Regional income inequalities and labour mobility in Hungary*

András Svraka†

July 26, 2021

Abstract

We analyse regional wage inequalities in the 2010s using administrative data sources at highly disaggregated regional levels, including commuting zones. The decline in national wage inequalities during this period is reflected at regional levels and we find convergence between regions in income levels and in the decreasing weight of between region inequalities as well. There are still large differences, and high income employees are concentrated in prosperous regions. Interregional mobility was also a driving force behind changes in income inequalities even in a country with low overall mobility rates. High income employees are much more likely to move, typically from less central, less developed regions to more central, larger labour markets. We find some evidence for a transitory mobility premium, although we cannot establish the causality of this relationship.

*I'm grateful to Katalin Horváth, Tibor Keresztély, Sebastian Königs, Péter Tóth, Bálint Ván and Anna Vindics 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

Regional differences in labour market, and other economic outcomes within countries are an often studied research topic. However, there's much less detailed research on the regional aspects of income inequality. This comes partly from the lack of sufficiently disaggregated data which link location to incomes on an individual's, or household's level. Household surveys' sample sizes are generally good enough to investigate differences between larger geographic regions. In a European context this typically refers to large administrative regions (see Castells-Quintana, Ramos, and Royuela 2015 for a recent overview). These regions can hide considerable heterogeneity, e.g. between urban and rural areas, or between areas with good access to prosperous labour market and more remote areas. Meanwhile divergences within these large regions, and the performance of laggard, or emerging smaller regions are highly relevant to policymakers. Administrative data source on income and granular residential data can help on filling these gaps.

Another motivation of this paper is the findings of Svraka (2021) on recent trends in Hungarian income inequalities. Administrative income data showed decreases in within region inequalities at national, and at a NUTS2 level and a significant drop in the share of between region inequalities relative to total inequality. However, it is hard to understand the driving forces behind these changes using a coarse regional classification, as NUTS2 level divides a country of almost 10 million into just 8 large regions. In this classification the capital city Budapest clearly stands out from the rest of the country with high average incomes and high inequality. However, these regions are very heterogeneous which warrants further research using more fine-grained location data. Fortunately, administrative datasets provide just the right information for that, and can go beyond simple administrative divisions. This allows us to look at not just pure geographic differences but differences in the socio-economic environments of various types of locations. Thus, one of this paper's new contribution will be a first look at income inequalities in Hungary using highly disaggregated regional data.

Another finding of Svraka (2021) was the importance of demographic factors in income inequalities. From a regional perspective, the question of how within country mobility can affect inequalities arises. Within country mobility in Hungary – similarly to other countries in the region – is low in compared to Western European countries, partly driven by the high home-ownership rates emerging from privatisations at below market rates during the transition to market economies. However, over a longer period even low mobility rates can amount the large shifts. Regional mobility is usually analysed from an aggregated perspective, as there's very little data connecting relocations to income. Administrative data can help with this too. Fine-grained information on location is especially important here, as it allows to account for the role of local labour markets that often reach across administrative boundaries. Thus another contribution of this paper is a first analysis of how interregional mobility can affect inequalities, in case of Hungary primarily through out-migration from less advantaged areas.

This paper build heavily on previous work by Svraka (2021) in approach, data, and methods but extends and adjusts it to highlight some important aspects of regional differences. Results show that the regional composition of income percentiles is equal along all analysed regional dimensions for the bottom seven deciles. In the top deciles the share of less affluent, less central regions, and smaller labour market decreases markedly. This leads to more unequal income distributions in the most developed regions compared to rest of the country.

We also find a strong positive association between income and interregional mobility. The probability of moving to a different region is similar in the bottom eight deciles but it increases throughout the top income brackets. This patterns holds for all regional dimensions, except for the most developed regions, where outflows are constant over the income distribution. There is a large mobility premium but we cannot draw an causal conclusions from this data. Regression analysis shows significant pre-move effects. This can suggest measurement errors (i.e. reporting a change in permanent residence is delayed) but along with the other findings, reverse causality is also likely, where people with better prospects are more mobile. Overall, interregional migration slowed the convergence between regions during the last decade but other economic factors still lead to significant convergence.

2 Data

This paper uses the full universe of taxpayers filing an annual personal income tax return for the years 2011 to 2017 to obtain wage data. Wage refers to annual wages and is not censored at either the bottom, or the top of the distribution. Although the tax returns contain information on all taxable income, and around 15% of the workforce is self-employed, in this paper we focus on employment income only, as changes in the tax system related to tax regimes for self employed businesses affected the reporting of some income sources. For a detailed discussion of the limitations see Svraka (2021).

For location we use residential address available from the tax office's records.[1] However, a few additional considerations have to made. In principle, Hungarian residents can have two types of officially registered addresses: a permanent address, or a temporary address. The demographic literature considers data on the former more reliable (Bálint and Obádovics 2018), as definitional changes are less frequent, and temporary changes are more likely to remain unreported. The latter is typically associated with student accommodations and renting. In general, the tax office's records are based on the permanent address, although there can be differences. The number of taxpayers by district aligns well with the total population of districts (Figure B.3) for prime age demographic groups, across different sized districts. The number of relocations in our data also closely tracks the number of total permanent relocations for the prime age demographic groups (Figure B.4). In both cases, the smaller gap relative to total population, and higher total migration for men compared to women could be explained by lower female labour participation rate. However, for some data points we observe more relocations in the tax data, suggesting we might capture some temporary changes as well.

Thus we will interpret our data as permanent address for residence and regional mobility purposes as well. However, it should be noted that besides missing many temporary relocations, changes in a permanent address can be biased in other ways as well. People might first move without changing any of their addresses (e.g. to a dormitory, or into a rental unit), and later follow it up by changing their permanent address, once they settle permanently and buy a home, or when they start a family (registering a local address is not mandatory for access to local services, like schools, or family doctors but it can make access easier). This introduces a time lag in our observed mobility rates but also a selection bias, as people who cannot afford buying a home

[1]The dataset contained only a postal code, which cannot be directly used for territorial classifications. We used a crosswalk from postal codes to municipalities, which can be aggregated in many different territorial classifications using the territorial code system (https://www.ksh.hu/tszJ_eng_menu). Postal codes are shared between nearby settlements thus the crosswalk is ambiguous but at most 0.2% in the total population could be misclassified with the aggregations used in this paper. Additionally we drop taxpayers whose postal code cannot be matched to a Hungarian commune, and for the mobility analysis we dropped observations where the change of address was from, or to such an unknown location. Note that this approach drops cases, where the postal code had a typo, as well as valid foreign addresses.

in the region where they moved to for work will show up as immobile in our data. With the time lag we might also capture people as mobile who moved to a university city for their studies, stayed there for work, and changed address well after graduating. Thus some education related mobility – which is likely an important aspect of overall labour mobility in the country, see Nyüsti and Ceglédi (2013) – might be captured, even though we cannot observe residence before someone files their first tax return. However, as income dynamics can be strong for young people, this can lead to overestimating any association between mobility and income premia. Although, given the high share of owner occupied housing in Hungary, at around 85%[2], and with only households with a head of household below the age of 25 having a significantly higher share of non owner occupancy than other cohorts,[3] these issues could be less severe, than in countries with much lower owner-occupancy rates.

There is also an increasing share of missingness in addresses beginning in 2015. This is possibly related to the introduction of pre-filled tax returns, which can be approved online (and even get approved automatically, if the taxpayer doesn't approve it by the deadline), therefore reducing the necessity having a postal contact information. This increase is the highest for the young men but it is present for all demographic groups (Figure B.2), and a missing address is correlated with filing electronically, therefore it could be highly non-random. Therefore we limit our analysis for the 2011-2017, despite the availability of income data for later year. For this period the share of missing addresses remains low, below 1%.

Another important caveat is that we cannot tell why a person disappears from our data by not filing a tax return. Thus we cannot distinguish between transitioning to inactivity and taking up a job abroad. This includes not just relocating to another country but also cross-border commuting, in which case a Hungarian resident might not have any taxable income in Hungary. Working abroad started increasing in the early 2010s among people who are closely related to a domestic household (Morvay 2015).

3 Methods

The paper follows a largely descriptive and visual approach to explore income distributions and the relationship between income and regional mobility

[2] https://www.ksh.hu/docs/hun/xftp/idoszaki/pdf/miben_elunk15_2.pdf
[3] https://www.ksh.hu/docs/eng/xstadat/xstadat_annual/i_zhc025a.html

across various regional units of Hungary. This includes basic measures of income levels, income growth and income inequality measures. The latter will focus on the Gini and Theil indices, with the Theil used to decompose inequality into within and between region components. The differences in regional wage distributions beyond single inequality measures will be investigated using Lorenz curves and population shares of various regions by income level.

The fundamental unit of location in our data is the settlement of an individuals' residence. During the analysis we will aggregate settlements into larger units more relevant to local labour markets and regional development. The primary units of our analysis will be districts, commuting zones, and settlement types described in Appendix A. In some analyses we will aggregate these units into larger groups based on size, or level of development for easier visualisations. In case of regional mobility, our primary definition for within country relocations will be a change of address between commuting zones. Will also look at other movements, namely between settlement types and between districts aggregated into quintiles based on per capita income. Note, that for commuting zones, mobility will always refer to mobility between actual geographic units, even when visualising aggregated groups of zones. The other two definitions refer to between group movements.

3.1 Exploratory analysis on regional mobility and income inequality

Establishing a causal relationship between mobility and income is difficult, and we do not attempt to draw any causal conclusions in this paper. Nevertheless, it is still useful to estimate the association between regional mobility and income from the perspective of income inequalities to see how large the effect of interregional mobility on within region, and between region inequalities could have been.

First, we'll use an event study framework to estimate the association between mobility and wages. Our data covers a fairly short period of six years, thus it could be enough to follow a sufficiently large number of mobile taxpayers over several years, although the number of observations after moving is relatively low due to the relatively low interregional mobility. For the event study analysis we estimate

$$\ln y_{it} = \sum_{e=-4}^{4} T_{eit} + \alpha_i + \alpha_t + \varepsilon_{it}$$

where y_{it} is an individual's annual income, T_e are event time dummies, indicating whether in year t person i moved to a different region e years ago (omitting the year of the move), and α_i and α_t are individual and time fixed effects. This model allows us to estimate the non-causal effect of moving on income not just at the time of moving but also in the years leading up to, and following the move, as interregional mobility could have different short, or long term effects.

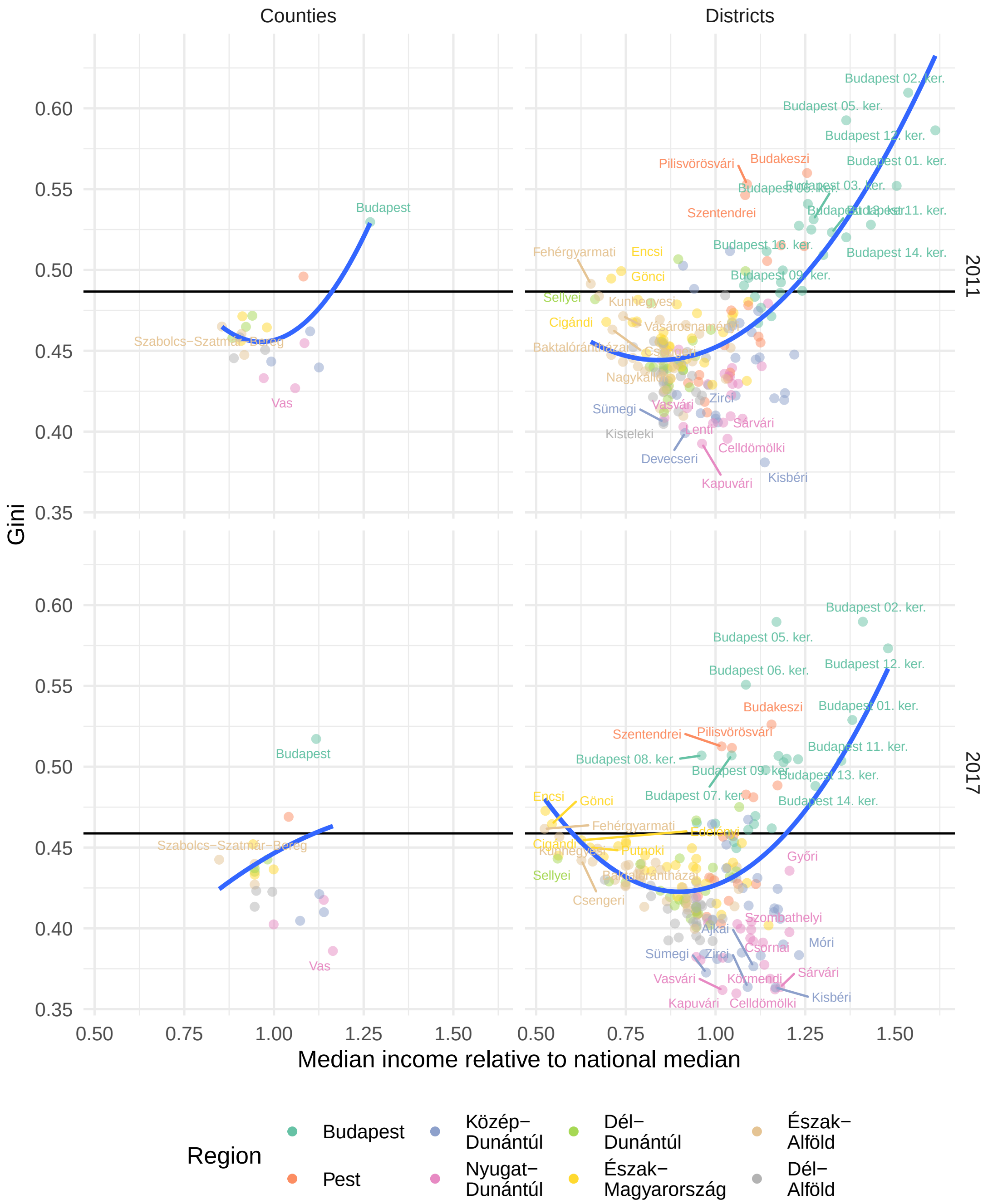
Then, we'll do a simple counterfactual exercise by comparing wages and inequality for 2017 using taxpayers' actual location, and the location where they resided in 2011. This approach assumes no returns to interregional migration. As a bracketing exercise we'll use the event study estimates to adjust observed wages for those, who moved to a different region. This is a very crude approach to construct a counterfactual but our aim is just to have a broad picture of which places are more affected by interregional mobility, and get a basic idea on the order of magnitude of this effect under various assumptions.

4 Results

4.1 Regional inequalities

Figure 1 shows the relationship between regional income levels – in this case using the median wage earners' income – and regional inequality. In 2011 there was a U-shaped relationship at both county, and district level. However, the strength of this relationship is not uniform. There is a strong positive correlation between the districts of Budapest, with the city core (districts 1, 2, 5, 6, 7, 12) having the highest median incomes and the highest inequalities as well. Meanwhile there is a large variation in inequality between districts around the national median, with some smaller districts in the Western, and North-Western part of the country (Közép-Dunántúl, and Nyugat-Dunántúl) much less unequal than other similarly developed places. These parts of the country are close to the Austrian, or Slovakian borders, thus there is relatively high cross-border commuting. This could distort our data, as we only observe wages earned in Hungary. However, not all of the least unequal districts are along the borders,

Figure 1: Regional median incomes and inequalities



Note: Areas in the top and bottom 5% by income, or by inequality are labelled. Horizontal lines show national Gini levels. Curves are fitted using a quadratic function weighted by population size.

thus the result of overall low inequality in these parts of the country should generally hold.

By 2017 we can see a significant convergence between counties and districts as well. Western and North-Western counties caught up with Budapest, and surpassed Pest county, which surrounds Budapest. This is not due to a decline of wages in the capital but thanks to catch-up (see Figures C.7–C.9). It is noteworthy though that mean wages in Budapest are sill much higher, partly explaining the higher inequality in the capital. There was a reduction in within region inequalities as well across the country (see Figures C.10–C.12) but relative positions mostly remained the same. With the large income gains of above-the-median regions, the strength of the relationship between income level and income inequality weakened, as now there's a 20 Gini point difference between regions with similar income levels: Budapest's centre (Budapest 5, and 6) and districts around Western towns (e.g. Kapuvár, Celldömölk, Kisbér).

The catch-up was unfortunately not universal, and some districts in the North-East fell further away from the national median. However, only some districts fell behind in these regions, as in the county-level analysis these districts' counties held their relative income position.

Between region inequalities fell as measured by the share of the between component of the Theil index as well (Figure C.13). Interestingly, the between component for districts fell significantly but the between component for district quintiles barely decreased, reflecting the broad convergence with some laggard regions on Figure 1.

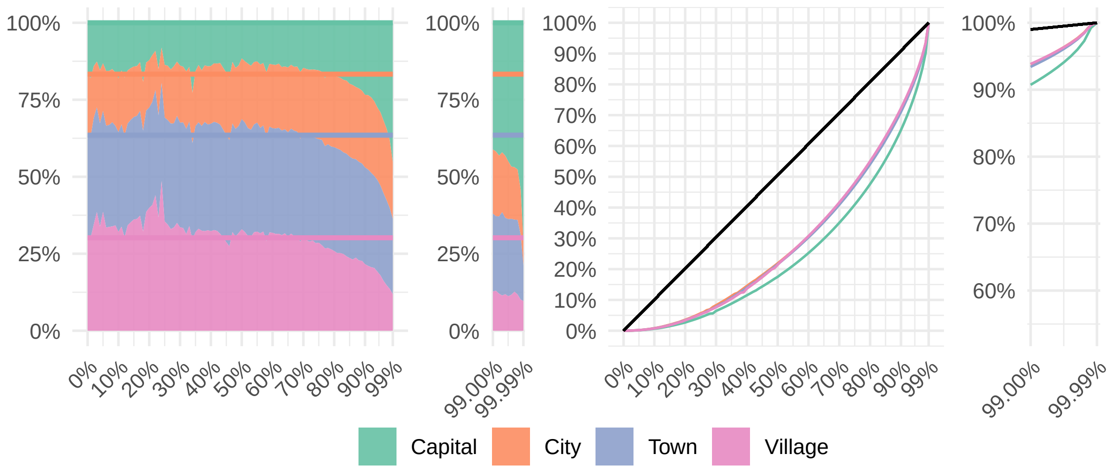
Next, we turn to the more detailed differences in regional wage distributions in Figure 2.[4] In terms of the composition of national income percentiles by regional categories, the share of each region corresponds roughly to their raw population shares in the bottom seven deciles. However, in the upper deciles the more urban, and more affluent areas start dominating. This trend is least stark in the breakdown by settlement type (due to affluent suburbs), while in terms of district level income, the share of high-income taxpayers in poor regions is especially low. In other words, lower income people are distributed evenly across the country, which considering the high variance of housing prices could mean very different living conditions. The declining share of less urban, less affluent, less central regions is true within the top percentile as well.

Within region Lorenz curves confirm these patterns. The high share of upper decile taxpayers in the more urban, more affluent, more central regions

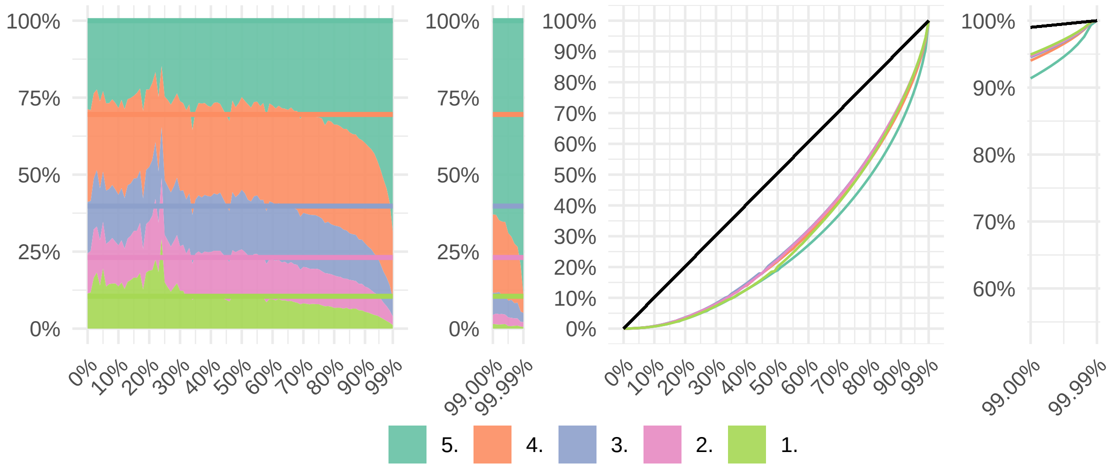
[4] For a similar analysis of statistical regions (NUTS2) see Svraka (2021).

Figure 2: Regional composition of national income fractiles and within region Lorenz curves, 2017

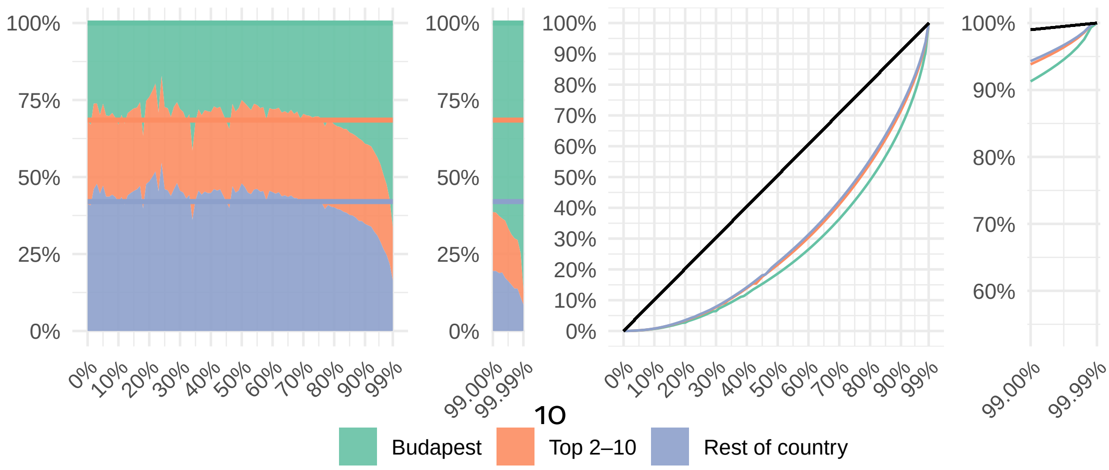
(a) By settlement type



(b) By district quintiles



(c) By commuting zones



Note: Horizontal lines on the composition figures show national population shares of each regional group. District quintiles refer to quintiles of district level per capita income, based on data for 2017 (5: highest, 1: lowest). Right panels for both types of figures zoom in on the top 1%, showing data points for fractiles 0.991–0.999 and 0.9991–1.

leads to higher inequality. The top region in all three classifications is significantly more unequal, while the other areas show very similar income distributions. Similarly to the patterns of regional composition, this trend is also true within the top percentile.

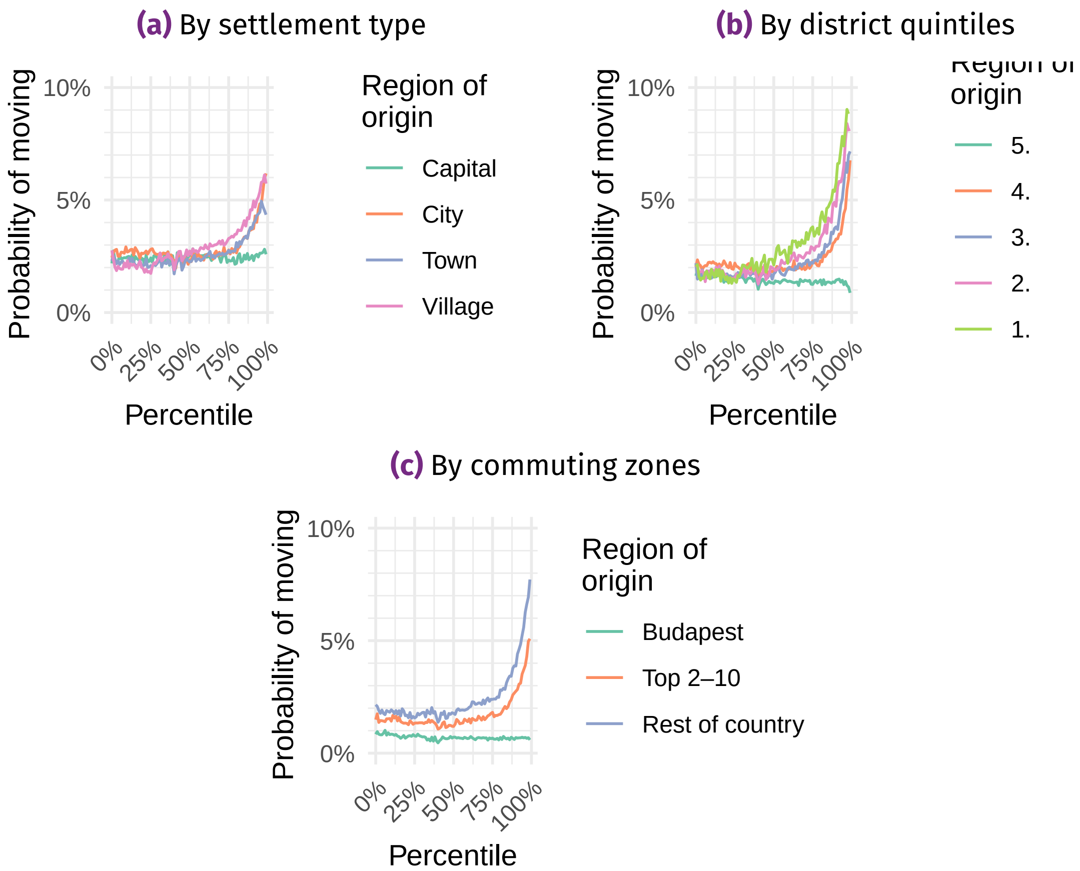
4.2 Regional mobility and income

While a detailed analysis of interregional mobility patterns is not the focus of this paper, we need to briefly look at what patterns emerge from our data, as it could differ from internal migration patterns based on the entire Hungarian population (see KSH 2018a; Bálint and Obádovics 2018 for overviews of recent trends), as opposed to our data in taxpayers with positive wage income (see Section 2). The incidence of moving to another regions peaks at around 30 years of age (Figure C.5), thus our main analysis will be based on the age group 25–39 (further results can be found in Appendix C.4). The direction of within country migration tends towards the more central, more developed, and bigger regions (Figure C.6). However, there's significant out-migration even from high-income regions towards the Budapest region (Table C.1) and very low out-migration from the Budapest region. Note, that this is based on commuting zones, and the largest population gains in the Budapest region happened in the suburbs, while the capital's population continued to decline.

There's an income gradient in the probability of moving to a different region (Figure 3). The probability of moving starts to increase already at around the median income and becomes very steep for the top two deciles. The gradients are present in all three regional classifications, by all regions of origin, except for the top regions, where we see very little out migration at any income level. This income gradient is steeper for the less developed regions and the more remote labour markets. These high income taxpayers are already represented in the less developed regions at a lower rate than the national average (Section 4.1), and out-migration patterns suggest a continued fall. For the youngest cohort – where data issues might be the most serious too – we get very noisy results due to the low number of them in the upper income percentiles but there is an increase in mobility for the top two deciles, and the oldest taxpayers also show a small income gradient (Figures C.14–C.16).

There are several possible explanations for this association between income and mobility. Moving to a different region could incur significant costs (higher living cost, search costs), and only higher income people can afford to look for a job, even if other regions offer much better job opportunities.

Figure 3: Probability of moving between regions by income



Note: Probabilities of moving, averaged over 2011–2017, by one year lagged national income percentile for ages 25–39. All year-pairs only taxpayers with positive income and a known address in two successive years were included. On panels (a) and (b) moving is defined as relocation between the aggregations showed. For panel (c) moving is defined as a relocation between individual commuting zones, and "Top 2–10" and "Rest of country" show the averages across commuting zones in these two aggregations. For all other age groups see Figures C.14–C.16. District quintiles refer to quintiles of district level per capita income, based on data for 2017 (5: highest, 1: lowest).

A related consideration could be agglomeration effects in knowledge industries, which cause a wage premium for high skilled, highly paid workers, where the premium is sufficiently high to counterweight increased living cost. Such premia could exist for low-skilled workers as well but in this case living cost differences might reduce the incentive to move. There could be a mismatch type explanation as well, where jobs in the more affluent regions require skills, qualifications, or experience that many people in less affluent regions do not have. Therefore the chance of finding a good match in a different location is low, and people will not permanently re-settle.

These results could be biased by the fact of having no data on individuals moving abroad, as they disappear from our database the same way as someone dropping out of the labour force. It is possible that mobile people with low income might move abroad rather than within the country if the income differential of a foreign job is higher. However, we don't see large differences in exit probabilities in the top six deciles (Table C.2), thus we expect the bias to be small, and it shouldn't affect the qualitative conclusions. Increasing the time lags severely limits our sample size, the qualitative findings are also robust for longer lags, although over the whole six years in our dataset the income gradient is less steep.

4.3 Mobility premia

As shown above, there is a strong correlation between income and mobility levels, which could have many – not mutually exclusive – causes. People with good earnings potential and an upward income trajectory in one region moving for better short, or long term opportunities to a different region, would be consistent with the observed higher mobility rates of high income people. There could be a negative association as well, if someone moves after job loss. Nevertheless, our data allows to highlight some aspects of this association.

Before analysing this association in an event study regression framework, we look at broad income dynamics of changes in individuals' deciles for a first overview. In Figure 4 the top row shows the overall persistence of income with the distribution of decile changes, and the bottom row shows the difference between movers and stayers for each step of the decile changes. The level of income is also important, as moving up from the upper, and moving down from the lower deciles is not possible, therefore the distributions are shown for three categories of lagged income. People who have recently moved are less likely to stay in the same decile as in the previous year. The highest in-

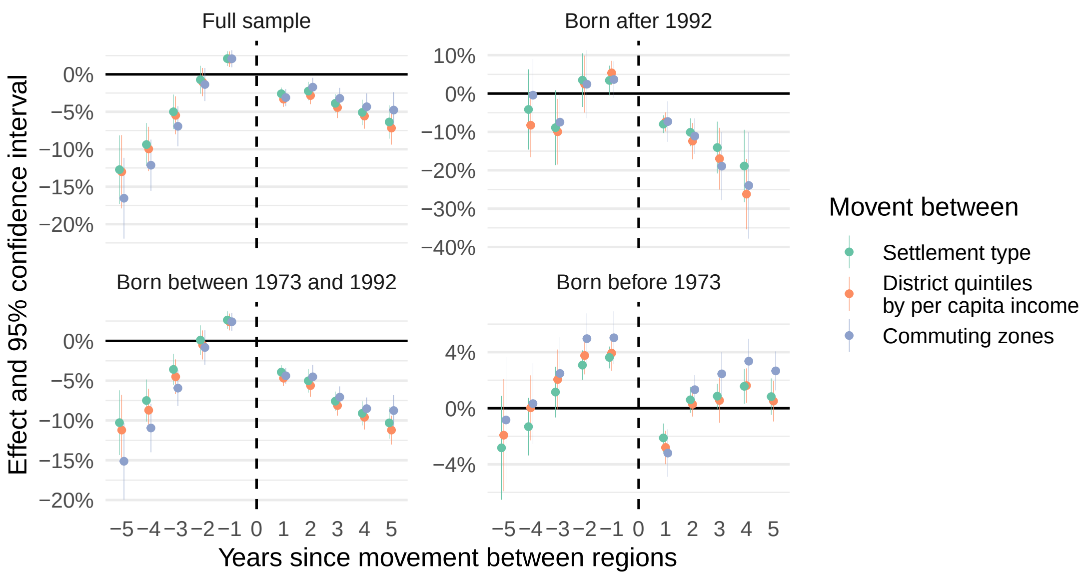
Figure 4: Distribution of change in relative income between 2011 and 2017 for prime age taxpayers who moved in previous year, annual averages

Note: Panel *A* shows the distribution of changes in deciles for all taxpayers. Panel *B* show the distribution of differences in changes of deciles between movers and stayers. See Figures C.17–C.19 for all age groups.

come tax payers are also less likely to move across the income distribution, while middle-income taxpayers experience significant changes in both directions but downward movement seems to dominate. The picture is similar for all regional classifications. See Figures C.17–C.19 for all age groups. The difference is lower for the youngest age group, who are most likely to move. For middle aged and older middle-earners there's a large peak for moving two steps down, and negative changes have a fatter tail.

The presence of upward and downward income dynamics can be interpreted as two different motivations for moving to a another regions. If people move for a better paying job, this would show up as an immediate bump in income. However, if people move for better opportunities, they might accept a short term dip in their earnings because better job prospects would make up for that in the longer term. Meanwhile, confounders and reverse causality could also be at play. An alternative explanation would be if the official change in address happens after the actual relocation, and there is some form of short term mobility premium (or just a signing bonus when changing jobs), while negative income dynamics could be due to regression to

Figure 5: Event study results for the effect of moving to a different commuting zone on annual income (% change)



Note: Confidence intervals are estimated using twoway standard errors at the individual and year effects. Vertical axis scales are different for each panel. The fifth leads and fifth lags were omitted from the plot for the cohort born after 1992.

mean. Moreover, a negative wage shock could be associated with job losses as well, either through spending time out of work, or through a lower paid post-severance job, that is a worse match for the worker.

The results of the regression analysis confirm potential confounding and reverse causality concerns (Figure 5).

The sample was split by cohorts. For the youngest cohort, the event time estimates leading up to the year of the move are relatively close to zero, after which we see a decline in wages. However, the generally strong income dynamics in these ages, combined with the suspected issues around the timing of a registered change in address could explain this pattern. Furthermore, university graduates are much more mobile than the rest of the population (Nyüsti and Ceglédi 2013), thus education related interregional mobility can also explain the large differences between young people who move, compared to their peers who don't. However, we cannot observe either ongoing studies, or educational attainment in our data, nor do we have data on students who don't declare income in a particular year. These biases can be amplified by the different income dynamics of similarly aged youth without a university degree, relative to students and early career university graduates. Additionally, the low number of observations, and especially the low number of observations with sufficiently long spells in our data make the estimates highly uncertain.

For prime age workers – who are represented here by people at least 25 years old in 2011, and at most 39 years old in 2017 – the estimates are large and significant, and we see a large continuous increase in wages before the move, and a gradual decrease after. The pre-trends strongly suggest the existence of reverse causality and confounding issues. The wage premium for movers decreases markedly, suggesting transitory effects of interregional mobility. If this is a true effect, it could help explaining low mobility rates, as in this case the mobility premium would not be sufficient to overcome the potentially higher living costs after moving.

For older workers we see evidence for different motives to move. Wages start declining before the move and continue to fall for one more year. This negative income shock seems to be transitory, as wages recover after the mode. This pattern could be explained by job losses but moving back to ageing parents requiring care is also a possible story.

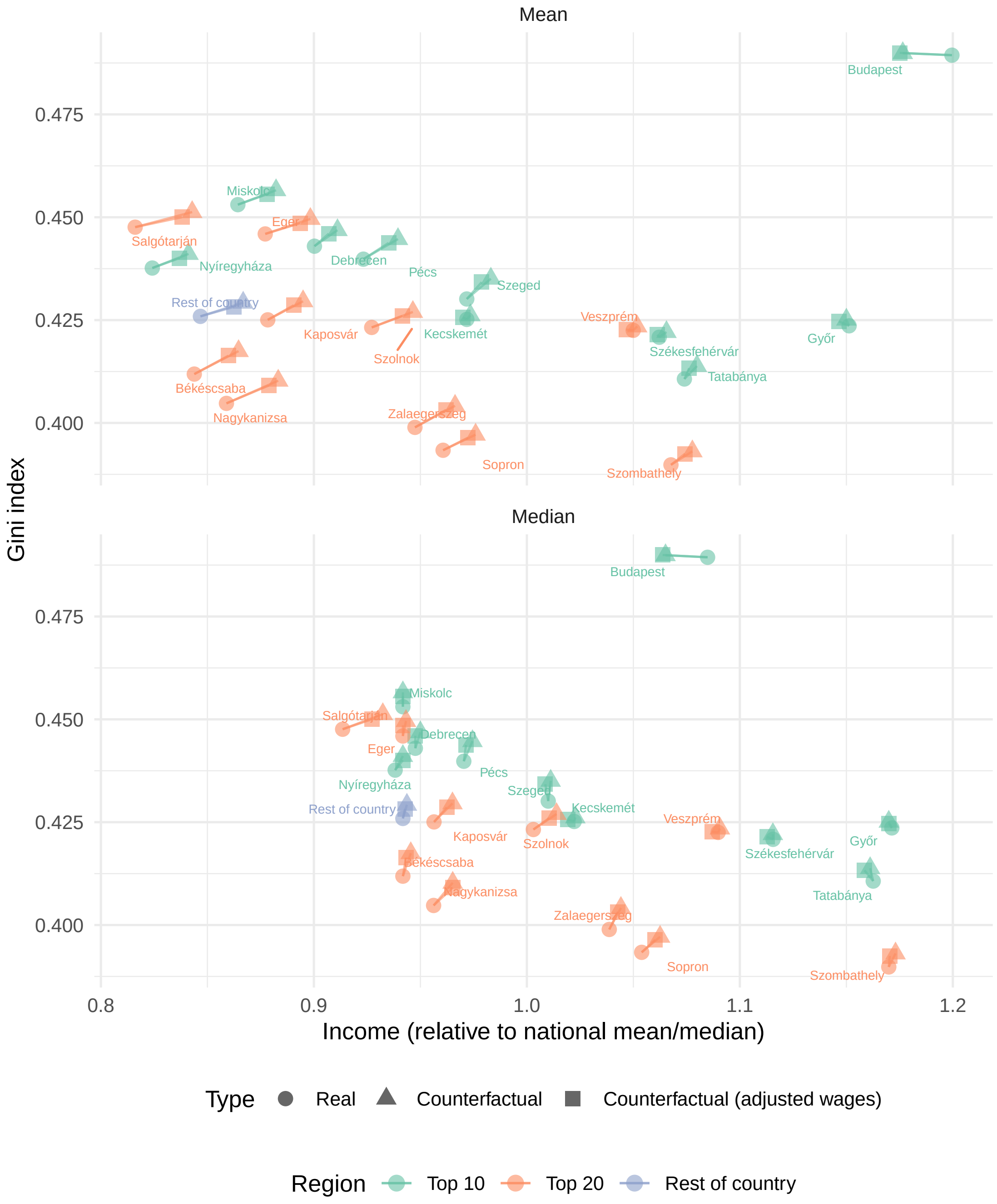
Even though these results cannot be interpreted causally, there is a large, and in most cases statistically significant but transitory association between interregional mobility and increases in wages.[5] However, the confidence intervals are wide, which, along with Figure 4 suggest that these estimated average effects hide considerable unobserved heterogeneity. Uncertainty is the highest for the commuting zone based estimates, where the number of relocations is lower than for the other two definitions. At the same time, people moving between commuting zones are the least like to stay in the same decile, thus their income dynamics could also be the highest, leading to to higher uncertainty in the estimates.

4.4 Interregional mobility and inequality

Finally, we summarise the findings thus far in a counterfactual exercise. Mobility rates are low for the bottom eight deciles and there is a steep income gradient above that. The sources of most of internal migration are low-income, non-central areas, and the targets are rich, central areas around urban centres. There is also a strong positive association between interregional mobility and

[5] The income gradient observed among the movers discussed in the previous section might include some of this mobility premium. However, this is not large enough to change the qualitative results on the income gradient.

Figure 6: A counterfactual exercise for the effect of internal migration between commuting zones on within region inequality



wage gains. Putting this together, we can expect our counterfactual regional wage distribution to have higher within region inequalities and lower between country inequalities. The results on Figure 6 largely confirm these hypotheses but there is some important heterogeneity in the differences between the observed and counterfactual states.

Due to the overall low interregional mobility rates, accounting for any potential mobility premia is mostly irrelevant. In the adjusted counterfactual case we subtracted the average of the event study coefficients from the 2nd lead to the last lag using the full sample estimates. Both income, and income inequality levels are very similar to the simple counterfactual case based on observed wages. Thus our conclusions on the relationship between interregional mobility and income inequality does not depend on whether we can assume a causal effect of moving on wages.

Mean income in the smaller, less central commuting zones would have been substantially higher if employees hadn't moved out. However, median income would have been almost identical. In the affluent large commuting zones – which experienced relatively high in- and out-migration rates as well – the real and counterfactual values are very close, and Budapest is the only commuting zone where both mean, and median incomes increased significantly compared to the counterfactual. Budapest is a major destination for internal migration, and the typically high income new residents arrived in sufficiently large numbers to pull the median wage up by around 2%.

In terms of inequality, the counterfactual case of high income people staying in less developed regions means roughly 1 Gini point higher within region inequality and no change in the larger regions. This is due to low mobility rates, and the Gini index's stability in case of such small changes in the population. Between region inequality, as measured by the share of the between component of the Theil index relative to the national Theil decreased slower due to interregional migration (Figure C.13).

5 Conclusions

This is the first paper to analyse regional income distributions and inequalities using administrative data sources at a granular geographic level. The research showed the importance of this approach, as we could identify laggard regions hidden by coarser classifications and highlight how regions with similar income levels can have widely different levels of inequality.

Interregional mobility also proved to be a significant factor in between region inequalities. The strong positive correlation between income and mobility also suggests that housing costs could be a constraint on better allocation of labour within the country. However, our data on mobility was limited to long term, permanent relocations. There could be a more broad based short term labour mobility but access to housing could be an issue for that as well.

References

Bálint, Lajos, and Csilla Obádovics. 2018. "Belföldi vándorlás." In *Demográfiai Portré 2018*, edited by Judit Monostori, Péter Őri, and Zsolt Spéder. KSH Népességtudományi Kutatóintézet.

Castells-Quintana, David, Raul Ramos, and Vicente Royuela. 2015. "Income Inequality in European Regions: Recent Trends and Determinants." *Review of Regional Research* 35 (2): 123–46. https://doi.org/10.1007/s10037-015-0098-4.

KSH. 2018a. "A belföldi vándorlás aktuális trendjei, 2012–2017." *Statisztikai Tükör*, 12. http://www.ksh.hu/docs/hun/xftp/stattukor/belfvand17.pdf.

———. 2018b. "Munkaerőpiaci körzetek Magyarországon." Központi Statisztikai Hivatal. http://www.ksh.hu/docs/hun/xftp/idoszaki/pdf/munkkor17.pdf.

Morvay, Endre. 2015. "Külföldi munkavállalás a közép-kelet-európai országokban az európai munkaerő-felmérések alapján." MPRA paper. University Library of Munich, Germany. https://mpra.ub.uni-muenchen.de/id/eprint/90631.

Nyüsti, Szilvia, and Tímea Ceglédi. 2013. "Vándorló diplomások, diplomáért vándorlók." In *Frissdiplomások, 2011*, edited by Orsolya Garai and Zsuzsanna Veroszta, 173–207. http://www.felvi.hu/pub_bin/dload/DPR_kotet_frissdiplomasok2011/DPR_nyomdai.pdf.

Svraka, András. 2021. "Recent Trends in Income Inequalities in Hungary Using Administrative Data."

A Regional classifications

We use three different classifications, shown on Figure A.1 to look at within region inequalities and between region mobility. All of these are aggregated from municipalities and communes as used in Eurostat's Local Administrative Units (in case of Budapest, the city's 23 districts are treated as separate settlements).

Settlement types according to the Central Statistical Office's classification can be used to investigate urban-rural differences. The country is divided into the capital (the 23 districts of Budapest), 23 cities (which are mostly county seats, and a few other large towns), 322 other towns, and 2809 villages.

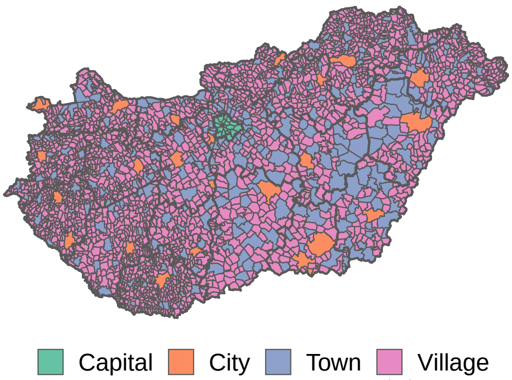
The 197 districts[6] are primarily administrative divisions below NUTS3 level (counties) but their definition is influenced by socioeconomic factors, such as district seats need to be accessible by public transit from within the district because many social services are organized by districts, and districts are also used in regional planning. This makes them useful as units of analysis. We also use district level per capita taxable income as a proxy for the level of development, primarily by aggregating districts into quintiles.

Commuting zones were introduced by KSH (2018b). These are defined based on commuting patterns observed at the last census in 2011 in a way to minimize the number of commutes between zones. Commuting zones – especially the zones around large cities – are substantially bigger than districts. This can hide some differences in income levels within local labour markets but a major advantage is that mobility between commuting zones should exclude most non-work related relocations within a single labour market, which might be captured as a significant change in location in the previous two classifications. There are 84 zones. For the analysis of mobility we'll consider all commuting zones as separate units but aggregate the results into three categories based on population size for the sake of tractable visualisations: the Budapest zone, which encompasses the entire central region (population of 3.1 million); the next nine zones (populations of 0.2–0.4 million, a total of 4.1 million), which mostly correspond to the largest cities and their suburbs; and the rest of the country. For inequality measures individuals were aggregated into these three groups.

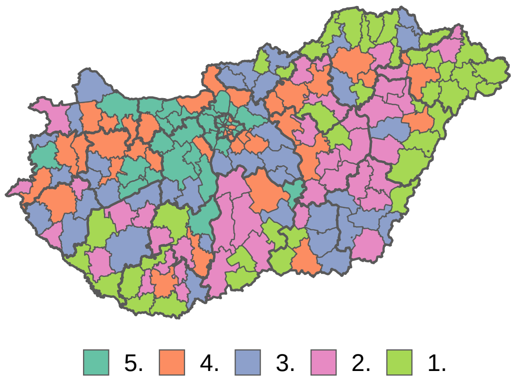
[6] The terminolgy is somewhat confusing in English. These micro-regions below NUTS3 aggragation are called districts (*járások*). The captital of Budapest is made up of 23 individual municipalities (*település*, in Eurostat's terminology Local Administrative Units (LAUs)), which coincidentally are also called districts in English (although *kerületek* in Hungarian). Budapest's districts are separate districts in the *járás* sense as well.

Figure A.1: Maps of aggregated regions used in the analysis

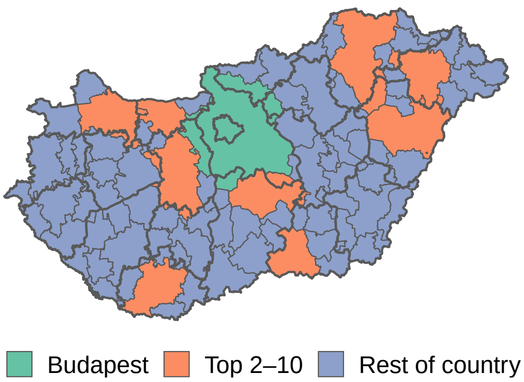
(a) Settlements



(b) Districts and quintiles of per capita income



(c) Commuting zones



Note: Thick borders denote counties (NUTS3), thin borders denote settlements, districts and quintiles of per capita income, commuting zones respectively.

B Construction of residential and mobility data

Figure B.2: Number of taxpayers without correspondence (in thousands)

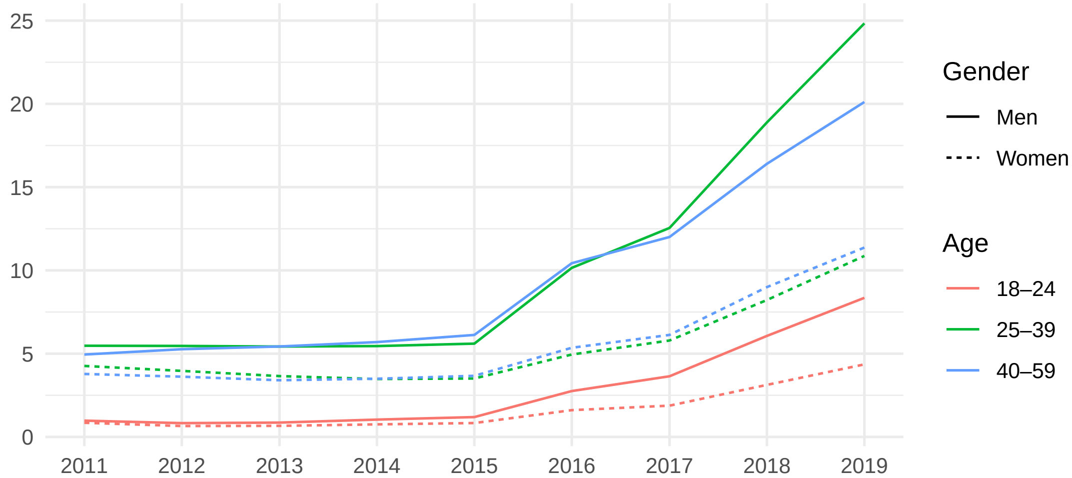
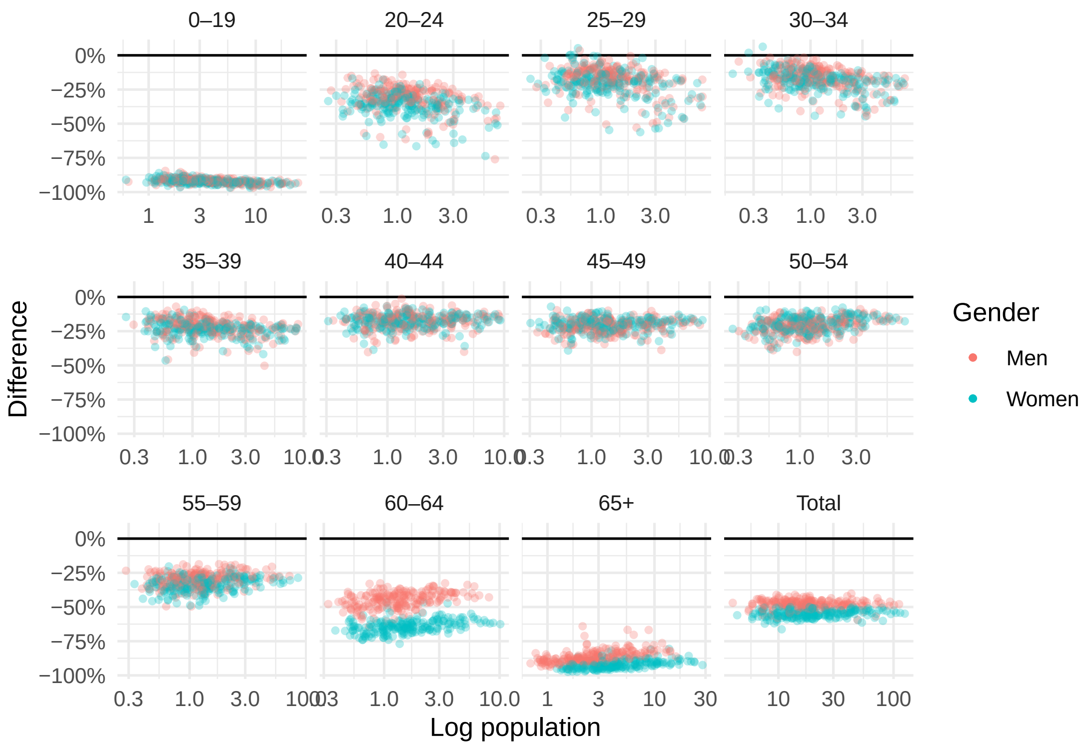
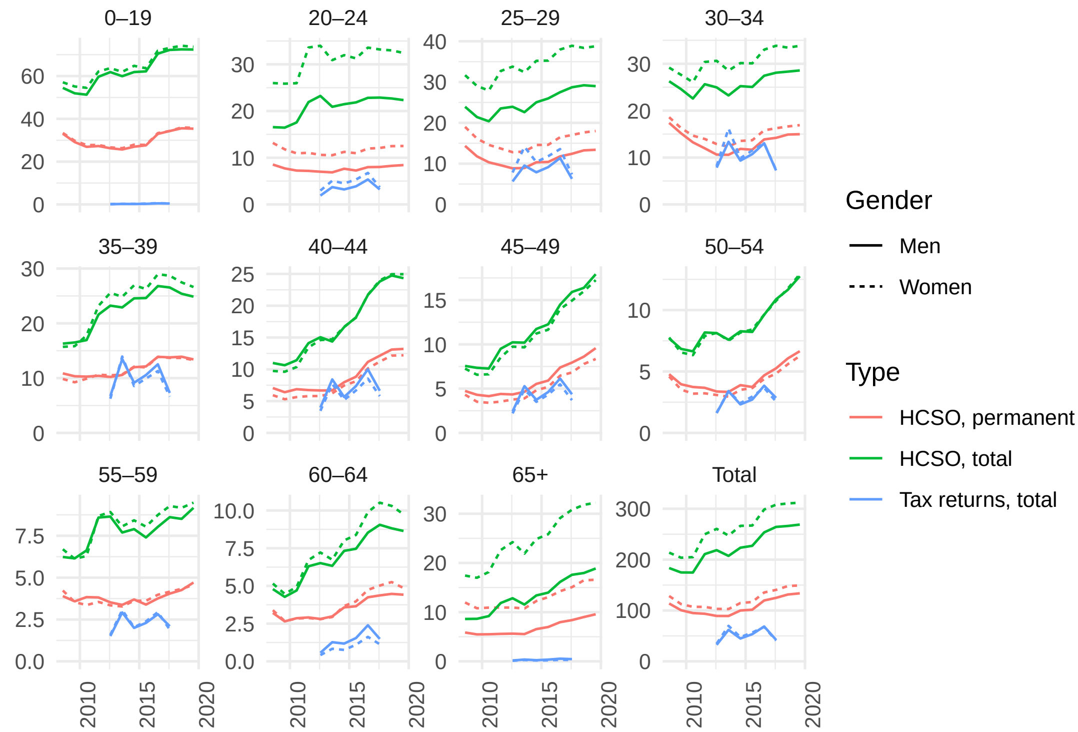


Figure B.3: Relative difference of district level taxpayers with employment income to total district population by demographic groups



Note: Horizontal axis scales are independent for each panel.

Figure B.4: Number of relocations between settlements (in thousands)



Note: HCSO refers to the Central Statistical Office's internal migration data using official residency records. Tax returns refer to internal migration observable in the tax returns. Vertical axis scales are not identical.

C Additional figures and tables

C.1 Interregional mobility

Figure C.5: Age profile of interregional mobility (annual averages for 2012–2017)

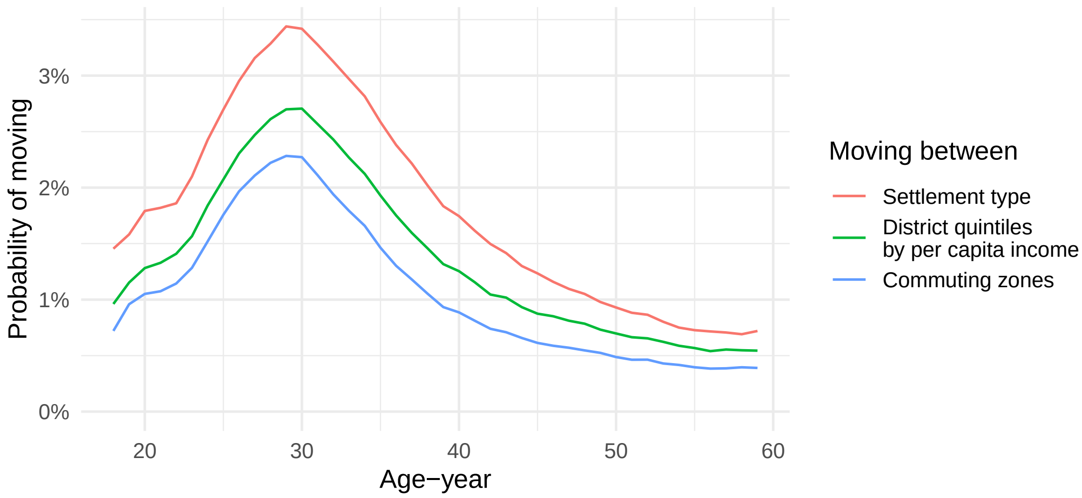
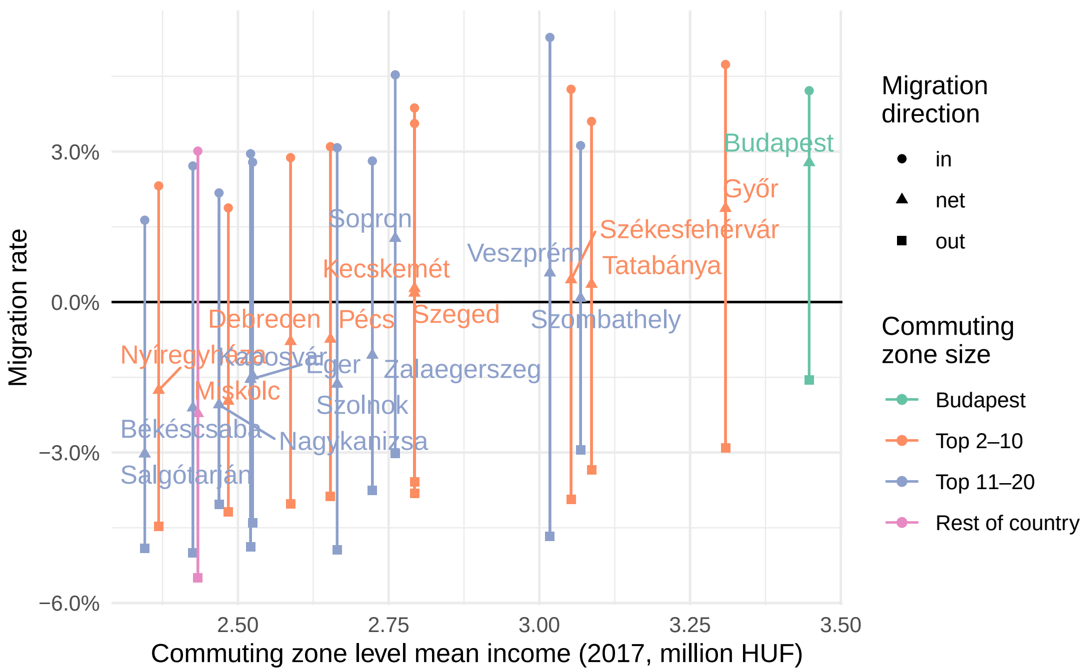


Figure C.6: 2012-2017 migration rates between commuting zones



Note: "Rest of country" denotes movements between different commuting zones outside the top 20 biggest zones.

Table C.1: Share of out-migration destinations (columns) from top 10 commuting zones (rows) between 2012–2017 (%)

	Budapest	Miskolc	Debrecen	Székesfehérvár	Nyíregyháza	Győr	Pécs	Szeged	Kecskemét	Tatabánya	Rest of country
Budapest		3.4	3.2	10.3	2.3	4.1	2.3	2.7	5.5	4.9	61.3
Miskolc	54.3		5.1	2.0	2.7	3.0	0.3	0.7	0.8	1.5	29.7
Debrecen	51.2	3.9		1.6	5.4	2.9	0.4	0.8	1.3	1.9	30.5
Székesfehérvár	51.1	0.6	0.5		0.4	2.8	1.6	0.9	1.0	3.4	37.6
Nyíregyháza	51.4	4.5	12.8	1.4		2.1	0.2	0.7	1.1	1.3	24.4
Győr	43.2	0.8	0.8	2.5	0.4		0.9	0.7	0.6	7.8	42.2
Pécs	45.5	0.5	0.7	2.6	0.2	2.5		1.2	1.3	1.3	44.2
Szeged	47.5	0.7	1.1	1.7	0.5	1.6	1.4		4.3	0.9	40.4
Kecskemét	54.3	0.6	0.6	1.7	0.5	1.2	0.9	5.2		0.8	34.2
Tatabánya	52.5	0.5	0.7	5.3	0.3	13.1	0.6	0.5	0.8		25.7
Rest of country	40.6	2.5	3.5	3.2	2.1	4.3	3.0	3.4	2.3	1.6	33.4

Note: All rows sum to 100%. Migration shares from “Rest of country” to “Rest of country” denotes movements between different commuting zones outside the top 10 biggest zones.

C.2 Income trends

Figure C.7: Mean and median taxable wages by settlement type

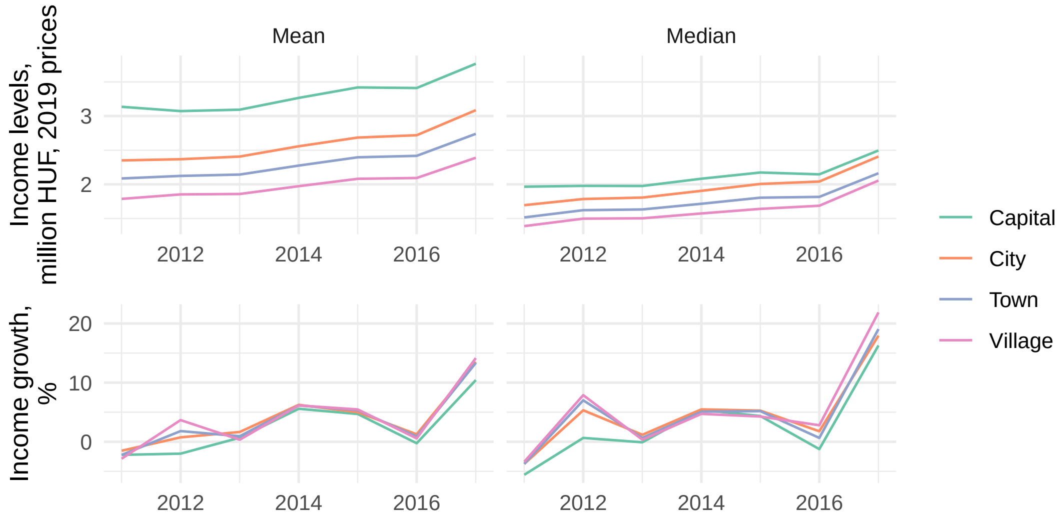
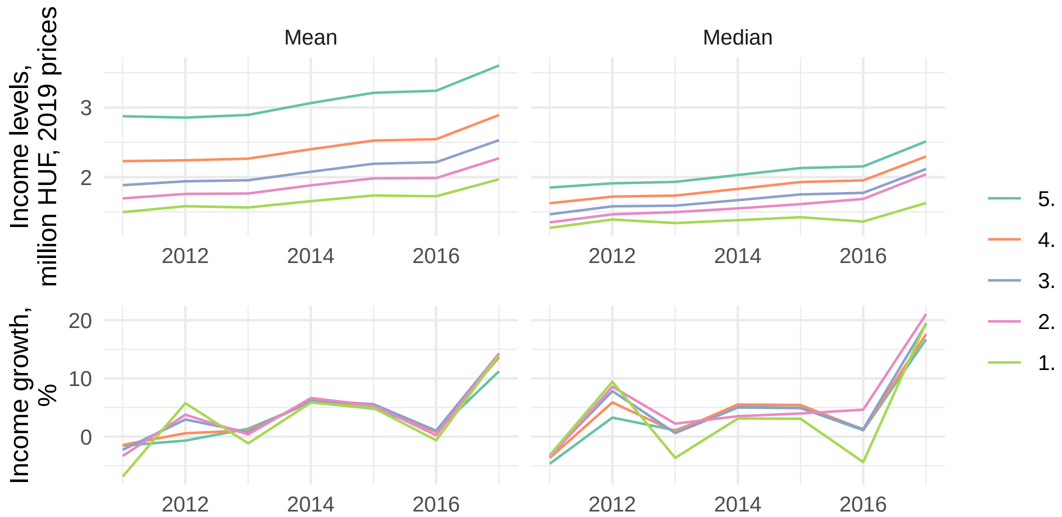
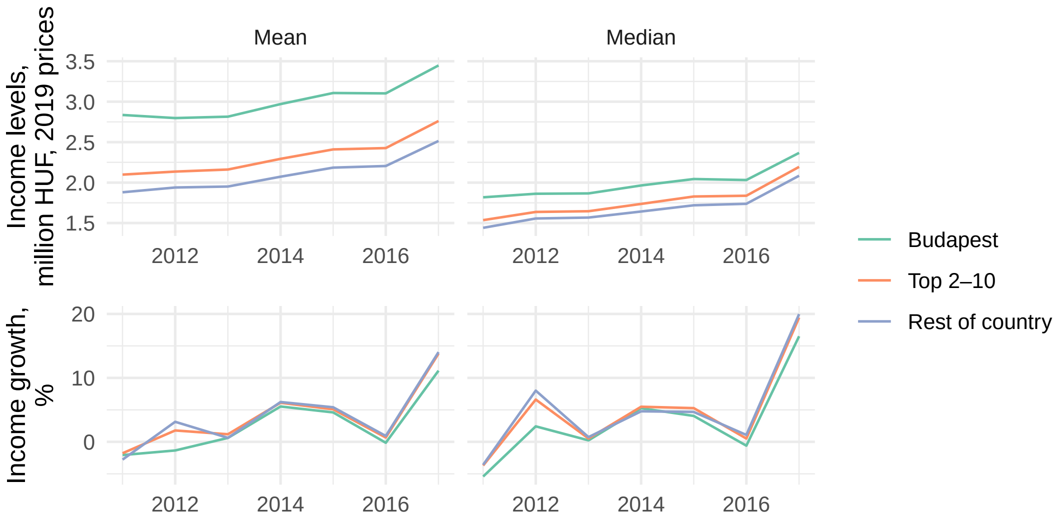


Figure C.8: Mean and median taxable wages by district quintiles



District quintiles refer to quintiles of district level per capita income, based on data for 2017 (5: highest, 1: lowest).

Figure C.9: Mean and median taxable wages by commuting zones



List of commuting zones is ordered by population.

C.3 Inequality measures

Figure C.10: Within group income inequality measures by settlement type

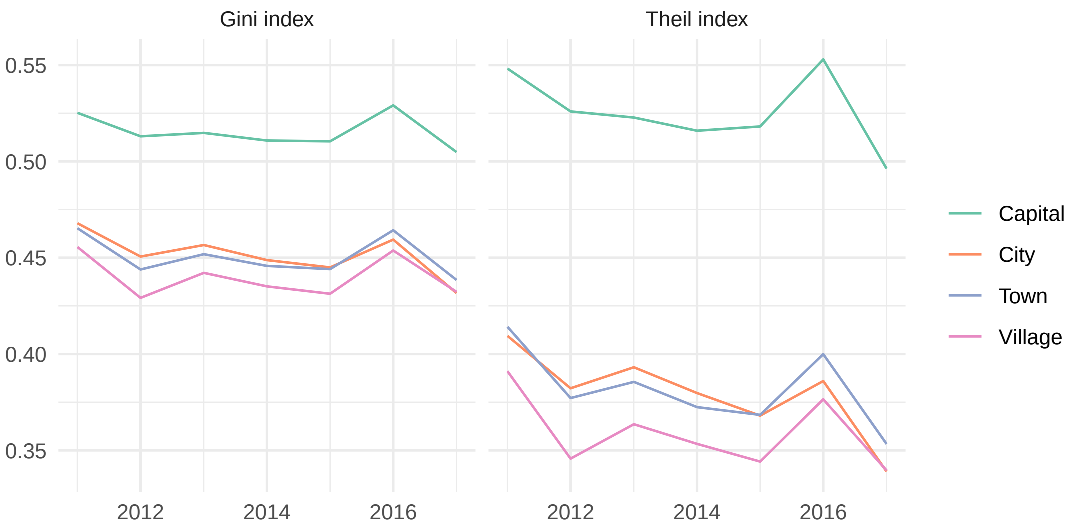
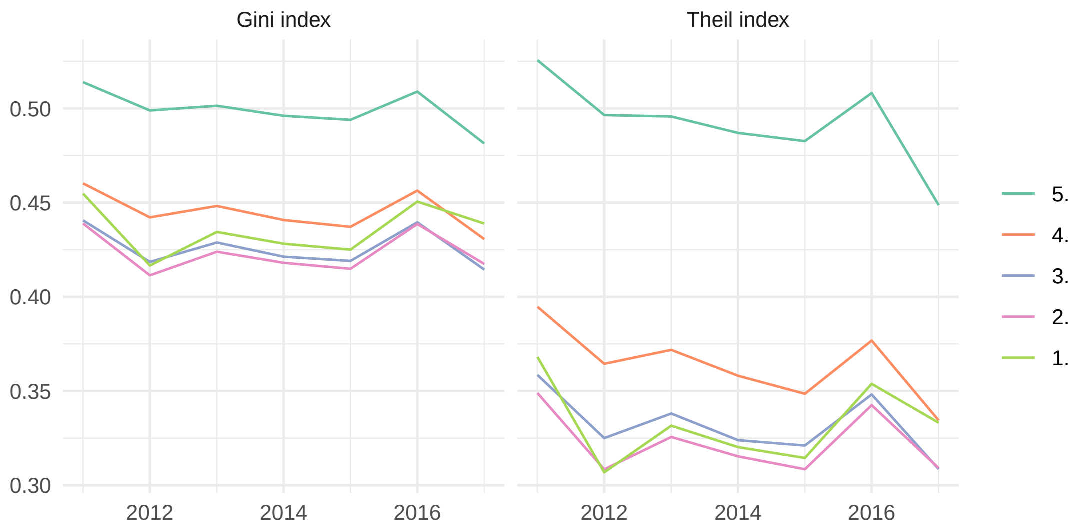
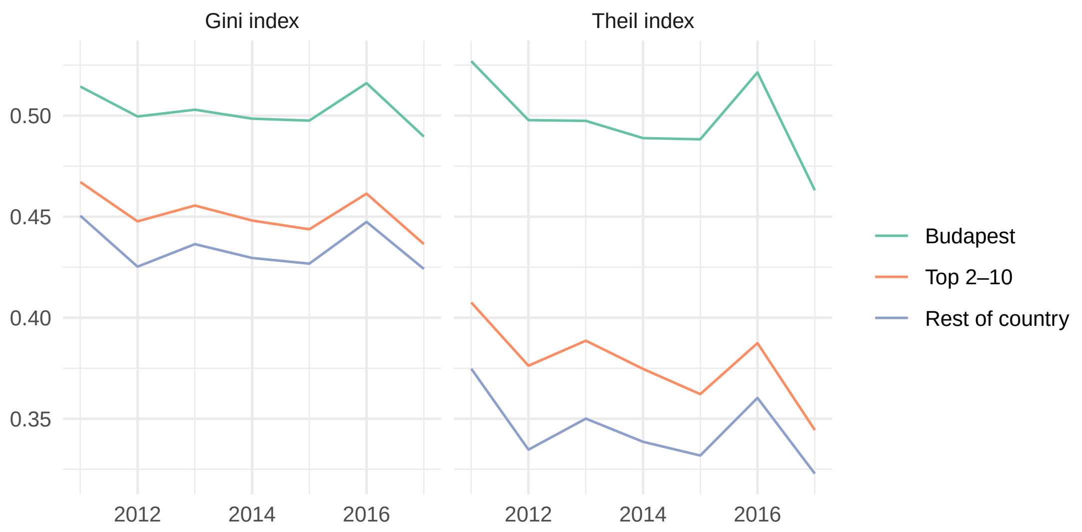


Figure C.11: Within group income inequality measures by district quintiles



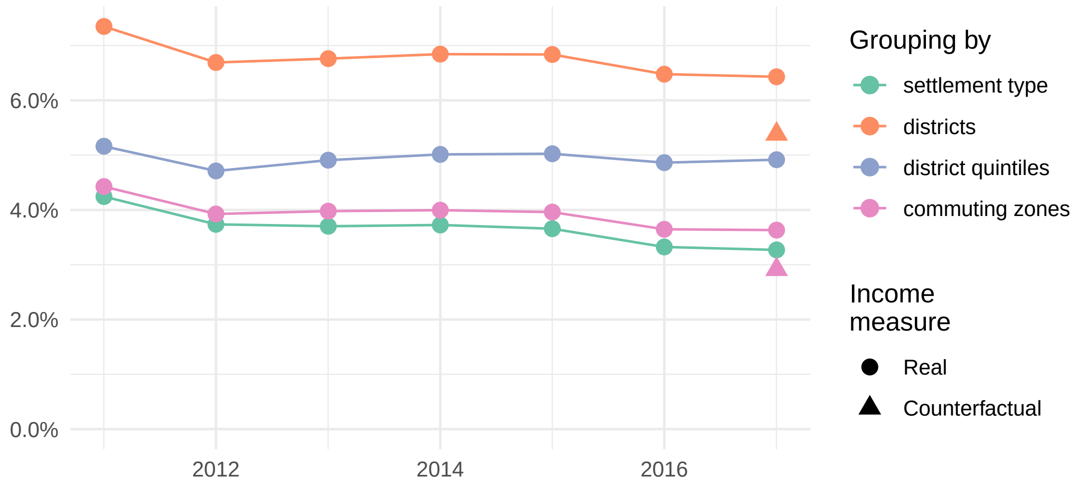
District quintiles refer to quintiles of district level per capita income, based on data for 2017 (5: highest, 1: lowest).

Figure C.12: Within group income inequality measures by commuting zones



List of commuting zones is ordered by population.

Figure C.13: Share of between group income inequalities measured by the Theil index



Note: Counterfactual values refer to the counterfactual exercise discussed in Section 4.4, in which inequality is measured on the 2017 wage distribution with 2011 geographic locations.

C.4 Regional mobility

Figure C.14: Probability of moving between regions, by settlement type

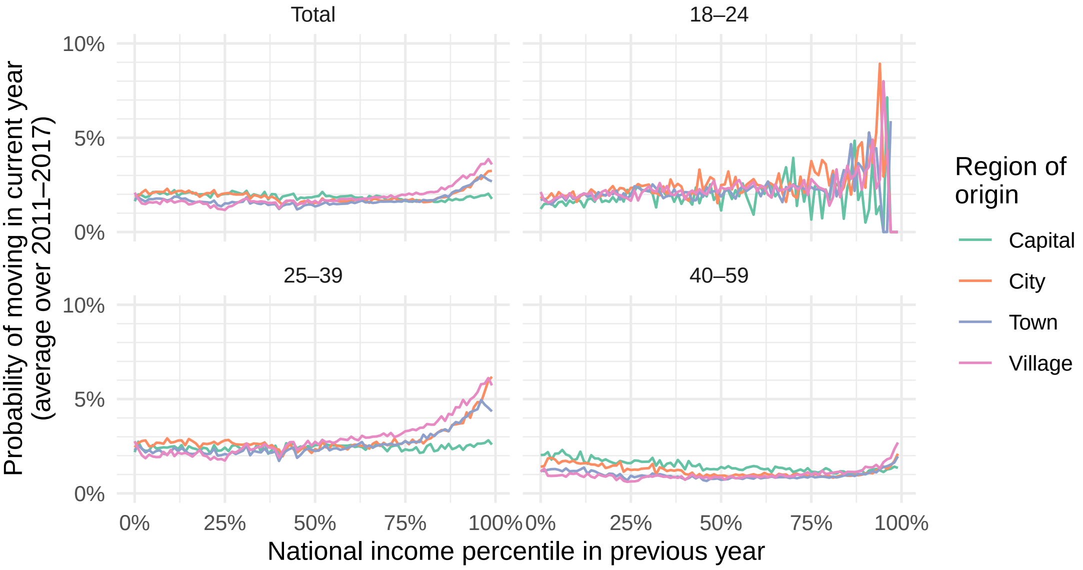


Figure C.15: Probability of moving between regions, by district quintiles

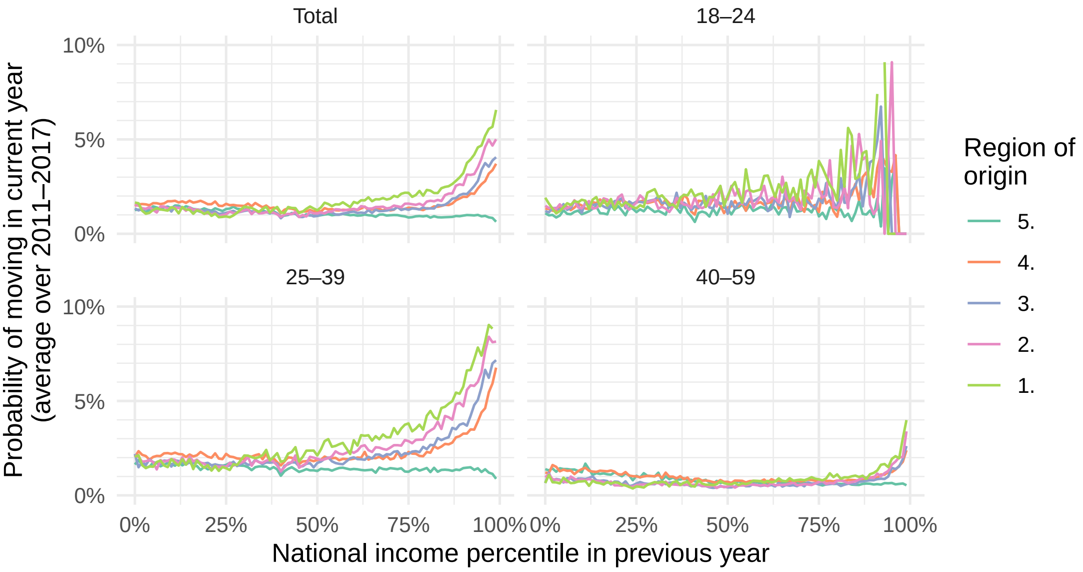


Figure C.16: Probability of moving between regions, by commuting zones

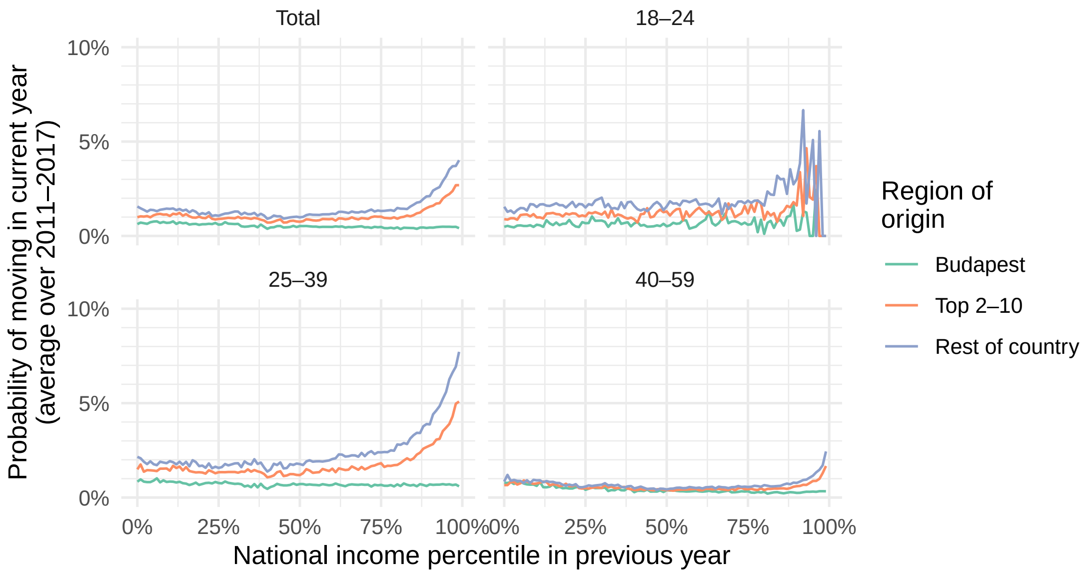
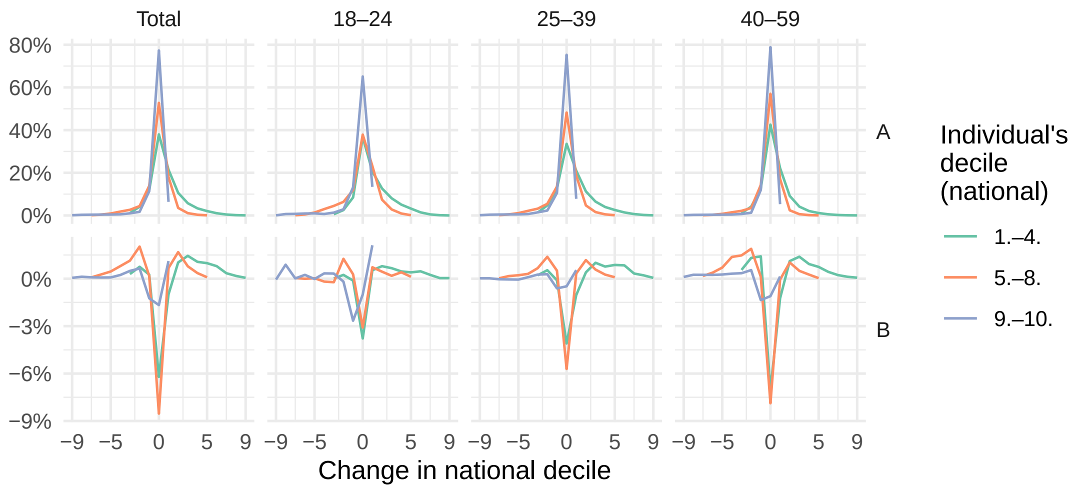
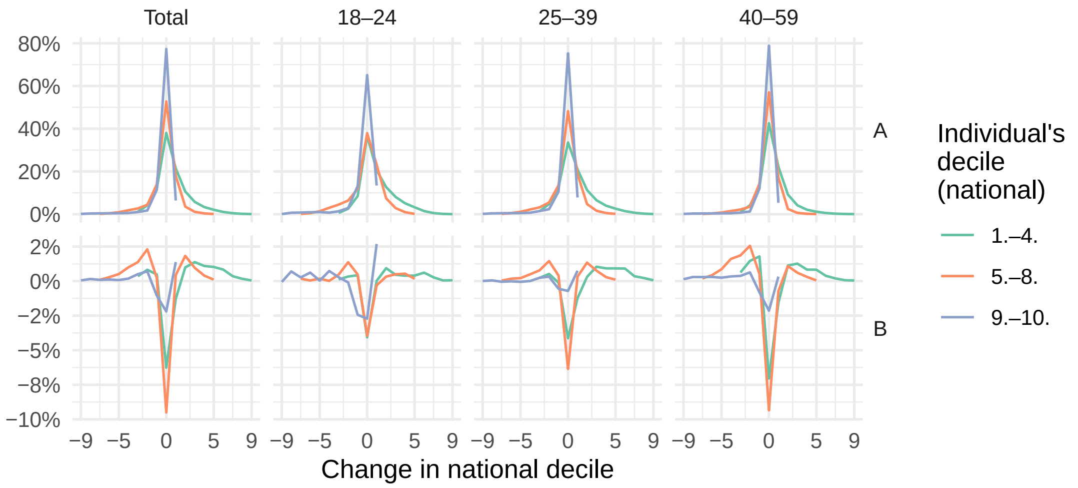


Figure C.17: Distribution of change in relative income between 2011 and 2017, annual averages, by settlement type



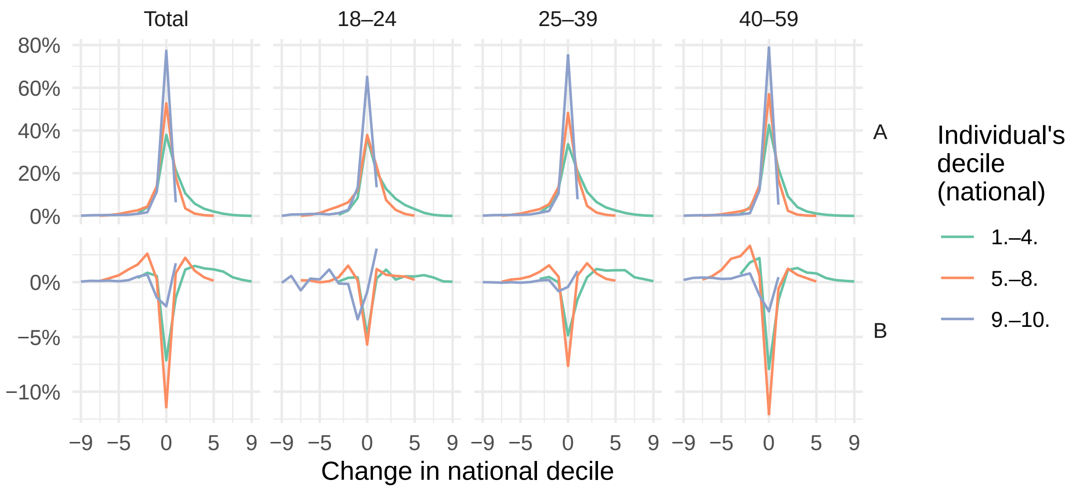
Note: Panel *A* shows the distribution of changes in deciles for all taxpayers. Panel *B* show the distribution of differences in changes of deciles between movers and stayers.

Figure C.18: Distribution of change in relative income between 2011 and 2017, annual averages, by district quintiles



Note: Panel *A* shows the distribution of changes in deciles for all taxpayers. Panel *B* show the distribution of differences in changes of deciles between movers and stayers.

Figure C.19: Distribution of change in relative income between 2011 and 2017, annual averages, by commuting zones



Note: Panel *A* shows the distribution of changes in deciles for all taxpayers. Panel *B* show the distribution of differences in changes of deciles between movers and stayers.

C.5 Robustness checks

Table C.2: Linear probability models for not filing a tax return by demographic groups

	Exit (1 year)						Exit (permanent)					
	18–24		25–39		40–59		18–24		25–39		40–59	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
Constant	0.417	0.399	0.366	0.374	0.391	0.348	0.057	0.060	0.122	0.116	0.164	0.143
1 year lagged wage decile (ref. category 1st)												
2	−0.160	−0.151	−0.142	−0.134	−0.168	−0.157	−0.031	−0.028	−0.054	−0.044	−0.080	−0.067
3	−0.212	−0.200	−0.215	−0.211	−0.248	−0.231	−0.043	−0.044	−0.080	−0.075	−0.117	−0.103
4	−0.241	−0.233	−0.264	−0.259	−0.292	−0.275	−0.052	−0.059	−0.097	−0.089	−0.135	−0.125
5	−0.265	−0.259	−0.297	−0.295	−0.321	−0.303	−0.059	−0.066	−0.110	−0.100	−0.148	−0.140
6	−0.271	−0.266	−0.299	−0.300	−0.327	−0.306	−0.060	−0.069	−0.110	−0.102	−0.150	−0.141
7	−0.281	−0.273	−0.310	−0.309	−0.333	−0.311	−0.064	−0.072	−0.114	−0.104	−0.153	−0.144
8	−0.286	−0.277	−0.318	−0.313	−0.340	−0.317	−0.067	−0.074	−0.117	−0.106	−0.156	−0.147
9	−0.291	−0.278	−0.323	−0.317	−0.344	−0.322	−0.070	−0.072	−0.119	−0.106	−0.157	−0.152
10	−0.283	−0.263	−0.325	−0.319	−0.344	−0.317	−0.065	−0.066	−0.120	−0.106	−0.157	−0.147
Year (ref. category: 2012)												
2013	−0.021	−0.031	−0.008	−0.012	−0.013	−0.024	0.000×	0.001×	0.002	0.002	−0.004	−0.014
2014	−0.037	−0.049	−0.015	−0.017	−0.020	−0.031	0.001+	0.001×	0.000×	0.002	−0.005	−0.016
2015	−0.040	−0.050	−0.016	−0.017	−0.022	−0.032	0.005	0.006	0.002	0.005	−0.005	−0.016
2016	−0.101	−0.095	−0.044	−0.036	−0.042	−0.048	0.009	0.012	0.001	0.006	−0.005	−0.018
2017	0.008	0.045	−0.015	−0.012	−0.018	−0.029	0.063	0.082	0.019	0.023	0.011	−0.007
County	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	−0.010	−0.013	0.000	−0.001	0.000	0.003	0.001	0.000**	0.001	0.000	0.001	0.004
N (millions)	0.968	0.873	4.536	4.208	5.366	5.823	0.968	0.873	4.536	4.208	5.366	5.823
R^2	0.111	0.106	0.113	0.112	0.121	0.111	0.032	0.036	0.042	0.040	0.056	0.060
Adj. R^2	0.111	0.106	0.113	0.112	0.121	0.111	0.032	0.036	0.042	0.040	0.056	0.060

Note: Exiting is defined as not filing a tax return with wage income in the current year (for 1 year exit), or not filing at all from current year until 2019 (for permanent exit), conditional on having filed in the previous year. Note that models only cover years 2012–2017, as we have reliable location data only for these years. However, exits are unrelated to location, and they can be tracked reliably until 2019. All models include county effects not reported here (ref. category is Budapest). Age is measured relative to age floor within age group. Coefficients are significant at $p < 0.001$, unless otherwise noted with ×~$p \geq 0.1$, +~$p < 0.1$, *~$p < 0.05$, **~$0.001 \leq p < 0.01$. Average one year exit rate from 2011 to 2017 was 9%, and average permanent exit rate was 3%.