different perspective on cities', *American Economic Review*, vol. 101, no. 5, pp. 2205-2225.
- Sinclair, R 2001, *Examining the growth model's implications: The world income distribution*, Working Paper. Department of Economics, Syracuse University, Syracuse, USA.
- Šolak, Z & Dobrić, N 2010, 'Zipf's law and cities', *Zbornik Matice srpske za društvene nauke*, vol. 131, pp. 149-154.
- Soo, KT 2005, 'Zipf's law for cities: A cross-country investigation', *Regional Science and Urban Economics*, vol. 35, pp. 239-263.
- Soo, KT 2007, 'Zipf's law and urban growth in Malaysia', *Urban Studies*, vol. 44, no. 1, pp. 1-14.
- Tabuchi, T & Thisse, JF 2006, 'Regional specialization, urban hierarchy and commuting costs', *International Economic Review*, vol. 47, no. 4, pp. 1295-1317.
- 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 principles of least effort*, Addison Wesley, Cambridge, MA.