

Hrvoje Jošić*
Berislav Žmuk**

JEL Classification: C23, R12.
Original scientific paper
<https://doi.org/10.32910/ep.71.4.1>

INVESTIGATION OF URBAN REGULARITIES FOR CROATIA IN THE PERIOD FROM 1857 TO 2011

Two main regularities in the field of urban economics are Zipf's law and Gibrat's law. Zipf's law states that distribution of largest cities should obey the Pareto rank-size distribution while Gibrat's law states that proportionate growth of cities is independent of its size. These two laws are interconnected and therefore are often considered together. The objective of this paper is the investigation of urban regularities for Croatia in the period from 1857 to 2011. In order to estimate and evaluate the structure of Croatian urban hierarchy, Pareto or Zipf's coefficients are calculated. The results have shown that the coefficient values for the largest settlements in different years are close to one, indicating that the Croatian urban hierarchy system follows the rank-size distribution and therefore obeys Zipf's law. The independence of city growth regarding the city size is tested using penal unit roots. Results for Gibrat's law testing using panel unit root tests have shown that there is a presence of unit root in growth of settlements therefore leading to the acceptance of Gibrat's law.

Keywords: Croatia, Gibrat's law, settlements, Zipf's law

* H. Jošić, PhD, Assistant Professor, University of Zagreb, Faculty of Economics and Business Zagreb (e-mail: hjosic@efzg.hr).

** B. Žmuk, PhD, Assistant Professor, University of Zagreb, Faculty of Economics and Business Zagreb (e-mail: bzmuk@efzg.hr). The paper was received on 20.12.2018. It was accepted for publication on 14.05.2019.

1. Introduction

Zipf's and Gibrat's laws are one of the most investigated phenomena in the field of urban economics that seems to hold. Zipf's law defines that the rank of city sizes is proportionate to its population sizes (Zipf, 1949). Gibrat's law affirms that city population growth is independent of its size, Gibrat (1931). Zipf's law is formalised in a static form while its dynamic process is associated with Gibrat's law. Zipf's law can be explained as a steady state distribution arising from Gibrat's law. Rank-size distribution arises naturally if the growth process is independent of its size. Therefore, both laws share the common source. The two laws are often considered together because of their complementarity.

The objective of the paper as well as the context of the investigation is the investigation of urban regularities in Croatia for the period from 1857 to 2011. Two research hypotheses will be tested. The first research hypothesis is whether Croatian urban population system follows Zipf's law or rank-size distribution in the historical period from 1857 to 2011. The method for validating the hypothesis will be the calculation of the Zipf's exponent using ordinary least squares (OLS) regression. The second hypothesis that will be tested is whether the population growth rate of Croatian urban system is proportional regardless of the size of settlements. This hypothesis will be proven using panel unit root approach.

The paper consists of six chapters. After the brief introduction the second chapter elaborates on regularities in urban economics described by Zipf's and Gibrat's laws. Literature review on both laws presented in section three while methodological issues and data are described in section four. The main results of the analysis and discussion are presented in chapter five. In the last chapter concluding remarks are given.

2. Regularities in urban economics

Zipf's law is striking regularity in urban economics which states that the distribution of city sizes follows the rank-size rule. The largest city gets the value of one, the second largest city gets the value two and so on. The term Zipf's law was coined after the work of George Kingsley Zipf (Zipf, 1949) although seminal work in this field was made by Felix Auerbach (Auerbach, 1913). In equation 1 the rank-size rule is presented:

$$R_i = \alpha S_i^{-\beta}, \quad (1)$$

where R_i is the rank of city i , α is a constant, S is city size and β is a Pareto coefficient. Zipf's law holds when the coefficient is equal to one. The Pareto coefficient is used as an indicator of a hierarchical degree of urban structure. It measures how unequal the city distribution is. Higher values of the Pareto coefficient are related to more unequally distributed city system. On the other hand, lower values of Pareto coefficient are related to a more even city system. For hypothetical values of $\beta = 0$ all cities are of the same size and if $\beta = \infty$ there is only one big city. Equation 1 is usually written using natural logarithms:

$$\ln R_i = \alpha - \beta \ln S_i + \varepsilon_i. \quad (2)$$

Gabaix and Ibragimov (2011) proposed a method to estimate and improve the OLS estimation of the tail exponents by using $Rank-1/2$ term by which leading order bias is reduced:

$$\ln \left(R_i - \frac{1}{2} \right) = \alpha - \beta \ln S_i + \varepsilon_i. \quad (3)$$

Gibrat (1931) was first to observe that the distribution of firm size could be approximated with the lognormal distribution with the growth process of firms independent of its size. Gibrat defined that the law of proportionate effect will imply lognormal distribution of variables. Gibrat's law started as an empirical regularity associated with the distribution of firms but since the 1990s it was empirically tested many times in the field of urban economics. It is generally assumed that the law is valid in the long run. The growth of city population did not depend on its size, as larger cities did not grow at a faster pace than the smaller ones.

Analytically, the growth process can be expressed using the following logarithmic equation (Modica et al., 2015):

$$\log P_{(t)} = \log P_0 + \varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(t) \quad (4)$$

where $P_{(t)}$ is the city population (size) at time t , P_0 is the initial city population at time 0 and ε is an independent and identically distributed random variable with mean μ and variance σ^2 . Gibrat's law supposes that the growth process should have a common mean and a common variance which have to be independent from the city size. Gibrat's law can be tested using the following equation (Soo, 2014):

$$\ln(P_{i,t}) = \beta + \theta_i + \gamma_t + \delta \ln(P_{i,t-1}) + \varepsilon_{i,t} \quad (5)$$

where $P_{i,t}$ is the population of a city i at time t , θ_i is a state fixed effect and γ_t is a set of year dummies. If $\delta = 1$ Gibrat's Law holds and the error term $\varepsilon_{i,t}$ is independently and identically distributed.

According to Cordoba (2008) Zipf's Law implies Gibrat's law which differs from Gabaix (1999b) who stated the opposite – that Gibrat's law implies Zipf's law. If different cities grow randomly with the same expected mean growth rate and variance then the distribution of city sizes converges to Zipf's law (Gabaix and Ioannides, 2003). In order to test Gibrat's law, the estimated mean and variance should be calculated based on the Nadaraya-Watson method (Luckstead and Devadoss, 2014):

$$\hat{\mu}(s) = \frac{\sum_{i=1}^n K\left(\frac{s-S_i}{h}\right) g_i}{\sum_{i=1}^n K\left(\frac{s-S_i}{h}\right)} \quad (6)$$

$$\hat{\sigma}^2(s) = \frac{\sum_{i=1}^n K\left(\frac{s-S_i}{h}\right) (g_i - \hat{\mu}(s))^2}{\sum_{i=1}^n K\left(\frac{s-S_i}{h}\right)} \quad (7)$$

where $K(\cdot)$ is a kernel weight, h is the bandwidth and $\mu(s)$ is the local average growth rate around point s .

3. Literature overview

The investigation about urban regularities can be divided into Zipf's law investigations, Gibrat's law investigations and both Zipf and Gibrat's laws investigations. Deliktas, Önder and Karadag (2013) examined the size distribution of cities and the determinants of city growth in Turkey. There is some evidence that Zipf's law holds in Turkey according to the rank minus one half rule. The main factors that affected urban population growth in Turkey were fertility rate, city location, migrations, agglomeration in services and specialization in manufacturing industry. Soo (2005) tested Zipf's law for 73 countries using OLS and Hill's estimator. Zipf's law was rejected more often for OLS method than for Hill's estimator. The preferred estimation method depended on sample size and theoretical priors. According to Gabaix (1999b:739) Zipf's law is "*one of the most conspicuous empirical facts in economics, or in the social sciences generally*". Knudsen (2001) examined the Zipf's law for 61 largest Danish cities for the year 2000 getting a value of Zipf's exponent close to 1 and consequently leading to the acceptance of Zipf's law. The existence of Zipf's law was again proved on a sample of 14,541

production companies with ten employees or more. Jiang and Jia (2011) provided a geospatial perspective on the validity of Zipf's law for all natural cities in the United States. Instead of using census data on metropolitan area, urban agglomerations are demarcated by clustering street nodes forming natural cities. Zipf's law held well for all natural cities in United States contrary to urban areas which do not hold stable for Zipf's law. Gabaix and Ibragimov (2011) suggested a new method for estimating the Pareto exponent. OLS regression is the most popular way to estimate this coefficient due to its simplicity and robustness. But its weak side is that it is biased in small samples. The authors proposed using $Rank-1/2$ instead of $Rank$ alone. This proposed approach performed well for heavy-tailed distributions.

In this section investigations on validity of Zipf's law in Balkans and Croatia specifically are presented. Romisch (2015) developed a model for estimating agglomeration in the European Union and the Western Balkan region. Agglomeration was estimated for 25 European countries in the years from 2000 to 2012. Agglomeration affected the growth of regional economic development; high income levels in developed countries are associated with lower agglomerated economies and vice versa. Dimou and Schaffar (2009) studied the evolution of urban hierarchies and city growth in the Balkans over the last thirty years. They estimated and recalled different methods to estimate the Pareto exponent in order to test the validity of Zipf's law for cities in the Balkans. The validity of Zipf's law was confirmed independently of the date, meaning before and after the political and social crisis in the nineties. Jošić and Nikić (2013) tested Zipf's law using population data for city proper and city settlements in Croatia in 2011. The results corroborated the existence of Zipf's law e.g. city proper and city agglomerations strictly followed Zipf's law based on the strong linear-relationship between the variables rank and size. Jošić and Bašić (2018) tested whether Zipf's law holds for Croatian settlements, cities and agglomerations using 2011 Census of Population Survey. The results have shown that Zipf's law for settlements holds for majority of city sizes. However, rank-size distribution did not hold for very small and large settlements.

Vining (1974) demonstrated some simple tests of Gibrat's law. He verified Simon's model of urban growth emphasizing variations in the growth of cities by births, deaths, rural-urban, urban-rural and urban-urban migrations. Resende (2004) proposed a direct panel unit root test for testing Gibrat's law for the growth of cities in Brazil in the period from 1980 to 2000. The results were supportive to the validity of Gibrat's law for city size. Gonzalez-Val, Lanaspa and Sanz (2011) tested empirically the validity of Gibrat's law for the twentieth century data for the United States, Spain and Italy. Using panel unit root tests Gibrat's law was confirmed in the upper tail of distribution. When using non-parametric methods there was an evidence in favour of weak Gibrat's law. Gonzalez-Val, Lanaspa and Sanz (2013) used untruncated city population data for six countries (the United Sta-

tes, Spain, Italy, France, England and Japan) in order to illustrate how parametric growth regressions can lead to biased results implying that non-parametric methods are the appropriate methods for testing Gibrat's law. Eeckhout (2004) used census 2000 data covering the entire distribution of cities in the United States. He showed that cities grew proportionally, independently of city size giving rise to a lognormal distribution.

In this section both Zipf's and Gibrat's law investigations are presented. Özden and Parsons (2014) examined Zipf's and Gibrat's laws when applied to international migration and their relevance for immigrant and emigrant numbers and densities and migration patterns for the high-skilled. Zipf's law seems to hold for 50 largest countries in the upper tail of the size distribution. They estimated parametric and non-parametric kernel regressions to test for the applicability of Gibrat's law. The results indicated to Gibrat's law holding. Soo (2014) investigated Zipf's and Gibrat's laws on the size distribution and the growth of populations in China, India and Brazil. Zipf's law was rejected for India but not for China and Brazil. On the other side, Gibrat's law held for Brazil. Luckstead and Devadoss (2014) examined whether the world's largest cities follow Zipf's and Gibrat's laws. The results of the parametric analysis for all sample sizes rejected Zipf's law with Zipf's exponent systematically declining as the sample size increases. On the other hand, Gibrat's law held so it can be expected for size distribution to converge to Zipf's law with a decline in the barriers to immigration. Modica, Reggiani and Nijkamp (2015) presented a comparative analysis of Gibrat's and Zipf's laws on urban population of four different countries (Botswana, Germany, Hungary and Luxembourg). If the mean is independent of city size and the Pareto coefficient is different from one, the variance is dependent on city size. Gabaix (1999a) argued that the prerequisite for applying the model of urban growth should be satisfying the validity of Zipf's law. Ioannides and Overman (2000) tested the validity of Zipf's law for cities on data for metro areas in the United States for the censuses from 1900 to 1990. They used non-parametric procedures in order to test Gibrat's law. Zipf's law held for a wide range of city sizes but the value of the Pareto exponent varied across city sizes. The results pointed to the acceptance of Gibrat's law. Giesen and Suedekum (2009) showed that Gibrat's law is satisfied at the national level and random and non-random samples of German cities. The mean and variance of annual urban growth rates in Western Germany were scale independent. Zipf's law was also satisfied for cities in the regions and the country.

4. Data and methodology

In the analysis data about settlement sizes in Croatia will be used. In order to have comparable and reliable settlement sizes, data from official population censuses will be observed (Croatian Bureau of Statistics, 2018a, b). Until now there were 16 population censuses conducted. The first population census was conducted in 1857 whereas the last one in 2011. The population censuses in Croatia are conducted every 10 years. However, there are some discrepancies from this rule, especially in the period before and after the Second World War. Due to Second World War, population census was not conducted in 1941, but in 1948 and shortly after this in 1953. Those time irregularities between censuses can be observed as the main data limitation. Consequently, when census data from different years are going to be compared, those time irregularities should be kept in mind.

First, the analysis will be conducted by taking all settlements into account. Because there are quite a lot of settlements which are very small, some of them do not appear in larger number of observed censuses. It has been decided that the additional analysis will be conducted by observing settlements with more than 5,000 inhabitants and by observing top 100 settlements according to their size.

The methodological approach to the analysis is the following one. In the first step, some basic descriptive results about settlements in the observed census years will be provided. After that Zipf's law will be tested and the results will be graphically presented. For this purpose the variables rank and size were expressed in natural logarithms. In addition, equation 2 will be used to estimate the Pareto coefficient (β). If Croatian urban population system follows Zipf's law or rank-size distribution in the historical period, the value of Pareto coefficient (β) should be around -1 . Also, according to Zipf's law the variable logarithmic settlement size should follow lognormal distribution. This will be inspected by applying the Kolmogorov-Smirnov, the Anderson-Darling and the chi-square goodness-of-fit tests. The Kolmogorov-Smirnov test observes the maximum difference between empirical cumulative distribution functions of observed and tested distributions. The test statistics of the Kolmogorov-Smirnov test is calculated as follows:

$$D = \max_{1 \leq i \leq N} \left[F(Y_i) - \frac{i-1}{N}, \frac{i}{n} - F(Y_i) \right] \quad (8)$$

where $F(Y_i)$ is the theoretical cumulative distribution of the specified distribution being tested. The Anderson-Darling test is quite similar to the Kolmogorov-Smirnov test but it gives more attention to the tails of distributions. The test statistics of the Anderson-Darling test is defined as:

$$A^2 = -N - \sum_{i=1}^n \frac{(2i-1)}{N} \ln F(Y_i) + \ln(1 - F(Y_{N+1+i})). \quad (9)$$

As opposed to the Kolmogorov-Smirnov test and the Anderson-Darling test, the chi-square test can be applied to discrete distributions as well. The chi-square test statistics is calculated as:

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - e_i)^2}{e_i} \quad (10)$$

where f_i are actual values from the observed distribution and e_i are expected values of tested distribution. In all three tests the null hypothesis states that the observed data distribution follows the tested data distribution (in this case lognormal distribution).

Gibrat's law assumes that the growth of settlement size does not depend on its size. In other words, the largest settlements should not necessarily have the largest growth rates and vice versa. In this way the focus in the paper will be on the analysis of settlements growth rates which will be calculated as follows:

$$s_{i,t} = \frac{y_{i,t} - y_{i,t-1}}{y_{i,t-1}} \cdot 100 \quad (11)$$

where $y_{i,t}$ is settlement size of settlement i according to census conducted in period t . The validity of Gibrat's law will be inspected by using following panel unit root tests: the Levin, Lin and Chu (2002) test; the Im, Pesaran and Shin (2003) test; the Augmented Dickey Fuller (1979) test and Phillips-Perron (1988) test. All of mentioned unit root tests have considered the following model as one of initial models for unit root presence inspection (the notation is adjusted to the observed case of growth rates):

$$\Delta s_{i,t} = \alpha_i + \beta_i s_{i,t-1} + e_{i,t}, \quad (12)$$

where $s_{i,t}$ is a growth rate of settlement i in census period t in comparison to census period $t - 1$. Because the unit root tests are out of the scope of this paper, more about them can be found in, for example, Barbieri (2005), Hlouskova and Wagner (2005), Hossain (2010).

Despite the difference in steps of analysis and assumptions, all four observed unit root tests have the same null hypothesis for the presence of unit root. If Gibrat's law stands, the unit root tests should result in the non-stationary unit root process or, in other words, the estimation $\hat{\beta}_i$ should be equal to zero (Lalanne and Zumpe, 2015). In the unit root analysis the panel of all settlements together will be first observed and after that unit root analysis will be conducted by taking into account settlements of the same size category. In order to do that, six settle-

ment size categories are introduced (1-500; 501-1,000; 1,001-5,000; 5,001-10,000; 10,001-50,000; 50,000 and more citizens).

5. Results and discussion

In the paper data about settlement sizes in Croatia in the period from 1857 to 2011 are used. For the purpose of analysis, three different samples are introduced. In the first case all settlements will be included in the analysis. In the second sample only settlements larger than 5,000 inhabitants will be observed whereas in the third case only the largest 100 settlements will be analysed. Basic descriptive results of those three groups of settlements are given in Table 1.

Table 1.

DESCRIPTIVE STATISTICS RESULTS OF SETTLEMENTS IN CROATIA, CENSUSES CONDUCTED FROM 1857 TO 2011

Census year	All settlements included				Settlements larger than 5,000				The largest 100 settlements			
	No. of sett.	Aver. sett. size	Coeff. of var.	Med.	No. of sett.	Aver. sett. size	Coeff. of var.	Med.	No. of sett.	Aver. sett. size	Coeff. of var.	Med.
1857	5,382	405	192	237	13	11,254	66	9,401	100	3,841	102	2,668
1869	5,128	468	198	270	17	11,358	73	9,564	100	4,450	104	3,095
1880	5,910	424	245	239	20	12,588	85	8,601	100	4,922	124	3,039
1890	6,238	457	267	264	22	13,691	100	8,704	100	5,504	140	3,369
1900	6,347	498	315	282	26	15,289	119	8,254	100	6,446	164	3,592
1910	6,345	545	364	306	30	17,196	132	8,898	100	7,586	182	3,784
1921	5,665	608	383	343	33	16,567	153	9,301	100	7,837	200	3,893
1931	5,829	649	520	360	33	20,987	190	8,701	100	9,277	260	3,868
1948	6,624	571	747	296	32	24,819	228	9,093	100	10,263	324	3,871
1953	6,665	590	803	295	38	24,520	238	8,548	100	11,496	323	4,203
1961	6,664	624	938	283	47	25,984	251	7,842	100	14,227	322	4,855
1971	6,673	663	1,161	248	61	27,914	273	8,097	100	18,624	325	5,940
1981	6,548	703	1,273	218	74	28,672	278	9,185	100	22,293	311	6,799
1991	6,533	732	1,320	199	81	29,299	281	9,250	100	24,552	304	7,691
2001	6,651	667	1,388	157	79	28,355	284	9,502	100	23,341	309	7,306
2011	6,560	652	1,413	143	78	28,029	286	9,999	100	22,829	313	7,167

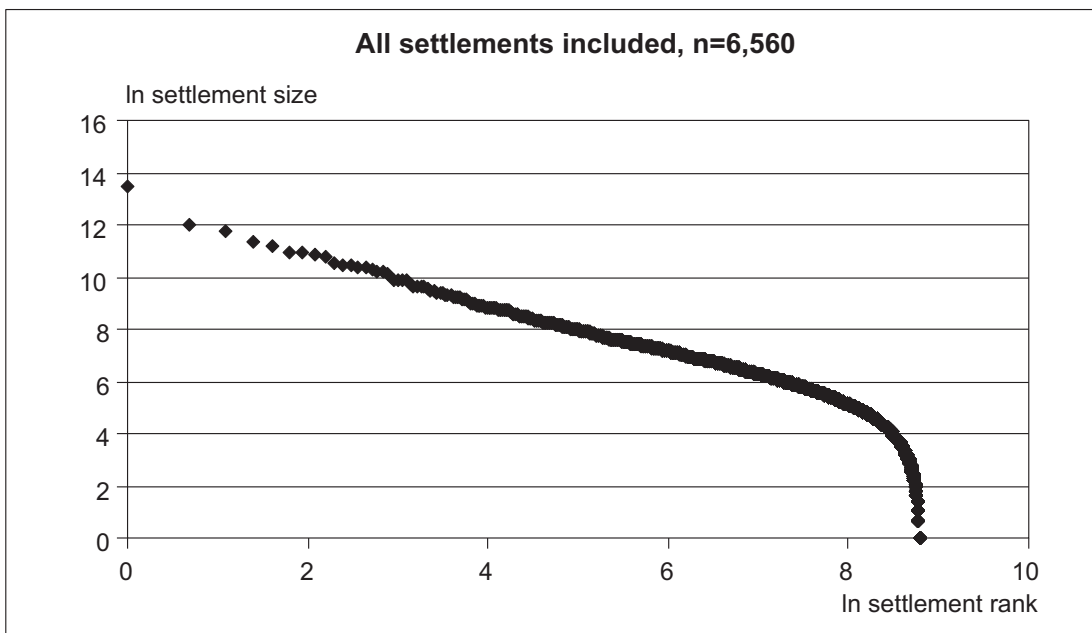
Source: Croatian Bureau of Statistics, 2018a, b, Authors.

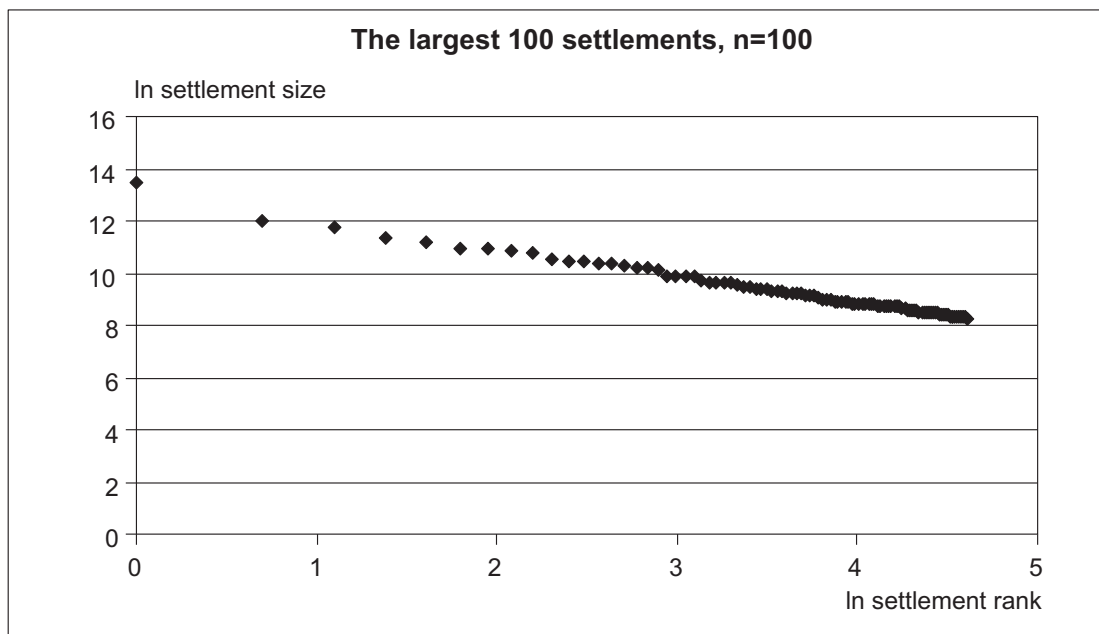
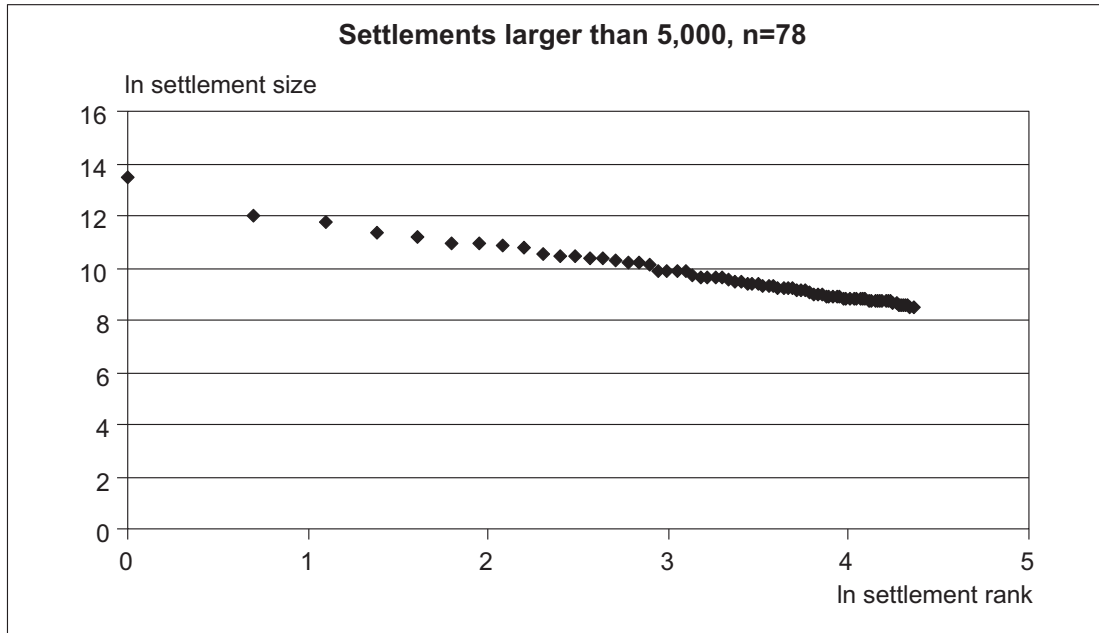
According to Table 1, the number of settlements in Croatia was in the range of 5,128 in 1869 to 6,673 in 1971. However, after the census conducted in 1948 the number of settlements was quite stable and stabilized at 6,600. On the other hand, the average settlement size showed higher deviation level between the censuses. Furthermore, the coefficients of variation showed that the average settlement size is not the representative indicator which, at the same time, points out groups of settlements. Therefore the values of median are provided in Table 1 as well. The variation level is lowered by observing only settlements larger than 5,000 inhabitants and only the largest 100 settlements but still it can be considered as very high.

In order to apply Zipf's law, the settlements were first ordered according to their size. The largest settlement was associated with rank 1, the second largest settlement was associated with rank 2 and so on. In the following step, both variables, settlement size and settlement rank, were transformed by applying natural logarithms on their original values. In Figure 1 scatter diagrams of those two new variables for different number of observed settlements are given. Due to paper length restrictions, only situations in 2011 are presented here whereas scatter diagrams, when all settlements are included, are shown in Figure A1 in the Appendix.

Figure 1.

SCATTER DIAGRAMS OF NATURAL LOGARITHMS OF SETTLEMENT SIZES AND THEIR RANKS, DIFFERENT NUMBER OF OBSERVED SETTLEMENTS, CENSUS CONDUCTED IN 2011



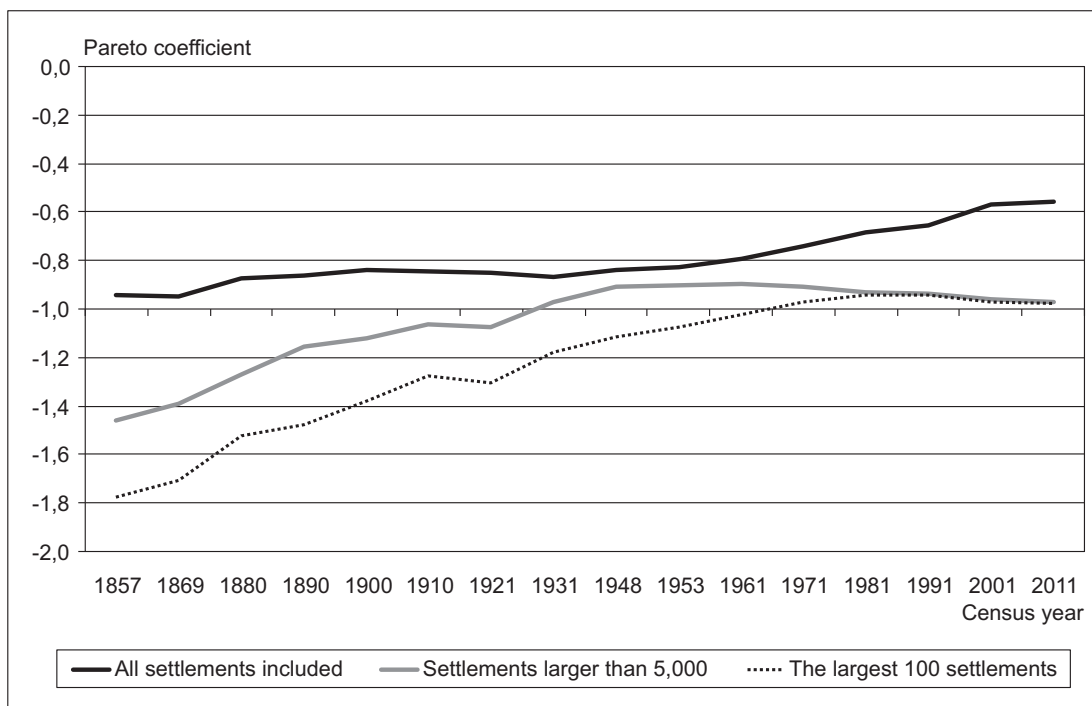


Source: Authors.

Scatter diagrams shown in Figure 1 suggest that there could be a linear relationship between the two observed variables (especially when all settlements are not observed). Therefore ordinary least squares (OLS) regression can be applied to estimate the Pareto coefficient (β) as defined in equation 2. The values of the Pareto coefficient for all three observed numbers of settlements including the whole observed period are given in Figure 2:

Figure 2.

THE PARETO COEFFICIENT (β) VALUES, DIFFERENT NUMBER OF OBSERVED SETTLEMENTS, FROM CENSUS CONDUCTED IN 1857 TO CENSUS CONDUCTED IN 2011



Source: Authors.

According to Figure 2, when all settlements are observed, the values of the Pareto coefficients are closer to the value of -1 at earlier censuses than in the later ones. On the other hand, if settlements larger than 5,000 inhabitants and the largest 100 settlements are observed, the values of the Pareto coefficient are considerably different from the value -1 at the earliest censuses. It can be concluded that if all settlements are observed Zipf's law is roughly valid until the census conducted in 1948. However, if settlements larger than 5,000 inhabitants and the largest 100 settlements are taken into account, Zipf's law can be considered as valid from the 1890 census and if the largest 100 settlements are observed, Zipf's law held since the 1948 census. In order to confirm the conclusion it will be inspected whether variable settlement size, given in natural logarithm values, follows the lognormal distribution which significantly differs from the Pareto distribution. In Table 2 the main results of conducted Kolmogorov-Smirnov, the Anderson-Darling and the chi-square goodness-of-fit tests are shown.

Table 2.

THE KOLMOGOROV-SMIRNOV, THE ANDERSON-DARLING AND THE CHI-SQUARE GOODNESS-OF-FIT RESULTS, OBSERVED VARIABLE LN SETTLEMENT SIZE, DIFFERENT NUMBER OF OBSERVED SETTLEMENTS, CENSUSES CONDUCTED FROM 1857 TO 2011

Census year	All settlements included			Settlements larger than 5,000			The largest 100 settlements		
	K-S	A-D	χ^2	K-S	A-D	χ^2	K-S	A-D	χ^2
1857	0.04***	21.48***	119.69***	0.21	0.42	N/A	0.20***	6.88***	81.44***
1869	0.03***	12.88***	70.78***	0.15	0.58	N/A	0.18***	5.18***	65.60***
1880	0.06***	38.93***	245.85***	0.15	0.70	N/A	0.16***	5.34***	36.80***
1890	0.06***	57.23***	346.89***	0.16	0.93	N/A	0.17***	5.80***	49.80***
1900	0.07***	72.68***	434.74***	0.17	1.23	1.27	0.18***	5.86***	42.20***
1910	0.07***	65.87***	421.15***	0.16	1.47	1.90	0.18***	5.87***	48.40***
1921	0.07***	58.98***	375.43***	0.17	1.43	4.55	0.21***	7.32***	81.40***
1931	0.07***	57.45***	385.19***	0.16	1.30	2.61	0.20***	7.07***	84.08***
1948	0.07***	73.06***	483.22***	0.15	1.21	7.19*	0.19***	7.17***	83.82***
1953	0.07***	71.41***	459.83***	0.18	1.91	13.19***	0.20***	7.05***	84.86***
1961	0.08***	81.64***	541.67***	0.21**	2.61**	26.22***	0.19***	6.64***	73.60***
1971	0.08***	99.86***	674.02***	0.22***	3.50**	50.15***	0.18***	5.70***	77.84***
1981	0.09***	120.50***	800.17***	0.19***	3.80**	57.68***	0.16***	4.36***	46.12***
1991	0.09***	113.75***	741.57***	0.16**	3.61**	35.00***	0.14**	3.97***	39.88***
2001	0.11***	177.01***	1,200.82***	0.15*	3.27**	32.65***	0.17***	4.17***	59.64***
2011	0.11***	178.96***	1,177.16***	0.14	2.94**	40.15***	0.16**	3.73**	39.88***

Notes: K-S – Kolmogorov-Smirnov test, A-D – Anderson-Darling test, χ^2 – chi-square test, statistically significant results at: * – 10% significance level, ** – 5% significance level, *** – 1% significance level.

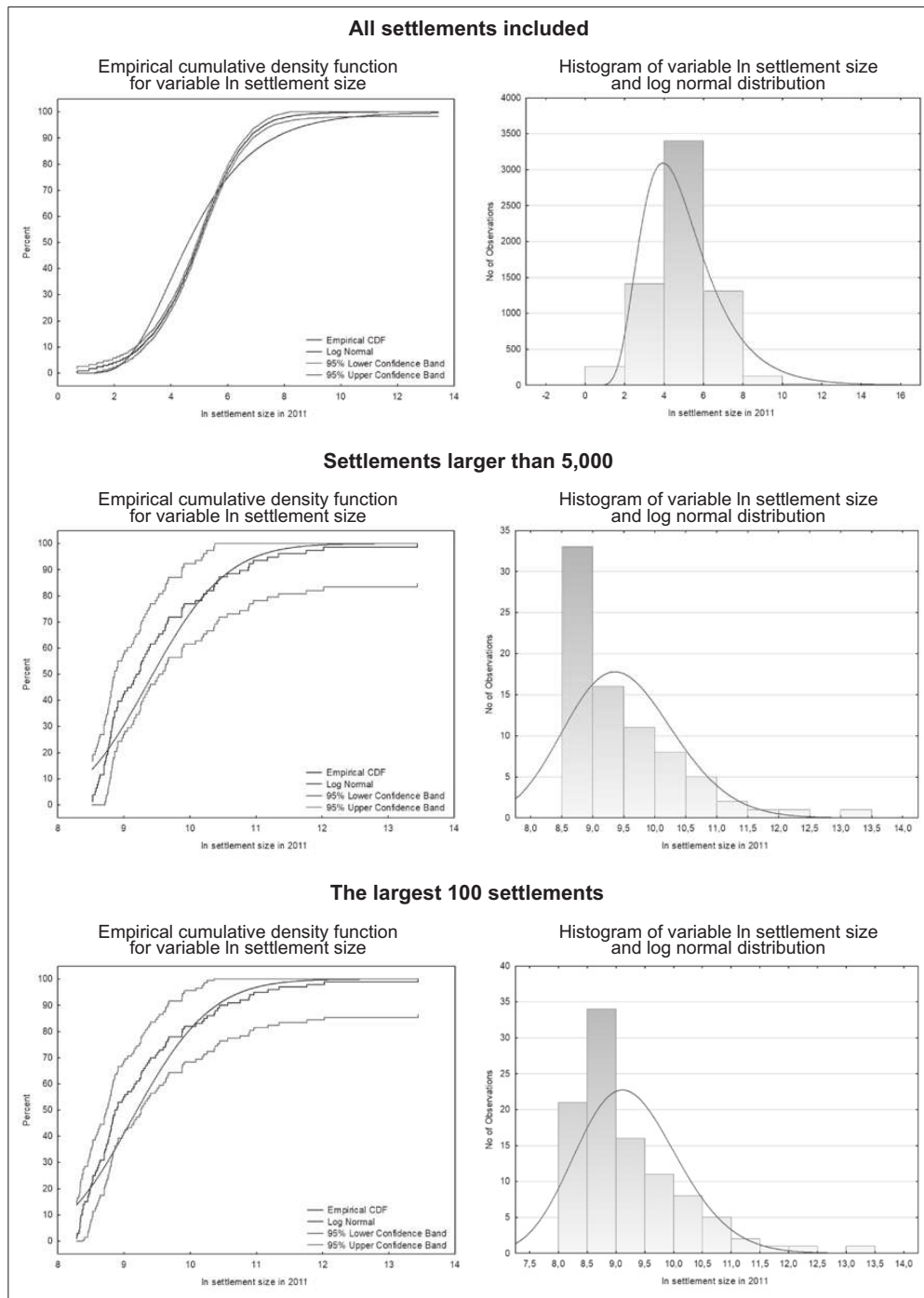
Source: Authors.

According to Eeckhout (2004:1432) „the only remaining issue to resolve is how it is possible that Zipf’s law is repeatedly confirmed in the literature, while the underlying distribution is lognormal. The Pareto distribution is very different from the lognormal, so it is obvious that if the true distribution is lognormal, the entire distribution can never be fit to a Pareto distribution at the same time...“. The answer to this intriguing question is which sample size is taken into consideration; the whole sample (untruncated distribution) or the truncated distribution at the very upper-tail where there is no significant difference between the lognormal and the Pareto distribution.

In all three applied tests the null hypothesis includes assumption that the variable settlement size, given in natural logarithm values, follows the lognormal distribution. The test results from Table 2 suggest in majority of observed cases that variable \ln settlement size does not follow the lognormal distribution. However, the results are very interesting. Namely, if settlements larger than 5,000 are observed, at most observed census years the null hypothesis cannot be rejected. This is very indicative meaning that both the truncated lognormal and the Pareto distribution trace the data relatively closely when the small number of the largest settlements is observed leading to the acceptance of Zipf's law. In order to support this conclusion comparison of variable \ln settlement size with the lognormal distribution for different number of observed settlements in 2011 is shown in Figure 3.

Figure 3.

COMPARISON OF VARIABLE LN SETTLEMENT SIZE WITH LOGNORMAL DISTRIBUTION, DIFFERENT NUMBER OF OBSERVED SETTLEMENTS, CENSUS CONDUCTED IN 2011



Source: Authors.

In order to inspect the validity of the Gibrat's law panel unit root testing is conducted. The variable of interest in the analysis is settlement growth rate. Four different panel unit root test have been conducted and the results are shown in Table 3. The panel unit root tests were conducted by testing for unit root in level and by including individual intercept in the test equation. Furthermore, the lag length was set to be four. First, all settlements were taken into account and after that settlements were observed according to their size. It has to be emphasized that an unbalanced panel approach was used because for some settlements data in all 16 censuses were not available.

Table 3.

THE RESULTS OF CHOSEN PANEL UNIT ROOT TESTS, VARIABLE
SETTLEMENT GROWTH RATE, 15 CENSUS BASED PERIODS
(FROM 1869 TO 2011)

Settlement size	Panel unit root test	Test statistic	P-value	Cross-sections	Observations
All	Levin, Lin and Chu	23.76	1.0000	5,172	50,999
	Im, Pesaran and Shin	9.18	1.0000	5,172	50,999
	Augmented Dickey Fuller	7,327.17	1.0000	5,172	50,999
	Phillips-Perron	35,288.18***	<0.0001	5,172	71,687
1-500	Levin, Lin and Chu	31.98	1.0000	2,938	28,521
	Im, Pesaran and Shin	8.02	1.0000	2,938	28,521
	Augmented Dickey Fuller	4068.84	1.0000	2,938	28,521
	Phillips-Perron	20,246.25***	<0.0001	2,938	40,273
501-1,000	Levin, Lin and Chu	2.10	0.9820	162	1,430
	Im, Pesaran and Shin	1.96	0.9752	162	1,430
	Augmented Dickey Fuller	237.00	0.9999	162	1,430
	Phillips-Perron	1,138.73***	<0.0001	162	2,078
1,001-5,000	Levin, Lin and Chu	8.93	1.0000	212	1,980
	Im, Pesaran and Shin	0.16	0.5627	212	1,980
	Augmented Dickey Fuller	388.22	0.8929	212	1,980
	Phillips-Perron	1,436.10***	<0.0001	212	2,828
5,001-10,000	Levin, Lin and Chu	-1.86**	0.0316	2	18
	Im, Pesaran and Shin	-1.21	0.1138	2	18
	Augmented Dickey Fuller	8.09*	0.0882	2	18
	Phillips-Perron	12.75**	0.0126	2	26

Settlement size	Panel unit root test	Test statistic	P-value	Cross-sections	Observations
10,001-50,000	Levin, Lin and Chu	-0.88	0.1884	2	17
	Im, Pesaran and Shin	-0.73	0.2328	2	17
	Augmented Dickey Fuller	5.44	0.2451	2	17
	Phillips-Perron	5.28	0.2594	2	25
50,001 and more	Levin, Lin and Chu	2.61	0.9955	1	8
	Im, Pesaran and Shin	1.05	0.8539	1	8
	Augmented Dickey Fuller	0.09	0.9577	1	8
	Phillips-Perron	1.34	0.5108	1	12

Notes: statistically significant results at: * – 10% significance level, ** – 5% significance level, *** – 1% significance level.

Source: Authors.

The majority of panel unit root tests have shown that at the significance level of 1% the null hypothesis cannot be rejected. It can be concluded that there is a presence of unit root for the variable growth rate in both cases when all settlements are included and when settlements were divided according to their size. Consequently, the conclusion is that Gibrat's law holds in all observed cases.

6. Conclusion

The goal of this paper was the investigation of two main urban regularities, Zipf's and Gibrat's laws, using data for Croatia in the period from 1857 to 2011. For this purpose two research questions or hypotheses were constructed. The first hypothesis investigated whether the Croatian urban population system follows rank-size distribution in the historical period from 1857 to 2011. The method for validating the hypothesis was the calculation of the Zipf's exponent using ordinary least squares (OLS) regression. The results have shown that the values of the Pareto coefficient for all settlements were closer to the value of -1 at earlier censuses than in the later ones indicating the validity of Zipf's law until census conducted in 1948. On the other hand, when settlements larger than 5,000 inhabitants and the largest 100 settlements were observed, the values of Pareto coefficients were considerably different from the value -1 at the earliest censuses. Zipf's law held from the 1890 census for settlements larger than 5,000 inhabitants and since 1948 census for 100 largest settlements. The second hypothesis investigated whether the population growth rate of settlements in the Croatian urban system is proportional

regardless of the size of settlements. This hypothesis was proven using panel unit root approach. The results for majority of the panel unit root tests have shown, at the significance level of 1%, that there is a presence of unit root in settlements growth rate in cases when all settlements were included and when settlements were divided according to their sizes implying that Gibrat's law holds in all cases. The limitations of the study are related to the time irregularities between censuses conducted and uneven number of settlements in censuses leading to an unbalanced panel due to the fact that some settlements cease to exist over time. Recommendation for future research in this field can go in the way of analyzing whether the main urban laws hold in the case of decreasing population in Croatia and understanding the complexity of the relationship between the two laws. It is highly recommended to repeat the analysis with the new data on population census planned to be conducted in 2021.

References

1. Auerbach, F. (1913). Das gesetz der bevölkerungskonzentration. *Petermanns Geographische Mitteilungen*, 59, 74-76.
2. Barbieri, L. (2005). Panel Unit Root Tests: A Review. *Serie Rossa: Economia – Quaderno*, 43, 1-53.
3. Cordoba, J. C. (2008). A generalized Gibrat's law. *International Economic Review*, 49(4), 1463-1468.
4. Croatian Bureau of Statistics (2018a). Census of Population, Households and Dwellings, 1857-2001. Available at: https://www.dzs.hr/App/PXWeb/PX-WebHrv/Menu.aspx?px_type=PX&px_db=Naselja+i+stanovni%C5%A1tvo+Republike+Hrvatske&px_language=hr&rxid=fc9d580f-2229-4982-a72c-cdd3e96307d3 [4 May 2019].
5. Croatian Bureau of Statistics (2018b). Census of Population, Households and Dwellings 2011. Available at: [http://www.dzs.hr/Hrv_Eng/publication/2011/SI-1441.pdf]. [4 May 2019].
6. Deliktas, E., Önder, A. Ö., Karadag, M. (2013). The size distribution of cities and determinants of city growth in Turkey. *European Planning Studies*, (21)2, 251-263. doi:10.1080/09654313.2012.722922.
7. Dickey, D. A., Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427-443.
8. Dimou, M., Schaffar, A. (2009). Urban hierarchies and city growth in the Balkans, *Urban Studies*, 46(1), 2891-2906.

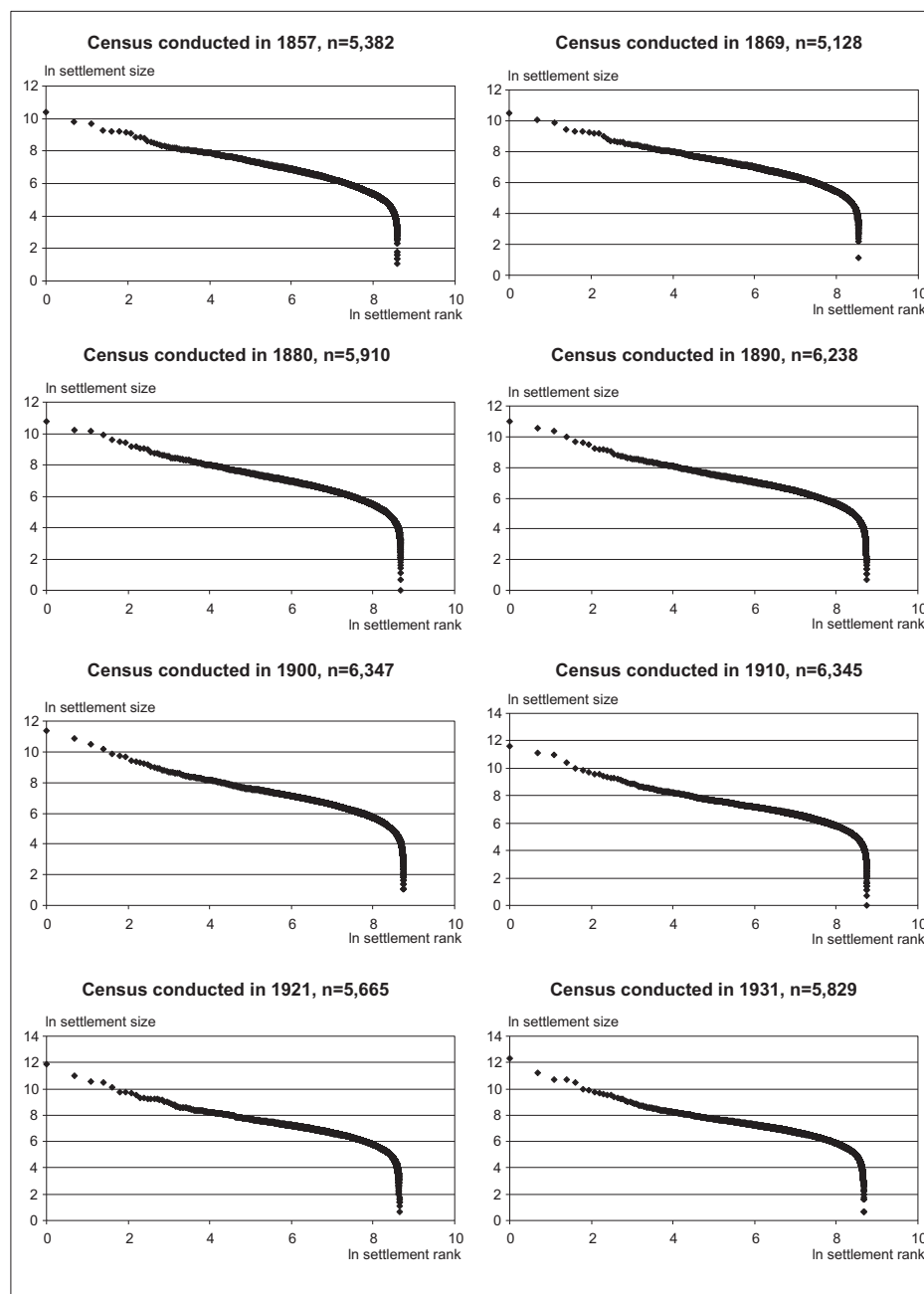
9. Eeckhout, J. (2004). Gibrat's law for (All) Cities. *American Economic Review*, 94(5), 1429-1451. doi: 10.1257/0002828043052303.
10. Gabaix, X. (1999a). Zipf's law and the Growth of Cities. *American Economic Review AEA papers and proceedings*, 89(2), 129-132. doi: 10.1257/aer.89.2.129.
11. Gabaix, X. (1999b). Zipf's law for Cities: An Explanation. *Quarterly Journal of Economics*, 114(3), 739– 767.
12. Gabaix, X., Ibragimov, R. (2011). Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents. *Journal of Business & Economic Statistics*, January 2011, 29(1), 24-39. doi: 10.2139/ssrn.881759.
13. Gabaix, X., Ioannides, Y. M. (2003). The Evolution of City Size Distributions. Chapter 53 in *Handbook of Regional and Urban Economics*, 4, 2341-2378.
14. Gibrat, R. (1931). *Les inégalités économiques; applications: Aux inégalités des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel*. Librairie du Recueil Sirey, Paris.
15. 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.
16. Gonzalez-Val, R., Lanaspa, L., Sanz, F. (2011). Gibrat's law for Cities Revisited, ERSA conference papers. European Regional Science Association. Available at: <https://econpapers.repec.org/scripts/showcites.pf?h=repec:wiw:wiwrsa:ersa10p199>. [4 May 2019].
17. Gonzalez-Val, R., Lanaspa, L., Sanz, F. (2013). Gibrat's law for cities, growth regressions and sample size. *Economics Letters*, 118(2), 367-369. doi:10.1016/j.econlet.2012.11.020.
18. Hlouskova, J., Wagner, M. (2005). The Performance of Panel Unit Root and Stationarity Tests: Results from a Large Scale Simulation Study. European University Institute working paper, 5, 1-38.
19. Hossain, A. S. M. (2010). A Panel Cointegration Analysis of the Euro area money demand. Master thesis, Department of statistics, Lund University, 1-30.
20. Im, K. S., Pesaran, M., Shin, Y. (2003). Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics*, 115(1), 53-74.
21. Ioannides, Y. M., Overman H. G. (2000). Zipf's law for cities: An empirical examination. *Regional Science and Urban Economics*, 33(2), 127-137.
22. 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*, 25(8), 1269-1281. doi:10.1080/13658816.2010.510801.

23. Jošić, H., Bašić, M. (2018). Reconsidering Zypf's law for regional development: The case of settlements and cities in Croatia. *Miscellanea Geographica*, 22(1), 22-30. doi: 10.2478/mgrsd-2018-0002.
24. 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 MC SER Publishing-Rome, Italy*, 2(9), 583-588.
25. Knudsen, T. (2001). Zipf's law for Cities and Beyond: The Case of Denmark. *American Journal of Economics and Sociology*, 60(1), 123-146.
26. Lalanne, A., Zumpe, M. (2015). Zipf's law, Gibrat's law and Cointegration. *Cahiers du GREThA*, n 2015-27. Available at: <http://ideas.repec.org/p/grt/wpegrt/2015-27.html>. [4 May 2019].
27. Levin, A., Lin, F., Chu, C. J. (2002). Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties. *Journal of Econometrics*, 108(1), 1-24.
28. Luckstead, J., Devadoss, S. (2014). Do the world's largest cities follow Zipf's and Gibrat's laws? *Economics Letters*, 125(2), 182-186.
29. Modica, M., Reggiani, A., and Nijkamp, P. (2015). A Comparative Analysis of Gibrat's and Zipf's law on Urban Population. *Quaderni – Working Paper DSE N°1008*, 25 pages.
30. Özden, C., Parsons, C. (2014). Zipf's law, Gibrat's law and international migration. *Fondation pour les études et recherches sur le développement international*, Policy brief number 101. Available at: <http://www.ferdi.fr/sites/www.ferdi.fr/files/publication/fichiers/br101-c.ozdenparson4.pdf>. [4 May 2019].
31. Phillips, P. C. B., Perron, P. (1988). Testing for Unit Roots in Time Series Regression. *Biometrika*, 75(2), 335-346.
32. Resende, M. (2004). Gibrat's law and the Growth of Cities in Brazil: A Panel Data Investigation. *Urban Studies*, 41(8), 1537-1549.
33. 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*, 25 pages.
34. Soo, K. T. (2005). Zipf's law for Cities: A cross-country investigation. *Regional Science and Urban Economics*, 35, 239-263.
35. Soo, K., T. (2014). Zipf, Gibrat and geography: Evidence from China, India and Brazil. *Papers in Regional Science*, 93(1), 159-182.
36. Vining, D. R. (1974). On the Sources of Instability in the Rank-Size Rule: Some Simple Tests of Gibrat's law. *Geographical Analysis*, 6(4), 313-329.
37. Zipf, G. K. (1949). *Human Behavior and the Principles of Least Effort*. Addison Wesley: Cambridge, MA.

Appendix

Figure A1.

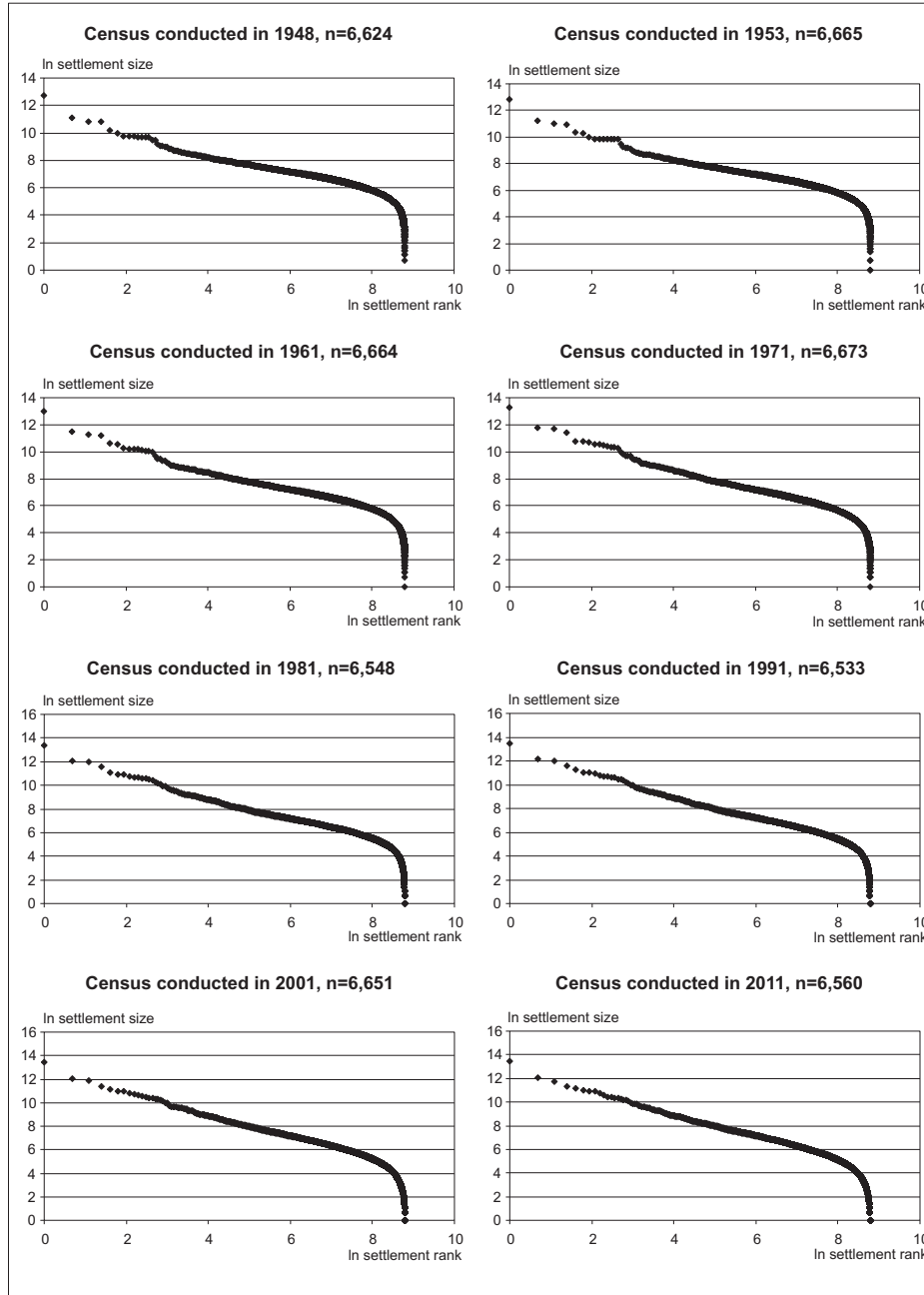
SCATTER DIAGRAMS OF NATURAL LOGARITHMS OF SETTLEMENT SIZES AND THEIR RANKS, ALL SETTLEMENTS INCLUDED, ALL OBSERVED CENSUSES



Source: Authors.

Figure A1.

SCATTER DIAGRAMS OF NATURAL LOGARITHMS OF SETTLEMENT SIZES AND THEIR RANKS, ALL SETTLEMENTS INCLUDED, ALL OBSERVED CENSUSES – CONTINUED



Source: Authors.

ISTRAŽIVANJE URBANIH ZAKONITOSTI ZA HRVATSKU U RAZDOBLJU OD 1857. DO 2011. GODINE

Sažetak

Dvije temeljne zakonitosti na polju urbane ekonomike su Zipfov zakon i Gibratov zakon. Zipfov zakon nalaže da bi distribucija najvećih gradova trebala slijediti raspodjelu prema Pareto rangu veličine dok Gibratov zakon navodi da je proporcionalni rast gradova neovisan o njihovoj veličini. Ta dva zakona su međusobno povezana i stoga se često razmatraju zajedno. Cilj ovog rada je istraživanje urbanih zakonitosti za Hrvatsku u razdoblju od 1857. do 2011. godine. Kako bi se procijenila i ocijenila struktura hrvatske urbane hijerarhije, Pareto ili Zipfovi koeficijenti su izračunati. Rezultati su pokazali da vrijednosti koeficijenata za najveća naselja u različitim godinama iznose približno 1, što upućuje na zaključak da hrvatski urbani hijerarhijski sustav slijedi pravilo veličine ranka odnosno Zipfov zakon. Neovisnost rasta gradova u ovisnosti o njihovoj veličini ocijenjeno je pomoću testa jediničnog korijena. Rezultati testiranja Gibratovog zakona uz pomoć panel testa jediničnog korijena su pokazali da postoji prisutnost jediničnog korijena za rast naselja što upućuje na prihvaćanje Gibratovog zakona.

Ključne riječi: Hrvatska, Gibratov zakon, naselja, Zipfov zakon