Sharing the gains of transition: Evaluating changes in income inequality and redistribution in Poland using combined survey and tax return data

Brzezinski, Michal; Myck, Michał; Najsztub, Mateusz

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-79663-5

Sharing the gains of transition: Evaluating changes in income inequality and redistribution in Poland using combined survey and tax return data

Michal Brzezinski [a,*], Michał Myck [b,c,d], Mateusz Najsztub [b]

[a] *University of Warsaw, Faculty of Economic Sciences, Poland*
[b] *Centre for Economic Analysis, CenEA, Poland*
[c] *University of Greifswald, Germany*
[d] *IZA-Bonn, Germany*

ARTICLE INFO

ABSTRACT

We use Pareto imputation, survey reweighting, and microsimulation methods applied to combined household survey and tax return data to re-evaluate trends in income inequality and redistribution in the follow up of the post-socialist transition in Poland. Our approach results in the first estimates of top-corrected inequality trends for real equivalised disposable incomes over the years 1994–2015, a period so far believed to represent a time of not only stable and but also equitable growth. The adjustments applied suggest that the Gini coefficient grew by 14–26% more compared to the uncorrected survey-based estimates. This indicates that over the last three decades Poland has become one of the most unequal European countries among those for which top-corrected inequality estimates exist. Looking at different centiles of the distribution shows that incomes at the top grew fastest during the post-transition years: the annual rate of growth of the 95-99th percentiles of income distribution exceeded 3.5%, while the median income grew by about 2.6%.The findings shed a new light on recent political developments in Poland.

1. Introduction

In recent years, many countries, including established liberal democracies, experienced the rise of populism, political polarization, right-wing illiberalism and democratic backsliding. These processes have been linked to long-term economic trends related to technological change, globalization, and to consequences of recent recessions, as well as to cultural backlash, attitudes to immigration, or the role of the internet and social media (for a detailed review, see, Guriev and Papaioannou 2020). Since 2016, Poland has been a leading example of a country which witnessed such a populist, illiberal turn. The right-wing Law and Justice party won the presidential and parliamentary elections in 2015 and since then implemented a series of institutional changes that weakened the rule of law and checks on executive power, attacked the judiciary, and politicised the public media. The degree of democratic backsliding in Poland has been perceived as somewhat surprising given the country's nearly universal reputation as the poster child of a successful transition to the market economy and a consolidated democracy. The country has also experienced strong economic growth since 1992 and was little affected by the global economic crisis of 2007 and other recessions. The reasons for the populist and illiberal turn in Poland

---

* Corresponding author.

E-mail addresses: mbrzezinski@wne.uw.edu.pl (M. Brzezinski), mmyck@cenea.org.pl (M. Myck), mnajsztub@cenea.org.pl (M. Najsztub).

remain unclear, but all authors – drawing on official statistics – argue that high or rising economic inequality could not have played an important role in bringing about this change. The official time series on income inequality suggest that Poland enjoyed relatively equal income distribution, and that the degree of inequality was improving in the 2010s (Markowski 2016; Tworzecki et al., 2019; Haggard and Kaufman 2021).

In this paper, we provide new evidence which shows that officially published income inequality levels, as measured in terms of post-tax post-transfer disposable household incomes, have been severely underestimated. While the standard view based on unadjusted survey data suggests that income inequality in Poland did not increase or increased only slightly from the mid-1990s to 2015 (Mitra and Yemtsov 2007; Tóth 2014; OECD 2015; Perugini and Pompei 2015; World Bank 2017; Piatkowski 2018), we find that the rise has been substantial and inequality in Poland converged to the levels observed in the most unequal countries in Europe. Moreover, we demonstrate that in the same period, income redistribution and tax progressivity have been significantly reduced. Our findings suggest that growing income inequality and declining income redistribution should be considered among the factors contributing to the rise of populism and illiberalism in Poland. This is consistent with survey evidence showing that citizens' demand for redistribution in Poland has been growing from the mid-1990s to the 2010s (Choi 2019; Bussolo et al., 2021). It is also notable that in the 2015 election campaign the Law and Justice party responded to the voters' demand for redistribution by proposing a generous, redistributive child benefit programme.[1] The programme has been implemented since 2016 and gained strong support among the general public. Our results suggest therefore that growing economic disparities might have been an important factor leading to the rise of populism and democratic backsliding in Poland.

There is growing literature which shows that inequality measures based on household surveys tend to underestimate the true degree of income disparity due to the under-coverage of top incomes resulting from relatively high rates of survey non-response and income under-reporting among high-income earners (see, e.g., Burkhauser et al., 2012, 2017; Bartels and Metzing 2019; Jenkins 2017). This has led many researchers to reach out to tax return data with better coverage of the upper tail of the income distribution.[2] Empirical studies have found that inequality estimated from administrative data such as individual tax returns (or aggregated income tax statistics) often displays significantly different levels and trends over time compared to survey-based estimates. For example, Jenkins (2017) reports that the Gini coefficient for individual gross income in the UK estimated using tax data rose by 7–8% between 1996/7 and 2007/8, while it decreased by about 5% in the same period when estimated using survey data. Top 1% income shares for fiscal incomes were found to be higher in tax data than in surveys by 3–6 percentage points (p.p.) in Germany (Bartels and Metzing 2019), by about 6 p.p. in the US (Burkhauser et al., 2012), and by as much as 12–14 p.p. in Russia (Novokmet et al., 2018). Over the last two decades, several methodological approaches have been developed in order to produce more reliable inequality estimates using the advantages of both household survey and tax return data (Alvaredo 2011; Jenkins 2017; Bartels and Metzing 2019). This literature produces top-corrected inequality estimates by integrating survey and tax data with harmonised income definitions and reconciling units of observations (households versus tax units) in both types of data sources.

The existing literature on measuring income inequality using integrated household survey and tax return data has so far mainly concentrated on developed counties. Much less attention has been devoted to post-socialist transition economies or emerging markets.[3] Some exceptions include Novokmet et al.'s (2018) work on Russia, Bukowski and Novokmet's (2017, 2021) papers on Poland and Piketty et al. (2018) study about inequality in China. In general, these studies show that survey-based inequality estimates for transition and emerging economies are substantially lower than top-corrected estimates which exploit additional information from tax data. However, most of these works estimate income inequality in terms of gross (pre-tax) income distributed among tax units or only among adult individuals. This income concept deviates considerably from the primary measure of the standard of living analysed in income distribution and welfare economics literature, which focuses on disposable (after tax and transfers) household income defined for the entire population and adjusted for differences in household size and composition using equivalence scales. In practice, trends in inequality of gross incomes distributed among the adult population may be significantly different from trends in inequality of equivalised disposable incomes for the entire population.

In this paper, we fill a gap in the literature by providing the first top-corrected inequality estimates for real equivalised disposable

---

[1] The child benefit (known as Family 500+ programme) in its initial form consisted of a transfer of PLN 500 (EUR 118) per month to every second and subsequent child in a family and every child in low-income families. The programme has been costly (about 1% of the GDP) and generous as PLN 500 was equivalent in 2016 to about 37% of monthly net minimum wage in Poland. The policy has significantly increased income redistribution and contributed to cutting by half the absolute child poverty rate in Poland over 2015–2017 (Brzeziński and Najsztub, 2017; Myck and Trzciński 2019). In July 2019, the programme has become fully universal for all children aged 0–17.

[2] Obviously, tax data also have their limitations. Usually the definitions of income and observation unit in tax data are different from those used in household surveys. For example register tax data in Poland do not capture income from farming, own production income, alimonies and other private transfers, all of which are reported in the Polish Household Budget Survey (PHBS) data. Tax data are also very sensitive to tax avoidance and evasion, as well as to legislative changes in the income tax law (Atkinson et al., 2011).

[3] This in part reflects the fact that tax-based data (individual tax returns or aggregated tabulations from tax returns) are either unavailable or rather scarce in transition and emerging countries.

household incomes in Poland over the period between 1994 and 2015.[4] We combine survey income data from the Polish Household Budget Survey (PHBS) with information from tax-based top income shares (Bukowski and Novokmet 2017) and use microsimulation modelling (Bargain et al., 2007; Morawski and Myck 2010) to reconcile differences in income concepts and observation units between the two data sources. Since Polish individual tax records are not regularly available for research purposes, we account for under-reporting and under-coverage of top incomes in the PHBS data by Pareto imputation of the highest income observations (see, e. g., Jenkins 2017) and survey reweighting techniques (Creedy 2004; Myck and Najsztub 2015).

Our paper contributes to the existing literature in two dimensions. On the substantial level, we provide first estimates of the top-corrected distribution of real equivalised disposable household income – the most common measure of the standard of living – of the entire Polish population for the period since early 1990s to 2015. Our main empirical result shows that contrary to the results based on unadjusted survey data suggesting no inequality increases, the top-correction procedures show that inequality of living standards as measured by the Gini coefficient increased substantially in Poland over this period – in the range from 14 to 26%. Our revisions imply that both the inequality trend and its level have been significantly underestimated. While according to the unadjusted data the Gini coefficient in 2015 is at a relatively moderate level (30.1), the top corrected estimate is 28% higher and equal to 38.4. We also show that the income growth was highest at the top end of the income distribution, suggesting that top earners benefited disproportionately from the post-socialist growth process compared to the middle class and the poor. For instance, the annual rate of growth for the 95-99th percentiles of income distribution has been within 3.5–5% range, while the median income grew by about 2.6% per year. Additionally, the microsimulation framework we use facilitates analysis of changes in progressivity of specific components of the tax and benefit system which links the distributional statistics to policy decisions. Our top-corrected estimates suggest a sharp decline in the progressivity of the social insurance and direct taxation system in Poland.

Second, we make a methodological contribution to the literature by showing that two techniques of adjusting income distributions (Pareto imputation of top incomes and survey reweighting) that have been so far applied separately can be implemented jointly to produce reliable top-corrected inequality estimates. To this end, we integrate the methodology of imputing to household survey data top incomes from Pareto models fitted on tax data (Bartels and Metzing 2019) with microsimulation-based survey reweighting approach using tax and other administrative information (Myck and Najsztub 2015).

The results also carry important implications for the political economy literature on the links between inequality, democracy and redistribution. The seminal work of Meltzer and Richard (1981) shows that under majoritarian democracy higher income inequality implies that the decisive voter (median voter) prefers more income redistribution. Recent empirical contributions which examined this relationship (see, e.g., Gimpelson and Treisman 2018; Kuhn 2019) have found that redistributive preferences relate more strongly to subjectively perceived (by individuals) inequality rather than to actual (objective) income inequality. Moreover, there is evidence that redistributive preferences are often unrelated to actual income redistribution (Niehues 2016; Choi 2019). Our results may shed a new light on these findings. If inequality measures used in these studies rely on unadjusted survey measures of income, which they almost always do, the relationship between these measures and preferences for redistribution may be blurred by measurement issues. We therefore believe that as more accurate indices of inequality emerge for an increasing number of countries, the empirical literature on the relationship between inequality and redistribution may deserve reconsideration.

The paper is organised as follows. Section 2 presents our income data from household surveys and tax records. In Section 3, we introduce the methodology of top-correcting survey-based income distribution using Pareto imputation, reweighting and micro-simulation modelling. Section 4 provides our empirical results on top-corrected income inequality in Poland over 1994–2015, as well as the comparison of top-corrected inequality levels and trends for Poland and other European countries. In this section, we also discuss how different segments of the income distribution in Poland benefited during the process of post-socialist transformation by looking at top-corrected rates of income growth by percentiles of income distribution and indicators of redistributive effect and progressivity of direct taxation. The last section concludes.

2. Data

2.1. Polish Household Budget Survey (PHBS) data

Our survey income data come from the Polish Household Budget Survey (PHBS) conducted annually by Statistics Poland since 1957. The PHBS is the main representative source of information on household incomes in Poland.[5] Since 1993, the methodology of collecting data in the PHBS is fairly consistent (Keane and Prasad 2002; Kordos et al., 2002). The sample size since 1993 is more than 30,000 households and 100,000 persons in each year. We use the PHBS data for 1994–2015 as the pre-1994 surveys do not contain data on individual incomes (required for our microsimulation modelling) and 2015 is the last year for which estimates of tax-based top

---

[4] This is the longest period for which consistent series of household survey data with individual incomes (required for microsimulation analysis, see section 2.1 and the Supplementary Appendix) can be constructed. Income data from the Polish Household Budget Survey (PHBS) for the first years of post-socialist transition are of limited reliability due to enormous economic uncertainty and volatility in this period, as well as because of the significant methodological changes in the survey design introduced in 1993. Keane and Prasad (2002) show that there is no evidence of an increase in income inequality in Poland over 1989–1996, when accounting for changes in the PHBS survey design and using equivalence scales in a consistent way.

[5] The PHBS has been previously used to study income distribution in Poland by, among others, Szulc (2000), Keane and Prasad (2002), Podkaminer (2003), Brzeziński and Kostro (2010), Myck and Najsztub (2016).

income shares are available (Bukowski and Novokmet 2017).[6] Kordos et al. (2002) provide comprehensive information about sample design and other features of the PHBS. The survey contains detailed data on monthly income from various sources for households as well as for individuals within households. Beside data on incomes, the PHBS provides information on household size and structure, economic activity of household members, housing conditions, detailed household expenditure, and others.

In order to account for survey non-response, Statistics Poland provides sampling weights that correct for inclusion of the households in the sample in accordance with sample design. The sampling weights computed as the inverse of selection probabilities are adjusted by post-stratification based on census data on place of residence (rural or urban) and size of the household. Thus, post-stratification does not use any information on sex, age or education of household members. We refer to these weights as 'baseline' PHBS weights. Myck and Najsztub (2015, 2016) show that using the baseline weights leads to significant under-representation of some population groups, and as a result over-representation of others. For example, compared to administrative statistics the PHBS results obtained with baseline weights overestimate the population of children aged 0–15 by more than 1.4 million in 2014 and underestimate the adult population by the corresponding amount. This divergence between the administrative population record and the population structure observed in the survey data using baseline weights has grown over time and has most likely been related to a growing rate of survey non-response, a challenge faced by many surveys (Kordos et al., 2002; Verma and Betti 2010). Similar discrepancies were found in the case of subpopulations defined with respect to the education level, employment type or type of household residential area. Following Creedy (2004) and Deville and Särndal (1992), Myck and Najsztub (2015) propose to address this problem by calibrating the PHBS baseline weights using information from various administrative sources. This reweighting approach leads to adjusted weights that allow for obtaining weighted PHBS estimates corresponding closely to values taken from administrative sources (such as the official number of children in the population or people with higher education).

In this paper, we follow Myck and Najsztub (2015) in using the reweighing approach to adjust the baseline PHBS weights. In particular, we use two types of weight calibration. The first of these adjusts the baseline weights by calibrating them to match the census-based number of males and females in several age groups (see Supplementary Appendix for details). We refer to these weights as 'population weights'. Secondly, we further calibrate population weights to match the number of PIT payers in each tax bracket based on official information from Ministry of Finance reports.[7] These weights are henceforth referred to as 'tax weights'. We expect that using tax weights will lead to significantly better coverage of the upper tail of the Polish income distribution as compared with raw PHBS data. However, since the number of taxpayers in the higher tax brackets is heavily underestimated even using tax weights (see Supplementary Appendix) it is necessary to adjust for missing top incomes using other methods. We do this by imputing top income values based on estimated Pareto distribution.

Our main income variable is real equivalised household disposable (post-tax, post-transfer) income, a measure most commonly used to reflect the material standard of living.[8] All incomes are expressed in 2015 prices. We use detailed individual incomes from the data and apply microsimulation methods to incomes which are subject to income tax to facilitate income adjustments using administrative tax data. We take advantage of a long series of tax systems programmed in the SIMPL microsimulation model (Bargain et al., 2007; Morawski and Myck 2010) and use it for two purposes. First we construct gross incomes within tax units on the basis raw net incomes reported in the PHBS data (see Supplementary Appendix for details). These grossed-up incomes are then set against the distribution of gross incomes as reported in administrative data, which is used to impute top (gross) incomes back to the PHBS data. The microsimulation model is then used once again to translate these gross incomes back to net values which are combined with other income sources giving us the adjusted disposable income series.[9]

2.2. Top income shares from tax return data

Bukowski and Novokmet (2017) provide top income shares series for Poland between 1892 and 2015.[10] In this paper, we focus on the period 1994–2015 for which we can construct reliable and consistent household survey data on income. For this period, Bukowski and Novokmet (2017) estimate top income shares using tabulations on the settlement of the PIT published annually by the Ministry of Finance.[11] The tabulations contain information on the number of taxpayers, the amount of income, and tax paid by income brackets as defined by the tax rate schedule. This information is highly grouped as limited progressivity of the Polish income tax system implies that the number of income brackets in the tax rate schedule is small (three for 1994–2008, and only two after 2008). For this reason, Bukowski and Novokmet (2017) estimate only income shares of those top percentile groups that are close to the percentage of tax-payers in the top income bracket (i.e. the top 5% and top 1% income shares). The top income shares, as standard in the literature, are calculated in terms of gross (pre-tax) income distributed among tax units. The income concept used covers income from employment, pensions, non-agricultural business activity, special departments of agricultural business activity, self-employment income, rental

---

[6] Note that the modern personal income tax was introduced in Poland only in 1992.

[7] The number of PIT payers is computed based on gross incomes simulated using the SIMPL microsimulation model.

[8] We use modified OECD equivalence scales.

[9] An alternative method to the net-gross and gross-net conversion which has been used in the literature relies on a simplified parametrised tax function (see for example Heathcote et al., 2017). Since our analysis covers a long period over which the tax code changed along many dimensions, and since microsimulation facilitates more detailed treatment of the complexity of tax regulations, we believe that microsimulation is a more suitable approach to be used in this case.

[10] Kośny (2019) provides top income shares estimated from individual tax returns for Lower Silesian Voivodeship in Poland.

[11] Information from individual tax returns from Poland is not at present available to researchers.

income, capital gains and income from other sources. Capital income is not included. The estimates are adjusted for the several changes in the tax code that were implemented since 1994. In order to estimate top income shares, Bukowski and Novokmet (2017) apply the standard Pareto interpolation techniques (see Atkinson 2007; Atkinson et al., 2011).

3. Top-correcting of income distribution using combined household survey and tax return data

Jenkins (2017) provides a recent comprehensive review of approaches to correct income distribution for the under-coverage of top incomes in survey data. Most of these approaches require access to individual tax return data covering at least the upper part of the population of taxpayers. The currently available tax-based income data for Poland are exceptionally sparse as the only accessible information is on the top 5% and top 1% income shares. Therefore, we are unable to apply the frameworks that rely on a heavy use of tax data. Instead, we exploit a flexible top-correcting methodology that was recently introduced by Bartels and Metzing (2019). The approach has three important advantages. First, it can be used when micro-data from tax records is unavailable and only aggregated tax-based statistics (such as publicly available top income shares) are at researchers' disposal. Second, it produces top-corrected distributional results for any distributional measure (i.e. inequality, poverty, middle class indices, etc.) as well as for any income (or well-being) definition. Third, Bartels and Metzing's (2019) framework produces distributional indices not only for the population of taxpayers, but also for the entire country's population (i.e. including also non-taxpayers). This allows for obtaining estimates of inequality and other distributional indices in terms of popular measures of the standard of living such as post-tax post-transfer equivalised disposable income for the whole population, which is a more comprehensive measure of individual well-being compared to individual gross incomes.

The methodology of Bartels and Metzing (2019) involves the following steps. First, survey and tax data are reconciled with respect to differences in income definitions, observation units, and the coverage of top incomes. Second, the appropriate share of top incomes in gross (pre-tax) household survey data is replaced with Pareto-imputed incomes estimated using information from tax-based top income shares. Finally, top-corrected gross income distribution is "netted down" to obtain the net (after-tax) equivalised household income distribution, which is used to compute final estimates of inequality and other measures.

In order to impute top incomes estimated from tax-based statistics, Bartels and Metzing (2019) follow most of the literature in using the Pareto Type I model (Atkinson et al., 2011; Alvaredo 2011).[12] The Pareto I distribution for income variable x can be defined through its survival function $S(x)$, which is equal to 1 minus the cumulative distribution function $F(x)$:

$$S(x) = 1 - F(x) = \left(\frac{x}{x_m}\right)^{-\alpha}, \tag{1}$$

where $x \geq x_m > 0$, and $x_m > 0$ is the threshold above which the data are Pareto distributed. Parameter α is known as Pareto tail index and describes the heaviness of the right tail of income distribution. The lower the tail index, the heavier the right tail and the more unequal Pareto distribution. Following Atkinson (2007) and Bartels and Metzing (2019), we estimate the tail index as:

$$\alpha = \frac{1}{\left(1 - \frac{\log(S_j/S_i)}{\log(P_j/P_i)}\right)}, \tag{2}$$

where P_i and P_j are the population shares of group i and j, and S_i and S_j are the income shares of these groups estimated using tax data. In practice, the indices i and j denote given fractiles of the population with i being a subgroup of j. Most of the literature uses 0.10, 0.05, 0.01, and 0.001 fractiles of the population. For Poland, the only existing top income shares are for the 0.05 and 0.01 fractiles (Bukowski and Novokmet 2017), which means that we estimate the Pareto tail index using population and income shares for the top 5% and 1% of the population.

After estimation of α, the threshold x_m can be obtained from equation (1) as follows:

$$x_m = x(1 - F(x))^{1/\alpha}, \tag{3}$$

where $F(x)$ and x are estimated using survey data. The value of x is implied by the proportion of survey incomes to be replaced with Pareto-imputed values. In this paper, we experiment both with replacing top 1% and 5% incomes from survey data and correspondingly set x to the 99th and 95th percentile of the survey income distribution, respectively.[13] In the final step of top-correcting gross income distribution, we replace the top 1% (or 5%) of tax unit incomes observed in our survey data with incomes implied by the Pareto distribution characterised by our estimates of α and x_m.

The methodological approach taken in this paper is as follows.[14] We use detailed year-specific information on the Polish tax-benefit

---

[12] Jenkins (2017) carefully studies the problem of fitting various Pareto models to income tax data. He finds that Pareto Type II (generalised Pareto) model provides a better fit do data than Pareto Type I model. According to his results, the threshold above which Pareto models fit income data well is the 99th or 95th percentile.

[13] Bartels and Metzing (2019) set x_m to the 99th percentile, while Jenkins (2017) finds that the appropriate value of the threshold is between the 99th and 95th percentile.

[14] The Supplementary Appendix provides a more detailed description of our top-correction procedure.

system parameters and the SIMPL microsimulation model to cross-walk from the PHBS household net income to gross income distributed among tax units (individuals or married couples). In the next step, we use data on top income shares from Bukowski and Novokmet (2017) to estimate the parameters of Pareto distribution (equations (2) and (3)) for gross income distribution in terms of tax units. Our estimates of Pareto distribution parameters are very similar regardless of whether we set x to the 99th or the 95th percentile.[15] The estimated Pareto tail index on average equals 1.89 and in general falls with time implying growing income inequality. The estimate of mean (median) income increases continuously over time from around 440 PLN (310 PLN) in 1994–2700 PLN (1750 PLN) in 2015. This reflects the robust and stable economic growth of the Polish economy over the period under study. Then, we replace the top 1% (or 5%) of tax units' incomes with incomes implied by the Pareto distribution characterised by our estimates of α and x_m.[16] The resulting imputed gross distribution is subsequently reweighted using population weights in order to ensure that our survey data are representative with respect to age and sex distribution in the population. In a variant of our analysis, we further reweight survey-based gross income distribution using tax weights in order to match the number of PIT payers in each tax bracket based on official information from Ministry of Finance data.[17] We then evaluate our approach by comparing the resulting top income shares with tax-based top income shares from Bukowski and Novokmet (2017). This is the only reliable validation check in our case as the tax-based top income shares are the only accessible external (to survey-based data) distributional information. After this imputation of top incomes, we again use the microsimulation model to compute top-corrected household net incomes by applying the appropriate tax schedule to gross incomes within the tax units in the data. The SIMPL tax-benefit model covers key details of the Polish tax code, which changed on numerous occasions over the period of analysis, including different treatment of incomes from various sources, numerous tax allowances and tax credits, joint taxation of couples, etc. The procedure is performed separately for each year between 1994 and 2015. The resulting top-corrected net incomes are then combined with other income sources to generate a measure of equivalised disposable household income, which is used for our final distributional indices.

Finally, it needs to be noted that our approach, while correcting incomes and weights, continues to rely on the composition of households as in the PHBS data. This means that if the actual composition of top income households is different, our distributional measures could be biased since top incomes would then be equivalised with incorrect household equivalence scales. Since generally household size falls with income even at the upper end of the distribution, our measures of inequality would then be biased downward.[18] We provide confidence intervals around the calculated measures derived from bootstrapping of the entire procedure which should account at least partly for this source of bias.

4. Empirical results

4.1. Estimating the under-coverage of high incomes in household survey data

We start by looking into the degree of under-coverage of high incomes in the PHBS. Fig. 1 compares estimates of top income shares from tax records (Bukowski and Novokmet 2017) and from the PHBS data (with population weights). The series are based on reconciled income definition (gross income) and refer to the distribution among tax units (individuals). We observe that there is relatively little discrepancy between the two series for the period 1994–2004. In case of top 5% income share, the discrepancy does not exceed 2 p.p. The top 1% income share is underestimated in survey data in this period more significantly by between 2 and 3.5 p.p. Since 2005 the gap between the estimates grows sizeably reaching in 2015 as much as 8.8 p.p. in case of top 1% income share and 6.5 p. p. in case of top 5% income share. The gap between top income shares estimated from household surveys and from tax data has also been found for other countries. For the top 1% share, the gap reaches 3–6 p.p. for Germany (Bartels and Metzing 2019), about 6 p.p. for the US (Burkhauser et al., 2012), and as much as 12–14 p.p. for Russia (Novokmet et al., 2018).

Income shares from tax records displayed in Fig. 1 suggest that the highest incomes in Poland grew particularly fast between 2003 and 2008. This spike in top shares is completely missing in household survey data. What may account for this large increase in the highest incomes? Using individual panel tax return data for 2002–2005, Kopczuk (2012) argues that this increase is related to the 2004 tax reform in Poland that introduced an optional flat tax for non-agricultural business income, which reduced the marginal tax rate for

---

[15] See Table A2 in the Supplementary Appendix for full details. The goodness-of-fit of Pareto models to data can be assessed using the so-called Zipf plots (which plot $\log(1 - F(x))$ against $\log(x)$). If data follow a Pareto model, the Zipf plot should produce a straight line with the slope equal to $-\alpha$. Figure A2 in the Supplementary Appendix shows that in general the unadjusted PHBS data do not follow Pareto distribution. For the top-corrected data, lines on Zipf plots are much flatter implying a lower tail exponent and more unequal income distribution.

[16] Blanchet et al. (2019) recently developed a top-correcting procedure that endogenously determines a valid merging point between survey data and tax data. However, their approach requires using individual or detailed tabulated tax data, which are not available for Poland.

[17] We have considered also an alternative procedure according to which we first reweight the PHBS data to match target statistics from official sources and in the second step we impute top incomes to the reweighted distribution. The two approaches are compared in the Supplementary Appendix (Figures B1-B2) in terms of discrepancy between top income shares estimated from top-corrected survey data and from tax records. We rely on the approach of imputing top incomes first and reweighting the imputed distribution next as it leads to estimates of top income shares that are closer to the tax-based estimates from Bukowski and Novokmet (2017).

[18] We verified the relationship between average household size and top incomes using the 2013 EU-SILC data for Denmark and Netherlands. Bartels and Metzing (2019) show that for these countries the top-correction of household income data in EU-SILC is not required as they provide reliable income information from registers (see section 4.2). Our calculations show that the average household size in subsequent income quintile groups is: 3.3, 3.2, 3.2, 3.2, 3.0. Moreover, it does not seem to change much between households in centile groups: 90–94, 95–99 and the top centile. The corresponding numbers for these three groups are: 3.0, 2.8, and 2.8.
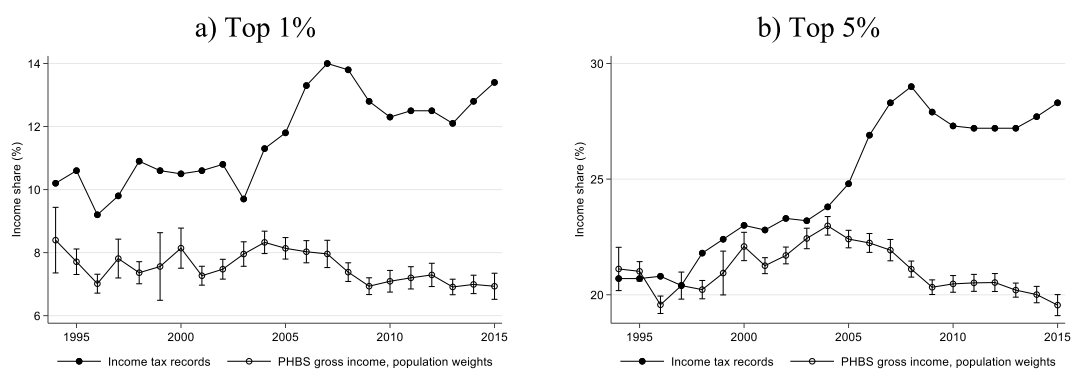
Fig. 1. Top income shares in income tax data and unadjusted household survey data for Poland, 1994–2015. *Note:* Vertical lines show 95% confidence intervals obtained using bootstrap. For both tax and survey data, income refers to gross income and unit of observation is tax unit. *Source:* Own calculations using PHBS data and Bukowski and Novokmet (2017).

the highest income taxpayers from 40% to 19%. Before 2004, business income was taxed according to the progressive scale with three marginal tax rates of 19%, 30%, and 40%. The reform introduced an option of taxing business income using the flat rate of 19%. Kopczuk (2012) shows that the reform was associated with a dramatic increase in the amount of reported business income in tax returns. Gross income reported by taxpayers affected by the reform grew by 48% over 2003–2004. Kopczuk (2012) suggests that although this increase may partly reflect the rise in real economic activity, it is largely driven by reduced tax avoidance or tax evasion. Fig. 2 plots year-to-year changes in the GDP per capita and total reported gross income between 1995 and 2015. The figure shows that for most of the 1995–2004 period reported gross income of taxpayers grew more slowly than the GDP per capita, which suggests that before the 2004 reform the problems of tax evasion and avoidance could have been more pronounced. It is therefore plausible that some of the top incomes in Poland before 2004 were unreported or under-reported. While under-reporting of top incomes can be addressed by adjusting survey data with imputation from Pareto model estimated on tax data, the under-coverage (the fact that some of the top incomes were likely not reported at all) cannot be properly corrected using this method. A more promising approach to address the under-coverage of top incomes is to adjust survey weights with tax weights. We explore these possibilities in the next section, which compares different methods of top-correcting PHBS data.

Fig. 2 shows also that reported gross income subject to flat tax grew exceptionally fast over 2005–2008, accompanied by strong growth in gross income taxed according to the scale over 2006–2008. It is rather unlikely that these sustained increases can be explained by increased tax compliance induced by the 2004 tax reform. They rather result from soaring real economic activity driven by fast economic growth during that period in the follow-up of Poland's entry to the European Union in May 2004.[19] This interpretation is shared by Bukowski and Novokmet (2017, 2021) who argue that since top shares grew strongly not only over 2003–2004 but also over the extended period up to 2008, the major underlying cause of this inequality spike is probably not related to the 2004 tax reform. They also show that the sharp rise in top income shares over 2003–2008 was exclusively due to the rise of business incomes and that since 2005 most of the top 1% income consists of business income.

Bukowski and Novokmet (2017, 2021) suggest several economic mechanisms to explain substantial growth in top income shares in Poland over 2003–2008, other than increased tax compliance. It could be driven by a cyclical expansion of the Polish economy (caused by the global economic boom of 2001–2007) that increased top business incomes disproportionately more than top labour incomes. On the other hand, the growth of top income shares over 2003–2008 could be a result of a long-term capital deepening and growing capital income share (declining labour share) in Polish national income (Growiec 2012; Gradzewicz et al., 2018).[20] A growing capital share could be in turn driven by capital-augmenting technical change or by globalization through trade-induced shift toward capital-intensive sectors. To these conjectures, we can add another one pointing out that monopolistic markups adjusted for cyclical effects have increased substantially in Poland over 2004–2009 (Hagemejer and Popowski 2014). The markups over this period were significantly higher for manufacturing and non-exporting firms. Nolan et al. (2018) suggest that higher markups and associated increased product market power could make income distribution more unequal through higher firm profits and (assuming low bargaining power of employers and unequal ownership of capital) higher incomes of the richest firm owners, or through negative impact on interest rates that leads to increased asset prices favouring richer individuals.

There is thus a number of potential explanations concerning the rapid rise of top incomes in Poland over 2003–2008, none of which can be tested empirically at present due to a lack of access to individual top income data.

---

[19] The GDP per capita growth amounted to 5.3% per year over 2005–2008, and to 6.7% over 2006–2007.

[20] The relationship between capital share and income inequality is complicated and depends, among others, on inequalities of capital and labour income and on the correlation between capital and labour income. Bengtsson and Waldenström (2018) show that empirically there is a strong positive link between capital share and top personal income shares, which is increasing for recent periods and in Anglo-Saxon countries.
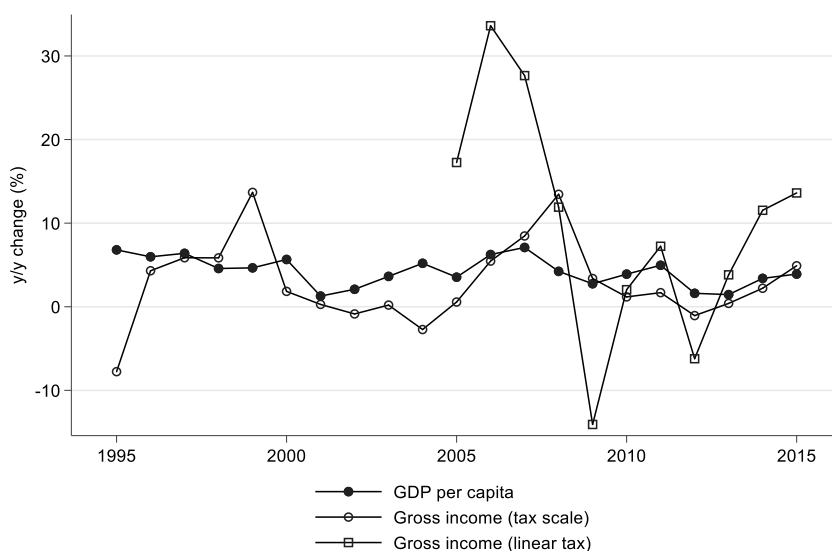
Fig. 2. Reported real gross income versus real GDP per capita for Poland, 1995–2015 (y/y change, %).
*Source:* World Development Indicators and Poland's Ministry of Finance data.

4.2. Income inequality levels and trends

Based on results from Fig. 1, we experiment with imputing either top 1% or top 5% of incomes in the PHBS data. We also use survey reweighting based either on population or on tax weights (see section 2.1). This gives four alternative ways of top-correcting the PHBS data – reweighting with population or tax weights and imputing either top 1% or top 5% incomes. Figure B1 in the Supplementary Appendix compares the four methods in terms of the gap between top income shares estimated from tax records and from top-corrected survey data. We assess goodness of fit of different top-correction approaches by calculating mean squared difference (MSD) between the top income shares' estimates implied by the approaches and tax-based top income shares. For the pre-2005 period, the MSD is minimised for top-correction procedure based on imputing top 1% incomes and reweighting with population weights. While this series tracks tax-based estimates of top income shares before 2005 quite closely, the discussion in the preceding section suggests these shares may be underestimated due to higher tax avoidance or evasion in this period. Therefore, for the pre-2005 period we consider also a top-correction procedure based on imputing top 1% incomes and using tax weights. In case of the post-2005 period, there are two approaches that lead to very similar and low MSD values: imputing top 1% incomes and reweighting with tax weights, and imputing top 5% incomes and reweighting with population weights. However, the confidence intervals associated with the former approach perform better after 2005 as they cover almost all of the tax-based top 5% income shares (see Figure B1 in the Supplementary Appendix). The alternative approach – imputing top 5% incomes and reweighting with population weights – significantly underestimates top 5% shares in each year between 2009 and 2015. For these reasons, our preferred top-correction procedure for post-2005 period is based on imputing top 1% and using tax weights. The 'hybrid' approach of imputing top 1% incomes with reweighting based on either population weights (up to 2005) or tax weights (after 2005) gives lower MSD than any alternative approach applied over the full period under study (1994–2015).

Fig. 3 provides estimates of top-corrected top income shares using our preferred approaches. The gap between tax-based estimates from Bukowski and Novokmet (2017) and survey-based estimates becomes much smaller for our top-corrected series (Fig. 1). The top-corrected series using tax weights overestimates the top 5% income share before 2005, but it can be considered as a plausible estimate of an upper bound on top income shares adjusted for higher top incomes underreporting in the pre-2005 period. The good performance of this approach in the post-2005 period reinforces our premise that this kind of correction can serve as a satisfactory upper bound on our estimates for the pre-2005 period. Point estimates for the top 5% income share are slightly higher than the tax-based figures since 2009, but our confidence intervals contain tax-based estimates for most of this period. Overall, our preferred method of top-correction significantly reduce the gap between tax-based and survey-based estimates, which confirms that the approach of combining Pareto imputation and survey reweighting can be a useful method for top-correcting the income distribution in Poland.[21]

Fig. 4 shows the evolution of income inequality in Poland over 1994–2015 as measured by the most popular inequality index, the Gini coefficient, using both unadjusted and top-corrected series. Until 2005, our two correction procedures show similar inequality

---

[21] Our estimates compare well to the very sparse evidence on top income shares calculated by the Polish Ministry of Finance using individual tax return data (Chrostek et al., 2019). These estimates are available only for 2016 and show that the top 1% income share was 15.2% and top 5% income share amounted to 29.9%. Our estimates (for 2015) are, respectively, 13.6% and 29.3%.

a) Top 1%

b) Top 5%
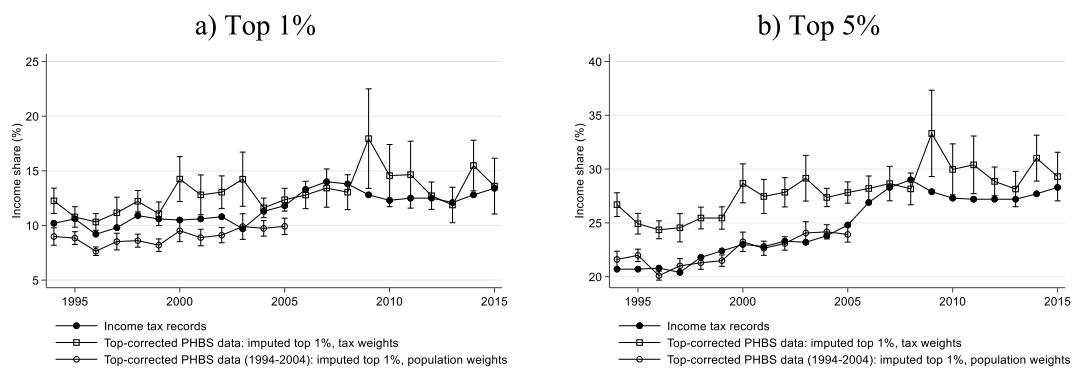
Fig. 3. Top income shares in income tax data and top-corrected household survey data for Poland, 1994–2015. *Note:* Vertical lines show 95% confidence intervals obtained using bootstrap. For both tax and survey data, income refers to gross income and unit of observation is tax unit. See main text for the details of the top-correction procedures.
Source: Own calculations using PHBS data and Bukowski and Novokmet (2017).
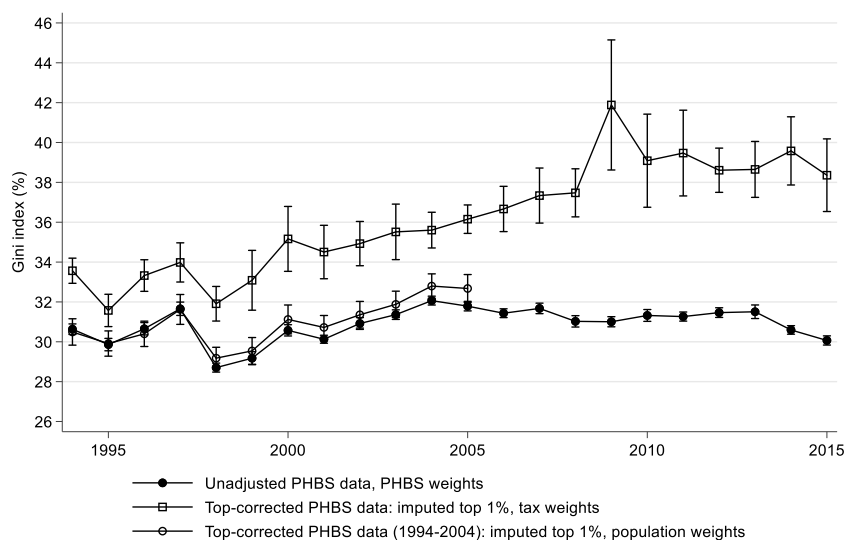
Fig. 4. The Gini index for Poland, 1994–2015: unadjusted vs top-corrected estimates. *Note:* Vertical lines show 95% confidence intervals obtained using bootstrap. The Gini index is computed for real equivalised disposable household income. See main text for the details of the top-correction procedures.
Source: Own calculations using PHBS data.

trends, but somewhat different levels. The correction using tax weights suggests that the Gini index is about 4 p.p. higher than that implied by the method using population weights. Since we are unable to determine which series gives more plausible Gini estimates for the 1994–2005 period, we interpret these as presenting upper and lower bounds on the "true" Gini. After 2005, our preferred correction method shows systematic and high divergence between unadjusted and top-corrected Ginis ranging from 4 to 8 p.p. The spike in the top-corrected series observed in 2009 results from a slight overestimation of top income shares in this year due to our procedure (see Fig. 3) and should be ignored. The point estimate for the Gini index for unadjusted data remains mostly stable over time and even declines somewhat after 2013. However, it surges significantly for the top-corrected series from about 30–34 p.p. in 1994 to around 38–40 p.p. over 2010–2015. The top-corrected Ginis increase in the range from 14 to 26% over the 1994–2015 period.

It should be stressed that the confidence intervals for the top-corrected Gini series are much wider than those for the unadjusted series. For example, in 2015 they range from 36.5 to 40.2 p.p. However, even accounting for the uncertainty associated with our estimates, the top-corrected Ginis suggest that at least since 2005 income inequality in Poland is substantially higher than previously thought.[22]

Compared to the top-corrected Gini coefficients of Bukowski and Novokmet (2018), our estimates of inequality for the period

---

[22] This result holds also for other inequality measures such as the Generalised Entropy indices. As they affect principally the upper end of the distribution the procedures we apply have much smaller implications for poverty measures (see Figures B5-B8 in the Supplementary Appendix).

between 2006 and 2015 are about 6–7 p.p. lower. It needs to be noted though, that these two series are not strictly comparable given the use of different concepts of income. While Bukowski and Novokmet (2018) use fiscal income (gross income before personal deductions and income taxes), our final estimates are for disposable (after taxes and transfers) incomes, additionally equivalised at the household level. The tax and transfer system is a fundamental mechanism translating fiscal into disposable incomes, and thus reflecting the redistributive preferences at each point in time. Using microsimulation to obtain corrected household equivalised disposable income allows us to combine the necessary adjustments to the distribution of gross incomes with expression of inequality trends using a more comprehensive and more intuitive measure of individual well-being compared to individual fiscal income.

We now turn to the implications of our results for the comparison of inequality trends and changes in Poland versus other countries. We focus mainly on the countries for which top-corrected inequality estimates have been obtained using methods similar to those used in this paper. Bartels and Metzing (2019) provide such results for a number of European countries using survey data from the German Socio-Economic Panel (SOEP) and EU Statistics on Income and Living Conditions (EU-SILC). Their results show that the gap between survey-based inequality estimates and top-corrected inequality estimates is negligible for countries (including Scandinavian countries, the Netherlands and Ireland) that have a long tradition of exploiting administrative sources in collecting income information in EU-SILC. The gap is somewhat larger, but still relatively small for the 'new register countries' that started using income data from register relatively recently (France, Italy, Spain, Switzerland). Not surprisingly, the largest gap is found by Bartels and Metzing (2019) for countries that collect income information in the EU-SILC using household surveys only (Germany and the UK). The top-corrected Ginis for disposable income distribution are higher than the Ginis for unadjusted net income distribution by 5–9% for Germany and 2–5% for the UK (see Figures B3-B4 in the Supplementary Appendix).

The comparison of our results with those of Bartels and Metzing (2019) for the EU-SILC 'survey countries' and Spain indicates that for the unadjusted data the Gini index for Poland takes only a moderately high value – it is higher than for Germany, but in general lower than for Spain and the UK.[23] However, the comparison of top-corrected estimates leads to a strikingly different conclusion. Since 2005, the top-corrected Ginis for Poland exceed substantially and significantly (with non-overlapping confidence intervals) those for the comparator countries. Thus according to our estimates, the standard view implying that income inequality in Poland is close to the average EU level, is probably wrong.

After 2005, the increase in the Gini index resulting from our top-correction procedure for Poland is substantially higher than the sizes of analogous top-corrections for other countries analysed by Bartels and Metzing (2019). It ranges from 15 to 30% of the Gini estimate, while the top-corrections for other EU countries are always smaller than 10%. This result suggests that the under-coverage of top incomes in the PHBS data may be significantly more severe than in household surveys used in other countries.[24] The substantially larger size of top-correction for Poland is probably related to the observation of Bukowski and Novokmet (2017, 2021) that in Poland most of the top 1% income consists of business income, while in most of the other countries the dominant source is labour income. Business owners may be harder to reach in household surveys as well as they may have more incentive to refuse to participate in surveys or to under-report their incomes.

Our top-corrected estimates suggest that the standard view maintaining that the post-socialist transition in Poland was associated with a modest rise of inequality of household net incomes has been overly optimistic. This view was widely accepted in academic publications and policy reports (see, e.g., Mitra and Yemtsov 2007; Tóth 2014; OECD 2015; Perugini and Pompei 2015; World Bank 2017; Piatkowski 2018). However, all these publications were based on income data from household surveys (usually from the PHBS). Our results call this previous literature into question by showing that between early 1990s and mid-2010s income inequality in Poland grew sharply. In particular, our top-corrected Gini index for net disposable equivalent incomes increased over 1994–2015 within a range from 5 to 8 p.p. (14–26%). Contrary to the standard view, our revised estimates suggest that income distribution has become much more unequal in Poland since the mid-1990s. Increasing income inequality had not been acknowledged and addressed by political parties in Poland until 2015 when Law and Justice pledged to introduce a generous redistributive child benefit programme. This was then followed by additional give-aways for families with children and for pensioners. The election results in 2015 and 2019 seem to suggest that there was a strong preference for redistribution among Polish voters, which would be consistent with inequality indices exceeding those officially reported.

4.3. Growth rates by percentiles

We now turn to the analysis of income gains for various segments of the Polish population using growth incidence curves (GICs) of Ravallion and Chen (2003). GICs provide graphical depiction of annualised growth rates for every percentile of income distribution between two points in time. These growth rates do not represent direct change in the material conditions of specific households conditional on their original relative position, since we examine a series of cross-sections and not data in a panel format, and also because, households may change their relative position in the distribution over time. However, to appreciate the degree of income levels and changes at different points in time across the distribution it is still useful to examine the evolution of incomes with reference to specific percentiles. Fig. 5 computes GICs for the unadjusted and top-corrected data between 1994 and 2015. As we can see the GIC

---

[23] We add Spain to the comparison despite the fact that it is a 'new register country' rather than 'survey country' since it has the highest level of income inequality (as measured by the Gini index) among the EU countries studied by in Bartels and Metzing (2019).

[24] The estimates of the Gini index for net incomes in Poland based on the unadjusted PHBS data are not significantly different in general from those based on EU-SILC data (results available upon request). Therefore, the severe underestimation of top incomes is not specific to the PHBS data but seems to be a universal feature of household surveys in Poland.
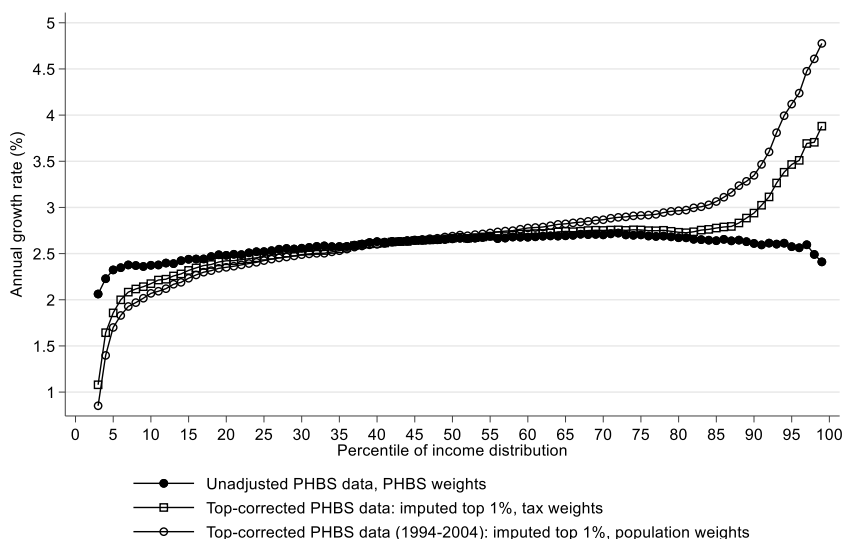
Fig. 5. Annual real income growth rates by percentile of income distribution in Poland, 1994–2015. *Note:* We do not show confidence intervals to avoid cluttering the figure. Percentiles expressed with respect to real equivalised disposable household income.
Source: Own calculations using PHBS data.

for unadjusted distribution is fairly flat suggesting that real incomes in Poland grew at a similar pace of about 2.5% per year across the income distribution.

The overall picture is strikingly different for the top-corrected GICs. Both have strictly positive slopes, implying that percentiles in the upper part of income distribution grew faster than percentiles at the bottom part. According to the corrected estimates, the poorer part of the income distribution (up to about 15th percentile) have experienced lower annual rates of income growth than the unadjusted estimates suggest, although our correction for the bottom decile group does not exceed one half of a percentage point on average. We observe a much bigger discrepancy between unadjusted and corrected data for the upper part of the income distribution (above 65th percentile). The top-corrected estimates suggest that annual income growth of the top percentiles of the distribution has been underestimated in the range from about 0.4–0.8 p.p. (at the 90th percentile) to 1.5–2.4 p.p. (at the 99th percentile). In other words, cumulative growth in real income over 1994–2015 for the top percentile reached 122–167%, while for the bottom 10% the corresponding number is at most 57%.[25] This is consistent with the calculated rise in overall income inequality as measured by the Gini index (see Fig. 4). Our estimate of the income growth rate for the 99th percentile (about 4.8% per year) is relatively high by regional standards. For example, Novokmet et al. (2018) have estimated that the annual rate of growth for the 99th percentile of pre-tax national income distribution in Russia in 1989–2016 was slightly lower than 3.5%.

With this in mind, we need to note that the top-corrected GIC for Poland in 1994–2015 is still positive at each percentile, indicating that incomes across the distribution in Poland have grown in absolute terms between 1994 and 2015.[26] The upward-sloping shape of our GIC for disposable income distribution is similar to that of the GIC for fiscal income distribution obtained by Bukowski and Novokmet (2018) for the period 1989–2015, with the exception that their estimates suggest negative rates of growth for the bottom few percentiles. This can be explained with reference to differences in income definitions and time coverage of the studies.

4.4. Redistribution and progressivity

This section investigates how our top-correction of household survey data affects measures of redistribution and progressivity of social insurance and direct taxation (income taxes, employees' mandatory social security contributions, and health insurance) in Poland. The system of social insurance has been reformed multiple times during our period under study. The modern social insurance system with social security contributions (SSC) paid partially by employees and employers has been introduced in 1999. Before that, all SSCs were paid by employers only. For these reasons, our analysis of redistribution and progressivity due to social insurance and direct taxes covers the period 1999–2015. In 2007 and 2008, the SSC rates were reduced by 7 p.p. Major reforms have taken place also in the system of personal income tax (PIT). Before 2009, there was a single progressive PIT schedule with three tax brackets and marginal tax rates of 19%, 30% and 40%. As described in Section 4.1, a reform in 2004 provided an option of choosing between the progressive tax rate schedule and the flat rate (19%) for non-agricultural business activity. Since 2009, the number of tax brackets and marginal tax rates in the progressive schedule was reduced from three to two (18% and 32%). About 95% of PIT payers were in the first tax bracket

---

[25] The cumulative income growth at the median income reached 75%.
[26] On Fig. 5, we do not show estimates for the two lowest percentiles which are slightly negative. However, this may result from measurement error and noisiness of income distribution at the very lower end.

in 2015 (see Table A1 in the Supplementary Appendix).[27]

To measure redistribution, we use the most popular redistribution measure, the Reynolds-Smolensky (RS) index of redistribution (Reynolds and Smolensky 1977), which in our context is defined as the percentage difference between the Gini coefficient for gross income (before income taxation, SSCs and health insurance) and the Gini coefficient for net income (after income taxation, SSCs and health insurance). Fig. 6 (panel a) presents the evolution of the RS index for the unadjusted and top-corrected survey data. Both series indicate that the redistributive effect diminishes over time, which reflects declining marginal rates of PIT and falling SSC rates. The top-corrected estimates show that the percentage reduction in the Gini index due to social insurance contributions and PIT has fallen from 19.2% in 1999 to 11.6% in 2015. Our estimates suggest that in 2005 Poland reached a comparatively low level of redistribution due to direct taxes and social insurance contributions. Verbist and Figari (2014) show that the average RS index for the 15 "old" EU members was 15.3% in 2008 with only three countries (Italy, Spain, France) having a lower level of redistribution (due to PIT and SSCs) than Poland in 2015.

Finally, we consider the problem of tax progressivity using the Kakwani (1977) index defined as the difference between the concentration coefficient of the tax and the Gini coefficient of gross income.[28] Fig. 6 (panel b) plots our estimates of the Kakwani index for the unadjusted and top-corrected income distributions. Without the top-correction, the progressivity in 2015 is overestimated by 2.3 p.p. (or by 40%). According to the top-corrected estimates, the Kakwani index fell from 7.5% in 1999 to 3.4% in 2015. The latter number ranks Poland as the country with lowest PIT and SICs progressivity in the EU (Verbist and Figari 2014; Mantovani 2018).

It is worth noting that declining actual income redistribution was not in line with Poles' preferences. Available survey evidence shows that citizens' demand for redistribution has been increasing over 1992–2009 (Choi 2019; Bussolo et al., 2021). This demand was not met with an appropriate policy response until in 2015 the Law and Justice party came forward with a generous proposal of a quasi-universal benefit for families with children, which was implemented swiftly in March 2016, shortly after its electoral victory.

5. Conclusions

This paper uses combined household survey and income tax data to re-evaluate trends in income inequality and redistribution during the post-socialist transition in Poland from early 1990s to the mid-2010s. We correct for the problem of survey under-coverage and under-reporting of top incomes by using Pareto imputation, survey reweighting and microsimulation techniques. We present first top-corrected trends in inequality of the standard of living measured in terms of real equivalised disposable household incomes. In contrast to the prevailing literature based on unadjusted household survey data, we find that inequality of living standards in Poland rose sharply between 1994 and 2015. The top-corrected Ginis in terms of our standard of living metric grew by 4–8 p.p. or by 14–26% over the period under study. Our estimates suggest that while Poland was already a relatively unequal country in the early 1990s, it has become one of the most unequal European countries among those for which comparable estimates exist. The top-correction to the Gini index that we estimate is 2–3 times larger than those obtained using similar methods for other (mostly advanced) European countries. We also find that income growth has been most pronounced at the upper end of the distribution. Top-corrected estimates show that progressivity of direct taxation and social insurance has fallen in 2015 to the lowest level among the EU countries. These results question the standard narrative about the Polish post-transition development, which maintains – on the basis of official indices – that growth in Poland has been relatively equitable. They also give weight to the arguments that there were good reasons for distributional tensions to intensify in Poland in the run up to the 2015 elections and that these should be considered as a factor paving the way to the rise of populism in the country.

Apart from the main results, due to the application of detailed microsimulation methods, we also presented analysis of the changes in redistributive effects of direct taxes and social security contributions. Part of the growth in inequality can be related to specific policy choices and adjusting the data for top incomes allows for a more complete picture of the actual redistributive effects of those reforms.

On the methodological level, our approach of combining Pareto imputation with survey reweighting techniques proves to be a successful method for top-correcting survey-based income information. We have shown that relying only on Pareto imputation without survey reweighting can substantially underestimate income inequality levels and changes. Survey reweighting with weights estimated using administrative tax data can be especially useful in recovering bounds on true inequality levels when top incomes, as our case of Poland before 2004 clearly shows, are significantly unreported or underreported. We believe that our methodology can be usefully applied in the context of top-correcting income distributions in other transition countries or emerging economies.

Our empirical results have important implications for the scientific literature assessing inequality trends in the follow up of post-socialist transformations or modernization processes in emerging countries. They indicate that using income tax data and imputation or reweighting techniques to account for the problem of missing top incomes in survey data can significantly alter the conclusions about income inequality levels and trends. Although the literature on correcting income distributions in transition and emerging countries (see, e.g., Bukowski and Novokmet 2017, 2021; Novokmet et al., 2018; Piketty et al., 2018) has been growing, clearly more data (especially individual income tax microdata) and research are needed.

Our results suggest also that survey-based estimates of income inequality indices such the Gini coefficient for many transition and emerging countries may be underestimated to a much larger extent than those for advanced economies. Such estimates are frequently

---

[27] Kopczuk (2012) and Bukowski and Novokmet (2017) provide more details of the Polish PIT system.

[28] The Kakwani progressivity index reflects the departure of tax system from proportionality to the gross (pre-tax) incomes. If the tax system is proportional, the concentration curve for the tax (showing the cumulative proportion of taxes versus position in the gross income distribution) should coincide with the Lorenz curve for pre-tax income. See, e.g., Lambert (2001) for more details.

a) Redistributive effect of direct taxes

b) Kakwani's progressivity index for direct taxes
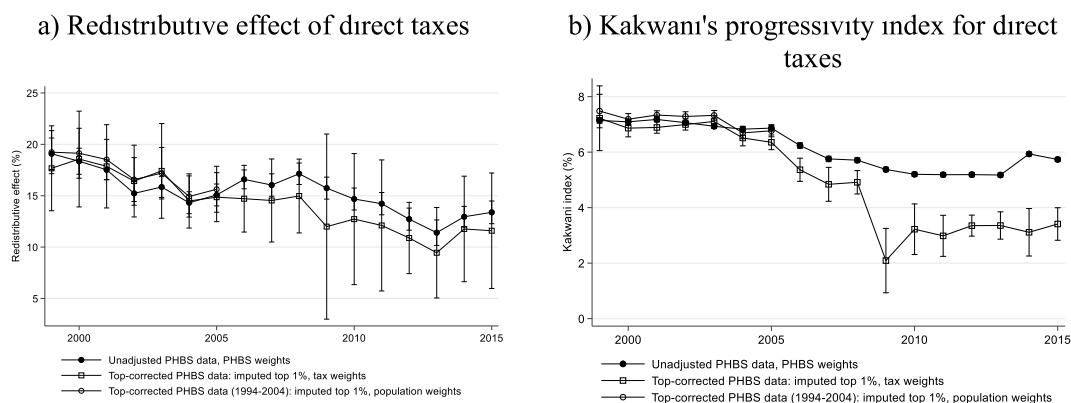
Fig. 6. Redistributive effect and Kakwani's progressivity index for direct taxes in Poland, 1999–2015. *Note:* Vertical lines show 95% confidence intervals obtained using bootstrap. Direct taxes include: Social Security Contributions, Personal Income Tax and Health Insurance. *Source:* Own calculations using PHBS data.

used in cross-country empirical studies exploiting inequality and redistribution data from such cross-national sources as the World Income Inequality Database (WIID) or Luxembourg Income Study (LIS). However, if those figures are distorted as much as the inequality numbers for Poland, the results of empirical studies on cross-country determinants or consequences of income inequality may be seriously biased or erroneous. In particular the relationship between inequality and redistributive preferences, which may be fundamental for our understanding of current political processes, may deserve a renewed closer examination. Finally, we should point out that our top-correction procedure for Poland is based on highly aggregated income tax statistics that allow only for computation of top 5% and top 1% income shares. Our estimates could have been more precise if access to comprehensive tax statistics or individual tax return data were granted to the research community. Studies based on disaggregated tax information could provide better inequality estimates, and in particular would allow for research on specific episodes of inequality spikes in the follow up of the post-socialist transformation.

Declaration of competing interest

We declare no conflict of interest.

Data availability

The authors do not have permission to share data.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejpoleco.2021.102121.

References

Alvaredo, F., 2011. A note on the relationship between top income shares and the Gini coefficient. Econ. Lett. 110 (3), 274–277.
Atkinson, A.B., 2007. Measuring top incomes: methodological issues. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes over the Twentieth Century: A Contrast between Continental European and English-speaking Countries. Oxford University Press, Oxford, pp. 18–42.
Atkinson, A.B., Piketty, T., Saez, E., 2011. Top incomes in the long run of history. J. Econ. Lit. 49 (1), 3–71.
Bargain, O., Morawski, L., Myck, M., Socha, M.W., 2007. As SIMPL as that: Introducing a Tax-Benefit Microsimulation Model for Poland. IZA Discussion Papers 2988. Institute for the Study of Labor (IZA).

Bartels, C., Metzing, M., 2019. An integrated approach for a top-corrected income distribution. J. Econ. Inequal. 17 (2), 125–143.

Bengtsson, E., Waldenström, D., 2018. Capital shares and income inequality: evidence from the long run. J. Econ. Hist. 78 (3), 712–743.

Blanchet, T., Flores, I., Morgan, M., 2019. The weight of the rich: improving surveys using tax data. WID. world Working Paper Series, 2018/12.

Brzeziński, M., Kostro, K., 2010. Income and consumption inequality in Poland, 1998–2008. Bank i Kredyt 41 (4), 45–72.

Brzeziński, M., Najsztub, M., 2017. The impact of "family 500+ programme on household incomes, poverty and inequality. Polityka Społeczna 1 (13).

Bukowski, P., Novokmet, F., 2017. Top Incomes during Wars, Communism and Capitalism: Poland 1892-2015. WID world Working Paper series No. 2017/22.

Bukowski, P., Novokmet, F., 2018. Inequality in Poland: Estimating the Whole Distribution by G-Percentile 1983-2015 (No. 731). LIS Cross-National Data Center, Luxembourg.

Bukowski, P., Novokmet, F., 2021. Between communism and capitalism: long-term inequality in Poland, 1892–2015. J. Econ. Growth.

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. Rev. Econ. Stat. 94 (2), 371–388.

Burkhauser, R.V., Hérault, N., Jenkins, S.P., Wilkins, R., 2017. Top incomes and inequality in the UK: reconciling estimates from household survey and tax return data. Oxf. Econ. Pap. 70 (2), 301–326.

Bussolo, M., Ferrer-i-Carbonell, A., Giolbas, A., Torre, I., 2021. I perceive therefore I demand: the formation of inequality perceptions and demand for redistribution. Rev. Income Wealth.

Choi, G., 2019. Revisiting the redistribution hypothesis with perceived inequality and redistributive preferences. Eur. J. Polit. Econ. 58, 220–244.

Chrostek, P., Klejdysz, J., Korniluk, D., Skawiński, M., 2019. Wybrane Aspekty Systemu Podatkowo-Składkowego Na Podstawie Danych PIT I ZUS 2016. Ministerstwo Finansów, Warszawa.

Creedy, J., 2004. Survey reweighting for tax microsimulation modelling. In: Studies on Economic Well-Being: Essays in the Honor of John P. Formby. Emerald Group Publishing Limited, pp. 229–249.

Deville, J.C., Särndal, C.E., 1992. Calibration estimators in survey sampling. J. Am. Stat. Assoc. 87 (418), 376–382.

Gimpelson, V., Treisman, D., 2018. Misperceiving inequality. Econ. Polit. 30 (1), 27–54.

Gradzewicz, M., Growiec, J., Kolasa, M., Postek, Ł., Strzelecki, P., 2018. Poland's uninterrupted growth performance: new growth accounting evidence. Post Commun. Econ. 30 (2), 238–272.

Growiec, J., 2012. Determinants of the labor share: evidence from a panel of firms. E. Eur. Econ. 50 (5), 23–65.

Guriev, S., Papaioannou, E., 2020. The political economy of populism. Forthcoming in Journal of Economic Literature.

Hagemejer, J., Popowski, P., 2014. The distribution of monopolistic markups in the Polish economy. E. Eur. Econ. 52 (6), 3–31.

Haggard, S., Kaufman, R., 2021. Backsliding: Democratic Regress in the Contemporary World. Cambridge University Press.

Heathcote, J., Storesletten, K., Violante, G.L., 2017. Optimal tax progressivity: an analytical framework. Q. J. Econ. 132 (4), 1693–1754.

Jenkins, S.P., 2017. Pareto models, top incomes and recent trends in UK income inequality. Economica 84 (334), 261–289.

Kakwani, N.C., 1977. Measurement of tax progressivity: an international comparison. Econ. J. 87 (345), 71–80.

Keane, M.P., Prasad, E.S., 2002. Inequality, transfers, and growth: new evidence from the economic transition in Poland. Rev. Econ. Stat. 84 (2), 324–341.

Kopczuk, W., 2012. The Polish Business "Flat" Tax and its Effect on Reported Incomes: a Pareto Improving Tax Reform. Columbia University (Working Paper).

Kordos, J., Lednicki, B., Zyra, M., 2002. The household sample surveys in Poland. Statistics in transition 5 (4), 555–589.

Kośny, M., 2019. Upper tail of the income distribution in tax records and survey data. Evidence from Poland. Argumenta Oeconomica 42 (1), 55–80.

Kuhn, A., 2019. The subversive nature of inequality: subjective inequality perceptions and attitudes to social inequality. Eur. J. Polit. Econ. 59, 331–344.

Lambert, P.J., 2001. The Distribution and Redistribution of Income. Manchester University Press, Manchester and New York.

Mantovani, D., 2018. Comparing Redistributive Efficiency of Tax-Benefit Systems in Europe. EUROMOD Working Paper EM12/18.

Markowski, R., 2016. The Polish parliamentary election of 2015: a free and fair election that results in unfair political consequences. W. Eur. Polit. 39 (6), 1311–1322.

Meltzer, A.H., Richard, S.F., 1981. A rational theory of the size of government. J. Polit. Econ. 89 (5), 914–927.

Mitra, P., Yemtsov, R., 2007. Increasing inequality in transition economies: is there more to come? In: Bourguignon, F., Pleskovic, B. (Eds.), Annual World Bank Conference on Development Economics – Regional 2007: beyond Transition. World Bank, Washington, DC, pp. 59–102.

Morawski, L., Myck, M., 2010. "Klin"-ing up: effects of Polish tax reforms on those in and on those out. Lab. Econ. 17, 556–566.

Myck, M., Najsztub, M., 2015. Data and model cross-validation to improve accuracy of microsimulation results: estimates for the Polish Household Budget Survey. International Journal of Microsimulation 8 (1), 33–66.

Myck, M., Najsztub, M., 2016. Distributional Consequences of Tax and Benefit Policies in Poland: 2005-2014. CenEA Working Paper Series WP02/16.

Myck, M., Trzciński, K., 2019. From partial to full universality: the family 500+ programme in Poland and its labor supply implications. ifo DICE Report 17 (3), 36–44.

Niehues, J., 2016. Ungleichheit: wahrnehmung und Wirklichkeit – ein internationaler Vergleich. Wirtschaftsdienst 96 (1), 13–18.

Nolan, B., Richiardi, M., Valenzuela, L., 2018. The drivers of inequality in rich countries. Forthcoming in Journal of Economic Surveys.

Novokmet, F., Piketty, T., Zucman, G., 2018. From Soviets to oligarchs: inequality and property in Russia 1905-2016. J. Econ. Inequal. 16 (2), 189–223.

OECD, 2015. In it Together: Why Less Inequality Benefits All. OECD publishing, Paris.

Perugini, C., Pompei, F. (Eds.), 2015. Inequalities during and after Transition in Central and Eastern Europe. Palgrave Macmillan, London.

Piatkowski, M., 2018. Europe's Growth Champion: Insights from the Economic Rise of Poland. Oxford University Press, Oxford.

Piketty, T., Yang, L., Zucman, G., 2018. Capital Accumulation, Private Property and Rising Inequality in China, 1978-2015. American Economic Review, forthcoming.

Podkaminer, L., 2003. A note on the evolution of inequality in Poland, 1992–99. Camb. J. Econ. 27 (5), 755–768.

Ravallion, M., Chen, S., 2003. Measuring pro-poor growth. Econ. Lett. 78 (1), 93–99.

Reynolds, M., Smolensky, E., 1977. Public Expenditures, Taxes, and the Distribution of Income: the United States, 1950, 1961, 1970. Academic Press, New York.

Szulc, A., 2000. Economic transition, poverty and inequality: Poland in the 1990s. Statistics in transition 4 (6), 997–1017.

Tóth, I.G., 2014. Revisiting grand narratives of growing income inequalities: lessons from 30 country studies. In: Nolan, B., et al. (Eds.), Changing Inequalities and Societal Impacts in Rich Countries: Thirty Countries' Experiences. Oxford University Press, Oxford, pp. 11–47.

Tworzecki, H., 2019. Poland: a case of top-down polarization. Ann. Am. Acad. Polit. Soc. Sci. 681 (1), 97–119.

Verbist, G., Figari, F., 2014. The redistributive effect and progressivity of taxes revisited: an International Comparison across the European Union. Finanzarchiv: Public Finance Analysis 70 (3), 405–429.

Verma, V., Betti, G., 2010. Data accuracy in EU-SILC. In: Atkinson, B.A., Marlier, E. (Eds.), Income and Living Conditions in Europe, pp. 57–78.

World Bank, 2017. Lessons from Poland, Insights for Poland: A Sustainable and Inclusive Transition to High Income Status. World Bank, Washington, DC.