Victor Kapitanov (JSC "Research Institute" Polyus" them. MF Stelmakh", Moscow); Ivanova Anna (SI "Institute of Applied Mathematics and Mechanics", Donetsk); Alexandra Maximova (SI "Institute of Applied Mathematics and Mechanics", Donetsk)

Issues of Inequality Representation in Conditions of Deep Social Stratification

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. Problems with the representation of inequality are caused by 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 last published Federal Tax Service (FTS) "Report on the declaration of income by individuals" (2015) only 22.053 trillion rubles. The principal difference between the Rosstat data and the FTS data is that the Rosstat estimates anonymous incomes. The FTS determines incomes with reference to specific taxpayers. 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 for big business and 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 incomes, as less information about them. Two important requirements for methods of inequality representation follow from this lack.

1. No data source can be considered trustworthy. Only the coincidence of the population distribution by income from several independent sources allow to consider the information as 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 representing inequality must be unconditionally sensitive to the width of the income range for any arbitrarily small filling of rich cohorts.

We know the incomes of some of the richest citizens of the country, published by the FTS and Forbes magazine. But these data are not sufficient to fully describe the existing economic structure of society (ESS) in Russia, since the number of rich people who have hidden their income and are not in Forbes and the FTS is unknown. In this situation, we have to admit that only information about the width of the income range (determined by the income of the richest person in the country) is available. There are no reliable data on filling rich cohorts, there are only data on how far to the right of the income scale are the rich cohorts. Histograms and frequency ranges are most known graphical method for presenting data on inequality. Frequency ranges show that the distribution of people by income is a lognormal with a so-

called heavy tail. It is radically different from the normal distribution of people according to individual conditions (weight, height, vocabulary etc). This difference implies an understanding of the origin of income inequality - it is not caused by individual physical or intellectual features, but only by exploitation, systematically practicing non-equivalent transactions in society, rank exchanges, during which the richer gets more than the poor.

But frequency ranges have a fatal lack, they are tied to data quantization. The same data on the prices of cars sold in the Russian Federation, but differently quantized (24 and 15 cohorts) give the different height of the peak of distribution polygons. Lorenz curves are not informative with incomplete data on the population of the higher distribution cohorts. The Lorenz curve, constructed according to Rosstat and taking into account the maximum income of 96 thousand rubles / month, is visually indistinguishable from the curve built on the combined data of Rosstat and Forbes, which takes into account the maximum income of 14.3 billion rubles / month.

We have computed using real data on income in the Russian Federation some common inequality indices. The decile coefficient of funds is fundamentally impossible to calculate with arbitrary quantization of data; it requires quantization strictly on deciles. In any case, this ratio, which contains data on the incomes of 10% of the richest people, is not informative in a situation where 99.98% of the population occupy 0.006% of the income scale.

The Pareto index was 1.22 ... 1.26 for data on the prices of cars, whereas for Rosstat data characterized by understated inequality, it was not less, but more - 1.36. The Pareto index is generally not able to definitely characterize inequality. The Gini index is the same (0.36) for both Rosstat and Rosstat+Forbes. The Gini index does not distinguish a society with maximum income of 96 thousand rubles / month from a society with maximum income of 14.3 billion rubles per month, it requires ideally complete data on the rich, and therefore is not suitable for revealing the true picture of inequality.

The first three indicators of total entropy (zero, first or Theil index and the second), and the ratio of maximum income to modal, showed significant sensitivity to quantization, and therefore unsuitable for comparing data from different sources. As an alternative, we propose the use of distribution functions for the graphic representation of inequality.

They are insensitive to quantization (and do not require it at all). Quantized data on prices for automobiles in different ways give completely identical curves. They don't require perfectly complete data on the filling of rich cohorts, the distribution function of the Rosstat data is radically different from the distribution function of the summary data of Rosstat+Forbes.

The distribution functions are clear and allow to distinguish the real ESS (a lognormal distribution with a heavy tail) from the normal distribution, for example. As quantitative

inequality representation, we suggest approximating real ESS by an exponential curve (the determination coefficient is quite high: 0.927 … 0.999), and using the intensity of the exponential distribution as an inequality index.

Our index can be calculated with any quantization, it doesn't depend on quantization, which allows comparing data from any sources, it doesn't require perfectly complete data on the filling of rich cohorts (for Rosstat data it is 2.43, while for summary data Rosstat and Forbes it is 485000).