IS INCOME INEQUALITY IN BULGARIA UNDERESTIMATED IN SURVEY DATA?

Petar Peshev, Kristina Stefanova, Ivan Bozhikin,

Radostina Stamenova, Ivanina Mancheva,

University of National and World Economy

Как да се цитира тази статия / How to cite this article:

Peshev, P. et al. (2022). Is Income Inequality in Bulgaria Underestimated in Survey Data? *Economic Thought Journal*, 67(3), 301-326.

To link to this article / Връзка към статията:

Published online / Публикувана онлайн: 29 July 2022

Submit your article to this journal / Изпратете статия за публикуване

https://etj.iki.bas.bg

Article views / Статията е видяна:

View related articles / Други подобни статии:

View Crossmark data / Вж. информация от Crossmark:

Citing articles / Цитиращи статии: View citing articles / Вж. цитиращи статии:

IS INCOME INEQUALITY IN BULGARIA UNDERESTIMATED IN SURVEY DATA?

Petar Peshev, Kristina Stefanova, Ivan Bozhikin,

Radostina Stamenova, Ivanina Mancheva,

University of National and World Economy

Abstract: The present study analyses wage income inequality in Bulgaria during the period 2010-2019, using data from the National Revenue Agency (NRA) on the actual incurred income of taxpayers and survey data on wage income from the Household Budget Survey of the National Statistical Institute (NSI). Key indicators of income inequality are derived under both approaches, which are compared and evaluated through descriptive analysis. The NSI survey data permanently underestimates the level of income and inequality in its distribution, as evidenced by the Gini coefficient, the decile and percentile ratios, as well as other indicators based on them.

Keywords: Bulgaria; inequality; income distribution; Gini coefficient; Palma ratio; S80/S20 ratio

JEL codes: D31; D63; E25; O15

Received 16 May 2022

Revised 8 June 2022

Accepted 28 June 2022

1. Introduction

Bulgaria is the country with the highest income inequality in the EU, and the trend is mostly upward. The Gini coefficient based on the Income and Living Conditions Survey (EU-SILC), which is adopted as an official indicator for international comparisons, has increased from 0.33 to 0.4 in the decade after 2010. In contrast to other countries, taxes and social transfers in Bulgaria do not contribute significantly to smoothing the differences in the distribution of net disposable income of households in Bulgaria. At the same time, the income concentration among households of the last decile is high, especially in the last percentile, where inequality is more pronounced.

One of the main sources of data on global income inequality is the World Bank's estimates in the World Development Indicators. According to them, income inequality in Bulgaria, as measured by the Gini coefficient, reaches 0.4 in 2019, with the upper decile accounting for 31% of income. Moreover, estimated and extrapolated data from the World Inequality Database (WID) show that Bulgaria is the EU country with the highest share of income of the highest-paid 1%, respectively the last decile (18.2% and 43.5% for 2019).

The problem of the accuracy of income inequality measurements is of interest to many researchers around the world working on this topic. The problem is related to missing or inaccurate data due to income concealment, as well as incompatibility of estimates and incomplete coverage, especially for data collected through surveys, different methodologies, etc.

Two main sources of the indicators of income inequality in Bulgaria exist, two important sources of data can be used – the Household Budget Survey of NSI and the EU Survey on Income and Living Conditions (EU-SILC). Accordingly, the corresponding values of the Gini coefficient for 2019 are 0.31 and 0.4, while the dynamics of the indicators also differ. Household budget monitoring is based on quarterly surveys (paper diaries) and the EU-SILC is also based on household surveys but uses a different methodology. The NSI only provides data on income distribution derived from the Household Budget Survey, while EU-SILC only provides calculated quintile ratios and Gini coefficients as well as absolute values of decile group limits.

According to the NSI Household Budget Survey, income from wages accounts for the largest share of total household income (over 56%), while the rest sources of income, e.g. income from property and economic activity, are not taken into consideration in this paper. Shifting the focus to wage income, the Gini coefficient calculated using NSI data amounts to 0.41 on average over the period 2010-2019, compared to 0.53 when using data from the NRA.

This analysis adds a third source of data on the distribution of household income – NRA summary data on the declared tax base of wage income in the period 2010-2019. The study assumes that the NRA data is characterised by a higher degree of reliability regarding the actual labour income paid and its distribution.

Based on these observations, the aim of the study is to analyse the inequality in the distribution of monetary wage income in Bulgaria and its dynamics depending on the type of data used for income distribution.

The main hypothesis to be confirmed or rejected by the study is that the NSI survey data persistently underestimate the size of income and inequality in income distribution, as evidenced by the Gini coefficient, decile and percentile variables and indicators based on them. In accordance with the formulated objective and hypothesis, the object of the study is income inequality in Bulgaria and the subject – the various indicators measuring wage income inequality and its deviations depending on the data source.

In particular, this study analyses income inequality in Bulgaria in the period 2010-2019, using data from the NRA for actually declared and paid income of employees and Household Budget Survey data from NSI. Using descriptive analysis, key indicators of income inequality are derived, and estimates and results on inequality in the distribution of wage income are compared and analysed according to both approaches.

The paper is organised as follows: The first section provides an overview of Bulgarian and international studies dedicated to measuring income inequality in recent years. The next section presents the existing problems with inequality data and explains the data used and the methodology of the present study. The third part analyses and compares the empirical data from the study of household budgets and the NRA data. The results of the study are summarised and interpreted in the

conclusion. The raw data of the National Revenue Agency on the distribution of wage income and the cumulative distribution by decile groups are presented in the appendix.

2. Literature Review

There are many studies on income inequality that analyse it from different perspectives and in different countries and regions. Below are some of the most significant research papers in this field in Bulgaria, as well as the most recent international studies on this topic from the last six years (2016-2022).

Tosheva et al. (2016) use the EUROMOD tax-benefit microsimulation model to assess the effects of socio-economic policy on the decile distribution of gross income and transfer payments. The authors include transfer payments and paid pensions, which contribute to the reduction of income inequality. However, this positive contribution of pensions is not observed when the focus is on reducing income inequality to the EU average (Tosheva et al., 2016).

Mihaylova and Bratoeva-Manoleva (2017) use NSI data covering all types of income from household budgets to analyse income inequality in Bulgaria over the period 1990-2014. The study focuses mainly on transfer payments and their role in income inequality. The researchers find that the effect of transfer payments on income inequality is positive, i.e. they contribute to reducing inequality. However, it is stated that the share of social transfer payments in GDP is lower compared to most EU countries.

Bratoeva-Manolova (2021) examines the impact of different income components on rising inequality according to the study of household budgets. The author concludes that while social transfers (especially pensions) contribute to a more equal income distribution, in terms of wage income, the unequal distribution in favour of high-income decile groups intensifies between 1993 and 2019. The differentiation of wage income is higher, as the Gini coefficient for it reaches 0.45 in 2019, compared to 0.31 for total household income.

Nenkova (2021) analyses the impact of fiscal transfers on income inequality based on data from the World Inequality Database (WID). She found that in Europe, on average, about 80% of the reduction in income inequality is achieved through social transfers, including pensions, and 20% through direct taxes. Bulgaria is one of the three countries with the lowest redistributive role of the state in reducing inequality.

Mintchev et al. (2010) parameterise inequality in the income distribution, with the number of children, the number of unemployed and the number of retirees being a factor for a household falling into a low-income group, while those living in the capital are more likely to fall into the higher income deciles. Mavrov (2021) argues that it is possible to maintain high economic growth while reducing inequality in income distribution. Of course, causality needs to be examined, whether high economic growth rates lead to a decline in economic inequality or vice versa. Most likely, economic growth and inequality are a function of a changing array of variables.

Alvaredo et al. (2018) analyse global income inequality over the period 1980-2016 using data from the WID and the corresponding distribution by region. In general,

incomes in the first five deciles show a long-term downward trend, but in Europe and Russia, this trend reversed in the mid-1990s, while in the US and Canada, the downward trend persisted throughout the period under review. At the end of the period, the first 50% of income in Europe corresponded to about 20% of the total, compared to about 17% in the mid-1990s. However, the income of the top decile is steadily increasing, especially in European countries, the United States and Canada. It accounts for about 35% of income in Europe and about 47% of income in the United States and Canada.

Morgan and Neef (2020) point out that income inequality in Europe in 2019 is determined more by divergence within countries than by inequality between EU countries. The trend for the share of the top decile to increase in Eastern European countries after the beginning of transition is particularly pronounced, including in Bulgaria.

Carvalho and Rezai (2016) conducted a theoretical and empirical study (for the period 1967-2010) for the United States to determine how changes in income distribution can affect aggregate demand. They found that the propensity to save increases significantly from the lower to the upper quintile of income groups. Moreover, the more equal distribution always leads to higher output, and there are conditions under which reducing income inequality among workers leads to a more dependent demand for wage income.

Papanikolaou (2021) focuses on progressive taxes and their impact on income inequality. The author finds that the US tax system was mostly regressive and, to some extent, progressive during the period from 1996 to 2011. Therefore, no tax progressivity can be identified for the personal income from the wages of the respondents. Iacono and Ranaldi (2021) examine inequality in the composition of income from labour and capital in Italy. The results show that over the period 1989-2016, the degree of income inequality has decreased due to an increase in the share of capital income accumulating at the bottom of the distribution and an increase in the relative weight of labour income at the top of the distribution.

Socci et al. (2021) analyse how a possible introduction of a flat income tax in Italy could affect the country's economic system. The main conclusion of their study is that different policy scenarios (with different tax rates and different hypotheses for policy financing) have a negative effect on real GDP and an unequal effect on household disposable income.

Lakner and Milanovic (2015) trace the evolution of global income inequality over the period 1988-2008 and attempt to correct for the underestimation of the highest incomes in household surveys. They create a database of national income surveys broken down by decile groups and identify the groups that gained and lost the most during the 20 years. At the same time, they estimate the missing end of the distribution due to refusals to participate in the surveys, i.e. the highest incomes, according to their own methodology. They also confirm that while differences in income levels between countries are declining over the period under review, national inequalities are increasing.

In another recent study on the topic, Ranaldi and Milanović (2022) focus on the relationship between capitalist systems and income inequality, using data for 47

countries over the period 1995-2018. Based on empirical analysis, they conclude that rich countries benefit from capital while poor countries benefit from labour, with income inequality higher in Latin America and significantly lower in Nordic countries.

Yonzan et al. (2022) summarise a number of studies in rich countries on the inability of household surveys to capture all incomes. They also examine the discrepancy between the highest incomes, where gaps in household surveys are most pronounced. For this purpose, the national surveys in the USA, Germany and France included in the Luxembourg Income Study (LIS) are compared with data from the tax authorities. For the United States, in particular, 3/4 of the differences between the two types of data were found to be due to non-earnings and, in particular, self-employment, income and incapacity of the surveys to capture the income of the richest 1% (Yonzan et al. 2022).

Many authors examine the difference in the results for the distribution of income and wealth when using survey data, such as data from the Household Budget Survey budgets of NSI, on the one hand, and administrative data, such as NRA tax data, on the other, including Bricker et al. (2016), Sabelhaus et al. (2015), Burkhauser et al. (2012). In the studies in question, administrative data show a higher concentration of income in the highest deciles, percentiles and 0.1 percentiles. Survey data traditionally do not cover both ends of the distribution because it is difficult to survey respondents from the high percentile and decile groups, or they underestimate their income and wealth when interviewed. It is also difficult to persuade respondents from the lowest percentiles to be interviewed. When analysing income in its entirety and not just salary income, it is possible that the administrative data underestimate the high deciles and percentiles which are, for example, unemployed or retired or receive a large share of property income.

In Bulgaria, two types of data are traditionally used to estimate the Gini coefficient and other measures of inequality. The first is the Household Budget Survey of NSI and the second is the EU-SILC survey, which the NSI conducts according to Eurostat methodology. The Household Budget Survey covers 3060 households divided into three rotating groups, while EU-SILC is conducted among 8815 households divided into six rotating groups. In 2019, the Gini coefficient measured on the basis of the quarterly household budget survey amounted to 0.31 (for total household income), while the coefficient calculated according to the European methodology (EU-SILC) is 0.4. The first study, however, in contrast to the Gini coefficient calculated according to the European methodology, does not show a clear trend towards increasing inequality. The explanation for these differences is largely related to the quality of the data.

Problems with the reliability of data from household budget surveys (non-stochastic errors) have worsened in recent years, partly due to redundancies within the NSI (Bogdanov, 2022). At the same time, there is a pronounced tendency to replace the initially selected households, for example, due to refusals, which affects the representativeness of the samples. Bogdanov (2020) concludes that household budget monitoring "loses its status of a representative sample survey" and suggests a number of changes in its implementation.

On the other hand, there are a number of differences between the two NSI surveys that make the data not comparable (Tsvetkov, 2020). These include the different

definitions of a person per household (the EU-SILC sample uses the so-called equivalent number of people, which is lower than the actual number of people in the household) and the higher scatter rate in both samples, such as from EU-SILC is larger and covers more people at both ends of the distribution. Tsvetkov (2020) believes that there is a systematic error and raises the question of a possible overestimation of inequality by EU-SILC.

International comparisons are mainly based on the EU-SILC survey with additional calculations and estimates. For example, WID data for EU countries combine microdata (from EU-SILC and, in some countries – tax registers) and national accounts data.

Given these problems and gaps, the addition of a new data source on inequality can be seen as a way to overcome the limitations associated with the raw data and obtain more accurate estimates. In this respect, the selection of indicators to measure inequality is crucial.

3. Data and Methodology

This study compares data on the distribution of cash income from wages paid to employees in the country. NSI sample data from the survey "Household Budgets in the Republic of Bulgaria" are used, as well as data from the National Revenue Agency for the total of all employees (employed under a labour contract and self-employed), obtained as a result of *Applications for access to public information with ent. № ЕП-94-П-1152 dated 04.12.2020 and ent. № ЕП-33-00-300 from 07.12.2020 according to the inventory of the Central Office of the National Revenue Agency, under University Research Project № НИД НИ-16/2020*. NRA provided us with data for declared remuneration in Declaration No 1 under the Personal income tax act and coupled with taxable incomes for self-employed, where data for more than 3 mln. taxable units being provided and with 90% and above of declared amount belonging to employed under a labour contract.

To calculate the missing data, the incomes of the first 5-7 deciles for each year, through the NRA data and the incomes of the last 5%, 1% and 0.1%, through the NSI and NRA data were obtained by calculating a polynomial of n- degree with the following formalised record:

$$F(x)=a_0*X^n+ a_1*X^{n-1}+ a_2*X^{n-2}+\dots a_{n-1}*X+ a_n, \text{ where:} \quad (1)$$

The sum of the product between the constant *a* and the variable *X* is called a polynomial of n-degree;

a_{n-1}– constant;

X^n – a variable corresponding to the cumulative decile variable, where the degree of *n* is equal to 6, corresponding to the largest possible degree of the polynomial used to estimate the arc of the Lorenz curve at both ends.

For the purpose of this study, different n-degree polynomials were used to estimate the function of the cumulative distribution at the beginning, when it is flatter, and at

the end, when it is steeper. The following indicators are used to compare the NRA and NSI data: the incomes of the highest yielding, i.e. the last 0.1%, 1%, 5% and 10%; decile ratios P90/P10; P90/P50; the quintile ratio S80/S20; Palma ratio and Gini coefficient.

The Gini coefficient is one of the most widely used measures of inequality, especially in international comparisons. Alternative indicators such as the Palma ratio, which measures the share of income of the upper decile divided by the share of income of the poorest 40% of the population, are also used. It is believed that it better reflects changes at both extremes, unlike the Gini coefficient, which is more sensitive to changes in average incomes (see Tsanov and Bogdanov, 2010). On the other hand, the various coefficients (deciles, quintiles, etc.) that measure by how much the income of the richest group exceeds the income of the poorest group do not provide complete information either. As Tsvetkov (2020) points out, they are an approximate measure of inequality because they ignore the redistribution of income both within the compared groups and between units not included in the compared groups, i.e. average income.

Palma ratio- the ratio between the last decile (the highest yielding decile group) and the first 4 deciles;

P90/P10 – a decile ratio between the last decile (highest yielding decile group) and the first decile (lowest yielding decile group);

P90/P50 – a decile ratio between the last decile (highest yielding decile group) and the first five deciles (lowest yielding five deciles)

S80/S20 – a quintile ratio between the last two decile groups (the highest two deciles) and the first two decile groups (the lowest two deciles).

Gini coefficient is estimated using the following formula:

$$G=1-\frac{2\sum_{j=1}^{n} Sj-1}{n}, \text{ in which:} \quad (2)$$

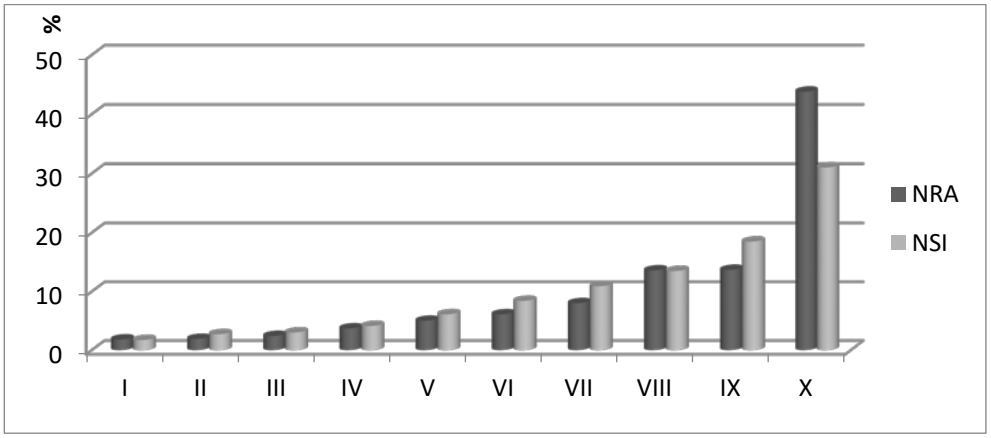
G – a designation for the Gini coefficient, taking values between "0" and "1";

S_j – cumulative share of income of percentile/decile/quintile/quartile group "j";

n – number of percentile/decile/quintile/quartile groups.

4. Descriptive analysis

The decile distributions of income calculated on the basis of the National Revenue Agency data on household income actually paid and according to data on income from the National Statistical Institute survey on the budgets of households show some differences. In 2019, the decile distribution of income, according to the NSI data, showed a higher share of income in the groups from the second to the seventh and ninth deciles. In the first, eighth and tenth decile groups, there is an opposite tendency – an underestimation of the share of income. The greatest distance in the values of the estimates is reported in the last two groups with the highest income (Figure 1).



Source: NRA, NSI, own calculations.

Figure 1. Decile distribution of income according to NRA and NSI data for 2019

To reveal the trends in the decile distribution of income, the dynamics of the income share of the highest and lowest income deciles are important. According to NRA data over the period 2010-2019, the highest income decile of the population receives between 41.1% and 45.2% of the income in the country, with a slight increase in 2019 compared to 2010 (Table 1). According to the NSI data, there is a higher and more diverse dynamic of this indicator. The highest value (31%) is reported at the end of the period under review, and the lowest (23%) – at the beginning (Table 2). Throughout the period, the share of income of the last decile group is higher according to the NRA data than according to the NSI data.

Table 1. Decile distribution of income according to NRA data (in %)

Decile	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
D1	3.0	2.9	2.8	2.7	2.6	2.5	1.9	1.8	1.8	1.8
D2	3.0	2.9	2.8	2.6	2.4	2.2	2.1	2.0	2.0	1.9
D3	2.7	2.6	2.9	2.8	2.6	2.5	2.7	2.6	2.7	2.4
D4	2.9	2.9	3.3	3.1	3.1	3.3	3.5	3.6	3.7	3.7
D5	3.7	3.8	3.9	4.1	4.1	4.5	4.7	4.8	4.9	5.0
D6	5.4	5.5	5.3	5.7	5.8	6.1	6.3	6.3	6.3	6.1
D7	8.2	7.9	7.7	8.0	7.9	8.1	8.1	8.1	8.7	8.0
D8	11.1	11.6	11.6	11.8	10.9	11.4	11.2	10.9	11.6	13.6
D9	15.9	14.7	15.4	15.8	16.5	15.8	16.0	16.3	15.0	13.7
D10	44.1	45.2	44.3	43.4	44.1	43.9	43.5	43.8	43.7	43.8

Note: Non-cumulative distribution, where *D1, D2 … Dn* stands for the nth decile.

Source: NRA, own calculations.

The lowest income decile group receives between 1.8% and 3% of income, according to the NRA data. The lowest value is reported in the last three years of the period and the highest – at the beginning of the period. Similar dynamics of decrease in 2019 compared to 2010 are reported according to NSI data. In the period under

review, the values of the indicator according to the NRA data exceed those calculated according to the NSI data, with the exception of 2017. Differences in the dynamics of income shares calculated according to the two databases are also reported for the other decile groups (Table 1 and Table 2).

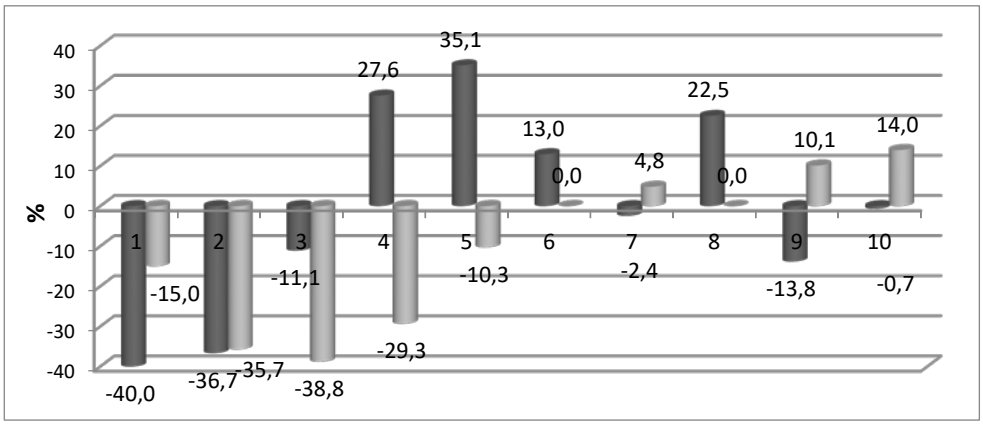
Table 2. Decile distribution of income according to NSI data (in %)

Decile	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
D1	2.8	1.5	1.5	1.7	1.6	1.6	1.7	1.9	1.5	1.7
D2	5.1	3.9	3.7	3.5	3.6	3.8	3.2	3.2	3.1	2.7
D3	6.4	4.4	4.2	4.1	4.7	4.6	4.6	4.0	3.7	3.0
D4	7.4	5.6	5.3	5.3	5.7	5.4	5.2	5.1	4.3	4.1
D5	8.3	6.9	6.6	6.4	6.8	7.0	6.7	6.7	6.4	6.1
D6	9.4	8.4	8.5	7.9	8.1	8.5	8.1	8.3	8.2	8.4
D7	10.7	10.4	10.4	9.9	10.1	10.6	10.5	11.1	11.1	10.9
D8	12.3	14.2	13.9	13.5	13.4	12.9	13.3	13.5	14.0	13.5
D9	14.7	17.3	17.6	17.3	17.2	17.4	17.4	17.2	18.5	18.5
D10	23.0	27.4	28.4	30.4	28.9	28.3	29.4	29.0	29.1	31.0

Note: Non-cumulative distribution, where *D1, D2 … Dn* stands for the nth decile.

Source: NSI, own calculations.

The analysis reveals certain peculiarities in the dynamics of income distribution by deciles according to the NRA and the NSI data. According to the NRA data, the rate of change in the income distribution by decile in 2019 decreased in six of the decile groups compared to 2010. This reduction is most significant in the first two decile groups with the lowest income. Such a trend can also be observed in the third, seventh and ninth groups, but with much lower values of the rate of change (Figure 2). At the same time, the decline in the share of the ninth decile group is more pronounced on a chain basis of the previous year only in 2019 (Table 1).



Source: NRA, NSI, own calculations.

Figure 2. Rate of change in income distribution by decile according to NRA and NSI data in 2019, compared to 2010 (in %)

The highest growth in income share (35.1%) is recorded in the fifth decile group, followed by the fourth and eighth decile groups, and the lowest – in the sixth decile group. The trends identified indicate that according to the distribution calculated on the basis of the NRA data, there is a significant increase in income in the intermediate income groups (fourth, fifth and sixth), which describe the middle-income population (Figure 2). When evaluating the results of the analysis, it is important to bear in mind that the dynamics of the first five to seven deciles (for different years) are assessed using polynomials of the fourth, fifth or sixth degree. For this reason, the dynamics and the data on their distribution cannot be absolutely accurate.

An opposite dynamic in the direction of the rate of change in 2019 compared to 2010 is observed in half of the decile groups according to the two databases considered. A one-way dynamic of the decrease is reported only in the three poorest decile groups (Figure 2), indicating a decrease in the share of their income. The observed differences between the decile distributions of income according to the NRA data and according to the NSI data suggest certain peculiarities in the indicators of income inequality that can be calculated on their basis and respectively in the conclusions of their analysis.

Table 3 and Table 4 present key indicators of income distribution calculated on the basis of NRA and NSI data, respectively. The NRA data is estimated for the missing deciles at the beginning of the cumulative income distribution (in the early sections of the Lorenz curve). The NSI and NRA data are estimated for the richest: 0.1%, 1% and 5% using the corresponding n-degree polynomials that best describe the arc of the cumulative income distribution function (Lorenz curve).

Table 3. Selected indicators for the distribution of income according to NRA data

%	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
0-50%	15.3	15.0	15.7	15.2	14.8	14.8	15.0	14.7	14.7	14.8
Top 20%	60.0	59.9	59.7	59.2	60.6	59.5	59.5	59.9	58.4	57.5
Top 10%	44.1	45.2	44.3	43.4	44.1	43.7	43.5	43.6	43.3	43.8
Top 5%	31.0	34.7	32.8	32.4	33.9	34.3	34.3	34.2	33.8	33.3
Top 1%	15.3	17.3	17.4	17.0	18.2	18.5	18.3	18.4	18.0	17.4
Top 0.1%	8.7	7.8	7.2	6.8	7.5	7.7	7.5	7.7	7.4	7.0
P90/P10	14.95	15.37	15.89	16.38	17.25	17.47	22.73	24.25	24.08	24.91
S80/S20	10.09	10.34	10.69	11.28	12.29	12.61	14.75	15.68	15.54	15.50
P90/P50	2.89	3.01	2.83	2.85	2.98	2.91	2.90	2.95	2.88	2.97
Palma ratio	3.81	4.01	3.76	3.89	4.15	4.14	4.25	4.35	4.27	4.47
Gini	0.520	0.525	0.520	0.521	0.530	0.525	0.529	0.533	0.525	0.531

Source: NRA, own calculations.

The first five deciles receive, on average, 28.5% less according to the NRA data than according to the NSI data. This difference gradually decreased from 36% in 2010 to 16.4% in 2019. It should be noted that regardless of the database used, there is an unfavourable downward trend in the income of the bottom 5 deciles over the period 2010–2019.

The National Revenue Agency data on actual remuneration paid reflect reality to a

greater extent, as they are based on the real distribution of income, albeit with a certain degree of expert judgement. As expected, in this regard, the NRA data for the richest 10% of employees exceeds the values of the indicator based on the NSI sample data by an average of 52% over the entire reporting period. However, the distance between the values calculated on the basis of the different sources has narrowed significantly over the reporting period (from 62% in 2010 to 41.2% in 2019). At the same time, the dynamic in this indicator differs depending on the two-income databases. While the indicator calculated using the NRA data is declining in 2019 compared to 2010, there is an increase according to the NSI data. It is this difference in dynamics that determines the convergence of the values of the indicators from the two databases.

Table 4. Selected indicators for the distribution of income according to NSI data

	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
0-50%	23.6	22.3	21.4	21.0	22.3	22.4	21.4	20.8	19.1	17.7
Top 20%	44.1	44.7	45.9	47.7	46.1	45.7	46.7	46.2	47.6	49.5
Top 10%	27.2	27.4	28.4	30.4	28.9	28.3	29.4	29.0	29.1	31.0
Top 5%	15.0	15.2	15.9	17.3	16.3	16.0	16.6	16.5	16.4	17.8
Top 1%	3.3	3.4	3.6	4.0	3.7	3.5	3.7	3.7	3.7	4.0
Top 0.1%	0.4	0.4	0.4	0.5	0.4	0.3	0.4	0.4	0.5	0.4
P90/P10	13.81	17.77	18.69	18.39	18.13	17.24	17.80	15.63	18.96	18.12
S80/S20	7.17	8.21	8.76	9.21	8.91	8.47	9.58	9.20	10.26	11.26
P90/P50	1.15	1.23	1.33	1.45	1.29	1.26	1.37	1.39	1.53	1.76
Palma ratio	1.62	1.77	1.92	2.08	1.86	1.84	1.99	2.06	2.29	2.69
Gini	0.375	0.391	0.404	0.418	0.399	0.394	0.409	0.410	0.430	0.450

Source: NSI, own calculations.

According to NRA data, the share of income in the last five percentiles exceeds the similar indicator calculated on the basis of NSI data by 106% on average, as this gap has narrowed over the period under review (from 107% in 2010 to 87% in 2019). At the same time, the values of the indicator calculated on the basis of the two databases show a uniform growth dynamic in 2019 compared to 2010.

The analysis of the percentage of income of the last percentile and the richest 0.1% of the population shows an even more significant discrepancy between the values calculated on the basis of different databases. According to the NRA, the share of income of the last percentile of the population exceeds the value of the similar indicator based on the NSI data by 381% on average, by 360% in 2010 and by 333% at the end of the period. For the highest income of 0.1% of the population, this excess is 2067% in 2010 and 1560% in 2019. These results indicate that the last percentile and the richest 0.1% of the population make the largest contribution to the higher estimates for the percentage of income of the highest income 5% and 10% of the population calculated on the basis of the NRA data, compared to the similar indicators calculated on the basis of the NSI data. In general, the NSI survey data significantly underestimate the share of income of the richest 10%, 5%, 1% and 0.1% of the

population compared to the NRA data.

The calculated values for the P90/P10 and P90/P50 decile ratios, the S80/S20 quintile ratio, the palm ratio and the Gini coefficient show significantly lower values when using the NSI data. In particular, the P90/P10 coefficient is, on average, 11.3% higher than with the NRA data compared to the NSI data. It is 8% higher at the beginning of the period and this difference increases to 38% in 2019. The decile coefficient P90/P50 calculated with the NRA data exceeds the indicator calculated with the NSI data by 115% on average. In 2010, the difference was 151% and gradually decreased to 69% in 2019.

The quintile coefficient S80 / S20 calculated using the NRA data exceeds the similar coefficient calculated using the NSI data by 41% on average. After an initial decrease in the gap to 22% in 2012, there is an increase to 71% in 2017 and another decrease to 37.7% in 2019. The Palma ratio, calculated according to NRA data, exceeds the similar coefficient calculated using NSI data by 107% on average, with an excess of 135% at the beginning of the period, compared to 66.3% in 2019. The Gini coefficient calculated using NRA data exceeds the similar indicator calculated using NSI data by 29% on average, as the excess is 39% at the beginning of the period and falls to 18.1% at the end of the period.

Despite the different values of the Gini coefficient according to different databases, there are grounds to conclude that the inequality of income distribution in society in Bulgaria has a relatively high value in 2019. According to the NRA data, the Gini coefficient is 0.531, and according to the NSI data, it is 0.450. The NRA data show a more realistic picture of income inequality due to the reasons described earlier, which supports the conclusion of high-income inequality in Bulgaria.

The dynamics of the Gini coefficient, calculated on the basis of data from the NRA and the NSI, also show an identical trend of increasing inequality in the distribution of income in society. According to the NSI data, the values of the indicator vary from 0.375 at the beginning of the period to 0.45 at the end of the period, with a clear and unambiguous upward trend (Table 4). The Gini coefficient calculated with NRA data shows a more limited upward dynamic, recording high values above 0.5 throughout the period when measured with real data. In 2010, the indicator took on a value of 0.52 and rose to 0.531 at the end of 2019 (Table 3).

The conclusion about the growth of income inequality in Bulgaria is also supported by the analysis of other key indicators of the income distribution. According to the NRA data, inequality in the distribution of labour income is increasing, but not as markedly as in the NSA data. At the same time, however, the level of inequality according to the first data source exceeds the values calculated on the basis of the NSI data. The income share of the top 5%, 1% and 0.1% of the population contributes most to the increase of inequality. According to NRA data, these groups in society increase their relative weight in income distribution. The top 1% received 15.3% of labour income in the country in 2010, as this value increased to 17.4% in 2019.

It is interesting to note that the share of income received by the top decile has decreased slightly during the period, from 44.1% to 43.8%, but the stratification is still evident – the richest 10% of the population receive, on average, 44% of labour income in the country. The richest 5% group increased its share of income from 31%

in 2010 to 33.3% in 2019, according to NRA data. According to the NSI data, the top decile has significantly increased its share of income from 27.2% in 2010 to 31% in 2019. The trend is similar and significantly upward in the highest income groups: 5%, 1% and not so much for the top 0.1%. The data for all indicators are subject to empirical evaluation by calculating an nth-degree polynomial that best describes the arc of the Lorenz curve in the upper section.

The other main reason for the increase in inequality in Bulgaria in the period 2010-2019 is the decline in the share of income of the poorest five deciles (bottom 50%), who, according to NRA data, receive between 15.3% and 14.7% of income in the period studied. In the 2017-2019 period, the values fluctuate between 14.7% and 14.8%. A continuing downward trend in the share of income received by the bottom five deciles of employees is also reported according to NSI data. Values range from 23.6% in 2010 to 17.7% in 2019.

The increasing inequality in the distribution of income is reflected in the growing decile and quintile ratios as well as in the growing values of the Palma ratio. The NSI data significantly underestimate the share of income of the last quintile, decile and percentile and overestimate the share of income of the first five deciles. However, the difference between the NSI and NRA data and the indicators calculated from them has narrowed over the period under review.

The dynamics of the decile and quintile coefficients and the Palma ratio, calculated from the NSI data (Table 4), are clearly upward, while the increase in the NRA data is not as strong as the corresponding values of the indicators under consideration are high at the beginning of the period. (Table 3). The Palma ratio assumes values of 1.62 at the beginning of the period and increases to 2.69 at the end of the period according to the NSI data, while the ratio assumes values of 3.81 in 2010 and 4.47 in 2019 according to the NRA data.

Table 5 presents the NRA data on persons for whom a personal income tax is declared by income groups according to the annual tax base in the period 2010-2019. The dynamics of the absolute values reveal a significant decrease in the number of persons in the lowest range of the income scale (<= BGN 6,720). Their number decreased by 772 thousand persons or by 33.3% in the period under review. The relative share of this group in the population structure has also declined significantly – from 73.9% in 2010 to 49.5% in 2019 (Table 5).

The increase in the minimum wage in the country during the period under study is important for the identified trend, which also affects the number of people in the near upper ranges of the labour income scale. These changes are expected to lead to a redistribution and an increase in the relative weight of these levels of the income scale. The trend observed is most pronounced in the fifth range (> BGN 12,000 and <= BGN 24,000), the share of which has increased from 7.3% at the beginning of the analysed period to 19.6% in 2019.

The redistribution after the seventh range is limited to less than 1 percentage point. At the same time, the changes in the structure of the cumulative population distribution follow the observed dynamics. In 2010, 90% of people declared an annual income of up to BGN 12,000, and 10 years later, their share has fallen to 71.9%, as can be seen in Table 6.

Table 5. Share of the population with declared incomes (in %)

	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
1.	73.9	72.7	71.4	69.3	67.2	64.7	61.3	58.1	54.0	49.5
2.	2.7	2.8	2.8	2.9	2.9	3.2	3.6	3.6	3.7	3.8
3.	2.0	2.0	2.0	2.1	2.2	2.3	2.4	2.4	2.5	2.4
4.	11.4	11.8	12.2	12.8	13.3	14.1	15.3	15.6	15.9	16.2
5.	7.3	7.7	8.1	9.2	10.1	11.0	12.1	14.3	16.8	19.6
6.	1.4	1.6	1.8	1.9	2.2	2.4	2.6	3.0	3.5	4.4
7.	0.8	0.9	1.0	1.1	1.3	1.4	1.6	1.8	2.1	2.4
8.	0.27	0.30	0.35	0.38	0.43	0.48	0.57	0.67	0.81	1.00
9.	0.07	0.08	0.09	0.10	0.12	0.13	0.14	0.16	0.20	0.24
10.	0.08	0.09	0.10	0.11	0.13	0.14	0.16	0.19	0.22	0.25
11.	0.03	0.04	0.04	0.05	0.06	0.06	0.07	0.08	0.09	0.10
12.	0.02	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.05	0.06
13.	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03
14.	0.03	0.04	0.04	0.05	0.06	0.07	0.07	0.08	0.09	0.10

Source: NRA, own calculations.

The highest income group (> BGN 360 thousand) is also characterised by a steady positive trend, as the number of people in this group increased by 224.5% compared to 2010, and its relative share reached 0.1% in 2019, while it was 0.03% in 2010 (Table 5).

Table 6. Cumulative distribution of income in accordance with the declared incomes and the cumulative distribution of the population by ranges (in %)

	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
1.	32.7	31.2	30.0	28.3	26.0	24.4	22.1	19.6	17.0	14.5
2.	35.7	34.1	32.8	31.1	28.7	27.0	24.9	22.2	19.4	16.7
3.	38.1	36.4	35.0	33.2	30.7	29.1	26.9	24.0	21.2	18.2
4.	55.9	53.7	52.0	50.1	46.9	45.3	43.2	39.4	35.4	31.4
5.	75.0	72.7	70.9	70.3	67.5	66.4	64.8	62.7	60.6	58.2
6.	81.6	79.7	78.3	78.0	75.5	74.5	73.2	71.5	70.0	68.7
7.	87.2	85.7	84.7	84.7	82.6	81.8	81.0	79.7	78.7	77.8
8.	90.5	89.1	88.4	88.5	86.6	86.0	85.6	84.7	84.2	83.9
9.	91.7	90.4	89.9	90.1	88.3	87.6	87.4	86.5	86.2	86.0
10.	93.5	92.4	92.1	92.3	90.7	90.1	90.0	89.2	89.1	89.1
11.	94.6	93.6	93.3	93.6	92.2	91.7	91.6	91.0	90.8	90.9
12.	95.5	94.5	94.2	94.5	93.3	92.8	92.8	92.1	92.1	92.1
13.	96.1	95.1	94.9	95.3	94.1	93.6	93.5	93.0	93.0	93.0
14.	100	100	100	100	100	100	100	100	100	100

Source: NRA, own calculations.

Table 6 presents the cumulative income distribution calculated in accordance with the declared incomes for the tax base under the personal income tax (see Table 12 of the

Annex) and the cumulative distribution of the population by ranges. All income groups show a decrease in their relative weight in the structure of the cumulative distribution of income in 2019 compared to the beginning of the period. The manifestation of this tendency gradually increases from the first to the fourth income group, after which it begins to slow down. In particular, when comparing 2019 to 2010, the share of the lowest income group has declined by 18.2 percentage points, the cumulative share of the first four groups has declined by 24.5 percentage points, while the cumulative share of the penultimate range has recorded a decline of only 3.1 percentage points (Table 6).

The calculated average annual declared incomes for the tax base in Table 10 of the Annex are relevant to the distribution of the population. For example, in 2010, 73.9% of the population had an average annual declared income for the tax base of BGN 2,730, from which the net income paid is obtained after the deduction of income tax. At the end of the analysed period, 49.5% of individuals declared income in the first range, forming an arithmetic mean of BGN 3,514, which is an increase of 28.9% compared to 2010. At the same time, the changes in the average annual incomes declared for the tax base in the higher income ranges vary between + 2.8% and -0.4%, with a significant increase of 8% only in the latter range.

Table 7. Gini coefficient calculated with NRA data

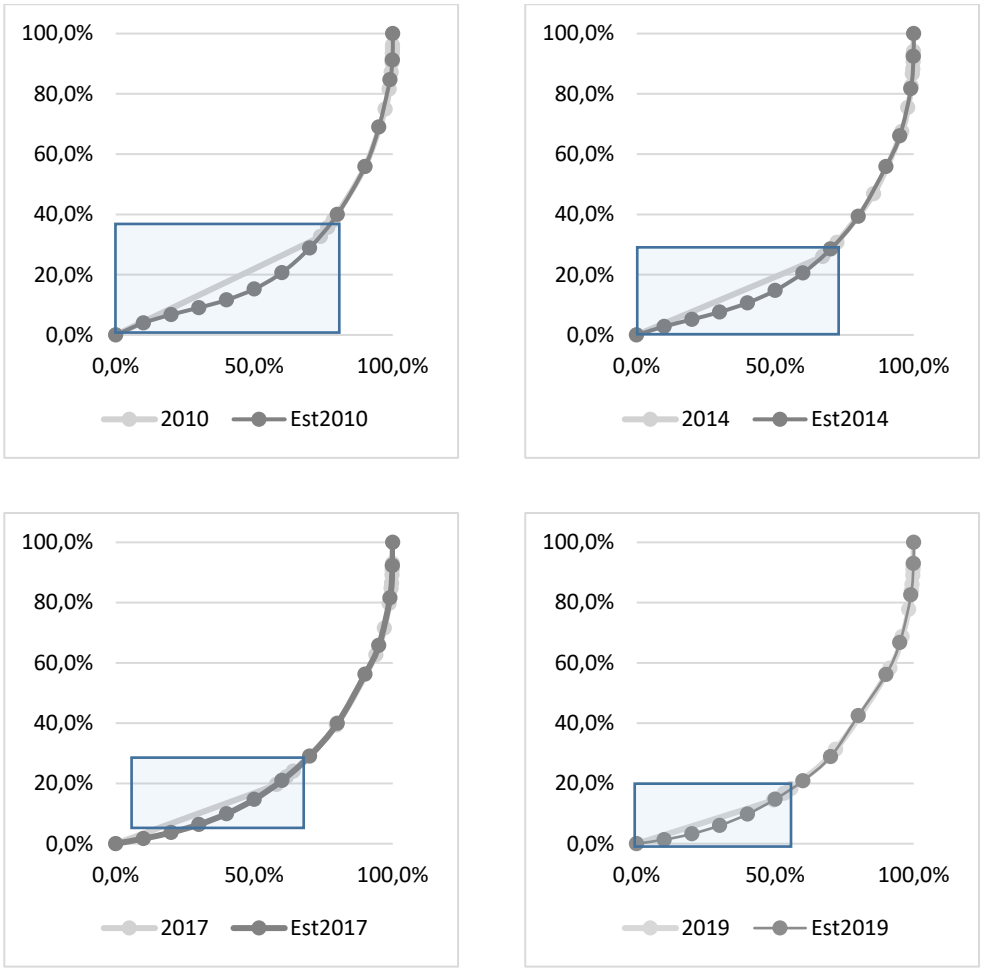
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Gini (decile)	0.520	0.525	0.520	0.521	0.530	0.525	0.529	0.533	0.525	0.531
Gini (NRA)	0.478	0.489	0.494	0.496	0.511	0.514	0.517	0.526	0.531	0.532

Note: Gini (decile) – The Gini coefficient with estimated decile data, based on NRA data, using n- degree polynomials for the part of the cumulative income distribution curve (Lorenz curve) that cannot be calculated directly from NRA data; *Gini (NRA)* **–** Gini coefficient calculated based on NRA data with a constant slope for the missing part of the cumulative income curve.

Source: NRA, own calculations.

The values of the Gini coefficient, revealed in Table.7, using the NRA data without estimating the missing deciles are lower than the Gini coefficients calculated with estimated data for the missing lower deciles, which are shown in Table 1 and Table 3. Using estimates for the first four or seven deciles for different years (for each year, the missing initial deciles are different) allows for a different slope of the Lorenz curve in its missing part, as shown in Figure 3. Table 7 clearly shows that the values of the Gini coefficient calculated with NRA data (second row) and the assumption of a constant slope of the missing part of the Lorenz curve are lower. The Gini coefficient with estimated decile data based on NRA data and using n-degree polynomials, has values corresponding to the first row in the table. It can be argued that the Gini coefficient calculated by estimating the missing part of the cumulative distribution curve with the real NRA data in the first row actually estimates the inequality of the income distribution more accurately. There is an upward trend in both Gini coefficients (Table 9), with the lowest values of the coefficient in the early years when a large proportion of the population (over 70%) earns an income in the first income range (see Table 6). During the period 2010-2019, the share of employees with incomes below BGN 6 720 per year has decreased, respectively, an increasingly smaller part of the Lorenz curve

remains without real data, so that an increasingly smaller part of it has to be estimated. The Gini coefficient data, calculated by evaluating the missing deciles and presented in the first row of Table 7, shows a high level of inequality that has hardly changed during the period under review.

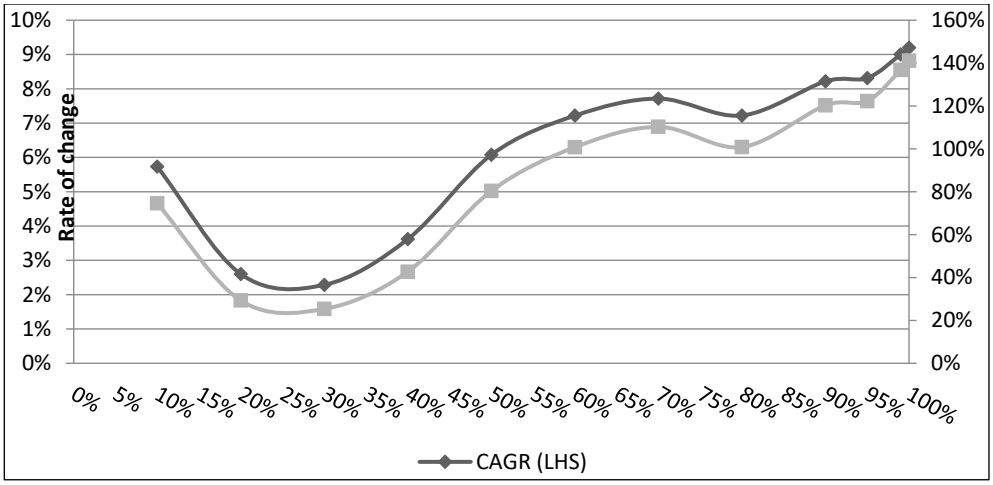


Note: With *2010, 2014, 2017, and 2019*, the cumulative distribution (Lorenz curve) data are presented without estimating the values for the missing initial deciles, allowing a proportional increase from the coordinate origin to the available first values; With *Est2014; Est2014, Est2017, Est2019*, the cumulative distribution (Lorenz curve) data are presented and the values for the missing initial deciles are estimated using n-degree polynomials.

Source: NRA, own calculations.

Figure 3. Cumulative income distribution curve (Lorenz curve), estimated with NRA data

Figure 3 shows the estimated values for the missing initial deciles of the cumulative income distribution obtained from the NRA data using n-degree polynomials, which gives the Lorenz curve a variable slope, respectively it is flatter at the beginning and steeper in the assessed section. The estimated part of the Lorenz curve is closer to reality, where it is normal that there is no proportional distribution with an equal share. In the outlined section of the diagrams for selected years, it can be seen that the inequality in the estimated function of the Lorenz curve is greater than with the purely mechanical connection of the first available values of the distribution with the coordinate origin and assuming that the function is linear.



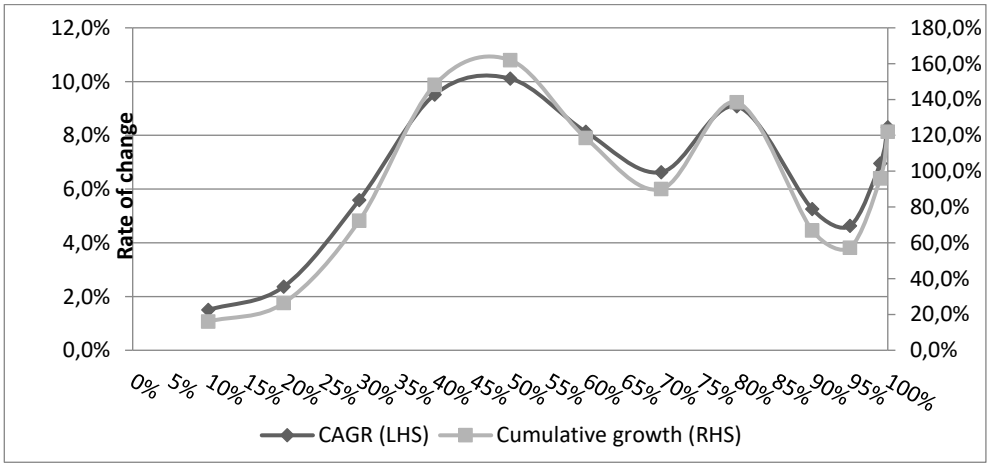
Note: CAGR – geometric mean of growth rate; Cumulative growth – Cumulative growth rate; RHS – right scale; LHS – left scale.

Source: NSI, own calculations.

Figure 4. Rate of change of the incomes of the individual deciles/percentiles according to NSI data

Figure 4 presents the average annual rate of change (geometric mean) in the nominal income of the first to ninth deciles, the income of the 90-95% percentiles, the next four percentiles and the last percentile over the period 2010-2019 using data from the NSI. The NSI data are used to calculate the geometric mean of the rate of change of the first nine deciles, as n-degree polynomials are used for the highest percentiles, which best describe the curve of the actual cumulative distribution (Lorenz curve) in its upper section.

The NSI data suggest that the nominal average annual rate of change for the deciles and percentiles studied ranges from 2 to 9.2%. The first decile increased its income on average by 5.7% annually, while the second, third and fourth deciles increased their income on average by 2 to 4%. The highest growth rate is recorded by the sixth, seventh, ninth, and above all, the tenth decile, especially the 95-99th and the last percentile. It is this dynamics among the individual decile and percentile groups that underlies the steady growth in income inequality as measured by the various decile/quartile ratios and the Gini coefficient.



Note: CAGR – geometric mean of growth rate; Cumulative growth – cumulative growth rate; RHS-right scale; LHS – left scale.

Source: NRA, own calculations.

Figure 5. Rate of change of the incomes of the individual deciles/percentiles according to NRA data

Figure 5 presents data on the average annual rate of change in the nominal incomes of the first to ninth deciles, the incomes of the 90-95% percentiles and the last two percentiles over the period 2010-2019, but using data from the NRA. Data for the first five to seven deciles in different years are estimated by calculating polynomials of n-degree that best describe the arc in the more sloping part of the Lorenz curve. Single polynomials of n-degree are used to calculate the function of the cumulative distribution in its steep part to obtain data on the corresponding decile and percentile values.

The rate of change is lowest in the first decile (1.5%), and from the second to the fifth decile, growth gradually accelerates to an average of 10.1% per year. In the sixth and seventh deciles, the income growth rate falls to 8.1% and 6.6%, respectively, and rises to 9.1% in the eighth decile. The growth rates for the ninth decile and for the five percentiles between 90 and 95% are lower, 5.3% and 4.6%, respectively, while the next 4 percentiles between 95 and 99% and the last percentile increase their growth rate – to 7% and 8.3% respectively.

The curve of the cumulative growth rate of income for each percentile and decile shown in Figure 5 largely corresponds to the cumulative growth curves of Figure 1a and Figure 1.b of Lakner and Milanovic (2015) and has the shape of the so-called "Elephant curve". On the two curves, Lakner and Milanovic (2015) plot global growth in inequality as a rate of change in income of each decile and percentile for the period 1988-2008, as well as for shorter sub-periods. According to the NRA data, the estimated first seven deciles in 2010 and the first five deciles in 2019 show a gradual increase in the growth rate (annual average and cumulative for the period 2010-2019), followed by a downward trend, except for the eighth and tenth deciles, with an increase in pace, particularly in the last 5 percentiles.

5. Discussion

Possible explanations for the significant difference in the measures of inequality in Bulgaria should be sought in three directions: first, based on differences in methodology and scope of statistical surveys; second, differences in income covered and, third, intentional deviations in the declared income to the NRA due to tax reasons. First of all, it is difficult to find representatives at both ends of the distribution, respectively, the lowest-income earners and the highest-income earners either do not want to take part in surveys or do not disclose real income data. In addition, the probability of a random cluster sample of 3 060 households covering the above 1% or 0.1% should be assessed. This problem is more pronounced with a strongly distorted distribution (Yonzan et al., 2022). In particular, the Household Budget Survey in Bulgaria has been calibrated to ensure representativeness in terms of indicators such as the number of persons in the household and regional disparities, rather than in terms of income inequalities.

Second, the comparison should take into account the fact that the NRA data refers to income from salary. Wage income inequality is growing faster than the average inequality in NSI surveys.

Third, the analysis is based on the assumption that the NRA data reflects to a greater extent the actually paid labour incomes and their distribution, as they are not based on a sample survey. However, there are also circumstances that lead to an inconsistency between the declared and the actually received wage income. The first is related to the influence of the grey economy, which is expressed in the receipt of tax-free income, both due to undeclared work and due to declaring lower wages than actually received. The second circumstance is related to the tendency of owners of profitable companies (ET, OOD, etc.) to declare labour income instead of dividend income, as in Bulgaria the effective tax-social insurance burden for wages significantly exceeding the ceiling of insurance income is lower than the sum of taxes on profit and income, distributed as a dividend, respectively – from the tax of the sole proprietor. The extent to which these circumstances affect the income near the minimum and the highest wages, may be the subject of additional studies, which are not currently conducted in our country.

The possible interpretation of the different dynamics of the declared to NRA labour incomes and the incomes estimated on the basis of the NSI surveys presupposes to assess the influence of factors such as lightening of the economy, growth of minimum and average wages, etc.

6. Conclusion

The performed analysis reveals that the NSI survey data underestimate the share of income in the highest income deciles and percentiles and overestimate the income of the first deciles, i.e., the lower-income groups. In the present analysis, it is assumed that with a high degree of reliability, the data of the NRA reflects the actually incurred and paid labour incomes and their distribution. The comparison shows that the inequality in the distribution of labour income is far greater than the NSI data suggest. During the period under review, the Gini coefficient, calculated with NSI data, assumes an average value of 0.41, compared to 0.53 with NRA data. The highest income

percentile receives an average of 3.7% of total income according to NSI data, compared to 17.6% for NRA data. The highest income decile receives an average of 27.5% of total income according to NSI data, while NRA data points to an average share of 44%. Palma ratio is, on average, 104% higher, calculated with NRA data, in comparison with NSI data. The P90 / P50 ratio is on average 41% higher, while the S80/S20 and P90/P10 ratios are on average 41% and 11% higher.

The findings of the present study confirm the hypothesis, as well as the results of international research, that survey data tend to underestimate the real income distribution, especially in the upper decile and percentiles.

The combination of survey data from the Household Budgets Survey in the Republic of Bulgaria and the EU-SILC survey, on the one hand, and NRA data on actually accrued and paid wages will allow to get a more accurate idea of the income distribution, in particular – wage income.

Conflict of interest

The authors declare no conflict of interest.

References

Bogdanov, B. (2022). Nablyudenie na domakinskite byudzheti v Balgariya: analiz, problemi, izvodi i perspektivi za razvitie. Statistica, No. 1 [Богданов, Б. Наблюдение на домакинските бюджети в България: анализ, проблеми, изводи и перспективи за развитие. *Статистика*, No. 1]. Available at https://www.nsi.bg/spisaniestatistika/page/download.php?articleID=283&lang=EN (*in Bulgarian*).

Bratoeva-Manoleva, S. (2021). Dohodnoto neravenstvo v Balgariya v perioda 1993-2019 g. – dekompozirane po osnovni iztochnitsi na dohod. *Dialog*, 1, 2021, 64-84. [Братоева-Манолева, С. (2021). Доходното неравенство в България в периода 1993-2019 г. – декомпозиране по основни източници на доход. *Диалог*, 1, 64-84] (*in Bulgarian*).

Nenkova, P. (2021). Podohodno neravenstvo i prerazpredelitelno vazdeystvie na darzhavnata namesa v evropeyskite strani. *Biznes posoki*, No. 1. [Ненкова, П. (2021). Подоходно неравенство и преразпределително въздействие на държавната намеса в европейските страни. *Бизнес посоки,* No. 1]. Available at https://bjournal-bfu.bg/uploads/posts/2021_1_25_42_ bg.pdf (*in Bulgarian*).

Tsanov, V., Bogdanov, B. (2012). Evolyutsiyata na neravenstvata kato predizvikatelstvo kam sotsialnata politika, *Statistica* No. 3-4. [Цанов, В., Богданов, Б. (2012). Еволюцията на неравенствата като предизвикателство към социалната политика. *Статистика*, No. 3-4]. Available at https://www.nsi.bg/ spisaniestatistika/page/bg/details.php?article_id=25&tab=en (*in Bulgarian*).

Tsvetkov, S. (2020). Statistichesko otsenyavane na podohodnoto neravenstvo na naselenieto. *Statistica*, No. 4. [Цветков, С. (2020). Статистическо оценяване на подоходното неравенство на населението. *Статистика*, No. 4]. Available at

https://www.nsi.bg/spisaniestatistika/page/download.php?articleID=250&lang=EN (*in Bulgarian*).

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., Zucman, G. (2018). The elephant curve of global inequality and growth. *AEA Papers and Proceedings*, Vol. 108, 103-08.

Bricker, J., Henriques, A., Krimmel, J., & Sabelhaus, J. (2016). Measuring income and wealth at the top using administrative and survey data. *Brookings Papers on Economic Activity*, 1, 261-331.

Brzezinski, M. (2018). Income inequality and the Great Recession in Central and Eastern Europe. *Economic Systems*, 42(2), 219-247.

Burkhauser, R. V., Feng, S., Jenkins, S. P., & Larrimore, J. (2012). Recent trends in top income shares in the United States: reconciling estimates from March CPS and IRS tax return data. *Review of Economics and Statistics*, 94(2), 371-388.

Carvalho, L., Rezai, A. (2016). Personal income inequality and aggregate demand. *Cambridge Journal of Economics*, 40, Issue 2, 491-505, DOI: 10.1093/cje/beu085.

Iacono, R., Ranaldi, M. (2021). The Evolution of Income Composition Inequality in Italy, 1989-2016. *Review of income and wealth*. DOI: 10.1111/roiw.12555

Lakner, C., & Milanovic, B. (2015). Global income distribution from the fall of the Berlin Wall to the great recession. *Revista de Economía Institucional*, 17(32), 71-128.

Mavrov, H. (2021). Dohodnoto neravenstvo v barzo razvivashtite se ikonomiki. *Nauchni trudove na UNSS*, No. 1, 401-411. [Мавров, Х. (2021). Доходното неравенство в бързо развиващите се икономики. *Научни трудове на УНСС*, No. 1, 401-411]. Available at http://unwe-research-papers.org/uploads/Research Papers/RP_vol1_2021_No17_H%20Mavrov.pdf (*in Bulgarian*).

Mihaylova, S., Bratoeva-Manoleva, S. (2017). Social transfers and income inequality in Bulgaria. *The South-East European Journal of Economics and Business*, 12(1).

Mintchev, V., Boshnakov, V., Naydenov, A. (2010). Sources of Income Inequality: Empirical Evidence from Bulgaria. *Economic studies*, 19(4).

Morgan, M., Neef, T. (2020). What's New About Income Inequality in Europe? (1980-2019). *World Inequality Lab Issue Brief*, 2020/04.

Papanikolaou, N. (2021). Tax Progressivity of Personal Wages and Income Inequality. *Journal of risk and financial management*, 14, 2, DOI: 10.3390/jrfm14020060

Ranaldi, M., Milanović, B. (2022). Capitalist systems and income inequality. *Journal of Comparative Economics*, 50, 1, 20-32, 10.1016/j.jce.2021.07.005

Sabelhaus, J., Johnson, D., Ash, S., Swanson, D., Garner, T., Greenlees, J., Henderson, S. (2015). *Is the Consumer Expenditure Survey Representative by income? Vol. 74*. Chicago: University of Chicago Press.

Socci, C., D'Andrea, S., Deriu, S., Pretaroli, R., Severini, F. (2021). *Does the Personal Income Flat Tax fit with Economic Growth and Inequality in Italy?* DOI: 10.1007/s40797-021-00149-0

Tosheva, E., Tasseva, I., Draganov, D., & Boshnakov, V. (2016). Effects of changes in tax-transfer system on households income distribution in Bulgaria: simulation analysis using EUROMOD for 2011-2015. *Economic Thought Journal*, (5), 51-71.

Yonzan, N., Milanovic, B., Morelli, S., & Gornick, J. (2022). Drawing a line: comparing the estimation of top incomes between tax data and household survey data. *The*

Journal of Economic Inequality, 20(1), 67-95.

Data Sources

Response to Applications for access to public information with ent. № ЕП-94-П-1152 dated 04.12.2020, ent. № М-94-П-1153 dated 04.12.2020 and ent. № ЕП-33-00-300 dated 07.12.2020 according to the inventory of the Central Office of the National Revenue Agency.

NSI. *Households Income, Expenditure and Consumption,* https://cutt.ly/UGjeYuY.

NSI. *Poverty and Social Inclusion Indicators – National Level,* https://cutt.ly/BGjeFX5.

World Development Indicators (WDI), https://data.worldbank.org/.

World Inequality Database (WID), https://cutt.ly/cGje86q, Methodology: https://wid.world/document/2020dina-update-for-europe-world-inequality-lab-technical-note-2020-04/.

Petar Peshev, PhD, is Senior Assist. Professor at the University of National and World Economy, Department of Economics, p.peshev@unwe.bg (corresp. author)

Kristina Stefanova, PhD, is Senior Assist. Professor at the University of National and World Economy, Department of Economics, k.petrova@unwe.bg

Ivan Bozhikin, PhD, is Senior Assist. Professor at the University of National and World Economy, Department of Economics, i.bozhikin@unwe.bg

Radostina Stamenova is a PhD student at the University of National and World Economy, Department of Economics, radostina_stam@unwe.bg

Ivanina Mancheva is a PhD student at the University of National and World Economy, Department of Economics, iva.mancheva@unwe.bg

How to cite this article:

Peshev, P. et al. (2022). Is Income Inequality in Bulgaria Underestimated in Survey Data? *Economic Thought Journal*, 67(3), 301-326.

Appendix

Table 8. Cumulative decile distribution of income, estimated with NRA data (in %)

MF	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
10	3.0	2.9	2.8	2.7	2.6	2.5	1.9	1.8	1.8	1.8
20	6.0	5.8	5.6	5.3	4.9	4.7	4.0	3.8	3.8	3.7
30	8.7	8.4	8.5	8.1	7.6	7.3	6.7	6.5	6.5	6.1
40	11.6	11.3	11.8	11.2	10.6	10.5	10.2	10.0	10.1	9.8
50	15.3	15.0	15.7	15.2	14.8	15.0	15.0	14.8	15.0	14.8
60	20.7	20.6	21.0	21.0	20.6	21.1	21.2	21.1	21.4	20.9
70	28.9	28.5	28.7	29.0	28.5	29.1	29.3	29.2	30.1	28.9
80	40.0	40.1	40.3	40.8	39.4	40.5	40.5	40.1	41.6	42.5
90	55.9	54.8	55.7	56.6	55.9	56.3	56.5	56.4	56.7	56.2
100	100	100	100	100	100	100	100	100	100	100

Source: NRA, own calculations.

Table 9. Cumulative decile distribution of income, estimated with NSI data (in %)

HHB	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
10	2.8	1.5	1.5	1.7	1.6	1.6	1.7	1.9	1.5	1.7
20	7.8	5.4	5.2	5.2	5.2	5.4	4.9	5.0	4.6	4.4
30	14.2	9.9	9.5	9.3	9.9	9.9	9.5	9.1	8.4	7.4
40	21.5	15.5	14.8	14.6	15.5	15.3	14.7	14.1	12.7	11.5
50	29.9	22.3	21.4	21.0	22.3	22.4	21.4	20.8	19.1	17.7
60	39.3	30.7	29.8	28.9	30.5	30.9	29.5	29.1	27.3	26.1
70	50.0	41.1	40.2	38.8	40.6	41.4	40.0	40.3	38.4	37.0
80	62.3	55.3	54.1	52.3	53.9	54.3	53.3	53.8	52.4	50.5
90	77.0	72.6	71.6	69.6	71.1	71.7	70.6	71.0	70.9	69.0
100	100	100	100	100	100	100	100	100	100	100

Source: NSI, own calculations.

Table 10. Average annual declared incomes for tax base for the respective year (ths. BGN)

	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
1.	2.73	2.81	2.91	3.00	3.07	3.17	3.28	3.34	3.42	3.51
2.	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0
3.	7.6	7.6	7.6	7.6	7.6	7.6	7.6	7.6	7.6	7.6
4.	9.6	9.6	9.6	9.7	9.7	9.7	9.7	9.7	9.7	9.7
5.	16.1	16.2	16.2	16.1	16.1	16.1	16.1	16.1	16.2	16.5
6.	28.9	28.9	28.9	28.9	28.9	28.9	28.9	28.9	28.8	28.8
7.	45.0	45.0	45.0	45.0	45.1	45.2	45.3	45.4	45.5	45.5
8.	74.1	74.1	74.2	74.1	74.0	74.1	73.9	73.9	74.0	73.8
9.	106.9	107.2	106.9	107.2	107.0	107.1	107.0	107.1	107.2	106.9
10.	145.2	145.5	144.9	144.7	145.3	145.0	144.7	144.9	144.6	144.7
11.	206.9	207.6	207.3	206.8	206.5	206.8	207.2	207.0	206.7	207.5
12.	267.1	268.3	267.8	269.3	268.0	268.8	266.9	266.6	266.6	267.6
13.	329.0	326.6	326.8	329.8	329.0	328.2	328.5	328.5	328.6	328.3
14.	775.1	796.4	792.3	767.3	813.1	821.9	840.6	859.8	844.5	837.4

Source: NRA, own calculations.

Table 11. Number of persons for whom personal income tax has been declared for the respective year by ranges

Tax year	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Range in BGN thousand	Number, in thousands of people									
1. >0 and <=6.7	2319.3	2268.4	2202.8	2136.7	2060.5	1989.6	1896.9	1822.7	1700.1	1547.2
2. >6.7 and <=7.3	83.6	86.6	86.2	89.1	90.3	98.8	110.4	112.5	115.8	119.8
3. >7.3 and <=7.8	62.1	62.1	62.8	63.7	66.2	71.4	75.2	76.6	79.3	75.7
4. >7.8 and <=12	358.9	367.6	377.3	395.0	406.3	434.7	474.2	488.6	501.1	505.3
5. >12 and <=24	229.1	240.0	250.6	283.1	310.9	337.9	376.2	448.7	529.6	611.7
6. >24 and <=36	44.2	49.2	54.5	60.0	67.0	73.2	81.8	95.2	111.4	136.9
7. >36 and <=60	24.4	27.3	30.7	33.7	38.5	41.7	48.2	55.9	65.3	74.6
8. >60 and <=96	8.4	9.4	10.8	11.8	13.2	14.8	17.7	20.9	25.5	31.3
9. >96 and <=120	2.2	2.6	2.9	3.2	3.7	3.9	4.5	5.1	6.3	7.4
10. >120 and <=180	2.5	2.8	3.2	3.4	4.1	4.5	5.1	5.8	6.8	7.9
11. >180 and <=240	1.1	1.2	1.3	1.4	1.7	2.0	2.3	2.6	2.9	3.3
12. >240 and <=300	0.6	0.7	0.7	0.8	1.0	1.0	1.2	1.4	1.6	1.8
13. >300 and <=360	0.4	0.4	0.5	0.5	0.6	0.6	0.7	0.8	0.9	1.0
14. >360	1.0	1.3	1.4	1.4	1.8	2.0	2.2	2.5	2.9	3.1

Source: NRA.

Table 12. Declared amounts for tax base by income ranges

Tax year	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Income ranges in BGN	BGN million									
1. >0 and <=6720	6322.7	6371.7	6421.2	6400.9	6325.0	6307.7	6224.9	6083.2	5810.7	5436.2
2. >6720 and <=7320	586.1	607.6	604.5	625.1	633.6	693.1	774.0	788.7	811.7	839.1
3. >7320 and <=7800	468.8	469.1	474.7	481.0	500.3	539.3	568.5	579.0	598.8	571.4
4. >7800 и <=12000	3442.5	3527.8	3628.5	3818.6	3922.3	4198.4	4592.4	4753.5	4870.8	4919.7
5. >12000 and <=24000	3692.6	3885.7	4058.3	4562.8	5010.9	5448.1	6068.8	7222.8	8586.0	10088.8
6. >24000 and <=36000	1279.2	1422.6	1576.2	1735.7	1933.2	2112.0	2360.4	2747.0	3214.9	3940.8
7. >36000 and <=60000	1095.6	1227.7	1383.8	1515.9	1734.6	1882.0	2186.0	2535.5	2968.4	3389.2
8. >60000 and <=96000	622.1	696.4	800.4	872.2	978.5	1097.6	1308.8	1546.9	1882.6	2306.7
9. >96000 and <=120000	234.1	278.6	311.6	342.9	393.6	415.2	479.6	549.4	670.5	787.8
10. >120000 and <=180000	355.5	400.2	462.2	495.8	595.3	644.3	738.7	842.9	984.3	1145.0
11. >180000 and <=240000	222.0	248.2	266.7	296.8	353.4	405.9	465.1	545.4	606.0	674.7
12. >240000 and <=300000	170.1	177.9	194.7	213.5	266.1	277.9	315.5	365.6	416.1	473.6
13. >300000 and <=360000	115.8	130.0	155.6	168.6	204.0	207.1	223.4	273.3	302.0	323.4
14. >360000	750.3	997.9	1083.8	1071.9	1431.9	1663.5	1814.0	2165.0	2405.9	2629.6

Source: NRA.