MEASURING MACRO-LEVEL INEQUALITY: THE ROLE OF THE TOP 1%

Ilona Cserháti, Tibor Keresztély

Department for Research, Office of Public Administration and Justice, Budapest, Hungary

Corvinus University of Budapest, Hungary

Abstract

Income inequality has been rising in the last decades. It is generally measured on micro-level by income surveys, but the richest 1% is partly missing from the survey samples. We used the Hungarian tax database for calculating 'macro-level' income distribution and the share of the top 1% for 2010-2014. We aligned the original EU-SILC questionnaire database, and compared the original and the updated inequality measures. We analysed the effects of the integration, with special regard to the role of the top 1%.

Key words: *income inequality, microsimulation, taxation*

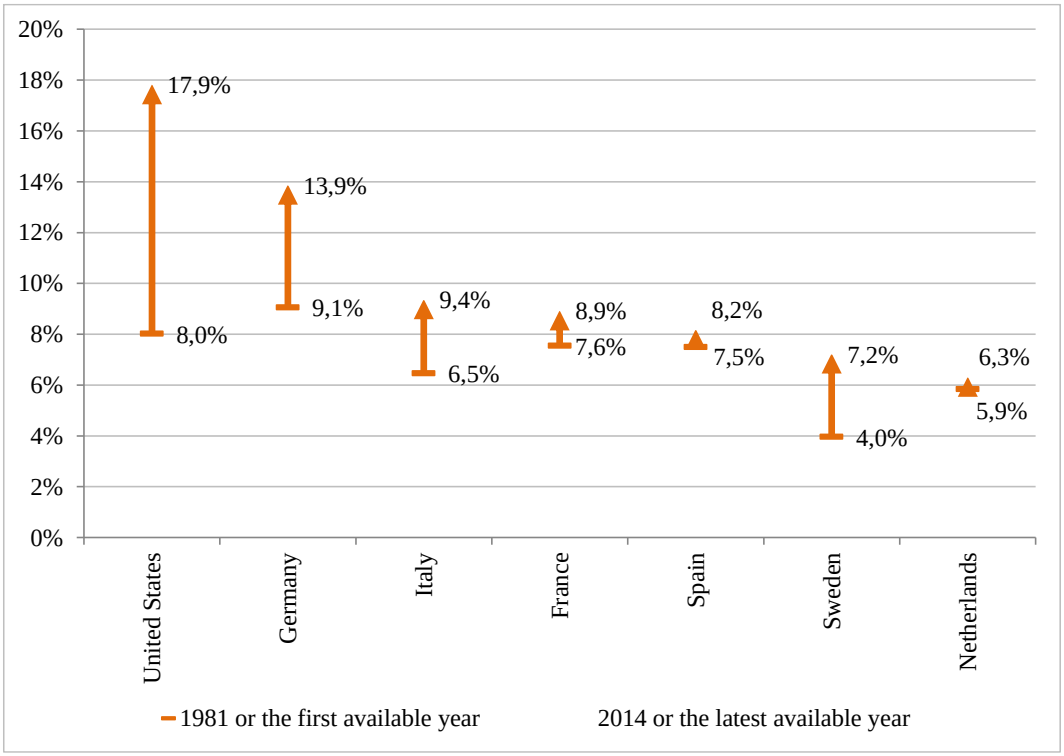
1. INTRODUCTION

The average disposable income of the richest 10% in OECD countries reached around 9½ times higher level than that of the poorest 10% in 2012, and the income share of the richest 1% is also sharply increasing. Inequality is generally measured on micro-level by income surveys, but the richest 1% is obviously almost completely missing from the survey samples. If one wishes to follow the real changes in the macro-level income disparities should integrate micro information with national accounts or tax return data. We used the Hungarian tax return database for calculating so-called 'macro-level' income distribution and the income share of the top 1% for the period of 2010-2014. Based on this information we aligned the original Hungarian EU-SILC (Survey of Income and Living Conditions for the European Union) questionnaire database for the same period. We compared the original and the updated income inequality measure trends; we analysed the effects of integration of macro-level information, with special regard to the role of the top 1%. Finally, we drew some conclusions and gave some future research plans in measuring macro-level inequality.

2. THE ROLE OF TOP 1%

The top percentile income share in pre-tax income has significantly increased in the last decades in most of the OECD countries as it was highlighted in Piketty (2014) and Förster-Llena-Nozal & Nafilyan (2014). By the latest information provided by World Top Income Database developed by Alvaredo-Atkinson-Piketty-Saez & Zucman (2016) it seems that - after a temporary halt due to the crisis - the share of the top income recipients in total gross earnings has been further increasing. In the United States the share the top 1% increased quite extremely: from 8% in 1981 to 17.9% in 2014. The trends are similar in Europe but the levels are much smaller and even in Sweden (that country is characterized by a widespread social welfare system and low level of inequalities) 3.2 percentage point increase can be seen for the analysed period on Figure 1.

Figure 1. Shares of top 1% incomes in total pre-tax income, 1981-2014 (or closest)



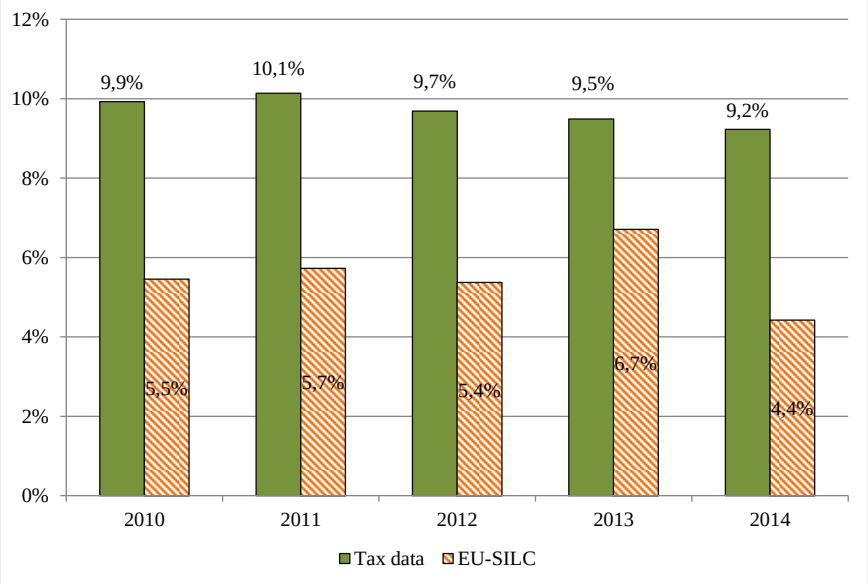
Note: Latest year refers to 2013 for Sweden, 2012 for France, Spain and the Netherlands, 2009 for Italy and 2008 for Germany.

Source: own calculation based on the World Top Income Database.

These trends are followed by the fact that the growth rate of income are tend to be much higher for the top 1% then the bottom 99%. Globalisation can be one of the most important factors behind these tendencies. Progressive tax systems could have contribute to make this accelerating process slower, but instead flat personal income tax rate systems was implemented in several European countries. Until the crisis a general decline in the top statutory rates was observable and capital income tax rates also tended to decrease. Based on all these factors one can conclude that fiscal policies in Europe usually did not change the stance of income distribution and contributed the acceleration in overall inequality in net disposable income of households.

Hungary is usually said to be a country characterized by relatively low inequalities (for example the published Gini index for Hungary (28.6) is significantly less than the EU average (30.9) for 2014). The pre-tax income share of the top 1% can be evaluated also as quite low based on the EU-SILC survey, but it almost doubled if the income tax return database was used - as can be seen in Figure 2.

Figure 2. Shares of top 1% incomes in total pre-tax income, 2010-2014



Source: own calculations based on EU-SILC and tax return databases for Hungary

Based on these facts it seems to be essential to take into consideration the most realistic available top income figures if we aim at following more accurate income distribution and income inequality trends in an economy.

3. METHODOLOGY: MEASURING INEQUALITY TAKING BETTER ACCOUNT OF TOP 1%

In the European Union and in the OECD countries survey data are the principal sources of measuring inequalities. The main question is how reliable and accurate these survey-based estimates can be in determining level and trends in inequalities. The main problems with the survey data is the under-coverage in general and especially at top incomes. There are at least two types of unreliability: underreporting of income at the top and not covering some part of top range at all. Regarding the first type of under-coverage you can reweight the observations and upright the level of income based on more reliable data sources. Regarding the second case the methodological issue is even more complicated, because you cannot simply modify the income levels of missing observations.

Evidence suggests that top 1% share has been on the rise in the developed world, but household surveys capture only the other 99%. It requires using alternative data sources, as well. Using information from income tax return data base can be a good alternative; because it provides a full picture of distribution of market incomes and personal top incomes are also mostly included (although incomes related to non-observed economy are missing).

There are several methodologies used in the literature to address under-coverage at the top incomes.

The first option is using them separately: calculating the usual income inequality measures based on the survey data while estimating the top income shares from the tax return data. Using this approach, you could avoid methodological problems with combining different database but you would probably underestimate the real inequality level at the same time.

The second approach suggested by Atkinson, Piketty and Saez (2011) and Jenkins (2016) is to calculate income inequality measures for poorest 99% based on the survey data and to do the same calculations for the top incomes based on tax data our other estimations. Even if you have only grouped top income data you can fit a Pareto distribution to them and based on that distribution Gini

measure can be calculated for the top incomes. Finally, you can combine Gini for the total population based on the decomposition formula developed by Atkinson (2007) that the total inequality equals to the weighted sum of inequality of top 1% and the bottom 99% plus the inequality between these two groups.

Having the full dataset for the top earners one can choose a third approach: adjusting survey data using tax data by cell mean imputation i.e. replacing the original survey data by gross incomes of the richest people in the tax return database (Burkhauser-Hérault-Jenkins & Wilkins, 2016). Following that methodology new inequality measures can be calculated based on the new combined database. The higher the top income share is the greater the differences in Gini coefficient based on the original survey and the imputed datasets are. Using this third approach you have to find the adequate cross walk between income definitions in survey and tax data.

Based on the World Top Income Database the role of the top 1% share is definitely increasing, i.e. under-coverage of top incomes by surveys tends to get more significant. The main question is that taking better account of top 1% changes only the level of inequalities or changes the survey based inequality trends also. In our research we followed the third approach (imputing survey data by the tax data) searching for the more accurate picture of inequality trends in Hungary.

4. ANALYSIS OF INEQUALITY TRENDS IN HUNGARY

We have a complete – anonymized – database about all the tax payers in Hungary, so we were able to apply the third approach suggested by Jenkins (2016) and Burkhauser (2016). But before starting this imputation we checked whether not only the distribution but also the level of income is different in our data survey regarding the tax data. We have found a relevant difference between the total means (more than 11%), so our first step was a complete imputation regarding the whole survey data. The second step was the imputation of the top 1% according the Jenkins' third approach. Finally, we have made a third step, which was a correction concerning percentiles 91-99 (the top 10%, excluding the top 1%). Hereunder we explain these three steps in detail.

Before the first step we had to define the scope of income category we would like to analyze. We have chosen the so-called market income, which is the sum of employee cash or near cash income, cash benefits from self-employment and unemployment benefit.

Then we identified the factors which explained best the difference between the survey (EU-SILC) and tax data regarding to the market income. We have found three (qualitative) variables: the region, the age group and the income decile. Tables 1-3 show the differences for each category of these variables.

Table 1. Market income by region (HUF/year or ratio)

	Tax data	**EU-SILC**	**Index**
Central Hungary	2 677 010	2 274 049	118%
Central Transdanubia	2 168 491	1 962 502	110%
Western Transdanubia	2 137 703	1 970 030	109%
Southern Transdanubia	1 855 041	1 729 050	107%
Northern Hungary	1 823 383	1 721 192	106%
Northern Great Plain	1 740 236	1 591 356	109%
Southern Great Plain	1 837 067	1 677 644	110%
Total	*2 139 255*	*1 919 681*	*111%*

Source: Own calculation based on Hungarian Central Statistical Office (HCSO) and National Tax and Customs Administration (NTCA) data

Table 1 shows that survey data underestimate the average of the market income. The total based on tax data is about 11% higher than that calculated from EU-SILC, and we can find noticeable variability by parts of the country. The largest gap can be observed in the central region, which is not independent from the fact, that this region shows the highest average market income.

Table 2. Market income by age group (HUF/year or ratio)

	Tax data	**EU-SILC**	**Index**
0-24	830 012	1 107 343	75%
25-29	1 814 529	1 727 611	105%
30-34	2 282 374	2 106 099	108%
35-39	2 431 179	2 170 095	112%
40-44	2 448 771	2 062 111	119%
45-49	2 429 261	2 031 189	120%
50-54	2 347 274	1 978 958	119%
55-59	2 270 154	1 941 552	117%
60+	1 978 079	1 672 014	118%
Total	*2 139 255*	*1 919 681*	*111%*

Source: Own calculation based on HCSO and NTCA data

From Table 2 we can see that the difference between survey and tax data is also influenced by the age of persons. It is very interesting that the survey data underestimate the average market income in nearly all of the age groups, only the youngest stratum produces overestimation. Its motive can be that young workers are overrepresented in the hidden economy, unfortunately they often prefer a higher wage to legal employment. The other reason can be that young persons are typically occasional workers, and these jobs are often offered by the black labour market.

Table 3. Market income by decile (HUF/year or ratio)

	Tax data	**EU-SILC**	**Index**
Decile 1	139 859	258 984	54%
Decile 2	497 476	696 505	71%
Decile 3	844 763	1 075 991	79%
Decile 4	1 207 565	1 346 814	90%
Decile 5	1 439 220	1 539 034	94%
Decile 6	1 684 680	1 790 879	94%
Decile 7	2 052 778	2 107 856	97%
Decile 8	2 603 743	2 508 898	104%
Decile 9	3 473 007	3 096 813	112%
Decile 10	7 450 064	4 780 683	156%
Total	*2 139 255*	*1 919 681*	*111%*

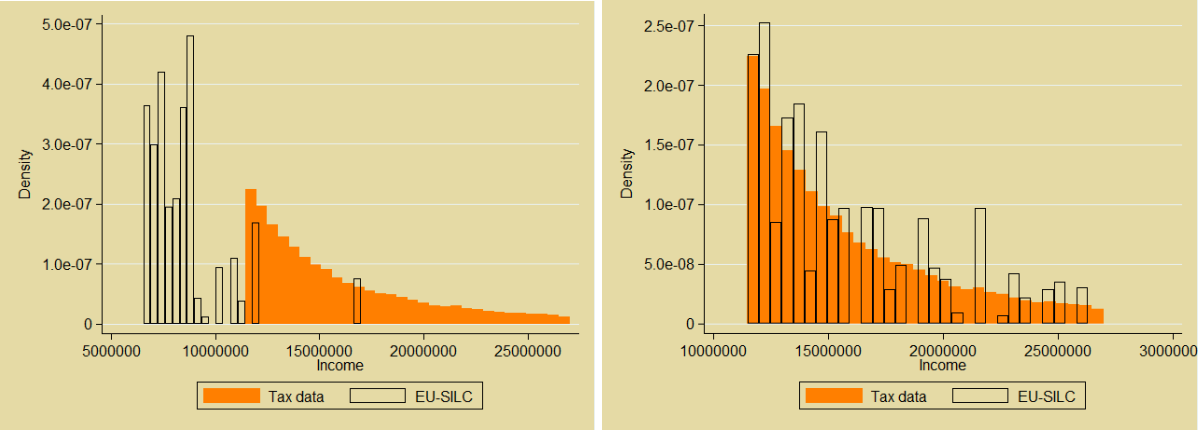
Source: Own calculation based on HCSO and NTCA data

Table 3 shows that we can find the highest dispersion in the gap between survey and tax data if we use the income decile categories as the explaining variable. In the first 7 deciles the EU-SILC database overestimates the tax data means, only the richest 3 deciles show underestimation. One can also see that the greater the income is the smaller the overestimation appears.

So we have used these three variables for the imputation (the first step). We used them together, so we composed 7*9*10=630 cells for the imputation. Each cell represented a multidimensional category, for example the first cell acted for persons from Central Hungary, age of 0-24 and from the poorest income decile. All the cells contained a correction factor, which were calculated as the indices in the tables above: the tax data mean divided by the survey mean concerned to the given category. We used this correction factor for all the persons who were included in this category.

After the first step the mean of the survey date converged to the tax data average, but the top 1% was still not perfectly characterized respected to both the average and the distribution. This problem is illustrated on left side of Figure 3. The solution was the imputation of the top 1% according the Jenkins' third approach (the second step). After correcting all the top 1% observations is the survey data, we obtained the distribution showed on the right side of Figure 3. The distribution of tax data is the same on both sides.

Figure 3. The distribution of top 1% in survey data before and after the second step of imputation



Source: Own calculation based on HCSO and NTCA data

In the second step we have increased the market income for all the observation in the top 1%. It caused an unwanted increase in the overall average. As a consequence, we had to make a third step to eliminate this problem. We did not want to change the decile structure obtained after step 1, so we decided to correct only the 10th decile.

5. RESULTS

In the previous section we presented, how we made the 3 steps of the imputation of market income received from EU-SILC survey data. Until this point we were concentrating only on the gross income category, this type of income was the target of the imputation. The income inequality measures are usually computed for net income, so we have also calculated the taxes and social contributions. After subtracting these amounts from the gross income we obtained the net market income for each observation. Table 4 shows the means of net market income for different percentile categories between 2010 and 2014.

Table 4. Market income averages before and after adjustment (HUF/year)

	Average market income	EU-SILC	EU-SILC-adjusted
2010	overall	1 033 333	1 058 833
	for percentiles 1-90	866 243	838 619
	for percentiles 91-99	2 356 435	2 626 072
	for top 1%	4 169 492	6 785 954
2011	overall	1 063 563	1 125 076
	for percentiles 1-90	876 302	853 584
	for percentiles 91-99	2 461 598	3 004 816
	for top 1%	5 362 138	8 690 568
2012	overall	1 057 748	1 110 569
	for percentiles 1-90	866 182	840 280
	for percentiles 91-99	2 468 800	2 978 459
	for top 1%	5 604 959	8 689 635
2013	overall	1 137 486	1 156 538
	for percentiles 1-90	912 097	869 096
	for percentiles 91-99	2 739 804	3 207 848
	for top 1%	7 048 027	8 580 846
2014	overall	1 171 130	1 285 240
	for percentiles 1-90	968 886	971 713
	for percentiles 91-99	2 727 622	3 468 835
	for top 1%	5 399 506	10 003 396

Source: Own calculation based on HCSO and NTCA data

One can see that the difference between the adjusted and unadjusted overall average in 2014 is around 10%, which is quite close to the gap we have observed between the survey and tax data for the gross income. This difference varies from year to year, but we can state that it is typically smaller in the beginning of the examined period. It is also obvious that the impact of adjustment is smaller is the poorer percentiles, the larger gap can be observed in the top 1%.

We also examined the impact of the adjustment process on the inequality measures. Tables 5-6 contain six different indicators including the income share of the top 1% which is in the focus of this paper. The results referred to years 2010-2014, and the first table shows the original situation before the adjustment, while the second one presents the measures after the three steps of the adjustment process.

Table 5. Market income inequality measures before adjustment

	GINI index	PALMA index	Divident point P90/P10	S80/S20	S90/S10	Top 1% share
2010	0,368	1,52	8,5	9,0	19,2	4,0%
2011	0,377	1,63	8,4	9,4	20,3	5,0%
2012	0,385	1,69	8,8	10,1	22,8	5,3%
2013	0,401	1,86	10,0	10,9	26,4	6,2%
2014	0,377	1,61	8,2	9,3	19,5	4,6%

Source: Own calculation based on HCSO and NTCA data

Table 6. Market income inequality measures after adjustment

	GINI index	PALMA index	Divident point P90/P10	S80/S20	S90/S10	Top 1% share
2010	0,417	2,05	10,4	11,8	27,4	6,4%
2011	0,442	2,39	11,0	13,3	31,2	7,7%
2012	0,451	2,51	12,5	14,9	36,6	7,7%
2013	0,459	2,63	14,6	16,0	41,9	7,4%
2014	0,455	2,56	12,6	15,1	35,9	7,6%

Source: Own calculation based on HCSO and NTCA data

Examining the results, we can detect a significant difference in the level of inequality between the adjusted and unadjusted data. The adjustment of the survey data causes a heavy increase in inequality, which is one of our most important messages. We can also see that not only the level but also the tendencies are changing because of the adjustment. The most remarkable change can be observed in 2014. According to the unadjusted data we see a relevant decrease in inequality this year, and therefore – after a constant rising until 2013 – the inequality returns to the level of 2010. The adjusted data shows a totally different trend in inequality: there is a slighter setback in 2014, so the level remains much above the 2010 rate.

Similar inferences can be drawn regarding the top 1% share. After the adjustment we can realize a significantly higher ratio, so our hypothesis turned to be right about underestimating the role of the richest income stratum in the EU-SILC dataset. The tendencies are different in the adjusted and unadjusted dataset as well. If we ignore the adjustment we could think that the role of the top 1% was constantly rising until 2013, and fell down in 2014. But in fact – taken into consideration the adjustment – we see that after a skip in the top 1% share in 2011, the income proportion of the richest persons remained stable from 2012.

6. CONCLUSIONS

We analysed the top 1% share in pre-tax income in Hungary for 2010-2014. We found a significant under-coverage in terms of gross earnings in Hungary. The under-coverage proved to be even higher in case of the top earners' income in the survey-based EU-SILC estimates, mainly because the richest 1% was partly missing from the survey samples. We adjusted the survey data using tax data by cell mean imputations suggested by Jenkins (2016) to get a more realistic picture of Hungarian income distribution. The main findings are as follows.

The average gross and net incomes and the levels of inequality are significantly higher in the modified database for the analysed period. Income inequality trends also differentiate sharply, especially in 2014. The most remarkable finding is the change in the share of top 1%. By the survey-based estimate the proportion of the top 1% went down sharply in 2014, while the adjusted database shows a further increase.

Based on these results we firmly believe that taking better account of top 1% can provide a more realistic picture about income distribution. In this paper we concentrated only on labour income, but we plan to cover also capital gains in our further research.

REFERENCES

1. Alvaredo, F, Atkinson, A, B, Piketty, T, Saez, E & Zucman, G, 'The World Wealth and Income Database', 29 June 2016, <http://www.wid.world>.
2. Atkinson, Anthony B, (2007), 'The Distribution of Top Incomes in the United Kingdom 1908–2000.' In *Top Incomes over the Twentieth Century—A Contrast between Continental European and English-Speaking Countries,* edited by A. B. Atkinson and T. Piketty, 82 –140. Oxford University Press.
3. Atkinson, A, B, Piketty, T & Saez, E, (2011), 'Top incomes in the long run of history', *Journal of Economic Literature*, 49 (1), 3-71.
4. Bach, S, Corneo, G & Steiner, V, (2009), 'From bottom to top. The entire income distribution in Germany', 1992-2003, *Review of Income and Wealth*, 55 (2), 303-330.
5. Burkhauser, R. V., Hérault, N., Jenkins, S. P., and Wilkins, R. (2016), 'What has been happening to UK income inequality since the mid-1990s? Answers from reconciled and combined household survey and tax return data', IZA Discussion Paper No. 9718, ISER Working Paper 2016-03.
6. Förster, M, Llena-Nozal, A & Nafilyan, V, 'Trends in Top Incomes and their Taxation in OECD Countries', *OECD Social, Employment and Migration Working Papers* No. 159.
7. Jenkins, S, 'Pareto distributions, top incomes, and recent trends in UK income inequality', invited plenary lecture at Conference on the 'Dynamics of Inequalities and their Perceptions', *Aix-Marseilles University, Marseilles*, 26–27 May 2016.
8. Piketty, T, "Capital in the Twenty-First Century", *The Belknap Press of Harvard University Press,* 2014.