Study of russian households income differentiation with various inequality measures

V P Sirotin\textsuperscript{1}, K V Alexeeva\textsuperscript{2}

Faculty of Economic Science, Higher School of Economics, 11 Pokrovsky Bulvar, Moscow, Russia

E-mail: \textsuperscript{1}vsirotin@hse.ru, \textsuperscript{2}ksenia.v.alexeeva@gmail.com

Abstract. The article reviews modern income inequality measure such that Gini index and its modification, Zenga index and Davydov, their advantages and disadvantages. The reviewed indices are applied to the household’s income data for different regions in the Russian Federation from 2014 to 2018. The indices studied showed different level of income inequality in regions compared, thus the tendency is the same – income inequality decreases for all type of inequality studied. Russian regions become more similar in terms of income inequality, while income inequality itself tends to decrease in most regions.

1. Introduction

Income inequality is one of the key issues discussed nowadays. The reasons for such attention to income inequality is that the value represents a measure of risk tension as excessive inequality or special unexplored forms of it are the factors of social development risk. Moreover, there is a variety of negative socio-economic consequences that income inequality exceeding the level economically justified and acceptable from the point of view of social justice entails. These negative consequences are not limited only to a low level of satisfaction of material goods and consumption of basic social services by the least well-off strata of the population – the concept of poverty varies significantly in economically developed and developing countries. Income inequality, often understood as social inequality and social stratification generates tension in society, strengthens separatist sentiments, and negatively affects the migration mobility of the population. Using only one method to evaluate income inequality may not be sufficient enough as not only the level of income inequality describes social development risks, but also the transformation of income inequality itself plays a great role. Therefore, several measurements should be used and compared in order to depict various changes in income inequality, which will describe the inequality level more accurately.

In modern Russia, the problem of income inequality, which is formed at the level of households and individuals, is exacerbated by inter-regional differences. Some of them are caused by climatic, resource, historical reasons. One of the most important determinant influencing the differences in income is modern regional economic policy. A priori, its goal is to revitalize the internal resources of regional development, create conditions for increasing the welfare of the population, and reducing interregional differences. The aim of the research is to review income inequality characteristics and study their values for Russian regions as well as to review the changes and tendencies in terms of income inequality. It is assumed that from 2014 to
2018 the income inequality in Russian regions has decreased due to two facts. The first one is that the economic crisis of 2014 has gone, and the economy emerged from it during these years. Secondly, special actions were taken by the Russian government in order to equalize Russian regions as a part of program “Russia – 2020” (Government of the Russian Federation. The concept of a long-term social period-economic development of the Russian Federation until 2020. URL: http://government.ru/info/6217/). Furthermore, it is assumed that the inequality level in Russian regions varies significantly due to geography and socio-economic features of the regions.

2. Review of literature on the topic of income inequality

When speaking about income inequality, we essentially mean the gap between incomes of poor population and the rich one. We can consider those differences by using absolute measure and relative one. And there are reasons for both as absolute measurements allow to understand the total difference in income between groups of population, thus it may complicate the comparison especially of changes in time as simple error of rounding for rich population may be several times the value of income of poor population. At the same time the relative measurement may show the changes in time, thus, it will not depict the scale, therefore, both types of comparisons should be used. There is empirical proof that income inequality has changed over time. One of the reasons for this change is globalization, which has influenced not only the level of income, job perspectives and everyday needs and availabilities, it has also imparted market rules as special organizations controlling for the production rate, environments protection and so on were created globally. Such organizations may impose certain restrictions on countries producing goods. Lakner C. found out that the gain in real income of households all over the world from 1988 to 2008 was the largest for people around the 50th percentile of the global distribution of income of households and among the richest population; at the same time the gain was the least for people of 80th percentile or the world income distribution of households which includes the lower middle class of the rich part of the world [1].

Scientists have established several methods to measure income inequality in the population. The most common measurement is the Gini coefficient [2]. Gini coefficient is just a numerical representation of the Lorenz curve [3]. It states the level of income inequality in the population. If Gini coefficient is close to one, this means that all income in population is concentrated in few rich individuals. At the same time, when Gini coefficient is close the zero this means that income is equally distributed among the population. The Gini coefficient is widely used by different organizations for example World Economic Review, World Bank, United Nations Development Program and in various research. The logic behind the widespread of Gini coefficient is quite simple – it is an easily computed measure as well as clearly interpreted and may be applied not only to economic sphere of life. At the same time, it is also quite obvious that the measure of inequality obtained by Gini coefficient may underestimate the real inequality of the population which may have various forms. One of the first advances of the Gini coefficient was made by Bonferroni. His idea was not to compare the partial sums, but to compare the mean of the k-percent of poorest population to the mean of the whole population. Researchers have pointed out that the correction of Gini to Bonferroni indices has made the estimation more accurate and sensitive for some poor groups of population [4]. Another drawback of using Gini coefficient is that the shift of income towards the rich proportion of the population which happened since 1980th to current times is underestimated in Gini coefficient since it increases both the numerator and the denominator [5]. Therefore, the median based Gini coefficient is proposed [6]. Thus, the advances presented may still underestimate the disparities between poor and rich population, therefore, a new index was promoted to overcome this gap.
The Zenga index compares incomes of the poorest population with richest one [7]:

\[ Z_Q(k) = \frac{\mathbb{E}[X \| X > Q^{-1}(k)] - \mathbb{E}[X \| X \leq Q^{-1}(k)]}{\mathbb{E}[X \| X > Q^{-1}(k)]}, \quad Z_Q = 1 - \int_0^1 Z_Q(k)dk, \]

where \( X \) is household’s income in ascending order, \( k \) is percent of population, \( Q^{-1}(k) \) is inverse quantile function, returning maximum household income from the stated percent of population. Zenga index in contrast to Gini coefficient compares two adjacent and exhaustive groups of population, while Gini coefficient compares the whole population and the poorest \( k \) percent of it. The researchers have compared both Gini and Zenga indices in terms of accuracy of estimation and have found out that Zenga index better estimated the inequalities in the population thus may be more complex to estimate [8].

Nevertheless, comparing \( k \) percent poorest population with \((1 - k)\) percent the rest population may not always be sensitive enough to stress on the noticeable contributions of the extremely poor population with those on the tails of the distribution, which may also be called as very rich population. Therefore, a new index (Davydov index) was proposed which compares equally sized and opposite groups of the population in terms of income. A new inequality curve which is also a Lorenz based inequality curve is constructed as follows [9]:

\[ D_Q(k) = \frac{(1 - L_Q(1-k) - L_Q(k))}{(1 - L_Q(1-k))}, \quad D_Q = 2 \int_0^1 D_Q(k)dk - 1, \]

where \( k \) is percent of population, \( L_Q(k) \) is the value of the Lorenz curve for the \( k \)th percent of population. The new Davydov index can not only compare the poorest population with the whole one (like Gini one) or use contrasting groups (like Zenga one), but any opposite groups of the chosen size. Furthermore, it seems that \( D_Q \) is easier evaluated that the Zenga index.

3. Study of income inequality in Russia
We will start reviewing income inequality in Russia using the Household Budget Survey (Household Budget Survey provided by the Federal States Statistics Service of the Russian Federation URL: https://obdx.gks.ru) from 2014 to 2018. Taking in account recent program by the Russian Government “Russia 2020” mean and median values of income in Russian regions should change since 2014 to 2018 so that the regions differentiate less. Firstly, the distribution of mean and median values of household income in Russian regions shows that from 2014 the total value of mean and median household’s income in regions has shifted to the right, which means that household income has increased in general. Secondly, the homogeneous groups of households with medium income (which is clearly seen in 2014 mean and median household income in regions distribution has divided into two homogeneous groups – households with lower medium income and higher medium income groups). At the same time, the distribution of mean and median household income of the regions with high household income does not seem to have a great change according to distributions reviewed (figure 1). This means that mean and median household income for regions with low and medium values of each have increased, thus, for rich regions nothing has changed in terms of income inequality. It seems that it has become smaller for all regions but rich ones.

To review income inequalities Gini, Zenga and Davydov coefficients were calculated. We will start by reviewing the distribution of Gini coefficients in Russian regions from 2014 to 2018. Gini index represents the proportion of inequality the following way – the higher the index – the more the inequality (this means that high income households receive more proportion of total wealth of the population). According the distribution of Gini index the inequality in Russian regions has become lower as the distributions shifts to the right almost each year. In 2018 we
Figure 1. KDE mean and median households’ income in Russian regions in 2014-2018, bandwidth 7500 Gaussian kernel (black – 2014, dark blue – 2015, blue – 2016, light blue – 2017, grey – 2018)

Figure 2. KDE Gini coefficient in Russian regions in 2014-2018, bandwidth 0.009 Gaussian kernel (black – 2014; dark blue – 2015; blue – 2016; light blue – 2017; grey – 2018)

can see the majority of regions with very small Gini coefficient and at the same time in 2018 there are fewer regions with high Gini coefficient (thus, this number is not the lowest).

Speaking about the rest regions with quite typical level of inequality in Russia, in 2014 they formed almost homogeneous group, then the distribution started to change and has visually separated in two homogeneous groups with medium low and medium high Gini coefficient measures correspondently. Since then, it started to change in order to form a more homogeneous distribution, which resulted in increasing of inequality level for those regions from medium low Gini coefficient group and at the same time decreasing the inequality level from medium high inequality level group.

It was discussed above that the Gini index may underestimate the inequality in the population, therefore, some other indexes should be checked whether the conclusions from reviewing their distributions coincide with those derived by reviewing Gini index. The Zenga index compares the gap between mean incomes of poorer and richer population to the mean income of rich population. As the results the larger the gap – the more inequality in the population. Initially the results, which may be derived from Zenga index distribution in terms of income inequality in Russian regions from 2014 to 2018 coincide with those derived from the Gini index distribution. This means that the tendency of changes or some top level conclusions may be derived from Gini index as well.

In order to compare the difference of income inequality level in Russian regions via Gini and Zenga indices, the distribution of both indices will be plotted for 2014 and 2018. It can be seen from the figure that Gini coefficient overestimates inequality for regions with high inequality, at
the same time it underestimates inequality for regions with low inequality; comparing to Zenga index the number of regions with low inequality is higher for Gini index distribution when the number of the medium inequality regions is much lower. Once again, we do understand that the calculation of indices is different, thus both of them represent the inequality level from zero to one, where one is the most inequality in population allowed, therefore in these terms they may be compared. There was promoted one more index which can measure inequality in the same way as Gini and Zenga do.

**Figure 3.** KDE Zenga index in Russian regions in 2014-2018, bandwidth 0.005 Gaussian kernel (black – 2014; dark blue – 2015; blue – 2016; light blue – 2017; grey – 2018)

**Figure 4.** KDE Zenga and Gini indices in Russian regions in 2014-2018 (black – 2014; blue – 2018; dashed – Gini index bandwidth 0.009 Gaussian kernel; solid – Zenga index bandwidth 0.005 Gaussian kernel)

Firstly, the tendency of income inequality measured by Davydov index in Russian regions in 2014–2018 is reviewed. It is expected that the tendency will be the same as discussed before. All in all, tendencies are the same, the average income inequality in Russian regions decreases overtime. At the same time the exact separation of medium income is inequal in regions of Russia in all years but 2016. Moreover, when comparing income inequality in Russian regions in 2018 to 2014 we can clearly see that not only income inequality in 2018 has decreased but also that the regions with medium income inequality has become more different, as the distribution is less sharp. It can be clearly obtained that Davydov index shows less inequality in the population than another two reviewed. It is worth mentioning that the distributions of Davydov and Zenga indices form the same shape, while the distribution of Gini index is quite different. Generally speaking, income inequality for most of regions decreases, while there are some regions, where income inequality acts differently. The income distributions of all Russian regions were studied, and three groups of household income distribution were found according to the shape of distribution. The first group represents regions, where there are few households with low and high income and the majority of households have medium income (the number of households in the gradation of medium income is almost the same). Such households’ income distributions usually have several subgroups and their specific characteristic is absence of peak and wide range. The second group of household income distribution has a more bell shape with few inconvexities which may be interpreted as separation of the middle class to poorer and higher one. And the uniqueness of the third group is presence of the distinguished subdistributions for rich and very rich households. The majority of Russian regions fall into group two, the examples of such regions are republic Yakutia, Tomsk region, Saratov region, Moscow region, Saint Petersburg and Moscow. The examples of regions falling into the first group are Krasnodar region, Lipetsk region and Republic Chevashia. And the third group’s regions examples are
Figure 5. KDE Zenga and Gini indices in Russian regions in 2014-2018 (black – 2014; blue – 2018; dashed – Gini index bandwidth 0.009 Gaussian kernel; solid – Zenga index bandwidth 0.005 Gaussian kernel)

Figure 6. KDE Zenga, Gini and Davydov index in Russian regions in 2014-2018 (solid – Gini index, bandwidth 0.009 Gaussian kernel; dashed – Zenga index, bandwidth 0.005 Gaussian kernel; dashed-dotted – Davydov index, bandwidth 0.005 Gaussian kernel)

Figure 7. KDE logarithm of household’s income distribution for Russian regions for three groups respectively, Gaussian kernel (1 – Republic Chevashia, bandwidth 0.1; 2 – Republic Tyva, bandwidth 0.1; 3 – Republic Ingushetia; bandwidth 0.075)

Republic Ingushetia, Republic Karachaev-Cherkessk.

It may be expected that the income inequality increases with the increase in the group number. Thus, as income inequality is measured through the relation of cumulative income,
Figure 8. Income distributions and Lorenz curve for groups of Russian regions (blue – Lorenz curve; (1 – Republic Chevashia, 2 – Republic Tyva, 3 – Republic Ingushetia)

the result may be quite different for the indices used in research – Gini, Zenga, Davydov. We will start by reviewing the Gini coefficient and the Lorenz curve. Suppose, it is necessary to construct Lorenz curve based on the income distribution. In the first step we take the poorest population indicated as grey dashed line in the graphs (logarithm of household’s income to the left of the grey dashed line is considered as poor population here as an example). Let us also suppose that this line represents 20% of poorest population for simplicity. As can be seen from the graphs and using some approximation we suppose that the proportion of cumulative income to the left of the grey dashed line and the cumulative households’ income to the right of the grey dashed line in each group is quite the same. Then let us review the poor middle class, which in this example will be denoted as the 40% of poorest population and it will be indicated in the graph as dark blue dashed line. Even though the cumulative sums of the logarithm of income to the left of dark blue dashed line may seem quite the same, in fact the cumulative sums of the second and third groups are larger. This happens because the distributions of the latter two have smaller households with high medium income, therefore the cumulative sum of logarithm of households’ income to the right of distribution is lower. This means that the value of the Lorenz curve for the first group is smaller (dark-blue dashed line) then the values for the second and third group for the poorest 40% of population, which are quite the same (dark-blue dotted line). The same happens when reviewing richest population (as in example, 10% richest population which is denoted as pink dashed and pink dotted lines). As the third group has a large homogeneous group of nearly rich households, the proportion for Lorenz curve is larger, then for the case when there are few
nearly rich households, as in the example of the second group. As Gini index is calculated as

\[ Gini = 1 - 2 \int_{0}^{1} LC(t) dt, \]  

where \( \int_{0}^{1} LC(t) dt \) is the square under the Lorenz curve, the larger the integral is – the smaller the Gini coefficient. This means that the larger the square under the Lorenz curve the less inequality is observed in the population. In terms of groups of regions studied the income inequality is the largest for group 1, then follows group 2 or group 3 as the Lorenz curves for them were almost the same (in terms of example reviewed and the slices made – the third group should have less income inequality, thus, in real data example and more precise calculations the data can change). The same order is observed for Zenga and Davydov curves, as all the three indices studied depict the major tendencies the same way. The distribution of these curves is presented in figure 9 for the regions used as examples of groups separated on the logarithm of income distribution.

The Zenga index represents the area above the Zenga curve, at the same time, the smaller the area the less inequality is in the population. From the graphs observed the smallest area under the Zenga curve is depicted by the group three, then comes the group two and group one has the most inequality. Speaking about the Davydov curve, the area between it and the curve of pure inequality represents the Davydov index, and once again the order of groups in terms of income inequality holds.

Table 1 presents the calculated income inequality in groups studied. The conclusions derived from graphs approval by the same numbers calculated. At the same time, the calculated indices show that for the regions with high income inequality Gini and Davydov indices depict rather various coefficients, while for other two examples, the difference is smaller. At the same time, the Davydov index for all cases shows less inequality in Russian population than Zenga and Gini ones. Speaking about the overall tendencies, it was found out that the majority of regions moves from the first group to the second one from 2014 to 2018, which means that income inequality decreases in Russia as high-income inequality regions are eliminated. At the same time, regions with lower inequality do not increase significantly. These results coincide with the average tendencies reviewed by distribution of income and income inequality measures in Russia. It may be generally expected that the tendencies of levels of inequality will be the same for regions close to each other. Generally, the same conclusions hold when comparing Zenga
and Davydov indices. At the same time, it seems that Zenga and Davydov indices depict less inequality in the population than Gini does, which was also mentioned above.

| Table 1. Inequality indices presented for the groups chosen. |
|------------------------------------------------------------|
| Gini 2014 | Zenga 2014 | Davydov 2014 |
| Group 1 (Republic Chevashia 2014) | 34% | 33% | 29% |
| Group 2 (Republic Tyva 2014) | 30% | 30% | 27% |
| Group 3 (Republic Ingushetia 2014) | 28% | 29% | 26% |

4. Discussion
The paper reviewed different works on inequality indices constructed on the base of comparing income of different groups of population, specifically Gini, Zenga and Davydov ones. Their basic characteristics and the algorithm to construct were described as well as the major drawbacks pointed by the researchers. The difference between the values of these indices applied to real data of other countries were mentioned. In general, it is considered that Gini index overestimated the impact of the middle class, and therefore provides the biased value of income inequality, Zenga index is used in order to eliminate this bias as it compares the cumulative income of poorer group not with income of the whole population but with cumulative income of the rest richer group. At the same time newly presented Davydov index may be reviewed as a more generalized concept as it compares the cumulative income gap between various poor and rich groups to the cumulative income earned by the whole population, but the richer group chosen. In terms of index construction, the Davydov index may be more general as it allows to correct for the proportion of rich and poor population. Then the indices described were calculated for Russian regions since 2014 and 2018 and their values were reviewed. The major conclusions are that household’s income in Russian regions generally is increasing, while number of very poor and very rich households decreases. And generally, the regions move to become more similar in terms of income inequality, while income inequality itself tends to decrease. Nevertheless, such conclusions may be majorly explained by the overall tendency of income inequality in Russia rather than features. In order to review the determinants of such allocation further analysis should be made in two levels. The first one will describe the macroeconomic characteristics of the region and may be studied by adding specific macroeconomic variables like the number of industrial and agricultural businesses, number of educational organizations, population density in the region and so on and analyzing such relations. While the second one is to review the impact of households of different types in income inequality level and to measure demanded income inequality level in Russian regions. In order to do so other databases should be used and some correction of data and the indices made in order to eliminate bias from ambiguous numbers of income provided by households during personal interview. These questions are some possible topics for further research.

Acknowledgments
Supported by a grant of Russian Foundation for Basic Research 18-010-00564 Modern Tendencies and Social and Economic Consequences of Digital Technologies Development in Russia. Supported by the Laboratory for Wealth Measurement, Faculty of Economic science, Higher School of Economics, Moscow, Russia.
References
[1] Lakner C and Milanovic B 2015 Global Income Distribution: From the Fall of the Berlin Wall to the Great Recession World Bank Econ. Rev.
[2] Gini C 1914 Atti del Reale Istituto Veneto di Scienze Sulla misura della concentrazione e della variabilità dei caratteri 73(2) 1203–48
[3] Lorenz M 1905 Methods of measuring the concentration of wealth Public. Am. Stat. Assoc. 9(70) 209–19
[4] Bonferroni C 1930 Elementi di statistica generale (Firenze: Libreria Seeber)
[5] Greselin F and Pasquazzi L 2010 Asymptotic confidence intervals for a new inequality measure Comm. Stat.: Simul. Comput. 38(8) 1742–56
[6] Atkinson A and Brandolini A 2011 On the Income Inequality: Identification of the Middle Class (CA: Stanford University Press) 77–100
[7] Prendergast L 2015 Exploiting the Quantile Optimality Ratio in Finding Confidence Intervals for Quantiles Stat. 5 70–81
[8] Zenga M 2007 Inequality Curve and Inequality Index Based on the Ratios Between Lower and Upper Arithmetic Means Statistica and Applicazioni 1 3–27
[9] Davydov Y and Greselin F 2020 Comparisons Between Poorest and Richest to Measure Inequality Sociological Meth. Res. 49(2) 526–61