2019

IARIW-HSE

Special IARIW-HSE Conference "Experiences and Future Challenges in Measuring Income and Wealth in CIS Countries and Eastern Europe" Moscow, Russia, September 17-18, 2019

Issues of Inequality Representation in Conditions of Deep Social Stratification

Victor Kapitanov

Anna Ivanova

Alexandra Maximova

Paper Prepared for the IARIW-HSE Conference
Moscow, Russia, September 17-18, 2019
Session 3B: Income Inequality
Time: 13:45 – 15:45, September 17

Issues of inequality representation

in conditions of deep social stratification

Kapitanov V.A.[1] kapitanov_v_a@mail.ru

Ivanova A.A.[2] ivanova.iamm@mail.ru

Maximova A.Yu.[2] maximova.alexandra@mail.ru

[1]JSC "Research Institute" Polyus" them. MF Stelmakh", Moscow

[2]SI "Institute of Applied Mathematics and Mechanics", Donetsk

Abstract. The aim is a criticism of generally accepted methods of graphical and numerical representation of inequality based on real, i.e. incomplete and distorted data, and a search for alternatives to generally accepted methods.

Problems with the presentation of inequality are caused by the specific incompleteness of data on incomes.

According to Rosstat, the total income of Russians in 2015 was 53.101 trillion rubles, and according to the Federal Tax Service (FTS) - only 22.053 trillion rubles. More recent data cannot be given, since at the moment the last published FTS "Report on the declaration of income by individuals" dates back to 2015.

The principal difference between the Rosstat data and the FTS data is that the Rosstat estimates anonymous incomes without binding them to the individuals of specific recipients. The FTS, by virtue of the need to fulfill fiscal duties, determines incomes with reference to specific recipients, from which the FTS collects taxes.

Therefore, it is very easy to hide real income, since hidden data on income is one and a half times more than opened.

The tools for concealing incomes are well known - a formal withdrawal abroad of a business located de facto in Russia. However, such tools are available only to big business, so, the information about incomes of only rich cohorts is hidden.

So the data on incomes are characterized by significant incompleteness. Moreover, this incompleteness is specific - as higher an income, as less information about them. Two important requirements for methods of inequality representation follow from this lack of data.

1. No data source can be considered trustworthy. Only the coincidence of data on the distribution of the population by income from several independent sources can give us the right to consider the information we have reliable. Since this kind of data is quantized in publication, the methods of representing inequality must be insensitive to quantization. This is nesses to compare the data obtained from different sources.

2. Methods of presenting inequality must be unconditionally sensitive to the width of the income range for any arbitrarily small filling of rich cohorts.

Such methods are developed in this work.

Introduction. The question of the real distribution of people by income level or the economic structure of society (ESS) is important because, having adequate data on the distribution of

people by income, the researcher can draw conclusions not distorted by any ideological blinders about the real mechanism of movement of goods between people and the real reason wealth and poverty.

The relevance of this work is due to the need to develop methods for presenting inequality in conditions of the absence of some information (most often about the incomes of wealthy members of society), since complete data on the distribution of the population by income are not available and it is not possible to obtain them.

According to Rosstat, the total income of Russians in 2015 amounted to 53.101 trillion RUB. [1], and according to the Federal Tax Service (FTS) it is only 22.053 trillion RUB [2]. The difference between Rosstat data and the Federal Tax Service data is that Rosstat estimates anonymous income without linking it to the personalities of specific recipients [3], while the Federal Tax Service determines the revenue with reference to specific recipients, from which the Federal Tax Service levies taxes.

From a comparison of the two numbers, it follows that it is very easy to hide incomes, since hidden incomes are one and a half times more than undisguised ones.

Tools to conceal income are not a secret - in the context of globalization, a formal withdrawal abroad of a business located de facto in Russia is not difficult: for example, a site offering similar services [4]. But such tools are available only to large businesses, that is, information only about the income of rich cohorts is hidden.

So, the data on income inequality at our disposal are characterized by significant incompleteness, and the incompleteness is rather specific - the higher the incomes, the less information about them. From here two important requirements to methods of representation of an inequality follow.

No source of data can be considered credible, and only the coincidence of data on the distribution of the population by income from several independent sources will allow us to consider the available information reliable. Since this kind of data is published quantized, divided into cohorts, the studied indices describing the inequality must be insensitive to quantization and computable for any quantization - otherwise it will not be possible to compare data obtained from various sources.

The incomes of certain rich and even the richest citizens of the country are known, published by Forbes [5] and the Federal Tax Service [6], but these data are insufficient for a complete description of the existing ESS in Russia, since the number of rich people, who concealed their income and were not presented in sight “Forbes” and the Federal Tax Service, is unknown. In this situation, we have to admit that we have access to information about the width of the income range (determined by the income of the richest person in the country), but nothing more. There is no reliable data on the content of rich cohorts; there is only data on how far to the right on the income scale rich cohorts are located. Therefore, the methods of representing inequality that we propose must be unconditionally sensitive to the width of the income range for any arbitrarily small filling of rich cohorts.

1. Polygons of relative distribution frequencies and the impossibility of comparing data from several sources

The polygon of distribution of relative frequencies of indicators is a broken line connecting the midpoints of the upper sides of the histogram columns. Frequency polygons are widespread as a method of graphical presentation of data on inequality, and this is not an accident - they are very demonstrative, have been used for a long time and widely [7,8] and allow you to visually compare the real ESS and some still existing ideas about social inequality (Fig. 1)

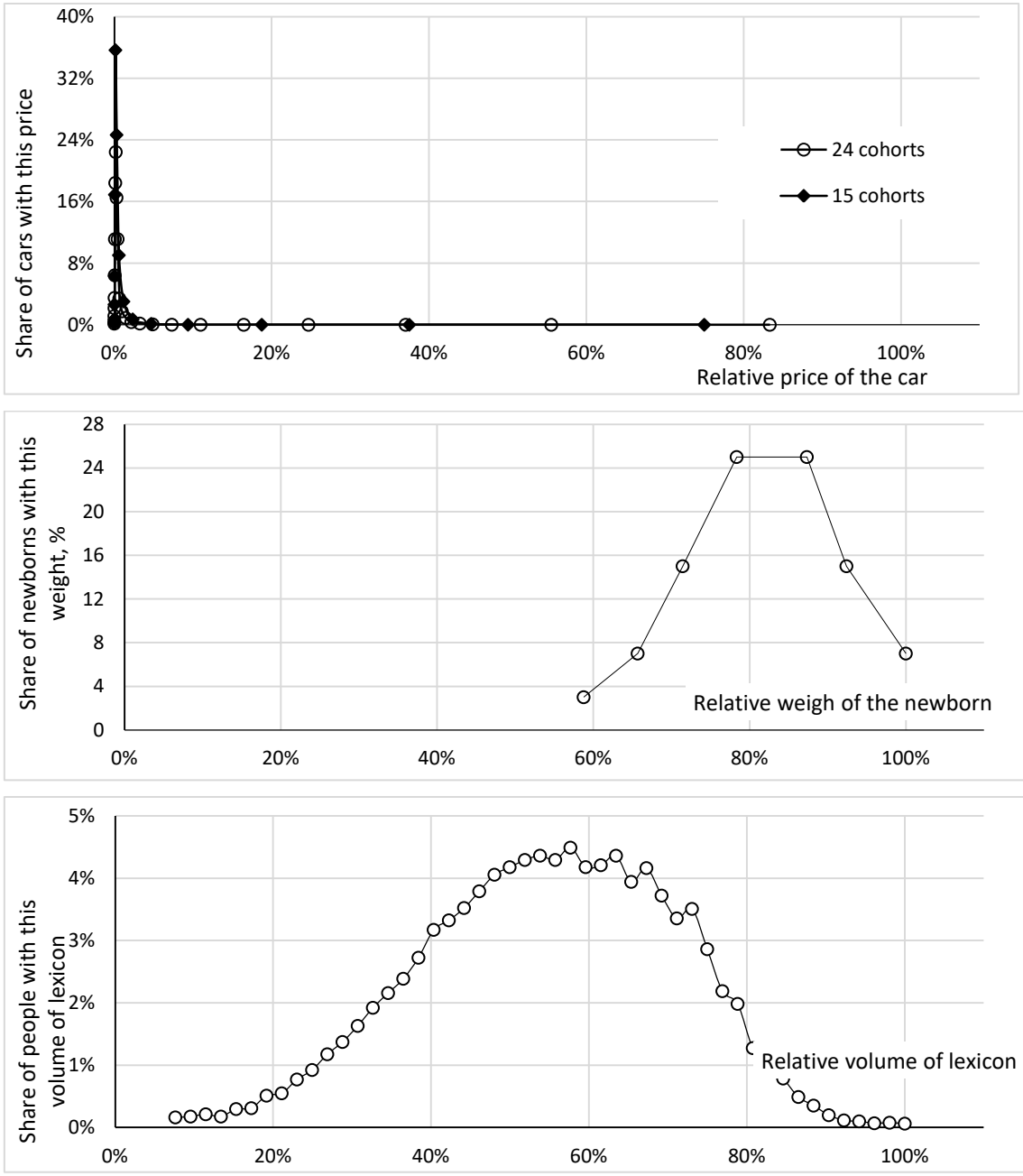


Figure 1. a - distribution of people by income, estimated by the cost of cars (with different quantization) [9]; b - distribution of newborns by weight [10]; c - distribution of people according to volume of lexicon (Russian language) [11, 12].

The type of frequency polygons representing the economic structure of society can confirm or refute the idea of the reasons for the emergence and intensification of inequality.

Unscientific notions of the causes of social inequality are widespread; modern society has inherited these misconceptions from the reactionary idealistic works of some thinkers of the past.

For example, here is an account of the foundations of social inequality as presented by the famous Russian philosopher I.A. Ilyin: “If people were truly equal, life would be extremely simple, and justice would be extremely easy to find and introduce into everyday life. It would only be worth saying: to the same people - the same share; or - equally to all of them ... In fact, people are not equal in body or soul or spirit” [13].

In the most odious versions, this belief system justifies inequality with “good genes” or “noble origin”, as in the works of the religious and political philosopher N. A. Berdyaev: “The existence of a “white bone” is not only an estate prejudice, it is also an irrefutable and indestructible anthropological fact ... The aristocracy was created by God and received its qualities from God ... No social revolution can destroy the qualitative advantages of a race” [14].

If the social-racist belief system reflected reality, then the income distribution of the population would be normal. However, people are really born (Fig. 1b) and live unequally, including intellectually (Fig. 1c), but only this inequality does not at all resemble that given in Fig. 1a income inequality.

In economic reality, nothing like the normal distribution is observed. The income distribution of the population is lognormal, with its maximum shifted so much to the left, towards the poorest cohorts, that it merges with the ordinate axis.

In [15], we showed that the cause of wealth and poverty is a rank exchange, i.e. systematic practice of non-equivalent transactions made by unequal partners. Transactions in which a rich counterparty shifts the market price in his favor, and the more wealth he has the more effective he acts. It is the rank exchange that gives the characteristic lognormal distribution with a “heavy tail”, presented in Fig. 1a.

But despite the indisputable advantages noted - first of all, the simplicity and visibility of identifying the causes of social inequality - frequency polygons are not without drawbacks due to the need to use only quantized data.

Therefore, it is not possible to compare frequency polygons from several sources with different splitting into cohorts on one diagram - the height of each column of the histogram is unambiguously related to the width of the cohort corresponding to this column, because the wider the income range of a given cohort, the greater the likelihood of a person to get into this cohort.

It is easy to see (see Fig. 1a) that a less detailed quantization gives a greater peak height of about one and a half times. Thus, for the same data, it is possible to construct radically different frequency polygons by changing the method of quantization of this data. Moreover, frequency polygons are not able to provide comparability of data from various sources.

2. Lorentz Curves

Fig. 2 shows visually indistinguishable Lorentz curves constructed according to Rosstat data (for a company with a maximum income of 96 thousand rubles / month) and according to a summary of Rosstat and Forbes (for a company with a maximum income of 14.3 billion RUB/month). The technique of compound data arrays is described in [10]). The addition of data to income information for the richest did not affect the shape of the Lorentz curve.

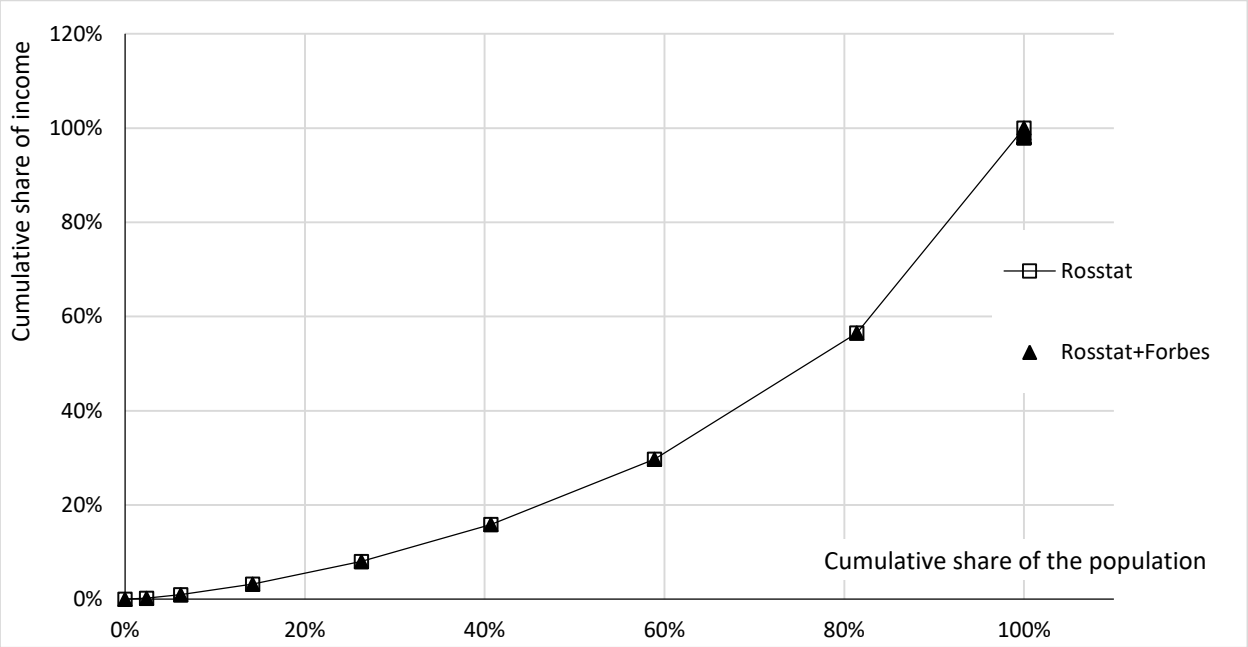
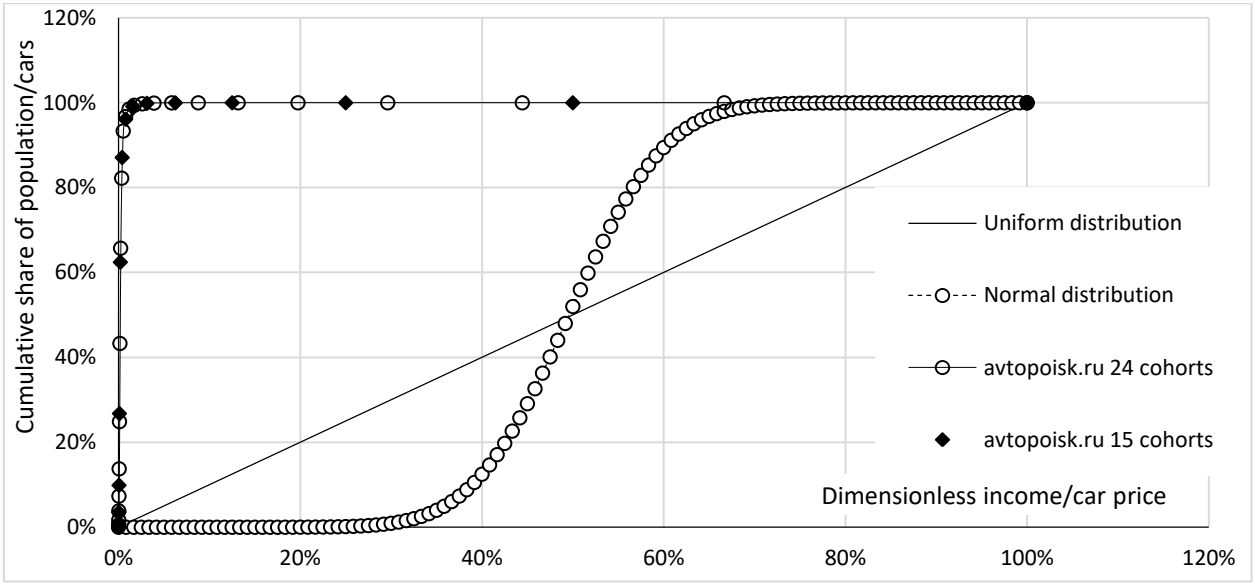


Figure 2. Lorentz curves. Rosstat data, summary data of Rosstat and Forbes [9].

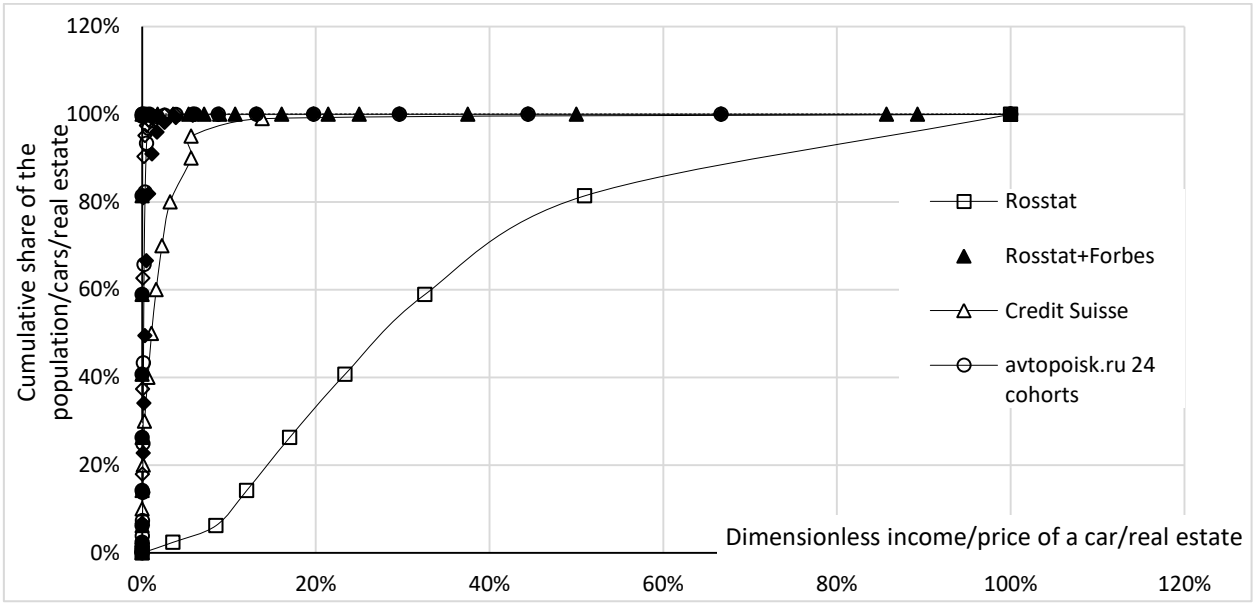
The fundamental problem with the use of Lorentz curves lies in the fact that economists do not have and will not have complete data on rich cohorts for the same reason as why the state does not have these data.

3. Distribution Functions

Distribution functions (Fig. 3) are already used as a tool for representing inequality [16], although they are not yet as popular as Lorentz curves or frequency polygons.



а



б

Figure 3 [9]. a - distribution functions for different quantization of data, as well as lognormal (real ESO), normal and uniform distribution; b - comparison of data on inequality in the Russian Federation from several independent sources, incl. with incomplete data on rich cohorts.

From fig. 3a it follows that differences in the quantization of data (avtopoisk.ru, 15 and 24 cohorts) do not affect the course of the curves, and therefore can be used to compare data from a variety of sources for any division into cohorts, and even in the absence thereof.

The distribution functions are clear. In fig. 3a visually, without any quantitative analysis, the fictitious distribution of the population by income (uniform and normal) and the real lognormal distribution based on car prices are perfectly distinguished.

The data from Rosstat and Rosstat + Forbes, thanks to the distribution functions, are easily visually distinguishable. Therefore, they clearly show the difference between data in which the maximum income is 96 thousand RUB/month. and data with a maximum income of 14.3 billion RUB/month. It should be noted that both the Rosstat data and the Rosstat + Forbes data do not contain complete information about the incomes of the richest members of society.

4. Numerical indices of inequality

Calculation of inequality indices (table) showed that the coefficients of the funds (decile, quintile, etc.) are not computable with arbitrary quantization, and therefore unsuitable for comparing data from sources differing in quantization (table, column 3).

The Gini index also does not provide an adequate representation of the degree of inequality, since its value does not respond at all to adding data on subjects that differ significantly in the maximum income (table, Column 4, rows 1 and 2).

The value of the Pareto indicator α is not unambiguously associated with inequality - the Pareto indicator is higher for Rosstat data that are obviously incomplete and characterized by underestimated inequality than for data on inequality determined by car prices (table, Column 5, rows 1, 4, 5, 6).

The first three indicators of the total entropy of GE (table, Column 6, 7, 8, row 9), as well as the ratio of the maximum income (wealth) to the modal Imax/mod (tab., Column 10, row 9), the most of all the considered inequality criteria depend on the quantization of the data, and therefore are not suitable for comparing data differing in quantization.

The intensity of the exponential distribution λ does not have the listed drawbacks - it is definitely greater with greater inequality, it is computable with arbitrary quantization and depends rather weakly on it, it is sensitive to the width of the income range and does not require completely complete data on rich cohorts.

Table. Comparison of inequality indices.

# #	Data	Inequality index								R^2 exponential approximation
		Fund ratio	Gini	α	GE(0)	GE(1)	GE(2)	Imax /mod	λ	
1	2	3	4	5	6	7	8	9	10	11
1	Rosstat	15,6[1]	0,38	1,36	0,187	0,241	0,256	8,03	2,43	0,9360
2	Rosstat + Forbes	Calculation impossible due to lack of decile quantization	0,38	1,54	0,283	0,451	2584	1,19E6	4,85E5	0,9775
3	Rosstat + FNS		0,48	1,38	0,436	1,284	4340	1,19E6	2,63E5	0,9706
4	avtopoisk.ru 24 cohorts		0,46	1,22	0,415	0,438	0,098	680	418	0,9915

[1] Adopted according to Rosstat, but the calculation of the ratio of funds on the published Rosstat ESS is impossible.

5	avtopoisk.ru 15 cohorts		0,48	1,25	0,468	0,491	1,125	589	434	0,9917
6	avito.ru cars		0,53	1,26	-	-	-	1290	217	0,9990
7	avito.ru real estate		0,55	1,37	-	-	-	1860	1010	0,9918
8	Credit Suisse	∞	0,91	1,4	-	-	-	-	60,8	0,9271
9	Discrepancy 4 and 5 lines	-	4,2%	2,4%	11,21%	10,85%	91,31%	14,33%	3,7%	-

5. Some limitations of the obtained results

The study of the behavior of the exponential distribution [17] on model examples showed that:

- only ESSs with high social inequality are well described by the exponential distribution function (the data of Rosstat and Credit Suisse, characterized by low inequality, see Fig. 3b are worst approximated by the exponential dependence, see table column 11);

- the values of the intensity of the exponential distribution can be used as an indicator of incomplete data (for $\lambda < 10$, we must admit that we do not have a complete description of the ESS);

Among the drawbacks of the distribution intensity, it is also possible to note some of its non-obviousness and computational complexity (the latter is not a problem with the modern development of computer technology), therefore, for a quick and clear, although more rough estimate of the inequality, you can use the Imax / mod index.

In addition, the distribution intensity is not additive, i.e. cannot be decomposed without a remainder into the sum of inequalities within individual groups and intergroup inequality [18], as, incidentally, the universally used Gini index. Therefore, if it is necessary to decompose inequalities (but only if unquantified data are available), it is advisable to use indicators of total entropy, including the Theil index.

In all other cases, it is better to prefer the intensity of the exponential distribution as the index of inequality.

However, it does not follow from what has been said that universally accepted indices for representing inequality should be unconditionally and forever abandoned.

In cases where the inequality is small and the data are complete (for example, when describing the inequality of the countries of the world in terms of per capita GDP), it is necessary to use the coefficient of funds, the Gini index etc.

In addition, the socio-economic sciences have already accumulated long-term trends of inequality, expressed using the Gini index or (less commonly) other generally accepted coefficients. And if the purpose of the study is to study the dynamics of inequality (it is the study of dynamics, not cross-country comparison, for example, it is impossible with the help of generally accepted indices), then the calculation of the Gini index in order to prolongate the trend is justified.

References

1. Federal Service of State Statistics. "Russia in numbers - 2016". "2. 7.6. Monetary incomes and relative density of expenses in monetary incomes of the population (as a percentage of the total volume). " Access mode http://www.gks.ru/bgd/regl/b16_11/IssWWW.exe/Stg/d01/07-06.doc.

2. Federal Tax Service of Russia. "Report on the declaration of income by individuals. Section 1". 2015 Access mode https://www.nalog.ru/rn77/related_activities/statistics_and_analytics/forms/#t1

3. "Information on the methodology for calculating the indicators used to monitor the implementation of instructions contained in the Presidential Decree No. 596-606 of May 7, 2012". P. 42. "Real disposable money incomes of the population (percent)". Access mode http://www.gks.ru/metod/metodika.htm

4. SAA Finance. "Registration of companies. Residence permit & residency & Citizenship." Access mode http://saa-finance.com

5. 200 richest businessmen of Russia. 2015 Forbes. Access mode http://www.forbes.ru/rating/200-bogateishikh-biznesmenov-rossii-2015/2015#all_rating.

6. Federal Service of State Statistics. "Russia in numbers - 2016". "7.9. Distribution of population by the value of per capita monetary incomes (as a percentage of the total). Access mode http://www.gks.ru/bgd/regl/b16_11/Main.htm, свободный.

7. *Davis H. T.* Political statistic. Evanston, Illinois: The Principia Press of Illinois; 1954: 365 c. https://babel.hathitrust.org/cgi/pt?id=mdp.39015065433917;view=1up;seq=386

8. *Kislitsyna O.A.* Inequality in the distribution of income and health in modern Russia. M .: RIC ISEPN. 2005. 376 p. P. 37. Access mode http://kislitsyna.ru/data/files/inequality/Glava2.pdf

9. *Kapitanov V.A., Ivanova A.A., Maksimova A.Yu.* Advantages of the distribution function as a method of graphical representation of the economic structure of society. "Statistics and economics". Volume 15, No. 1 (2018), p. 4-16. Access mode https://statecon.rea.ru/jour/article/view/1188

10. *Kuzmichev Yu.G., Orlova M.I., Burova O.N., Gurenko S.P., Lazareva E.P.* Assessment tables for the physical development of full-term newborns in the city of Nizhny Novgorod. Doctor-graduate student. 2013, t. 59, No. 4.3, p. 494-498. Access mode http://vrach-aspirant.ru/articles/pediatrics/13334/

11. “How many words do you know?” Access mode http://myvocab.info

12. *Golovin G.V.* Measuring the passive vocabulary of the Russian language. 2015 Socio- and psycholinguistic research. 2015. Issue. 3, p. 148-159. Access mode http://splr.psu.ru/wp-content/uploads/2015/12/Головин_2015.pdf

13. *Ilyin I.A.* I look into life. Book of reflection. M .: "Eksmo"; 2007: 528 p.

14. *Berdyaev N.A.* Philosophy of inequality. M .: Institute of Russian Civilization; 2012: 624 p.

15. *Kapitanov V.A., Ivanova A.A.* Inequality and rank exchange. Russia: trends and development prospects. RAS INION: Yearbook. 2017, no. 12, p. 1, p. 661-668.

16. *Romanovsky M.Yu., Romanovsky Yu.M.* Introduction to econophysics: statistical and dynamic models. - Izd.2-e, corrected. and additional .- M.-Izhevsk: Institute for Computer Research, 2012. -340 p.

17. *Kapitanov V.A., Ivanova A.A., Maksimova A.Yu.* Problems of numerical estimates of inequality. Statistics and Economics. 2018; 15 (4): 4-15. Access mode https://doi.org/10.21686/2500-3925-2018-4-4-15