Recent trends in income inequalities in Hungary using administrative data*

András Svraka†

July 26, 2021

Abstract

This paper uses 2009–19 income tax data to investigate trends in Hungarian market income inequalities. Wage inequality decreased, total inequality increased but these changes differ across the income distribution. Income share of the bottom 40% decreased slightly due an inflow of marginally attached workers. Wages above this threshold – corresponding to full time employment – significantly flattened, including in the top 1%, while total income's share remained mostly unchanged. Similar patterns hold within groups for gender, age, and regions but between group inequality rose for age and decreased for regions due to changes in taxpayer population. Growth was broadly shared for non marginally attached, although in the top 0.1% all growth came from capital income. Persistence of income during this period was low in the bottom half of the income distribution, suggesting an increased inequality in this segment might not lead to higher inequality in the long run.

*I'm grateful to Katalin Horváth, Tibor Keresztély, Péter Tóth, and Bálint Ván for comments and suggestions, and to the staff of the National Tax and Customs Administration of Hungary for providing access to the anonymized tax microdata. I thank the support of Norbert Izer, and Benedek Nobilis. The views expressed herein are those of the author and do not necessarily reflect the views of the Ministry of Finance.

†Senior researcher, Ministry of Finance, andras.svraka@pm.gov.hu

1 Introduction

The rise of income inequality became a widely discussed topic over the last decade, as across many countries inequality rose substantially. In Hungary, in particular, income inequality has been low even compared to the relatively low inequalities in the Central Eastern European region. However, post tax inequality rose after a major tax reform in the early 2010s, introducing a flat personal income tax rate and abolishing an in-work tax credit (European Commission 2019), redistributing income upwards (Tóth G. and Virovácz 2013). These reforms, along with other measures are likely to have contributed to a large increase in employment rates (Benczúr, Kátay, and Kiss 2012). The actual rise in inequality in the 2010s though (Eurostat 2021) remained below simulation based predictions, in fact pre-tax inequality decreased by some measures (see e.g. Blanchet, Chancel, and Gethin 2019, whose approach is based on the concept of distributional national accounts).[1]

This observation is the main motivation of this paper. How did pre-tax market incomes change over the last decade, and what factors could have counterbalanced the rise in inequality due to changes in redistributive policies? In this paper we give an overview of the recent trends in the distribution of market income using descriptive methods.

While these questions can be investigated using household surveys, which are used in the literature cited above, our approach will deviate from them, and we'll use administrative data in the form of personal income tax records. In some ways this leads to a more limited view of income inequalities, as Hungarian tax records have no information on household members, thus our analysis will be limited to individual income. As our aim in this paper is to look at changes in the distribution of market incomes, this should not be a major limitation. Meanwhile, the richness and completeness of administrative data offers several advantages. Data is available until 2019, the last year of a strong business cycle before the Covid-19 pandemic, which is not the case of microdata from household surveys, especially household surveys augmented using administrative data. Administrative data allows us to look at the very top of income distribution, where income growth outpaced growth in the rest of the income distribution during the last decades (Blanchet, Chancel, and Gethin

[1] Some of the difference might be explained by dynamic effects, as higher employment could reduce inequality. The results of behavioural simulations in Benczúr, Kátay, and Kiss (2012) show a lower increase than the static, non-behavioural results of European Commission (2019) but this cannot account for the whole gap.

2019). Additionally, with administrative data attrition is not an issue, thus we can look at long run income mobility in more detail. Nevertheless, the results of this paper have a limited comparability to research based on household data. Therefore this paper's contribution is to offer a complementary perspective, while noting the importance of household level analysis on inequality in living conditions, or poverty.

Previous research on Hungarian income inequalities using administrative data includes Tóth G. and Virovácz (2013), who used tax return data to assess the distributive effects of the flat tax reform of 2011. Kovács (2011) looked at income inequality using tax data from the late 1990s until the late 2000s, comparing results from pure tax return based inequality to results form household surveys, finding large differences in the levels of inequality, but similar trends from both data sources.

Therefore this paper is the first detailed analysis of income distributions and income inequality in administrative data sources for a later period. Our approach also extends previous work in other areas. We go beyond income decile based disaggregation of the distribution, and use percentiles, and even more fine grained fractiles to see how incomes changed in the top 1%. We also disaggregate our results by demographic and regional groups to see whether composition effects also drove inequality dynamics. Finally, we present first results on longer term income mobility and persistence.

Our results show that Hungarian taxpayers' real incomes grew substantially over the last decade. Wage inequality decreased, primarily due a flattened income distribution above the full time minimum wage, including in the top 1%. However, along with the rising employment rate, a large number of previously inactive taxpayers entered the labour market – including older workers after raising the retirement age but also among the youth –, some of who work part time, or who are only marginally attached to the labour force. This resulted in a fall in average incomes below the minimum wage. However, within group inequalities fell for these age groups, while within group inequalities in other disaggregations showed the same pattern, as the national wage distribution.

Inequality in total taxable income rose slightly. While income share in the top percentile increased a little, most of the total increase was driven by the bottom half of the income distribution, partly due to the aforementioned expansion of taxpayer population but also due to measurement issues, as a large number of self-employed stopped reporting taxable income after the introduction of a specialized tax scheme for small businesses. Therefore we consider results for total income somewhat less reliable, when looking at the entire distribution. An interesting result emerged for the top of the distribution, as

wage income growth in the top 0.1% was close zero, while total income grew at a much faster rate than anywhere else in the income distribution, suggesting a shift in income sources.

Although the descriptive methods used in this paper are not sufficient to identify causes other than the demographic and regional factors, some results strongly point to the importance of large minimum wage rises – supported by cuts in payroll taxes – throughout the 2010s in the falling wage inequality. Growth was the highest around the full time, full year minimum wage, and slowly tapered off above. Additionally, the strong economic recovery of the late 2010s, and the following tight labour market also looks like an important factor, as most of the flattening happened during this period.

Nevertheless, the rise in marginally attached workers could be worrying, if it leads to permanently low income. It is certainly an issue that requires close monitoring but our results on income mobility showed that persistence is the lowest in the bottom half of the income distribution.

2 Data

This paper uses the full universe of taxpayers filing an annual personal income tax return for the years 2009 to 2019. Income data refers to annual income and is not censored at either the bottom, or the top of the distribution.[2] Due to a consistent anonymized ID, individuals can be followed over time.

2.1 Income sources

Personal income tax is assessed and paid on a individual basis in Hungary, and generally everybody with taxable income has to file an annual tax return declaring all taxable income. We use data from these tax filings in this paper. However, some types of income are not declared by individuals, and some sources of income are taxed outside the personal income tax system. With these caveats we will use two measures of income in this analysis. The primary outcome will be employment income, which is broadly comparable across the period analysed in this paper. The other will be total income de-

[2] Although tax avoidance can lead to an underestimation of income at the top end of the distribtution (Alstadsæter, Johannesen, and Zucman 2017).

clared on the tax returns,[3][4] which has significant biases due to some changes in the tax system during the period analysed.[5] All income levels were adjusted to 2019 Hungarian Forints using Eurostat's Harmonised Consumer Price Index.

2.1.1 Limitations

Income taxed at source is declared by the entities who pay out those income streams, thus these are not included in individuals' tax returns. These income sources include fringe benefits (paid by employers) and taxes on some capital income, e.g. interest income on savings accounts and treasury bills, mutual and pensions funds (paid by financial institutions managing those assets). This omission is likely to reduce the observed income levels and inequalities relative to the true total income.

Some atypical forms of employment, and self-employment regimes are not subject to personal income taxes, and income from them is generally missing from the tax returns. Changes in the tax system had an effect on how these income sources are reported, which introduced some bias in the observable non-employment income.

The first major type is simplified employment.[6] This scheme is aimed at low-wage temporary employment, including day labourers, with lump sum taxes and contributions set at a daily rate. The usage of simplified employment is limited by the number of days a person can work for the same employer. Wages in simplified employment can fall below the minimum wage. However, wages above the daily minimum wage are liable to personal income tax, and are reported on tax returns, although they are indistinguishable from

[3] Personal income taxes were levied at company cars until 2009 but since 2010 they are taxed at source, paid by businesses. As this was declared in a separate field in 2009 we do not include it in the total. Until 2010 some non-taxable incomes, like pensions and scholarships had to be declared if the person had other taxable income as well. These non-taxable incomes had to be included in the tax base calculations but they were refundable, meaning they affected marginal rates. Since the introduction of a flat rate personal income tax in 2011 these previously refundable items are not decalred on the returns, and are not included in the totals for 2009 and 2010 either.

[4] Total income covers some non-market incomes as well, including some childcare benefits. Reporting of these incomes didn't change, this shouldn't bias our results.

[5] Although the tax return data has information on more detailed income sources, we will limit the analysis to these two. The breakdown of non-wage income is hard to interpret, due to the wide variety of tax planning considerations around capital and self-employemnt income (capital gains, dividends, passthrough via self-employed business, foreign income etc.).

[6] Employers have to declare wages paid in simplified employment on a per person basis but this data was not available for this paper.

regular employment income. The number of people employed in this scheme increased throughout the period analysed in this paper from a monthly average of around 150 thousand to around 250 thousand (with large seasonal variation, mainly due to the seasonality of the agriculture and hospitality sectors). It is an open question how much of that comes from higher demand during the strengthening labour market of the 2010s, and how much could be tax avoidance, by substituting regular employment in some low wage sectors. Either way, it could lead to some working people completely missing from our data, or – more likely – the unavailability of income from simplified employment leads to downward biased observed incomes at the lower sections of the income distribution, as workers in simplified employment could also have other sources of income.

An optional lump sum tax for self-employed (KATA) was introduced in 2013 with a steadily increasing number of entrepreneurs opting for this form of taxation (close to 300 thousand by the end of 2019). KATA offers a low tax rate that replaces all income taxes and social contributions by a single tax, and many self-employed businesses switched to KATA from being a sole proprietor paying personal income tax, or small partnership paying corporate income tax. They always had income subject to personal income tax before the switch (typically as dividend income, and if they were working full time as self-employed entrepreneurs, some income is typically declared, usually at the rate of the minimum wage due to some peculiarities of the Hungarian tax and social security system) but KATA is an entirely separate tax. KATA subjects report business revenue but as it is a lump sum tax, and costs cannot be deducted, they are not reported, making the imputation of KATA subjects' income difficult. At the same time another simplified tax regime (EVA) saw significant outflows due to increased tax rates (partly into KATA, partly into other corporate taxes), from around 100 thousand EVA subjects in 2009, to around 20 thousand in 2019. EVA subjects paid some personal income tax on what could be described as their labour income (a monthly income drawn form their business, similarly to the above-mentioned system) but most of their business income was declared on EVA forms, which couldn't be merged into personal tax returns for this paper. This means that full time self-employed businesses can disappear from the personal income tax records, while part-time self-employed taxpayers' business income will be omitted from personal income tax returns. As a third of KATA businesses are not the main job of the owner, these trends can cause serious biases in our data.

2.2 Demographic and geographical variables

We will augment the analysis of inequality with some background variables available in our dataset.

Basic demographic variables include birth year and gender. Both are generally available but with some errors. Date of birth is used for identifying individuals by the tax office, therefore it is missing only for a handful of observations, most of which can be imputed from earlier and later observations. However, primarily due to tax returns for 2009 and 2010 12% of all individuals covered in all tax returns have multiple values for year of births. Gender is more incomplete. Its values come from an optional field on the tax returns, and the tax office runs a name based classification algorithm. This could fail on uncommon, foreign, or mistyped names. There's an increasing missingness for gender data in the latest years, potentially due to more foreign taxpayers (we have no data on nationality). Additionally, 3% of individuals had multiple values for gender across the years. Both year of birth and gender were harmonized for every individual using the most common, or in case of a tie, the latest value they provided. After harmonisation an additional 9,913 individuals were dropped from the dataset due to no gender, or age data in any year.

Geographical information comes from two sources. Each individual is assigned to a regional directorate of the tax authority. Additionally, the tax authority has information on residential data for most taxpayers.[7] The tax directorate is available for every person-year observation, while the latter is not available for 2009 and 2010, and has an increasing rate of missingness beginning in 2015. This is possibly due to the higher number of electronically filed tax returns, driven by the introduction of pre-filled tax returns. Data from the two sources are usually identical but the self-declared address could be more up-to-date than the tax directorate based location, and some taxpayers are always assigned to a central directorate regardless of their actual location. Therefore we will combine the two sources, to create regional (NUTS2) classifications, primarily relying on the self-declared address but imputing the county of the tax directorate if the former is missing.

[7] This includes mistyped and foreign addresses as well but those cannot be distinguished. Any postal code that cannot be correlated with a Hungarian address is considered missing.

3 Methods

This paper follows a descriptive methodology to explore trends in income inequalities in the past decade.

First, we look at some well-known inequality measures, the Gini coefficient and the Theil index for the entire population and also for demographic and geographic subgroups. As shown in Section 4, there are large differences in mean and median incomes across various groups. Additionally, the composition of the workforce also changed significantly throughout the period analysed in this paper. Therefore changes in overall inequalities can come from changing inequality within groups but also from a composition effect. While both measures of inequality satisfy most requirements discussed in the literature (anonymity, scale invariance, population principle, principle of transfers, see Cowell 2016), the Gini coefficient cannot be easily decomposed into within and between group components. The Theil index has the property of additive decomposability which we use to measure changes in the share of between group inequalities. Nevertheless it is useful to look at both for overall and within group inequalities as a robustness check but also for the intuitive meaning of the Gini coefficient. It can be thought of as either the average interpersonal difference in incomes across all pairs of individuals, or the area between the Lorenz curve – which plots the cumulative income share by relative income – and the diagonal line representing an income distribution of full equality, while the entropy based Theil index doesn't have a similar intuitive meaning.

The indicies compress the entire income distribution into single numbers that don't show how the shape of the income distribution looks like. We'll use Lorenz curves, population shares of subgroups by income level, and growh incidence curves – describing changes in quantile averages – to see changes along the income distribution.

Finally, taking advantage of the panel nature of our dataset, we will explore the persistence of income. Longitudinal analysis of income inequality is increasingly common (see e.g. OECD 2018). Our dataset covering twelve years is not sufficient to look at changes in lifetime income between generations but even shorter term movements across the income scale can highlight some important features of income dynamics. For this purpose we'll look at annual and multi-year transition probabilities between deciles, and the frequency of an individual occurring at a particular position of the income distribution during the entire period analysed.

It is important to note that all these measures of inequality indices will

have very limited comparability to other studies, as the population and income sources used in this analysis are specific to the context of the Hungarian tax system.

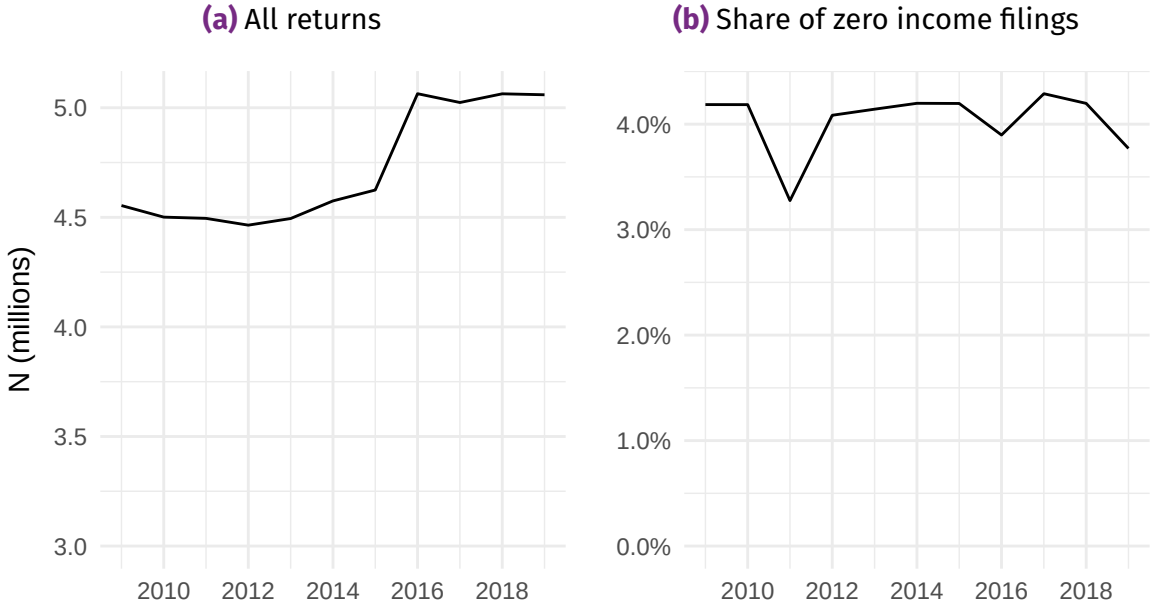
4 Descriptive statistics

The number of taxpayers grew over the period analysed. This is partly due to the increasing employment rate of the 2010s but there is also a large break in the time series in 2016 (Figure 1). Although we don't have direct evidence, this was most likely caused by the introduction and widespread adoption of pre-filled electronic tax returns. In this system the tax office prepares a tax return for everybody for whom they had third party income information, typically from employers' filings. These pre-filled forms become the final tax returns even if taxpayers don't approve them in any way. The trends suggest, that before 2016 many taxpayers neglected to send in their tax returns.[8] This probably didn't affect their tax liability, as the pre-filled forms were compiled from employers filings, for which advancements were already paid.

This change in the observed taxpayer population could bias our results. We cannot distinguish between those, who were truly new taxpayers in 2016 (e.g. young people entering the labour force for the first time, or parents returning from parental leave), and those who might have had taxable income before but didn't file their returns. Therefore we cannot create a counterfactual population to deal with this potential bias. Descriptive statistics suggest this change in taxpayer population did not have a major impact on the demographic and regional composition of the taxpayer population (Figures 2 and 3). Meanwhile the income distribution of new taxpayers in 2016 differs from other years. Most of them are low income – which is expected, as many non-regular tax filers are probably only marginally attached to the labour force – but prime age new taxpayers are relatively uniformly distributed across the entire income range (Figure A.4). Additionally, the income distribution and the share of those, who have taxes due, or refunds claimable at the time of filing is similar to those whose tax advancements matched their final tax liability (Figure A.4). This suggests the increased number of new taxpayers was mainly driven by the pre-filled tax returns, and not by higher willingness of voluntary

[8] Our datasets were compiled around three monhts after the filing deadlines for every year. This should include late filers as well but with the automatic acceptance of pre-filled forms, late filing is virtually impossible. Therefore some portion of the jump in 2016 could be explained by the different levels of completeness of our datasets.

Figure 1: Number of tax returns filed



See also Figures A.4 and A.5 for further break down of the newly appeared taxpayers in 2016.

filing. Such changes in data collection can be a downside of administrative sources, as definitions and measurement follow the requirements of the tax system, not those of statistical purposes.

Meanwhile, similarly to the ageing Hungarian population, the age distribution of taxpayers also shifted (Figure 2). Besides this cohort effect, the age and gender composition also changed due to the changing pension system, and due to the increasing employment rates in previously underemployed groups, like the younger cohorts. Most early retirement schemes were phased out, and the retirement age was raised from 62 to 65 in several steps. However, women can still retire after 40 years of work. The regional composition of taxpayers also changed noticeably (Figure 3). The Hungarian workforce has relatively low geographical mobility, thus a lot of these changes may be attributed to the varying strengths of regional labour markets but over the 12 years analysed in this paper labour mobility could play a role too. Due to demographic and geographic changes, it will be important to disaggregate changes in income inequalities along these dimensions.

After a dip during the recession of 2009–2011, taxable income grew over the period analysed (Figure 4). Mean and median incomes showed similar rates of

Figure 2: Density functions of taxpayers' age by gender between 2009 and 2019

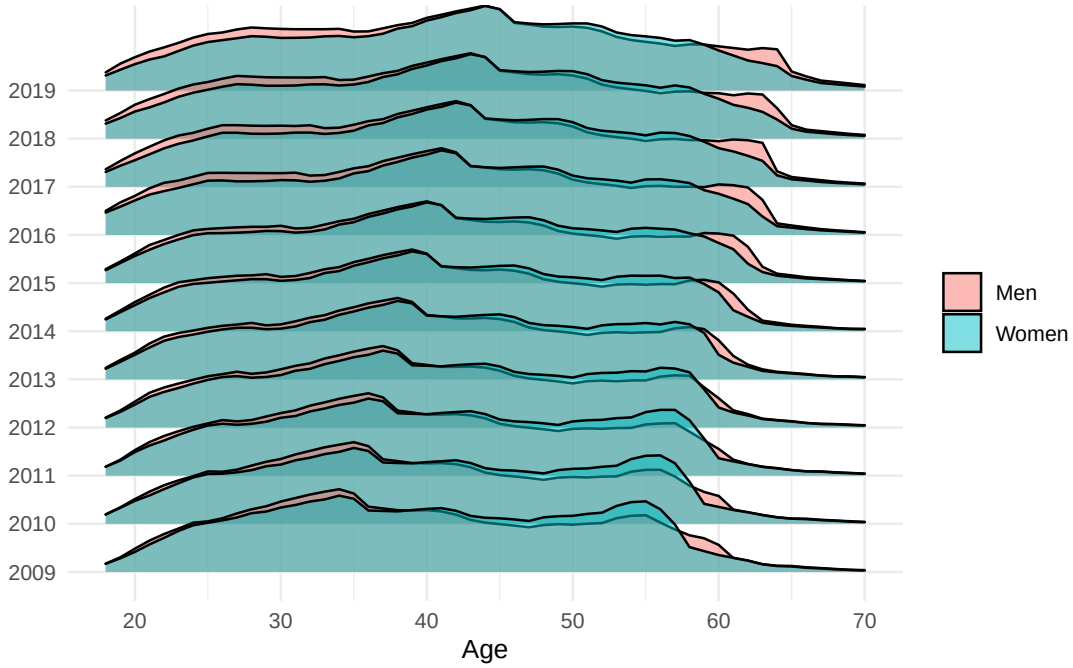


Figure 3: Taxpayers' population size and population share across regions

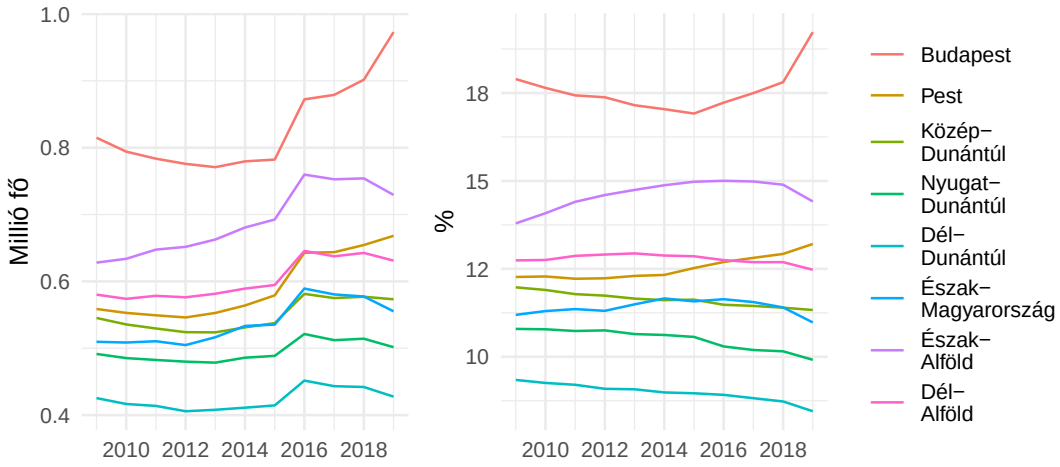
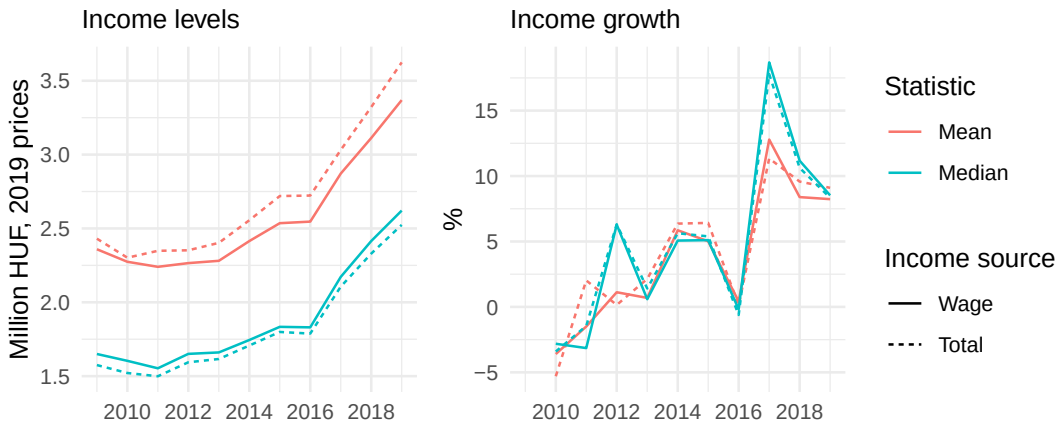


Figure 4: Mean and median taxable income



growth in most years. However, in two years with large rises in the minimum wage – 2012 and 2017 – median incomes grew more. The introduction of the new pre-filled filing system in 2016 can be seen as a break in the upwards trend. Its effect needs to be investigated in more detail.

The same patterns hold for demographic and geographic subgroups as well but we can see large disparities between groups. (Figures A.1–A.3). The average man earned more than the average woman but the medians were very close for the two groups. Young taxpayers saw smaller income rises than older groups. Differences between means and medians for the prime age groups were smaller than between genders but due to low employment rates, incomes for the youngest and oldest cohorts came disproportionately from non-wage sources, for which means were much higher than medians. Mean incomes were the highest in the central regions (Budapest and the surrounding Pest county) by a wide margin but median incomes were much closer across the country.

Growth rates in mean and median incomes were similar for all groups in all breakdowns, thus existing differences mostly persisted in the last decade.

5 Results

5.1 Inequality measures

Both the Gini and Theil indicies show a steady decrease in wage income, and a steady but smaller increase in total income inequality since 2009 (Figure 5).

What could explain the different trends in wage and total incomes? As discussed in Section 2.1.1, due to the continuous outflow of a large number of self-employed businesses (mostly small-scale businesses, like tradespeople) from personal income taxation, the middle could have hollowed out, leading to increased inequality in total income but not in wage income. Loose monetary policy can also cause higher inequality, especially at the top of the income distribution (Andersen et al. 2021; Amberg et al. 2021). In particular for Hungary, during the 2010s real estate prices rose sharply (MNB 2021), during which investors could reap large capital gains. Rental markets also boomed due to short term rentals on online platforms which could have contributed to increasing inequalities as well. The labour share in the economy decreased since the pre-recession levels,[9] which along with the higher inequalities of capital income could also lead to diverging trends. Another factor could be the measures enacted globally against international tax evasion, such as automatic exchange of information requests (Alstadsæter, Johannesen, and Zucman 2017; Beer, Coelho, and Leduc 2019). This could have lead to Hungarian residents booking more of their taxes at home, instead of in tax havens. Meanwhile the economic recovery and the following tight labour markets, along with minimum wage rises supported by payroll tax cuts[10] are plausible explanations for decreasing inequality in wage incomes. Additionally, measures to improve tax compliance[11] could have lead to an increase in reported incomes, especially for employees reporting the minimum wage but earning more than that in "envelope payments" (Elek et al. 2011).[12] Teasing out the relative importance of these factors would require extensive additional data (e.g. property and business records).

The effect of the electronic tax filing system reform in 2016 is ambiguous on inequality. Both the Gini, and the Theil indicies satisfy the population principle, which states that two income distributions are identical, if one is formed by replications of the other (Cowell 2016). As shown in Section 4, new taxpayers in 2016 are different from other years, and at least for some subpopulations the inflow of these new taxpayers could be described as a replication. Never-

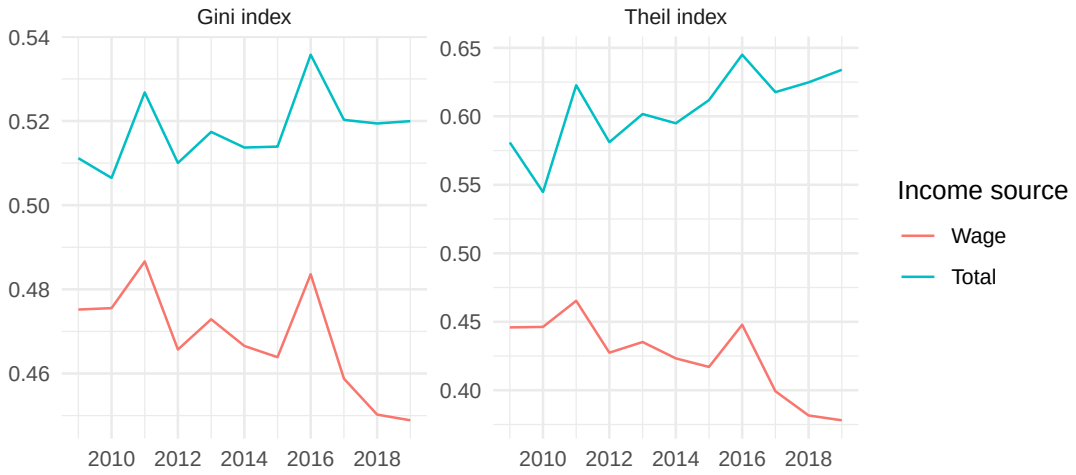
[9] https://dashboard.tech.ec.europa.eu/qs_digit_dashboard_mt/public/single/?appid=667e9fba-eea7-4d17-abf0-ef20f6994336&sheet=2f9f3ab7-09e9-4665-92d1-de9ead91fac7

[10] See Table A.2 for minumum wage levels. The raises in 2017 and 2018 were enacted along large cuts in the payroll tax (from 27% to 22%, then to 20%), which was followed by another 0.5% point cut.

[11] E.g. the introduction of online cash registers, or more detailed VAT declarations.

[12] The aforementioned anti-evasion measures lead an increased turnover (see Lovics et al. 2019), which indirectly could have lead to businesses also declearing previously unreported wages.

Figure 5: Inequality measures



theless, we see a sudden increase for both measures and both income sources in 2016. If most of the taxpayers who filed for the first time in 2016 kept filing in later years year as well, the peaks could suggest a permanent shift in the level of inequality. If this is the case, a counterfactual in which the newly appeared taxpayers were already regularly filing before 2016, would mean wage inequality might have decreased at a faster pace, while the increase in total income inequality might have been smaller, or even reversed.

5.2 Within and between group inequalities

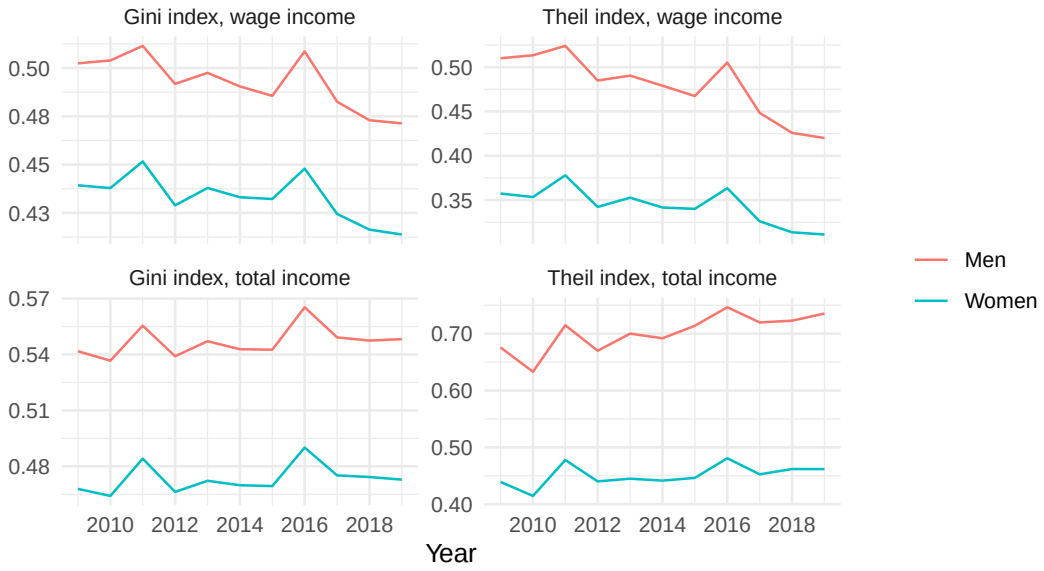
Figures 6–8 show within group inequalities by gender, age, and region.

Both genders show the same patterns, as the overall trends and men's incomes are more unequal in all categories.

In the age breakdown the prime age groups have similar patterns both in levels and trends, while the other groups' incomes are much more unequal. However, their trends diverge. Income inequality for taxpayers above 60 declined: due to the rising retirement age more people stayed in the workforce, where previously typically only higher income individuals chose to keep working in their 60s.[13] The below 17 group is small, and shows a very noisy pattern, outlier years are most likely related to inheritances. However, inequal-

[13] Ageing societies are generally associated with increased inequality but this is not the case for Hungary according to OECD (2017). Although it is imporant to note that our data includes only market income, and we ignore the effects of (delayed) pensiones in inequality.

Figure 6: Within group income inequality measures by gender



ity within the 18–24 group didn't decline. A possible explanation could be a compositional change, as the share of university graduates in this age group increased during the 2010s,[14] which could lead to higher inequality through a graduate wage premium.

The two central regions stand out with significantly higher inequalities relative to the rest of the country but wage inequality declined in all regions, and total inequality increased slightly.

There's heterogeneity in the inequality of all analysed subpopulations. Figure 9 summarises this finding in the share of the Theil between group component. Between gender inequality grew by a small amount from a low base. The importance the between group component was similar for age and region in 2009 but they diverged sharply.

In case of the Theil index the between component represents the income share weighted sum of the logarithmic differences of the group average from the population level average. Average incomes grew at a similar rate in all subgroups (Figures A.1–A.3) thus the change in between group inequalities mostly came from changing income shares, i.e. the changing composition of the taxpayers population along these dimensions.

[14] https://www.ksh.hu/stadat_files/okt/en/okt0005.html

Figure 7: Within group income inequality measures by age

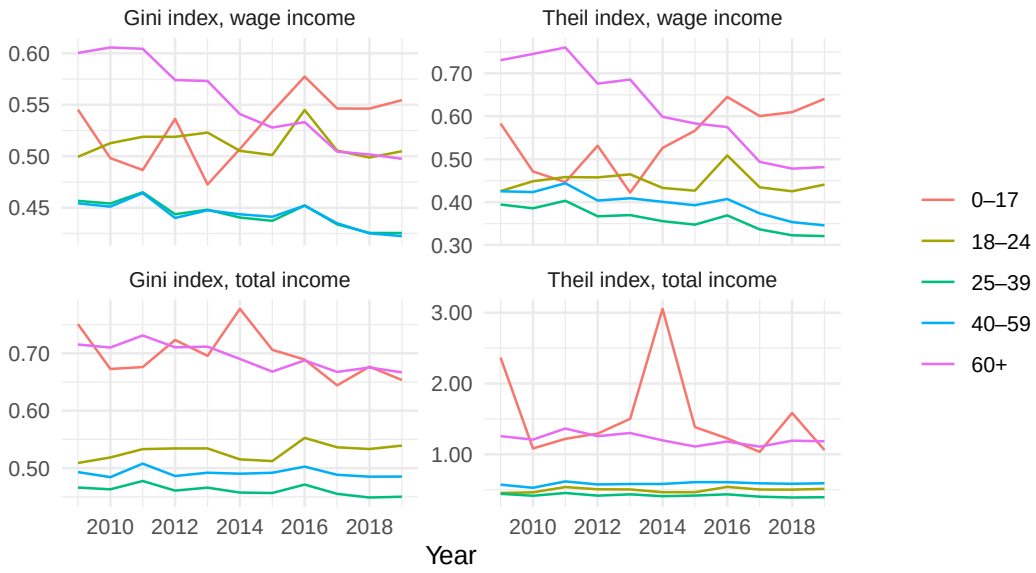


Figure 8: Within group income inequality measures by region

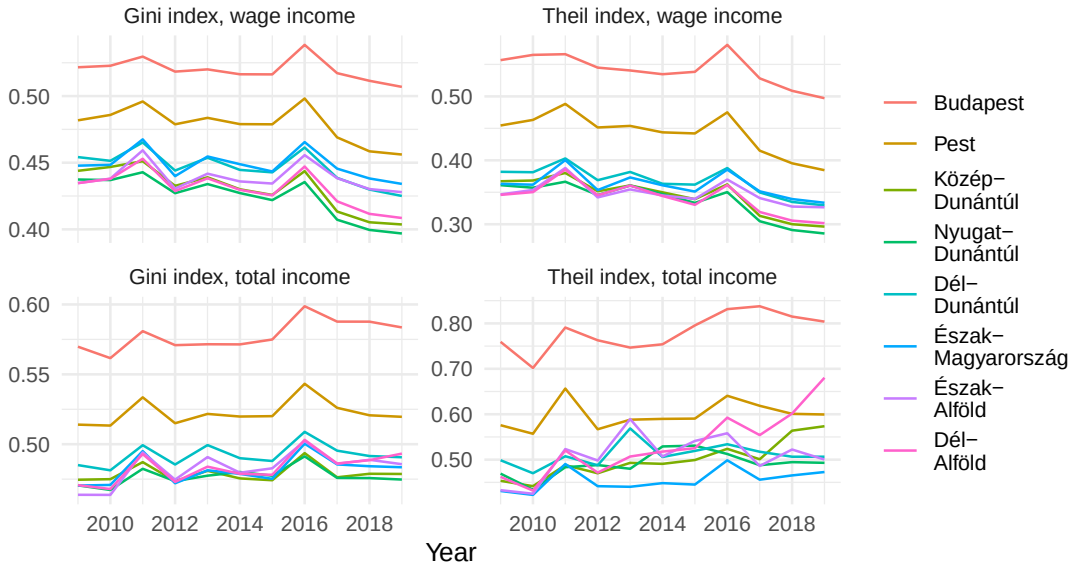
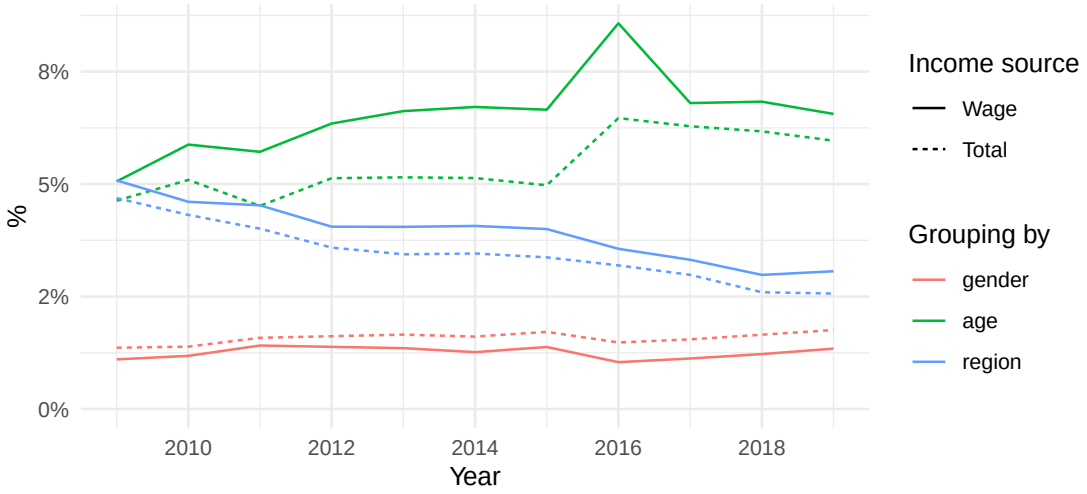


Figure 9: Share of between group income inequalities measured by the Theil index



5.3 Lorenz curves and income shares

A detailed look at the shape of income distributions using Lorenz curves confirms the inequality measure based findings, and also highlights some important trends.

Between 2009 and 2019 wage income shares moved towards a more equal distribution (Figure 10). In the bottom deciles the 2019 distribution tracks the 2009 distribution closely but at around the 40th percentile – representing income from a full year, full time minimum wage job, see Table A.1 – income shares moved substantially closer to the diagonal line representing full equality. For total income we see the reverse. The bottom of the distribution became more unequal but income shares changed very little above the median. We can see a flattening in the top 1% as well: the share of the top 1%'s wage income decreased, while the overall income share remained unchanged, and same patterns holds within the top 1%.

As discussed previously, multiple causes are possible, and the Lorenz curves do not rule out any of those. The increased number of taxpayers form younger and older demographics, and the expansion of tax filers through pre-filled forms are both consistent with the unchanged bottom half of the distribution. Especially the latter factor could have even lead to decreasing percentile thresholds (in real terms) in the lower parts of the income distribution (Figure 14 and Table A.1) but not on the upper half. Minimum

wage rises could have played a role in the increasing income share around the median but wages equalized even more in the 60th–90th percentile range. The disappearance of self-employed businesses from personal income taxation could have contributed to a more unequal total income distribution in the bottom deciles. Meanwhile, capital income's distribution became more unequal, counteracting the flattening trends in wage income. The same trends can be observed for both genders (Figure A.7), as well as for wage income within age groups (Figure A.8). However, it is noteworthy that total income follows almost exactly the same distribution for younger prime age tax payers (25–39), as wage income, while for older prime age taxpayers (40–59) the two diverge (this can be seen in raw inequality measures too in Figure 7). Regional trends are similar to national trends (Figure A.8).

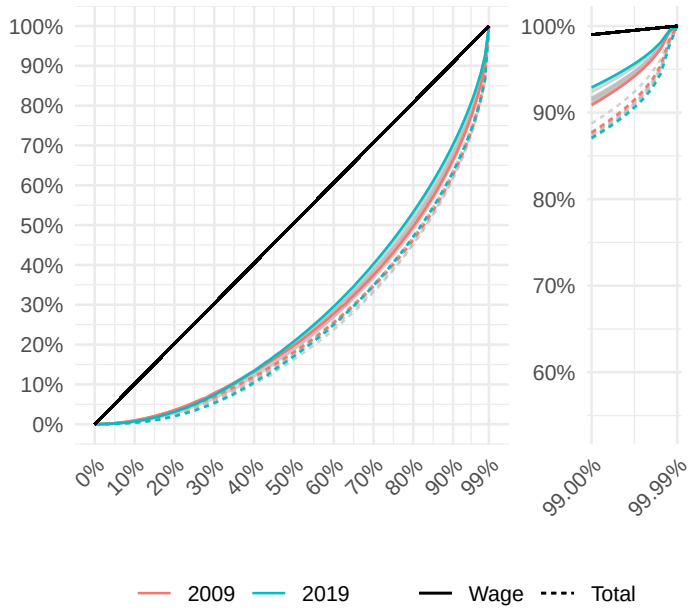
These findings put important caveats on results coming from inequality measures. The Gini and Theil indices increased for total income but crucially, this rise didn't come from an increased income share in the upper tail of the distribution.

Lorenz curves show within group inequalities but just based on simple descriptive statistics, subpopulations' income distributions have very different moments. Thus, the total composition of the taxpayers population can be highly skewed towards particular groups, even if the groups have similar within group inequalities.

In case of gender differences, men's higher income inequalities are primarily the result of their higher share in the top deciles (Figure 11). Workers marginally attached to the labour force have an even gender breakdown, women are more likely to be employed in minimum to median wage earning jobs but in the top deciles men's share increases sharply. This gap in the top 30% even widened from 2009 to 2019. The gap is even wider in the top 1% but it didn't increase there during the last decade. While we don't have industry level data to confirm this hypothesis, there could be sectoral differences, offering a partial explanation. Women are more likely to work in the public sector (including education and health care) and most of those jobs are likely to fall in the 60%–90% range. Average and median wage growth was similar across genders (Figure A.1) and there were wage rises in the public sector (e.g. for teachers around 2013, for healthcare workers other than doctors starting in 2018) but public sector wage dynamics generally lagged behind the private sector, which employs relatively more men.

Age group shares (Figure 12) reflect the life cycle earnings, with a lower share of prime age taxpayers in the lower deciles. Changes in the composition happened evenly across the income distribution, mostly due to the changing

Figure 10: Lorenz curves of taxable income



Note: Right panel zooms in on the top 1%, showing data points for fractiles 0.991–0.999 and 0.9991–1. See Table A.1 for percentile threshold values. Data for 2010–2017 in unhighlighted grey.

age composition of taxpayers.[15]

Higher income taxpayers are disproportionally more likely to live in Budapest, and to some extent in the surrounding Pest county, although Pest's share only increases significantly in the top decile (Figure 13). While the regional composition of the taxpayer population changed significantly, this had no major impact on percentile shares. The only notable change is the increasing share of low income taxpayers in Budapest, suggesting a large number of those marginally attached workers, who entered the labour force – and contributed to increased inequality in the bottom half of the income distribution – live in the capital. This increase is similar for wage and total income, therefore it cannot be explained by the outflow of self-employed from personal income tax returns.

[15]Although older men started staying in the workforce more than women, these changes are unlikely to explain the large decrese in women's share in the upper deciles. There is a gender wage gap across the income distribution but it is the narrowest for older taxpayers (see Figure A.6).

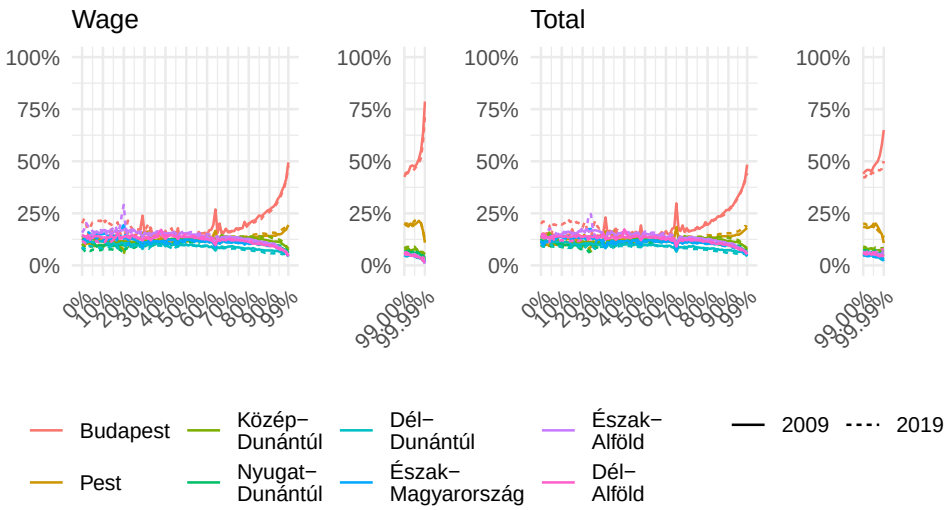
Figure 11: Percentile shares of taxpayers by gender

Note: Right panel zooms in on the top 1%, showing data points for fractiles 0.991–0.999 and 0.9991–1. See Table A.1 for percentile threshold values. Percentiles refer to overall percentile rank.

Figure 12: Percentile shares of taxpayers by age

Note: Right panel zooms in on the top 1%, showing data points for fractiles 0.991–0.999 and 0.9991–1. See Table A.1 for percentile threshold values. Percentiles refer to overall percentile rank.

Figure 13: Percentile shares of taxpayers by region



Note: Right panel zooms in on the top 1%, showing data points for fractiles 0.991–0.999 and 0.9991–1. See Table A.1 for percentile threshold values. Percentiles refer to overall percentile rank.

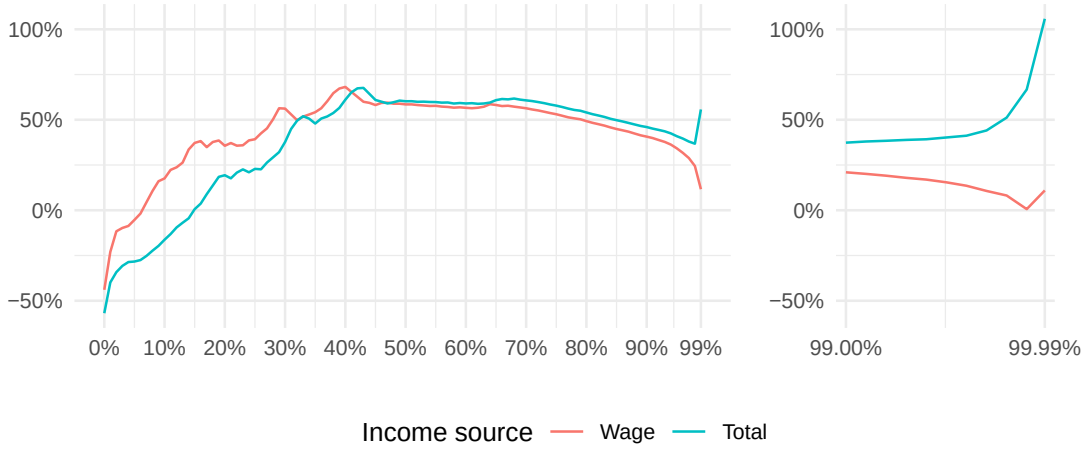
5.4 Growth incidence

We already discussed some trends in the growth rates of income percentiles regarding the effect of the inflow of marginally attached workers and the outflow of self-employed businesses. The growth incidence curve summarizes these findings for the entire income distribution (Figure 14).

There was a real income decline at the bottom of the wage income distribution, most likely due more people having short term, one-off jobs. However, real incomes grew from the 5th percentile, steadily increasing for part time employees. The highest wage gains were at around the full year, full time minimum wage, and then growth steadily tapers off for higher earners. This confirms all factors discussed in Section 5.1 – minimum wage rises, payroll tax cuts, the strong recovery, tight labour markets, and a decrease in wage underreporting – as potentially important contributors.

In total income we see a decline until the 15th percentile, which is also likely linked to measurement issues around changes in the taxpayer population. In the upper part of the distribution – where the measurement of total income should be more reliable – total income grew slightly faster than wage income, and the gap between the two widened somewhat in the highest percentiles

Figure 14: Cumulative growth of real average income by percentile between 2009 and 2018



Note: Right panel zooms in on the top 1%, showing data points for fractiles 0.991–0.999 and 0.9991–1. See Table A.1 for percentile threshold values.

outside the top 1%. In the top 1%, and especially in the top 0.1% we see very little wage growth but a large increase in total income. While the big rise in capital income can ben explained by the factors discussed above (e.g. real estate prices, or monetary policy), the flat wage income suggests some form of tax planning. Top marginal rates on capital were always lower than on wages during the period analysed in this paper, and with the flat tax reform in 2011 the difference decreased. The decrease in tax rates could have contributed to foreign income moving to Hungary, although the shift from labour to capital income in the top 0.1% suggest other tax policies could have played a role as well. This chart does not rule out the aforementioned hypothesis of Hungarians' offshore income moving to Hungary due to measures related to international tax avoidance.

5.5 Persistence

The inflow of marginally attached workers, and its effects on the bottom half of the income distribution raises the question how these trends can effect inequality in the longer term. While we do not attempt to make any predictions, it would be useful to see how relative income positions change over shorter,

and longer terms in general on an individual level.

Figure 15 summarises transition probabilities between income brackets. Due to the measurement issues with total income, we limit this exercise to wage incomes. In year-over-year changes differences between cohorts are relatively small, jumping more then two deciles is rare in any direction from any income level, even for the youth. The probability of staying in the same decile roughly halves over 11 years compared to annual probabilities for the bottom nine deciles. The highest persistence occurs in the top decile both in the short, and in the long run. Additionally, long run persistence decreases by only 10–20% points compared to the short run in the top decile. The persistence within percentiles is less strong but still noticeable. It is higher in the top percentiles, which could be due to the wide income bands, which are further away in absolute terms, thus even large income shocks (either permanent, or transitory) won't move someone into a different percentile.

We don't have a long enough time series to look at changes in longer term persistence but when breaking our data into two five year intervals, persistence in the 5th–9th deciles is somewhat lower in the second half of the period, while it is somewhat higher in the lower deciles. However, it is difficult to tease out the effect of the larger number of marginally attached taxpayers in the second five year period.

Table 1 shows a longer term view, highlighting how long prime age taxpayers, who had high employment rates throughout the period analysed stayed in a particular spot of the income distribution.

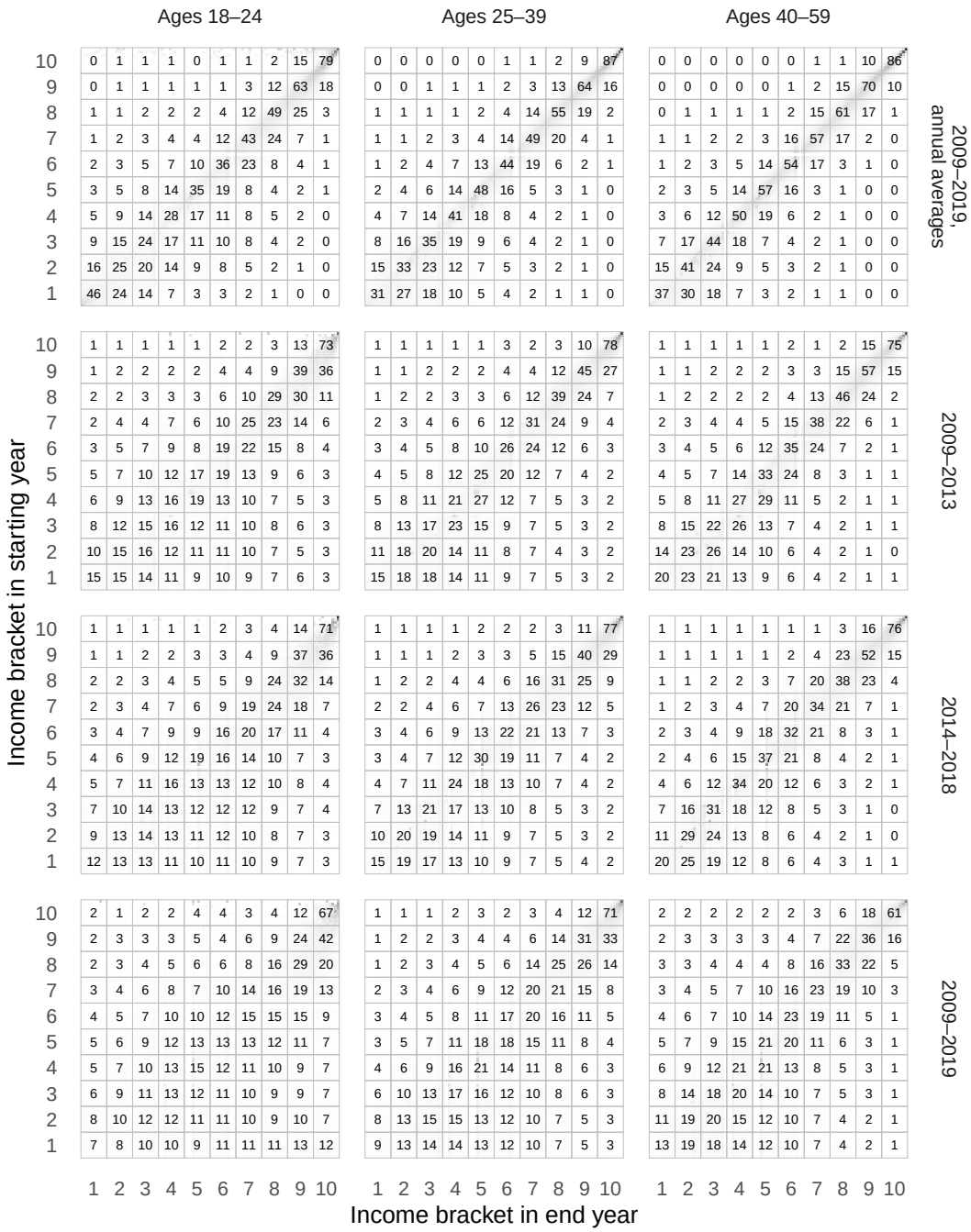
It is rare to stay in the bottom quintile for more than a couple of years but more than half of the taxpayers drop out of the Hungarian labour force at least once. Our dataset has no information on what these people do during their inactive years. They could be unemployed, studying, taking care of family members (either children, or older relatives), or moving abroad.

The minimum wage is the modal annual income, and the skilled minimum wage[16] is also very common, but only around 20 and 28% of the prime age taxpayers respectively ever report an annual income close to these values. Those who report minimum wage at some point typically do not keep earning at the same level. This could be due to lifecycle-related income mobility but it might be due to the narrow band, that can be considered as a minimum wage job in our annual dataset if even smaller one-off payments, like bonuses, or even a short sick leave[17] can move someone away from this range. Temporary unem-

[16] For jobs requiring some vocational training.

[17] Sick pay is generally 70% of base pay.

Figure 15: Short and long term transition probabilities between wage income brackets (%)



Note: Tables describe transition probabilities from income decile in starting year (in rows) to income decile in end year (in columns) by period and cohort for individuals with positive income in both the starting and end years. All rows sum up to 100%. E.g. in the top left panel, a 18–24 year old earning the 6th decile was on average between two years 36% likely to stay in the same decile, while 10% , and 23% likely to move down, or up one decile respectively. Squares in the background show the same transition probabilities between percentiles, with darker squares denoting higher probabilities.

Table 1: Probabilty of occuring in an income bracket between 2009 and 2019 (%)

	Number of occurances between 2009 and 2019											
	0	1	2	3	4	5	6	7	8	9	10	11
No income	46.7	10.0	6.5	5.3	4.7	4.4	4.2	4.3	4.1	4.4	5.3	
Bottom decile	67.4	18.9	7.7	3.4	1.5	0.7	0.3	0.1	0.1	0.0	0.0	0.0
Bottom quintile	51.3	18.9	10.6	6.9	4.7	3.0	2.0	1.2	0.7	0.4	0.2	0.1
Mimimum wage	80.4	11.6	3.8	1.8	0.9	0.6	0.4	0.2	0.2	0.1	0.1	0.1
Skilled mimimum wage	72.4	12.0	5.7	3.5	2.2	1.4	1.0	0.7	0.5	0.3	0.2	0.2
Top quintile	69.5	4.4	2.9	2.5	2.2	2.0	1.8	1.7	1.7	1.8	2.3	7.3
Top decile	83.3	3.1	1.8	1.4	1.2	1.0	0.9	0.9	0.9	1.0	1.1	3.5
Top percentile	97.6	0.6	0.3	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.2

Note: All rows sum to 100%, showing the distribution of the number of occurrences in an income bracket for individuals born between 1958 and 1979 with at least one tax return in 2009–2019 (e.g. 46.7% of taxpayers in the sample worked in all years, and 3.5% were continuously in the top decile). (Skilled) minimum wage refers to full year, full time (skilled) mimimum wage ±5%. See Table A.2 for minimum wage percentile ranks.

ployment can also be relatively high for lower skilled people who tend to earn close to the minimum wage.

While these basic results without any benchmarks as potential comparison are difficult to interpret, they nevertheless highlight the importance of life-cycle elements, and longer term aspects in income inequality.

6 Conclusions

In this paper we offered an overview of recent trends in market income inequality taking advantage of the richness of administrative data on personal income taxes. Wage inequality decreased significantly in the upper half of the income distribution, while inequality of all taxable income increased a little. However, at the bottom of the distribution income shares fell, partly due to the inflow of previously inactive populations but partly also due to an outflow of self-employed business income form personal income taxation after the introduction of a new tax regime for self-employed.

Both of these factors require further research using more detailed data.

Our research highlighted a relatively low income persistence in the lower half of the income distribution. The interaction of part time work with poverty, and other social indicators is an important area, especially with the recession following the Covid-19 pandemic that hurt youngest and oldest workers the most (Köllő and Reizer 2021) but this was beyond our current research, which focused on market income, using individual level data. Similarly, the effect of changes in the taxation of self-employed should be explored using linked data. Overall, the demographic and regional factors explain some of the changes in inequality but structural changes, e.g. the reallocation of labour between sectors, or between firms might be more important. Finally, the decreasing inequality between regions along with the significantly more unequal capital region also warrants further research at a more granular geographic level.

References

Alstadsæter, Annette, Niels Johannesen, and Gabriel Zucman. 2017. "Tax Evasion and Inequality." w23772. National Bureau of Economic Research. https://doi.org/10.3386/w23772.

Amberg, Niklas, Thomas Jansson, Mathias Klein, and Anna Rogantini Picco. 2021. "Five Facts about the Distributional Income Effects of Monetary Policy." 9062. CESifo Working Paper Series. CESifo. https://ideas.repec.org/p/ces/ceswps/_9062.html.

Andersen, Asger Lau, Niels Johannesen, Mia Jørgensen, and José-Luis Peydró. 2021. "Monetary Policy and Inequality." 1761. Economics Working Papers. Department of Economics and Business, Universitat Pompeu Fabra. https://ideas.repec.org/p/upf/upfgen/1761.html.

Beer, Sebastian, Maria Coelho, and Sebastien Leduc. 2019. "Hidden Treasure: The Impact of Automatic Exchange of Information on Cross-Border Tax Evasion." *IMF Working Papers* 18 (286). https://doi.org/10.5089/9781513519975.001.

Benczúr, Péter, Gábor Kátay, and Áron Kiss. 2012. "Assessing Changes of the Hungarian Tax and Transfer System: A General-Equilibrium Microsimulation Approach." MNB working paper 2012/7. Magyar Nemzeti Bank (the central bank of Hungary). https://ideas.repec.org/p/mnb/wpaper/2012-7.html.

Blanchet, Thomas, Lucas Chancel, and Amory Gethin. 2019. "How Unequal Is Europe? Evidence from Distributional National Accounts, 1980-2017." hal-02877000. Working Papers. HAL. https://ideas.repec.org/p/hal/wpaper/hal-02877000.html.

Cowell, Frank. 2016. *Inequality and Poverty Measures*. https://doi.org/10.1093/oxfordhb/9780199325818.013.4.

Elek, Peter, János Köllő, Balázs Reizer, and Péter A. Szabó. 2011. "Detecting Wage Under-Reporting Using a Double Hurdle Model." 6224. IZA Discussion Papers. Institute of Labor Economics (IZA). https://ideas.repec.org/p/iza/izadps/dp6224.html.

European Commission. 2019. *Labour Market and Wage Developments in Europe: Annual Review 2019*. LU: Publications Office. https://doi.org/10.2767/399955.

Eurostat. 2021. "Living Conditions in Europe – Income Distribution and Income Inequality." 2021. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Living_conditions_in_Europe_-_income_distribution_and_income_inequality.

Köllő, János, and Balázs Reizer. 2021. "A koronavírus-járvány első hullámának hatása a foglalkoztatásra és a vállalatok árbevételére." *Közgazdasági Szemle* 68 (4): 345–74. https://doi.org/10.18414/KSZ.2021.4.345.

Kovács, Ilona. 2011. "A jövedelemeloszlás és jövedelemegyenlőtlenség a személyijövedelemadó-bevallási adatok tükrében." *Statisztikai Szemle* 89 (3): 294–312. http://www.ksh.hu/statszemle_archive/2011/2011_03/2011_03_294.pdf.

Lovics, Gábor, Katalin Szőke, Csaba G. Tóth, and Bálint Ván. 2019. "The Effect of the Introduction of Online Cash Registers on Reported Turnover in Hungary." 2019/137. MNB Occasional Papers. Magyar Nemzeti Bank (Central Bank of Hungary). https://ideas.repec.org/p/mnb/opaper/2019-137.html.

MNB. 2021. "Lakáspiaci jelentés, 2021. május." Magyar Nemzeti Bank. https://www.mnb.hu/kiadvanyok/jelentesek/lakaspiaci-jelentes/lakaspiaci-jelentes-2021-majus.

OECD. 2017. *Preventing Ageing Unequally*. OECD. https://doi.org/10.1787/9789264279087-en.

———. 2018. *A Broken Social Elevator? How to Promote Social Mobility*. https://doi.org/10.1787/9789264301085-en.

Tóth G., Csaba, and Péter Virovácz. 2013. "Nyertesek és vesztesek. A magyar egykulcsos adóreform vizsgálata mikroszimulációs módszerrel." *Pénzügyi Szemle* 58 (4). https://www.penzugyiszemle.hu/penzugyi-szemle-folyoirat-archivalt-cikkek/nyertesek-es-vesztesek-a-magyar-egykulcsos-adoreform-vizsgalata-mikroszimulacios-modszerrel.

A Additional figures and tables

Figure A.1: Mean and median taxable income by gender

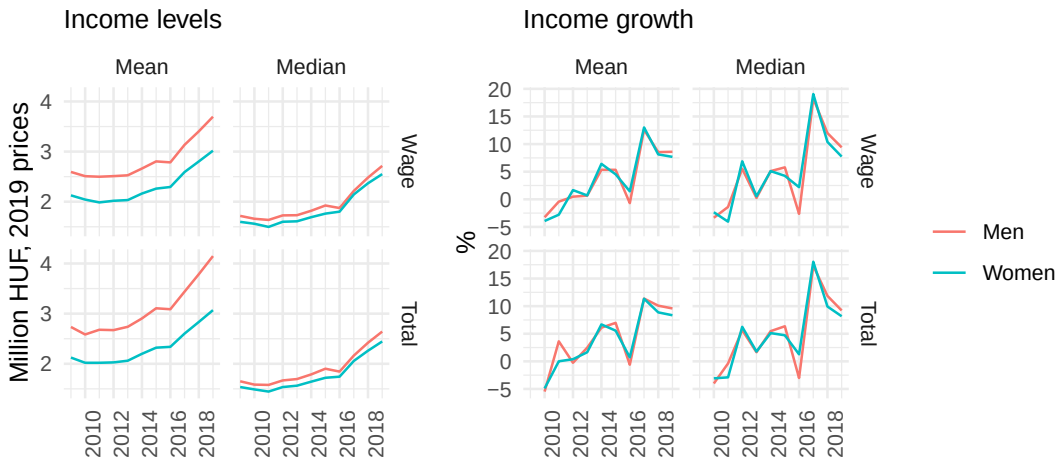


Figure A.2: Mean and median taxable income by age

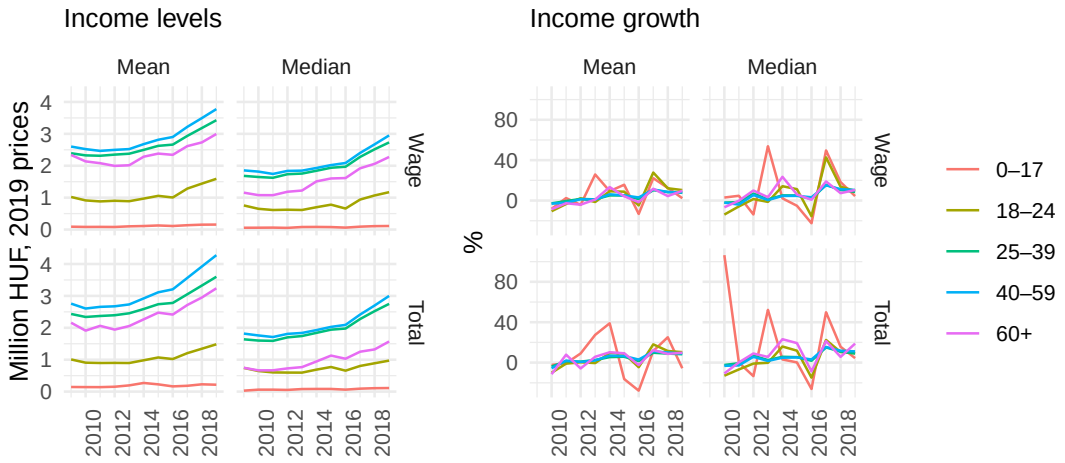


Figure A.3: Mean and median taxable income by region

Figure A.4: Percentile ranks of taxpayers without a tax return in the previous year by age group

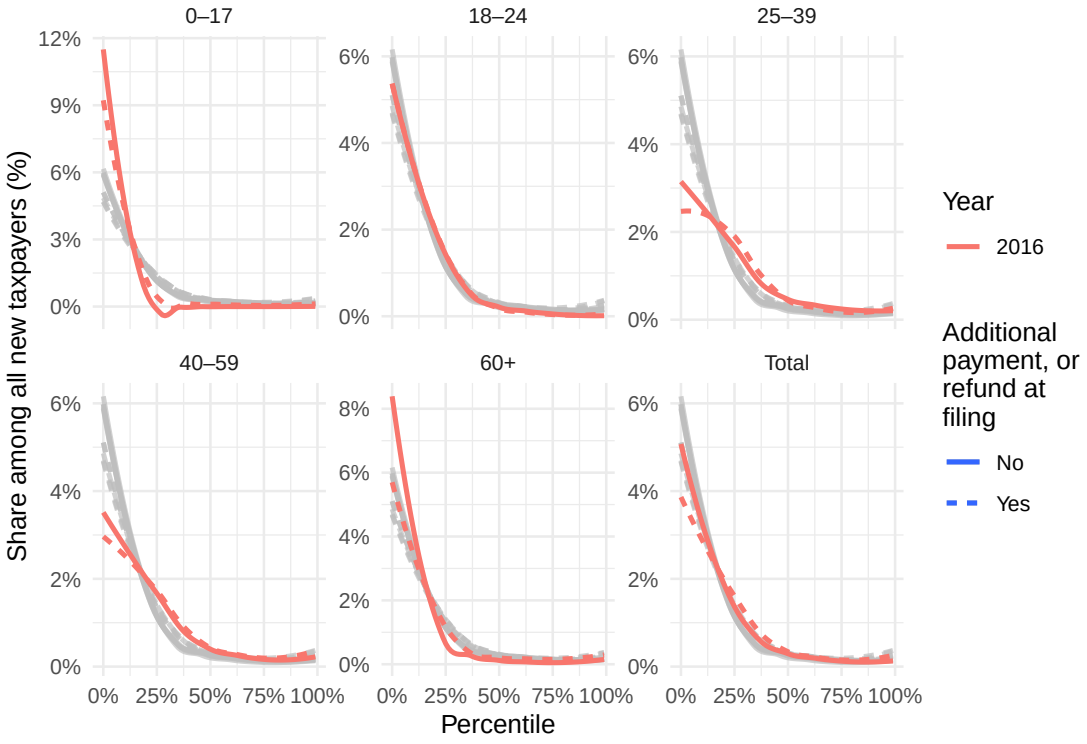
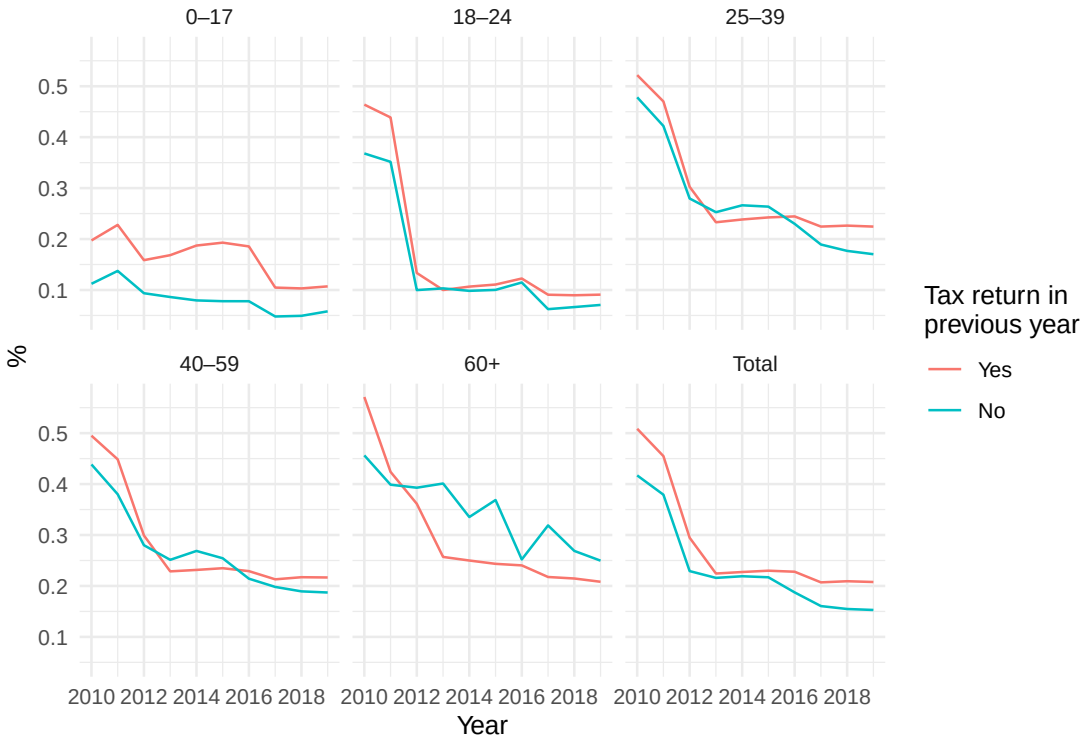


Chart shows data for years 2009–2019, with 2016 – the first year of widespread usage of the pre-filled electronic forms – highlighted. Due to bunching around percentile thresholds (especially around the minimum wage, see Table A.2), LOESS-smoothed curves of percentile rank shares are plotted. If advanced payment (typically deducted by employers) differs from the true tax liability, taxpayers have to pay the remainder of their tax liability, or can claim their refunds at the time of filing for differences above HUF 1,000.

Figure A.5: Share of taxpayers with tax payable, or refund claimable at the time of filing by age group

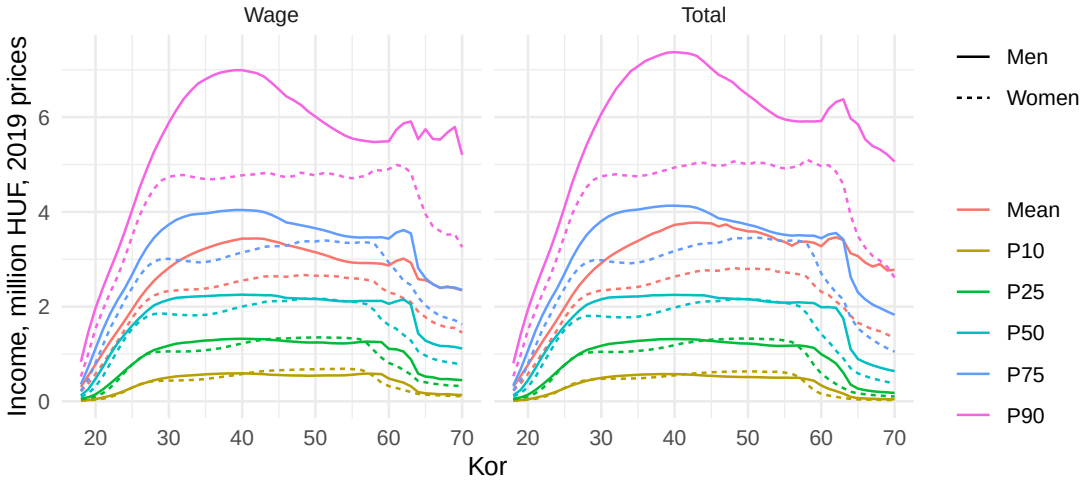


See notes on Figure A.4 for explanation on payable, or refundable tax.

Table A.1: Upper thresholds of income deciles (million HUF, 2018 prices)

Decile	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Wage income											
1.	0.30	0.29	0.24	0.30	0.27	0.34	0.37	0.25	0.37	0.41	0.45
2.	0.60	0.59	0.56	0.66	0.64	0.71	0.74	0.67	0.85	0.95	1.05
3.	0.87	0.88	0.91	1.02	1.00	1.07	1.14	1.01	1.32	1.55	1.75
4.	1.08	1.08	1.13	1.30	1.33	1.41	1.46	1.44	1.78	2.05	2.31
5.	1.29	1.31	1.32	1.48	1.52	1.59	1.68	1.68	2.04	2.34	2.62
6.	1.60	1.62	1.64	1.80	1.84	1.93	2.04	2.07	2.46	2.82	3.20
7.	1.94	1.98	2.01	2.19	2.27	2.40	2.53	2.60	3.04	3.44	3.89
8.	2.50	2.54	2.60	2.74	2.86	3.10	3.26	3.37	3.82	4.26	4.79
9.	3.61	3.65	3.78	3.90	4.04	4.25	4.50	4.68	5.26	5.82	6.53
10.	1011.31	1298.53	919.76	760.08	836.03	1119.58	1757.40	1461.32	1609.28	1534.50	1474.85
Total income											
1.	0.26	0.24	0.21	0.23	0.22	0.27	0.30	0.18	0.24	0.26	0.27
2.	0.56	0.54	0.52	0.59	0.58	0.63	0.70	0.58	0.70	0.78	0.86
3.	0.85	0.84	0.85	0.93	0.91	0.99	1.06	0.94	1.11	1.28	1.47
4.	1.05	1.06	1.10	1.25	1.27	1.36	1.44	1.37	1.63	1.87	2.12
5.	1.23	1.24	1.27	1.43	1.48	1.56	1.64	1.64	1.98	2.25	2.52
6.	1.54	1.56	1.59	1.75	1.80	1.90	2.01	2.03	2.39	2.74	3.14
7.	1.89	1.92	1.98	2.16	2.25	2.39	2.53	2.58	3.00	3.40	3.88
8.	2.47	2.51	2.60	2.74	2.89	3.12	3.31	3.40	3.84	4.31	4.89
9.	3.66	3.68	3.88	4.00	4.18	4.42	4.71	4.87	5.44	6.06	6.86
10.	1750.80	1671.77	7199.19	4959.66	11048.51	4347.29	9758.93	4993.87	26531.25	13922.33	25981.82

Figure A.6: Distribution of taxable income by gender and age-year cohorts between 2009 2019



Note: The chart shows the joint distribution of income across all years using fixed price level.

Figure A.7: Lorenz curves of taxable income by gender

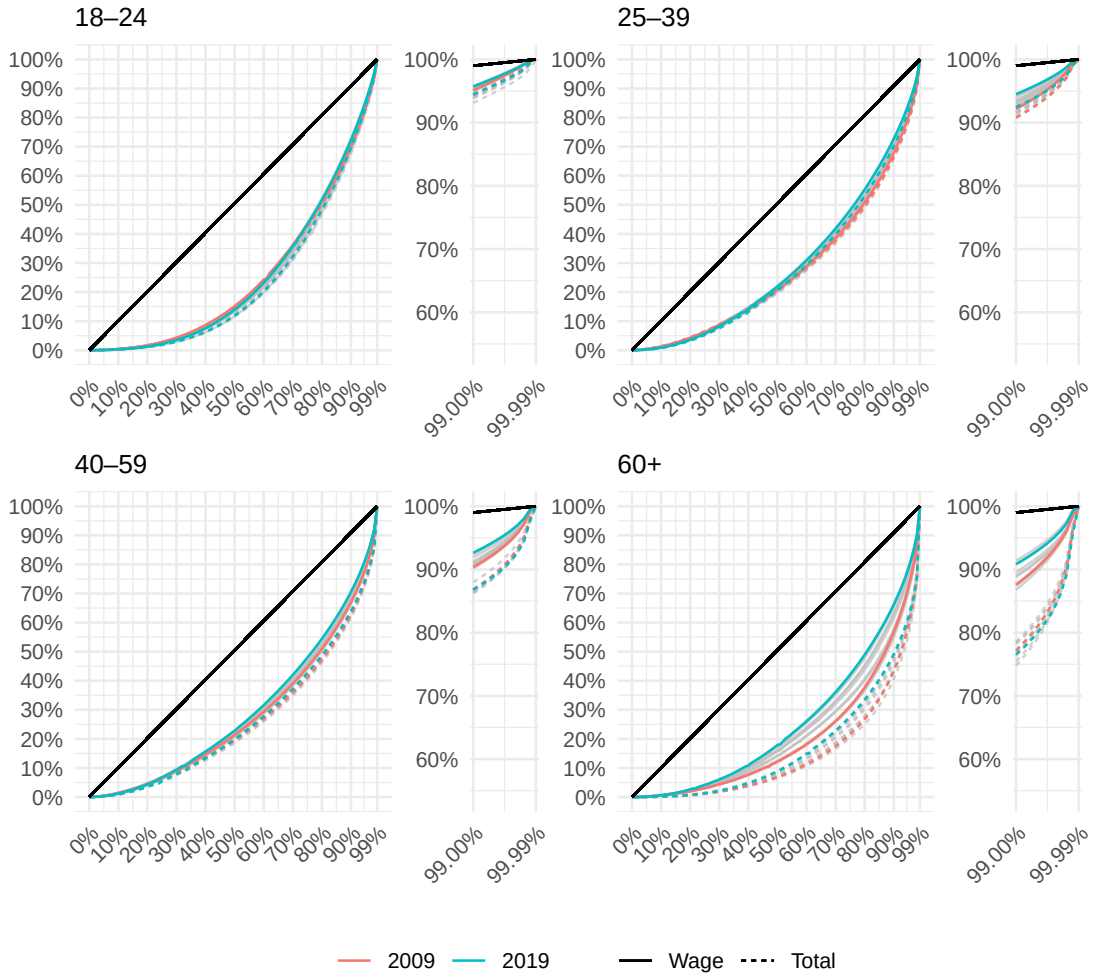
Note: Right panel zooms in on the top 1%, showing data points for fractiles 0.991–0.999 and 0.9991–1. See Table A.1 for percentile threshold values.

Table A.2: Full year, full time monthly minimum wage and skilled minimum wage and their rank in the wage income distribution

	Minimum wage			Skilled minimum wage		
Year	Value (current prices)	Value (2019 prices)	Percentile	Value (current prices)	Value (2019 prices)	Percentile
2009	71,500	91,552	29--30	87,000	111,399	37--38
2010	73,500	89,888	30--31	89,500	109,456	38--39
2011	78,000	91,811	31--33	94,000	110,644	40--42
2012	93,000	103,564	33--34	108,000	120,268	40--42
2013	98,000	107,307	35	114,000	124,827	42--43
2014	101,500	111,140	34--35	118,000	129,207	41--43
2015	105,000	114,857	33--34	122,000	133,453	40--43
2016	111,000	120,937	37--38	129,000	140,548	43--46
2017	127,500	135,658	34--35	161,000	171,302	44--47
2018	138,000	142,692	33--34	180,500	186,637	43--46
2019	149,000	149,000	31--33	195,000	195,000	41--44

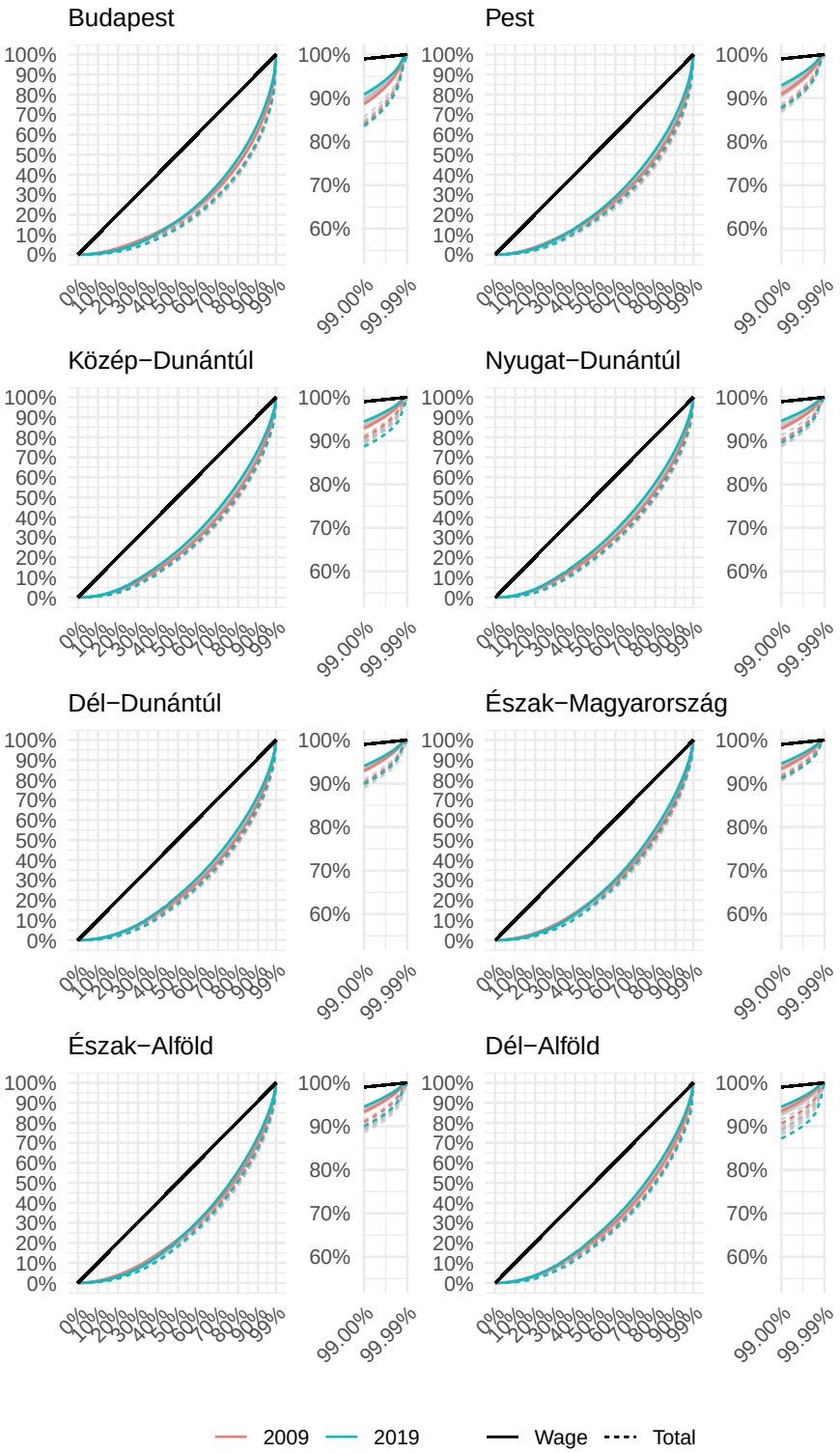
Note: Skilled minimum wage refers to the minimum wage payable in jobs requiring vocational training. Values of minimum wages can span several percentiles due to bunching. Bunching occurs at some "round" values but without extending beyond percentiles.

Figure A.8: Lorenz curves of taxable income by age



Note: Right panel zooms in on the top 1%, showing data points for fractiles 0.991–0.999 and 0.9991–1. See Table A.1 for percentile threshold values.

Figure A.9: Lorenz curves of taxable income by region



Note: Right panel zooms in on the top 1%, showing data points for fractiles 0.991–0.999 and 0.9991–1. See Table A.1 for percentile threshold values.