# ZIPF'S LAW AS ASSESSMENT TOOL OF URBAN INEQUALITY<sup>1</sup>

#### Inna MANAEVA

World Economy Chair -Belgorod State National Research University, Russia, http://www.bsu.edu.ru In.manaeva@yandex.ru

#### Svetlana RASTVORTSEVA

World Economy Chair -Belgorod State National Research University, Russia, http://www.bsu.edu.ru Srartvortseva@gmail.ru

### Abstract

The paper is concerned with the topical issues of regional economics – urban inequality in the Russian Federation. Empirical investigations of Zipf's law were studied in the foreign and Russian literature. Application of this law for assessment of urban inequality using the method of least squares was substantiated. Assessment of urban inequality within the boundaries of the RF federal districts by the indices of population, volume of own production of goods and services is carried out in the paper. The authors used the data of the Federal State Statistics Service for 2014, the investigation included the settlements with the status of a town and with the population over 100 thousand people. Zipf's law displays over the entire territory of Russia. By the population index in the federal districts, Zipf's factor varies within the range from - 0.7 (Northwestern Federal District) to - 0.9 (North Caucasian Federal District). As a result of the performed analysis of the Russia's cities by the population index, Zipf's factor is within the range from -0.3 (Northwestern Federal District) to -1.2 (Central Federal District). Analysis of the volume of production of goods and services determined the range of Zipf's factor from -0.26 (North Caucasian Federal District) to -0.7 (Central and Volga Federal Districts). By the index of population and volume of production of goods and services the following "primate cities" are determined: Moscow and Saint Petersburg, Yekaterinburg (population), which allows to draw a conclusion on their dominance in urban system and high differentiation of cities by these indices. The obtained empirical estimators prove that Russia has no intermediate group of cities macroregional centers. The results of the investigation can be used for creation of methodological tools to develop the mechanisms of smoothing of interregional inequality, program of economic and social development of cities.

**Keywords:** city, spatial inequality, Zipf's law, population, population density **JEL classification:** R12

# 1. Introduction

Russia ranks first in the land area and 181th – in population density in the world, which defines a unique feature of distribution of its cities and economic activity in the territorial area. The main feature consists in confrontation of large capital cities and basic mass of cities. Interregional inequality creates a range of problems for modern Russia; high income differences, concentration of competitive advantages within the same territories and their deficit within the other ones are worsened by social inequality. The problems of unequal access to employment market, education, public health service intensify, that threatens the integrity of the country and social and political stability. Different mechanisms of smoothing the interregional differentiation are put into practice: budget interregional transfer deeds, support of "priority development areas", implementation of modernization of monotowns (single-industry towns). In the course of the mentioned measures, understanding of concentration of resources, population, companies in the cities as a source of economic

<sup>&</sup>lt;sup>1</sup> The investigation is supported by the President's grant, project No. MK- 4493.2016.6

growth is essential. Russia's cities are a special subject for investigation of interregional differentiation. The range of urban population varies from 1 thous. people to 12 108.3 thous. people.; population density in cities – from 9.1 per km² to 7,165 per km²²; production volume – from RUB 10.8 mln. to RUB 5,653,126 mln³. Geographical and climatic conditions which strongly differ within the Russian Federation territory have effect on the economic development of a city.

The investigation objective, the results of which are given in this paper, is analysis of urban inequality with application of Zipf's law by a variety of the indices: population, population density, volume of own production of goods and services.

The authors chose Zipf's law as a tool for assessment of urban inequality due to applicability of this empirical law for a number of economics spheres that is proved in the papers of foreign economic researchers: distribution of profits of companies testifies that income statistics complies with Zipf's law (Ramsden & Kiss-Haypal, 2000); assessment of hierarchy of city sizes (Blank & Sorin, 2000); "Pareto index – a measure of spatial inequality" (Soo, 2005);

## 2. Review of Investigations

To achieve the stated objective, let us consider the known empirical studies on these topics, found in foreign economic literature. Vilfredo Pareto brought up a concern with income distribution inequality as early as in 1896 (Flux& Pareto, 1896). The scientist considered wealth distribution among regions and determined that it was utterly unequal: 20% of the population has 80% of the benefit (wealth), and 80% of the population – only 20% of the benefit - the rule "the wealthy become wealthy" (Flux& Pareto, 1896). F. Auerbach (1913), having proved the power-series distribution for city sizes, offered the hypothesis of empirical dependence between a city size (its population) and its rank in the hierarchy of cities of a region or country (Auerbachu, 1913). Thereafter, the provision of F. Auerbach was improved by D. Zipf. The term "Zipf's law" provides that within the territory distribution by city size is subject to Pareto's distribution with the index equal to a unit value. The other definition of Zipf's law consists in the fact that if large cities are ranked in descending order of their population, then relation of population of two cities will be inversely proportional to relation of their ranks. M. Naldi studies four indices (Gini, Bonferroni, Amato, Herfindahl-Hirschman) for their application under the conditions described by Zipf's law (Naldi, 2003). The Herfindahl-Hirschman index displays flexible behavior and, consequently, is capable to solve (expanding) variations in distribution of economic magnitude described by Zipf's law. T.Rosen and M. Reznik (1980) were the first who carried out complex investigations by the example of 44 countries. Approbation of Zipf's law showed that maximum Pareto index takes place in Australia (1.9), minimum one is typical for Morocco (0.8). The researchers state that the index is overrated in Australia, and relate this case to exceptions. If Australia is removed from the sample collection, then Nigeria becomes the leader by Pareto index – 1.5 (Rosen & Resnick, 1980). Kolomak (2014), analyzing Zipf's law in relation of Russian cities concluded: from stability of distribution of city size it follows that there may be population fluctuations for separate cities, however, this is not accompanied by increase (or decrease) of irregularity in urban system and processes of urban population concentration (or deconcentration) (Kolomak, 2014). Validity of Zipf's law is affirmed for Japan and France: urban structures remain stable over time (Zipf's factor is not changed) (Eaton & Eckstein, 1997).

For many countries and periods Zipf's factor significantly differs from 1 (it is often revealed that Zipf's factor is < 1), and is not constant over the long term (Soo, 2005; Nitsch, 2005). The urban system where growth of cities does not depend on their size is able to generate Zipf's law in a steady state (Gabaix, 1999). In the context of expanding economics, Zipf's law requires average growth of cities, in order not to depend on their size (Córdoba, 2008). Zipf's law properly describes urban systems of countries which had no control of sizes, location of cities and internal migration of population (Kolomak, 2014).

Foreign scientists introduce the term "principal city" (primate city) a large city which is a financial and political center of the country. Availability of the "primate city" in the country is indicative of spatial development imbalance "progressive core and lagging periphery" (Baker & Phongpaichit, 2009).

### 3. Methodology

Zipf's law concerning distribution of city size is represented by a nonlinear relation between city ranks (r) and its size (s), sometimes this rule is called "rank-size". It may be recorded in the form of the following formula:

$$s = r^{-1} \tag{1}$$

Respectively, city size s will be equal to 1, 1/2, 1/3, etc. if city rank r is equal to 1, 2, 3. The largest city is twice as large as the second one, etc. and thrice as large as the third one, etc. This equation is typical for economic phenomena in distribution of income and sizes of companies (Soo, 2005).

Power expression of Zipf's law is represented as follows:

$$y = kx^{-\alpha} \tag{2}$$

where x is quantity, k is a constant,  $\alpha$  is exponent of power law. This law is known as a Pareto distribution.

Plausible methods were suggested by M. Goldstein (2004), M. Newman (2005), who base on the Kolmogorov-Smirnov test for determination of distribution as per the power law. The methods are used not only to comply with the data (or a part of the data) as per the power law but also to determine to what extent these data are suitable in comparison with the other types of distribution. Te index is set by the following formula (Jiang & Jia, 2011):

$$\alpha = 1 + n \left[ \sum \ln \frac{x_i}{x_{\min}} \right]^{-1}$$
(4)

 $\alpha$  means a assessable exponent, xmin is minimum value which power function distribution achieves. In Zipf's law, the exponent equals to a unit.

The Kolmogorov-Smirnov test modified by A. Clauset (Clauset et al., 2009), allows to obtain maximum compliance: city size complies with the power law distribution. The main idea consists in maximum distance ( $\delta$ ) between data on cumulative density function and the model:

$$\delta = \max f(x) - g(x)|_{,x} > x_{mim}$$
 (5)

where f(x) is cumulative function of synthetic data with the value not less than  $x_{min}$ , and g(x) is cumulative function of power law distribution which complies with the condition  $x > x_{min}$  in the best way.

Zipf's law or Pareto distribution is an expression of the power law. In the empirical literature for assessment of the power law function exponent the method of least squares is used. Its advantage consists in the fact, that it provides visual criteria according to the law. For large-scale samples such as financial data it is precise enough (Gabaix & Ioannides, 2004):

$$\ln \operatorname{rank} = A - \operatorname{Kln \, size}$$
 (6)

Where:

In<sub>rank</sub> is a logarithm of city rank;

In<sub>size</sub> is city population;

K is a parameter of distribution, Zipf's rating which inclines linear dependence between city size and city rank. Zipf's law is observed subject to K=1, that is the largest city  $\kappa$  times as large as  $\kappa$ -th city by size. At K<1 sizes of large cities more predicted by Zipf's law; at K>1 distribution of cities of more uniformly predicted by Zipf's law.

During investigations using Zipf's law, determination of sampling size is important (Brakman, et al., 2009):

- to use fixes quantity of cities (for example, 100 cities per sampling);
- to determine threshold level of index (for example, cities with population more than 100 thous. people).

# 4. The data

For analysis of urban inequality using Zipf's law we used the data of the Federal State Statistics Service for 2014. The investigation included settlements with status of a city and population from 100 thous. people, that was imposed by absence of the data in the statistics digests for small towns. Testing of Zipf's law was carried out using the method of least squares.

# 5. Results

Let us classify Russia's cities by population as follows: small towns – population up to 20 thous. residents, medium-size cities – from 20 thous. to 100 thous. residents, big cities – from 100 thous. to 250 thous. resident, large cities – from 250 thous. to 1 mln. residents. Considering this classification let us analyze the RF cities in 2014 r. (Table 1)<sup>4</sup>.

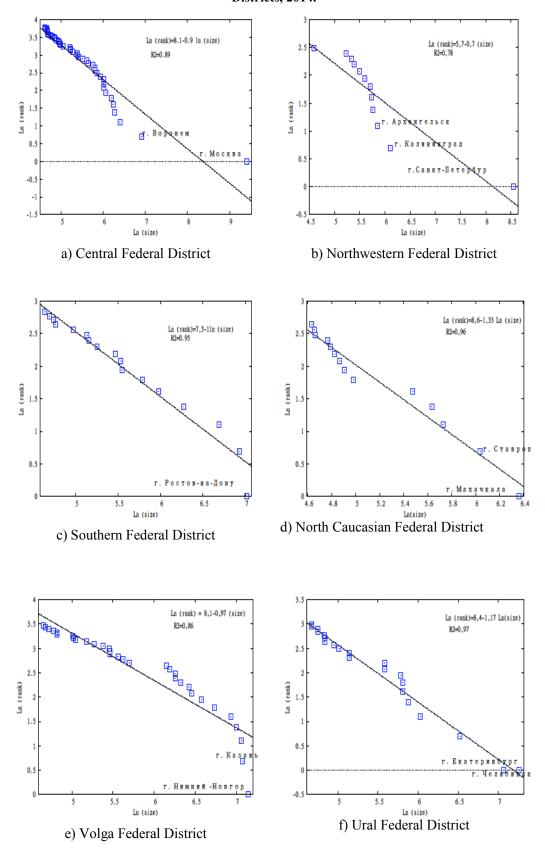
Item No.	Federal District	Small towns up to 20 thous. people		Medium-size towns, 20-100 thous. people		100-250 thous.		thous. people	
		qty,	Ratio,	qty,	Ratio,	qty,	Ratio,	qty,	Ratio,
		units	%	units	%	units	%	units	%
1	Central	139	45	124	40	27	9	17	6
2	Northwestern	84	56	53	36	4	3	7	5
3	Southern	19	24	43	53	9	12	8	11
4	North Caucasian	7	13	35	62	9	16	5	9
5	Volga	71	36	95	47	15	8	17	9
6	Ural	32	23	86	62	11	8	10	7
7	Siberian	44	34	65	50	11	8	10	8
8	Far Eastern	30	46	26	39	6	9	4	6
9	Total over Russia	426	38	527	47	92	8	78	7

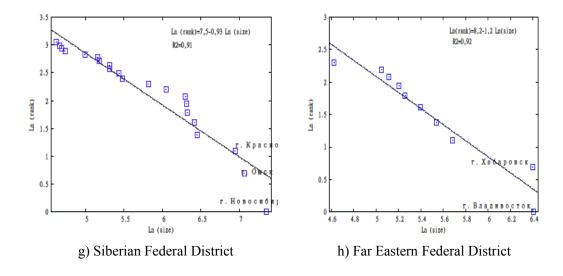
Table 1 Classification of Russia's cities by population in 2014.

Within the Russia's territory a majority of cities is of medium size. In a number of districts small towns prevail: in the Central Federal District -45% of total number, in the Northwestern Federal District -56%. The analysis results show that on average the number of small and medium towns is five times the number of big and large cities.

The diagrams reflecting display of the rank-size rule (Zipf's law) by the index of population in the cities at the regional level of Russia are given in Figure 1<sup>5</sup>.

Figure 1. Rank-size relationship by population index, calculated for Russia's cities by Federal Districts, 2014.





Legend: + - observed; - predicted.

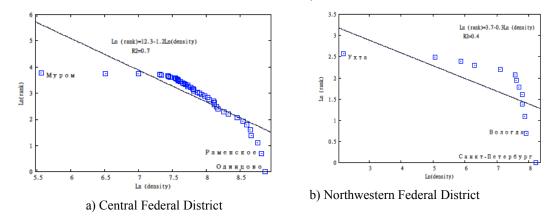
In the Central Federal District the population distribution corresponds to Zipf's law except for six largest cities (490,500-1,014,600 people). As the diagram shows, Moscow is a "primate city", where 52 % of sampling population is concentrated. In the Northwestern Federal District Zipf's law is not fulfilled: the estimated data coincide with linear ones in Murmansk (299.1 thous. people, Petrozavodsk (272.1 thous. people), Ukhta (99.2); Saint Petersburg – "primate city" within the Northwestern Federal District. The estimated data show validity of Zipf's law within the territory of the Southern Federal District. Rostov-on-Don is an exception: its population is lower than one predicted nu Zipf's law. The analysis of city sizes in the North Caucasian Federal District revealed the deviation from Zipf's law. It is particularly remarkable that city sizes are more uniform than they are predicted by the law.

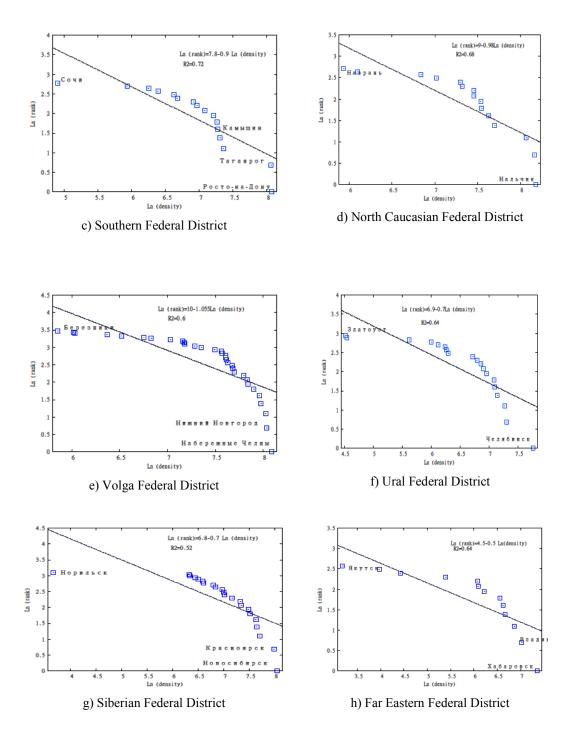
Within the territory of the Volga Federal District large cities from 1,169,200 to 1,263,900 people are exception to Zipf's law. (Nizhny Novgorod, Kazan, Samara).

The Ural Federal District has a "primate city" that is Yekaterinburg. Zipf's factor obtained at calculations is more than a unit that is indicative of more uniform city sizes, than they are determined by Zipf's law. In the Siberian Federal District Novosibirsk (1,547,900 people), Omsk (1,166,100 people) do not fall within the scope of Zipf's law. Within the territory of the Far Eastern Federal District the estimated data are lower than predicted ones in Vladivostok, Artem, and higher than predicted ones in Khabarovsk.

The population density analysis results for the cities with population more than 100 thous. people using Zipf's law are given in Figure 2 <sup>6</sup>.

Figure 2. Rank-size relationship by population density index, calculated for Russia's cities by Federal Districts, 2014.





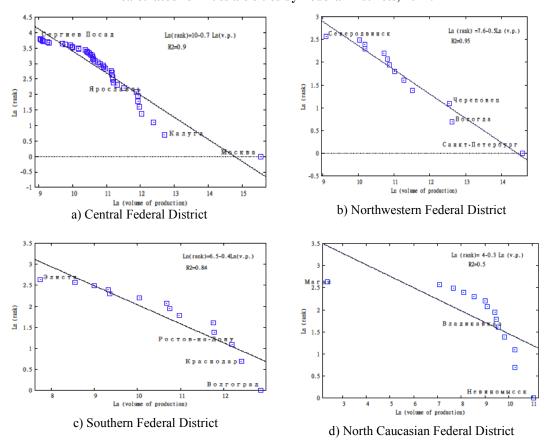
Legend: + - observed; - predicted.

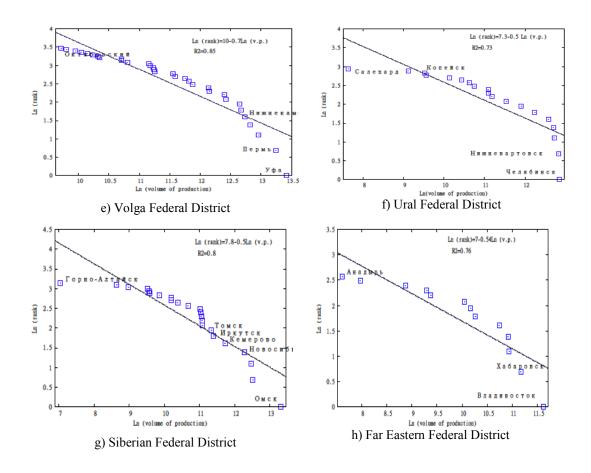
The city with maximum population density is located within the territory of the Central Federal District (Odintsovo, the Moscow Region), in the Central Federal District distribution of urban population partially corresponds to Zipf's law. According to the diagram Zipf's law is not applied to the cities with low population density (261.4 per km²– 1098.6 per km²) and cities of the Moscow Region, population density of which is higher than 5724 per km². It is worthy of note that the regional centers and Moscow fall within the scope of the "rank-population density" rule within the boundaries of the Federal District. In the Northwestern Federal District relation between city population density in and rank is weak (R²=0.4). Zipf's factor has low value (0.3), that is indicative of high differentiation of the analyzed index, inequality of population concentration in the cities within the boundaries of the Federal District. In the Southern Federal District the "rank-population density" relationship is moderate. Incompliance with Zipf's law is observed in Sochi, Taganrog, Rostov-on-Don. In Novorossiysk and Kamyshin the observed values correspond to the predicted ones. In the

North Caucasian Federal District Zipf's factor is maximum close to 1, that is indicative of uniform population concentration in the cities. The cities with low density, such as Magas, Nazran; with high density, such as Nalchik, Khasavyurt may be an exception. In the Volga Federal District the Zipf's law validity is observed in the cities the density range of which is within 583.9-2,855.7 people per km<sup>2</sup>. In the Ural Federal District the "rank-population density" relationship is moderate. The diagram shows that Zlatoust, Miass (low population density), Chelyabinsk, Pervouralsk (high population density) do not fall within the scope of Zipf's law. The obtained data are indicative of nonuniform distribution of population within the Ural Federal District boundaries. The analysis of Zipf's law in the cities of the Siberian Federal District shows wide differentiation of the analyzed index. Population density of Novosibirsk (the leader in the index within the Siberian Federal District boundaries – 3,095.8 per km<sup>2</sup>) exceeds the lowest index several dozens times (Norilsk – 39.3 people per km<sup>2</sup>). Within the territory of the Far Eastern Federal District nonuniform concentration of city population is observed that confirms low Zipf's factor of 0.48. The estimated data coincide with those predicted in Magadan, Artem. In Khabarovsk and Vladivostok (the highest population density) 30 % of city population of the Federal District is concentrated.

The volume of production of goods and services is an important economic index of the city. Figure 3 shows the results of the analysis of this index within the boundaries of the Federal Districts of the Russian Federation<sup>7</sup>.

Figure 3. Rank-size relationship by index of volume of production of goods and services, calculated for Russia's cities by Federal Districts, 2014.





Legend: + - observed; - predicted.

The Central Federal District displays strong relation of "rank-volume of production of goods and services". As the diagram shows, high differentiation on the analyzed index is in Moscow ("primate city"). In Serpukhov, Novomoskovsk, Yaroslavl the estimated data coincide with predicted ones. The Northwestern Federal District shows the similar situation: the production volume in Saint Petersburg ("primatecity") is widely differentiated within the Federal District. As the diagram shows, in Pskov, Naryan-Mar, Cherepovets the estimated data coincide with the predicted ones. The analysis of the sampling of the Southern Federal District cities showed relative uniformity of the production volume in the cities: the estimated data coincide with those predicted in Rostov-on-Don, Kamyshin, Maikop. Within the territory of the North Caucasian Federal District Zipf's law is not fulfilled, and wide differentiation of the analyzed index is observed. In Vladikavkaz the estimated data coincide with the predicted ones. Taking into account the analysis results, it is possible to conclude that uniformity by the analyzed index is observed in the Volga Federal District, according to R<sup>2</sup> =0.85 the "rankproduction volume" relation is strong. The estimated data coincide with those predicted in Nizhnekamsk, Syzran. In the Ural Federal District own production in the cities is nonuniform: Zipf's factor is 0.47. The diagram shows wide differentiation of the index within the boundaries of the Federal District; in Kopeisk and Kamensk-Uralski the estimated data coincide with the predicted ones. In the Siberian Federal District the volume of production in the cities is nonuniform: the estimated data coincide with the ones predicted in Biysk, Novosibirsk, Irkutsk, Kemerovo. In the Far Eastern Federal District low value of Zipf's factor allows to conclude that the production is concentrated in separate cities and wide differentiation of the analyzed index takes place. The estimated data coincide with those predicted in Magadan, Komsomolsk-on-Amur.

By the index of population and volume of production of goods and services the following "primate cities" are determined: Moscow and Saint Petersburg, Yekaterinburg (population), which allows to draw a conclusion on their dominance in urban system and high differentiation of cities by these indices.

Summary statistics of the obtained investigation results is given in Table 2.

Table 2 Results of Zipf's law check in the Federal Districts of the Russian Federation in 2014.

	Quantity of										
Federal District	observations in a federal district, units	Minimum index value	Maximum index value	Zipf's factor for district cities							
Population, thous. people											
Central	44	101.9	12,108.3	- 0.9							
Northwestern	13	99.2	5,132	- 0.7							
Southern	16	104	1,109.8	- 1							
North Caucasian	15	103.1	578	- 1.33							
Volga	31	105.1	1,096.7	- 0.97							
Ural	20	107.5	1,169.4	- 1.17							
Siberian	22	101.7	1,547.9	- 0.93							
Far Eastern	13	102.4	603.2	- 1.2							
Population density, people per km <sup>2</sup>											
Central	41	261.4	7165	-1.2							
Northwestern	13	9.1	3,657.8	-0.3							
Southern	16	135	3,180	-0.8							
North Caucasian	15	377.5	3,558.9	-0.98							
Volga	31	349.5	3,244.3	-1.1							
Ural	19	92.1	2,334.6	-0.75							
Siberian	22	39.3	3,095.8	-0.7							
Far Eastern	13	24.1	155.7	-0.48							
Volume of production of goods and services, mln. RUB											
Central	44	8,521.7	5,653126	-0.7							
Northwestern	13	9,030.2	2,160,129	-0.5							
Southern	14	2,340.6	8,060.2	-0.4							
North Caucasian	14	10.8	6,1834.6	-0.26							
Volga	32	16,603	668,254.5	-0.7							
Ural	19	2,089.7	376,257.3	-0.47							
Siberian	23	1,151.5	601,058.3	-0.5							
Far Eastern	13	2,040.6	110,497	-0.5							

We investigated the displays of Zipf's law by the indices of population, population density, volume of production of goods and services within the boundaries of the RF federal districts. This allowed us to determine the following aspects. At large, Zipf's law displays over the entire territory of Russia. By the population index in the federal districts, Zipf's factor varies

within the range from -0.7 (Northwestern Federal District) to -1.33 (North Caucasian Federal District).

As a result of the performed analysis of the Russia's cities by the population index, Zipf's factor is within the range from -0.3 (Northwestern Federal District) to -1.2 (Central Federal District). The obtained data allow to conclude that the urban population is distributed unifromly within the territory of the North Caucasian Federal District. Analysis of the volume of production of goods and services determined the range of Zipf's factor from -0.26 (North Caucasian Federal District) to -0.7 (Central and Volga Federal Districts).

## 6. Conclusion

The analysis performed allows to draw the following conclusions: The processes of spatial concentration of economic activity take place in Russia. Availability of primate cities by the indices of population and volume of production of goods and services (Moscow, Saint Petersburg, Yekaterinburg (population)) highlights a worrying situation in social and economic space of the Russian Federation. Remarkable separation of Moscow and Saint Petersburg from the subsequent group of cities and high population density of the Moscow Region determine them as centers of interregional attraction. The obtained empirical assessment proves that Russia has no intermediate group of cities macroregional centers. The results of the investigation can be used for creation of methodological tools to develop the mechanisms of smoothing of interregional differentiation, program of economic and social development of cities.

#### **ENDNOTES**

2–7 Regions of Russia. Basic social and economic indices of cities 2014 // Federal State Statistics Service. [Electronic resource]

URL:http://www.gks.ru/wps/wcm/connect/rosstat\_main/rosstat/ru/statistics/publications/ca talog/doc 1138631758656 (accessed date 07.04.2016).

# **REFERENCES**

- Alperovich, Gershon. (1984). "The Size Distribution of Cities: On the Empirical Validity of the Rank-size Rule. Journal of Urban Economics, 16.2, 232-39.
- Auerbachи Felix. (1913) "Das Gesetz Der Bevolkerungs konzentration." Petermanns Geographische Mitteilungen, 49, 73-71.
- Baker, Chris, and Pasuk Phongpaichit. (2009). "A History of Thailand."
- Blank, Aharon, and Sorin Solomon. (2000) "Power Laws in Cities Population, Financial Markets and Internet Sites (scaling in Systems with a Variable Number of Components). "Physica A: Statistical Mechanics and Its Applications 287, 1-2, 279-88.
- Brakman, Steven, Harry Garretsen, Charles Van. Marrewijk, and Steven Brakman. (2009) The New Introduction to Geographical Economics. Cambridge, UK: Cambridge UP, 568.
- Clauset, Aaron, Cosma Rohilla Shalizi, and M. E. J. Newman. (2009)"Power-Law Distributions in Empirical Data." SIAM Rev. SIAM Review 51.4, 661-703.
- Córdoba, Juan-Carlos. (2008) "On the Distribution of City Sizes." Journal of Urban Economics, 63.1 177-97. Web.
- Eaton, Jonathan, and Zvi Eckstein. (1997) "Cities and Growth: Theory and Evidence from France and Japan." Regional Science and Urban Economics 27.4-5, 443-74.
- Flux, A. W., and Vilfredo Pareto. (1896) "Cours D'Economie Politique. Tome Premier." The Economic Journal 6.22, 249.
- Gabaix, X. (1999) "Zipf's Law for Cities: An Explanation." The Quarterly Journal of Economics 114.3,739-67.

- Gabaix, Xavier, and Yannis M. Ioannides. (2004) "Chapter 53 The Evolution of City Size Distributions." Handbook of Regional and Urban Economics Cities and Geography, 2341-378.
- Goldstein, M. L., S. A. Morris, and G. G. Yen. (2004) "Problems with Fitting to the Power-law Distribution." The European Physical Journal B Eur. Phys. J. B 41.2,255-58...
- Jiang, Bin, and Tao Jia. (2011)"Zipf's Law for All the Natural Cities in the United States: A Geospatial Perspective." International Journal of Geographical Information Science 25.8, 1269-281.
- Kolomak Ye. A. (2014) Development of urban system of Russia: tendencies and factors. "Problems of Economics", 10, 82-96.
- Naldi, M. (2003) "Concentration Indices and Zipf's Law." Economics Letters 78.3, 329-34.
- Newman, Mej. (2005) "Power Laws, Pareto Distributions and Zipf's Law." Contemporary Physics 46.5,323-51.
- Nitsch, Volker. (2005) "Zipf Zipped." Journal of Urban Economics 57.1, 86-100.
- Ramsden, J.j., and Gy. Kiss-haypal. (2000) «Company Size Distribution in Different Countries.» Physika A: Statistical Mechanics and Its Applications 277.1 -2, 220-27.
- Rosen, Kenneth T., and Mitchel Resnick. (1980) "The Size Distribution of Cities: An Examination of the Pareto Law and Primacy." Journal of Urban Economics 8.2, 165-86.
- Soo, Kwok Tong. (2005) "Zipf's Law for Cities: A Cross-country Investigation." Regional Science and Urban Economics 35.3, 239-63.