PÁLMA FILEP-MOSBERGER | ÁDÁM REIFF

INCOME TAX EVASION ESTIMATION IN HUNGARY

MNB WORKING PAPERS | 4

2022
MAY

INCOME TAX EVASION ESTIMATION

IN HUNGARY

MNB WORKING PAPERS | 4

2022

MAY

MNB Working Papers 2022/4

Income tax evasion estimation in Hungary *

(Jövedelemeltitkolás becslése Magyarországon)

Written by Pálma Filep-Mosberger**, Ádám Reiff***

Budapest, May 2022

Published by the Magyar Nemzeti Bank

Publisher in charge: Eszter Hergár

Krisztina körút 55., H-1013 Budapest

www.mnb.hu

ISSN 1585-5600 (online)

*For helpful comments we thank Gergely Baksay, Balázs Csomós, Péter Elek, Marianna Endrész, Péter Gábriel, Gábor P. Kiss, Krisztina Szabó, Mihály Szoboszlai and seminar participants at the Central Bank of Hungary. We thank Beáta Kassainé Nagy, Krisztina Kollár, Márton Papp and Antónia Éva Schwartz (Hungarian Central Statistical Office) for consulting on questions about the Hungarian Household Budget Survey database. The views in this paper are solely those of the authors and do not necessarily reflect the views of the Central Bank of Hungary.

**Central Bank of Hungary; mosbergerp@mnb.hu.

***Central Bank of Hungary; reiffa@mnb.hu.

Contents

Abstract

This paper studies labour market tax avoidance in the 2010s in Hungary, following major labour market tax reforms in the beginning of the decade. First we show that aggregate time series are broadly consistent with a ”whitening” process, in which a higher fraction of incomes are declared. However, as aggregate developments are driven by several, often unobservable factors, we cannot conclude that the observed phenomena are indeed caused by a whitening process in the labor market. Therefore in the second part of the paper we use several micro datasets to shed light on the nature of the whitening process. By comparing the consumption pattern of entrepreneurs (who might have undeclared incomes) and state sector employees (who are unlikely to have undeclared income), we show that income underreporting of entrepreneurs did decline in the 2010s. On the other hand, we find that the number of illegal employees – e.g. of those who work without any work contract – only temporarily declined in the aftermath of the financial crisis and seems to follow a procyclical pattern.

JEL: H26, J21, J31 .

Keywords: labour market tax avoidance, illegal employment, income underreporting .

Összefoglaló

Tanulmányunkban a munkaerőpiaci adóelkerülést vizsgáljuk a 2010-es évek elején bevezetett főbb munkaerőpiaci reformok utáni évtizedben. Először bemutatjuk, hogy az aggregált makró idősorok szerint a jövedelem egyre nagyobb hányada került bevallásra, ami ”fehéredésre” utalhat. Azonban, mivel a makro folyamatokat számos más tényező befolyásolja, köztük nem megfigyelhetők is, így nem vonhatjuk le azt a következtetést, hogy a munkaerőpiacon megfigyelt stilizált tényeket valóban a fehéredés okozta. Ezért a tanulmány második részében mikroadatbázisok segítségével vizsgáljuk a fehéredés kérdését. Összehasonlítjuk az egyéni vállalkozók (akiknek lehet eltitkolt jövedelme) és közszférában alkalmazottak (akiknek nem valószínű, hogy eltitkolt jövedeleme van) fogyasztási mintáját, és azt találjuk, hogy az egyéni vállalkozók jövedelemeltitkolása csökkent a 2010-es évek során. Másrészt azt találjuk, hogy a nem bejelentett foglalkoztatottak száma - pl. akik munkaszerződés nélkül dolgoznak - csak ideiglenesen csökkent a pénzügyi válság után és változása prociklikus mintát követ.

1 Introduction

Labour market tax avoidance is a global phenomenon. According to the estimate of the International Labour Organization, in 2016 two billion people – or 61% of the global working population – earned they daily living in the informal sector (ILO (2018)). According to the estimates, illegal employment is daily routine in the less developed (low income) countries, where the employment share of the informal sector can exceed 90%; but with its estimated 18% employment share it is also very widespread in developed (high income) countries.[1]

Naturally, informal employment – besides being financially beneficial for those concerned – has its drawbacks for both the states, and also for the individuals (although in a more indirect way). Informal employment means forgone labour market tax revenues which have to be collected from other types of taxes, moreover it also hurts fair competition between and within sectors. For the individuals concerned, informal employment means much lower formal and informal legal protection in their workplaces, and in addition these individuals fail to have full access to the social security system (e.g. they might not be entitled for paid leaves and/or sick leaves, they might not be covered by health insurance, and might not accumulate pension rights in their countries' state pension systems). Overall, these disadvantages for the employees might even outweight the financial benefits (lower taxes and contributions) that they enjoy from being employed in an illegal way.

Therefore most countries are fighting against the illegal or undeclared employment. One obvious way of this is labour market inspections, in which authorities can discover any attempts for labour market tax avoidance. In recent years, the technological development made it possible to strengthen the effectiveness of these inspections. But besides these, several countries try to provide positive incentives as well by implementing tax systems that also provide less incentives for illegal employment. Empirical estimates from Hungary suggest that lowering marginal and average labour tax rates might lead to decreasing tax avoidance in the labour market. In particular, Bakos et al. (2008) and Kiss and Mosberger (2015) showed that the taxable-income elasticity with respect to the marginal tax rate is negative, especially for high-income earners – which indicates that people do respond to tax decreases, although it is unclear whether they do so by declaring a higher fraction of their total earnings and/or they just simply work more. Further, Benczúr et al. (2014) demonstrated that the number of legally employed also increases when the tax burden decreases. We can also assume that tax avoidance is less common in simpler tax systems, where it is more difficult to hide (some part of) the labour income.

In Hungary, at the beginning of the 2010s there have been numerous tax reforms that pointed into these directions. In 2010-2013, Hungary gradually implemented a flat Personal Income Tax regime, which significantly decreased the marginal tax rate for relatively high earners, and at the same time it also decreased the average tax rate. In parallel, the government continuously decreased the social security contribution rates. The motivation behind the labour market tax reform was to provide incentives for labour market whitening, and to decrease the extent of labour market tax avoidance.

In this paper we use two different data sets between 2001 and 2017 to evaluate the labour market effects of these policy changes. Our central question is whether we can indeed detect positive developments in the declaration of labour market incomes in the 2010s. We first study the aggregate time series related to the labour market, and find that their evolution in the 2010s is consistent with a labour market "whitening process". Then in a next step, we try to find the possible causes of this macro phenomenon, and investigate separately the changes in the number of employed ("extensive margin"), and changes in income underreporting ("intensive margin"). Our finding is that income underreporting of entrepreneurs did decrease significantly in the 2010s, relative to the mid-2000s level. On the other hand, we also find that the observed decline in the incidence of illegal employment around 2010 was only temporary, and probably a result of the recession in aftermath of the financial crisis.

In Section 2 we present a simple framework for studying the labour market whitening process, then we also present the evolution of the relevant aggregate macro time series since 2001. Then in Section 3 we estimate the extent of illegal employment by

[1] Informality is estimated at 90% on average in low income countries, and at 67% and 18% in emerging and developed countries. When we exclude agriculture, where informal employment is most widespread, the percentages are 73%, 59% and 17% in the same three country groups, respectively. (See Bonnet et al. (2019)).

comparing two micro data sets – one which contains illegal employment also, while the other contains only legal employment. In Section 4 we use another micro data set and present an estimate of income underreporting. Section 5 concludes.

2 Labor market tax evasion trends based on aggregate macro data

In this chapter we analyse labor market tax evasion and compliance trends based on a simple model showing the connection between the following available aggregate macro time series: declared gross labor income, total labor income, disposable net income and consumption expenditure. Figure 1 displays a stylized model of connections between these macro series.

The declared gross labor income (Y_g^d) is on the top of the graph, minus taxes equals the (legal) net labor income (Y_n^d): $Y_n^d = Y_g^d(1-t)$, where t contains both personal income tax and employee social security contribution rate. Beside net wages and salaries (Y_n^d), net disposable income also includes the undeclared "illegal"net labor income (Y_n^u) and other net non-labor incomes (Y_n^e). Net disposable income (I_n) is the sum of these:

$$I_n = Y_n^u + Y_n^d + Y_n^e = Y_n^u + Y_g^d(1-t) + Y_n^e = I_L - Y_g^d t + Y_n^e,$$

where I_L equals the sum of undeclared and declared labor income ($I_L = Y_n^u + Y_g^d$). Finally, consumption (C) is a function of net disposable income. Assuming constant aggregated consumption ratio (γ), then $C = \gamma I_n = \gamma(I_L - Y_g^d t + Y_n^e)$, where C is the aggregated household consumption expenditure. Macro statistics usually contains the time series of declared "legal"gross labor income (Y_g^d), aggregated labor income including both declared and undeclared (I_L) and household consumption expenditure (C).

The so-called "whitening"process, when part of the undeclared net labor income is transformed to declared gross income, is represented by an arrow on the top left part of Figure 1. The following process evolves in case if a unit of undeclared income is whitened: $\Delta Y_n^u = -1$, $\Delta Y_g^d = 1$, $\Delta I_L = 0$, hence the undeclared "illegal"labor income (Y_n^u) decreases by a unit, the declared gross labor income (Y_g^d) increases by a unit, while the total labor income (I_L) is unchanged. In case of whitening both disposable income and consumption expenditure decreases assuming unchanged tax rates (t), consumption ratio (γ) and other net non-labor income (Y_n^e) as follows: $\Delta I_n = \Delta I_L + \Delta Y_n^e - t\Delta Y_g^d = -t$ and $\Delta C = \gamma \Delta I_L + \gamma \Delta Y_n^e - \gamma t \Delta Y_g^d = -\gamma t$. The result suggests that if some part of the previously undeclared labor income becomes declared to the tax authorities, then – other factors unchanged – the disposable income drops by the amount of paid taxes and the household consumption also decreases.

We can conclude that in case of the whitening process the following stylized facts are expected: 1) consumption expenditure (C) increases at a lower rate than the aggregate labor income (I_L), 2) aggregate labor income (I_L) increases at a lower rate than the declared gross labor income (Y_g^d). To check these stylized facts empirically, we compare the following aggregated macro time series: declared gross labor income (Y_g^d), aggregated labor income (I_L) and consumption expenditures (C). These comparisons might be suggestive on the extent of labor income tax evasion, and time trends in tax evasion might suggest whether labor market whitening took place recently. First, we present the time series and compare their evolution over time, then we analyze the growth dynamics of the time series.

Figure 2 compares the time series of declared gross labor income (Y_g^d) and aggregated labor income (I_L). We consider the wages and salaries (D.11) in the national accounts estimated by the Hungarian Central Statistical Office (HCSO) as a proxy for the aggregated labor income. The declared labor income is proxied with the employees labor income declared under the Personal Income Tax (PIT) to the Hungarian Tax Authorities. The fundamental difference between the two series is that the HCSO estimates income from the "grey" and "black" economy (e.g. envelope and under-the-table payments, undeclared tips and income from illegal activities) [2] and includes the amount in the national accounts, while the PIT only includes income declared to the Tax Authorities. Computing the difference between the two series can shed light on the grey/black labor income amount estimated by the Central Statistical Office.[3] The difference between the two series, i.e which is the proxy for the estimated

[2] To ensure the exhaustiveness of the Hungarian national accounts the HCSO makes certain adjustments in national accounts data in line with Eurostat's Guidelines. They correct for the income underreporting and also for the cost over-reporting behaviour of small-sized enterprises. They also include an estimate for tips (hairdressing and other beauty services, taxi operation, restaurants, bars), for gratitude money in the health care system, wages in kind, and illegal activities (prostitution, drugs and smuggling). Szabó and Pozsonyi (2011)

undeclared labor income, decreased after 2011, which might be explained by the labor market whitening. However, it can be also observed that the relative difference between the two time series also decreased in 2002 and in the years before 2008, so the decrease after 2010 and the possible whitening are not unique even in the analyzed 18-year period.

Possible explanations behind the decreasing discrepancy between the estimated total and the declared labor income series could include the gradual introduction of the flat PIT rate in 2010-2012 and hence the significant reduction in the top marginal tax rate (MTR). Another explanation could be the continuous increase of the minimum wage, as Tonin (2011) shows that the large-scale minimum wage hike in 2001 increased compliance among high-skilled workers in Hungary. Figure 3 presents also the MTR [4], the percentage change in the minimum and average wage ratio and the estimated share of declared (legal) income to total labor income. However, based on this Figure, it is not unambiguous whether these economic measures caused labor market whitening, although there might indeed be connection between the significant decrease in the top MTR and the gradual increase in the ratio of declared income after 2010.

2.1 COMPARISON OF THE YEARLY CHANGES IN THE AGGREGATE MACRO TIME SERIES

Based on the stylized model presented earlier, in case of a whitening process – when part of the undeclared labor income is declared to the tax authorities, ceteris paribus – the following stylized facts are expected.

1. The declared (legal) labor income changes at a faster rate than the aggregate labor income.
2. The aggregate (gross plus net) labor income changes at a faster rate than both the net disposable income and final consumption expenditure, due to the increased tax compliance and tax burden.

We should emphasise again that these stylized facts can only be expected if other things (e.g. tax rates, consumption and savings rates, the relative importance of non-labor incomes within disposable income) are unchanged. In the period of 2012-2017, tax rates, consumption rates and savings rates are relatively stable when calculated from macro time series; however, the proportion of non-labor income within disposable income shows a gradual decline (from 34% in 2011 to 26% in 2017). This latter might be a consequence of relatively quick real wage growth in the 2010s, with which other types of incomes (and most importantly, inflation-indexed pensions) could not keep up.[5]

Nevertheless, in order to test these “ceteris paribus”hypothesises we compare the yearly growth rate of consumption expenditure (blue bar), aggregate labor income (dark grey bar) and declared labor income (light grey bar) time series on Figure 4. The Figure shows that in seven years of the 17 years analysed – and importantly, in four years of the 6 years between 2012-2017 – the expected stylized facts under a whitening process are realized. Declared labor income generally grew faster than the total labor income calculated by the HCSO in the national accounts, which includes both ”grey” and ”black” incomes; and also household consumption expenditure generally grew more slowly than both types of labor income series.[6]

As we have seen, the evolution of the aggregate time series is in many respects consistent with what we would expect in case of a gradual whitening process of the labour market incomes; but in some cases there are alternative explanations for these

[3] Wages and salaries (D.11) in the national account contains the total remuneration in cash and in kind of employees, without employers’ social security contribution. It includes items such as wages and salaries, allowances, bonus, 13th month wages, exceptional payments to employees leaving the company, tip and gratitude money and in kind benefits. The Tax Authority reports yearly the summary statistics containing the aggregated values of each cell of the PIT tax declaration form. In order to create a comparable declared labor income series we include wages and salaries of employment, employment-related reimbursed expenses, severance payment, and wages and salaries and reimbursed expenses from other non-autonomous activity. Both series include paid personal income tax and employees’ social security contribution, and exclude employers’ social security contribution.

[4] The MTR includes the personal income tax, the employee social security contribution rate, and also the 4 percent solidarity tax for the periods 2007-2009 and is calculated for the top income earners above the pension contribution ceiling.

[5] For the comparison of wages and consumption it is important to keep in mind that consumption data also includes pensioners and non-wage earners, as the consumption series is only calculated at the national level – contrary to income series which only refers to wage and salary earners.

[6] This relatively slow growth of household consumption expenditures might simply be a consequence of increasing saving rates. However, in the early 2010s households underwent a quick de-leveraging process, after which it is reasonable to assume that the savings rates did not increase. We indeed see in macro time series relatively stable saving rates – fluctuating between 6.4% and 8.4%, with an average of 7.3% – in the period of 2011-2017. Another reason of the slow growth of household consumption expenditures might be the relatively slow growth rate of non-labor incomes (e.g. pensions). As discussed earlier, this is indeed the case in the 2010s.

macro stylized facts (e.g. the relatively slow growth rate of non-labor incomes and pensions). So these aggregate time series might be influenced by several other factors, and hence from these observations alone we cannot conclude that the observed phenomena are indeed caused by the whitening process. By comparing the available micro databases, we can get a much more accurate picture of whether we are indeed observing signs of whitening in the labour market in the 2010s.

Figure 1
Aggregate macro data

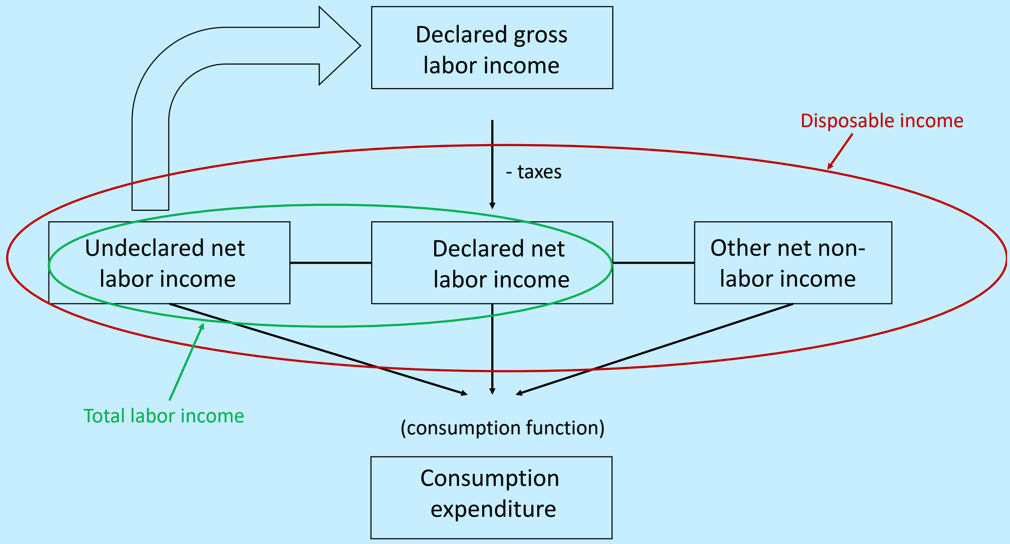
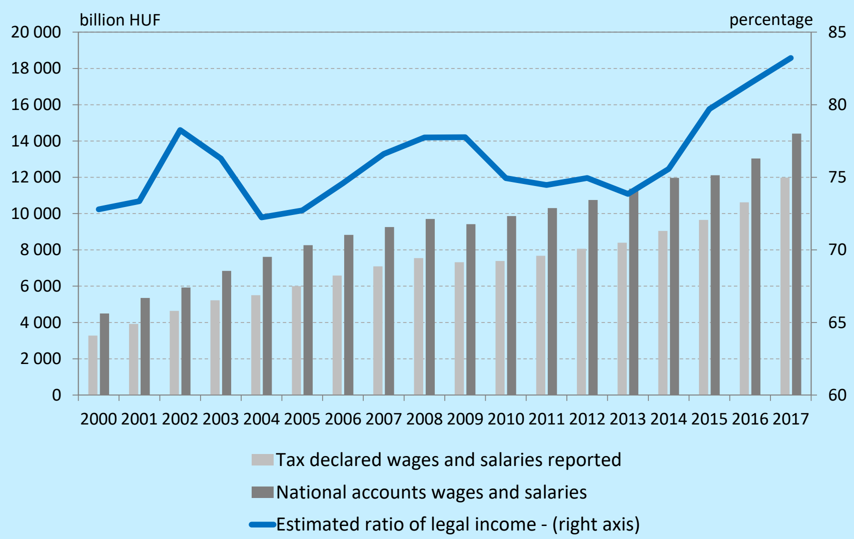
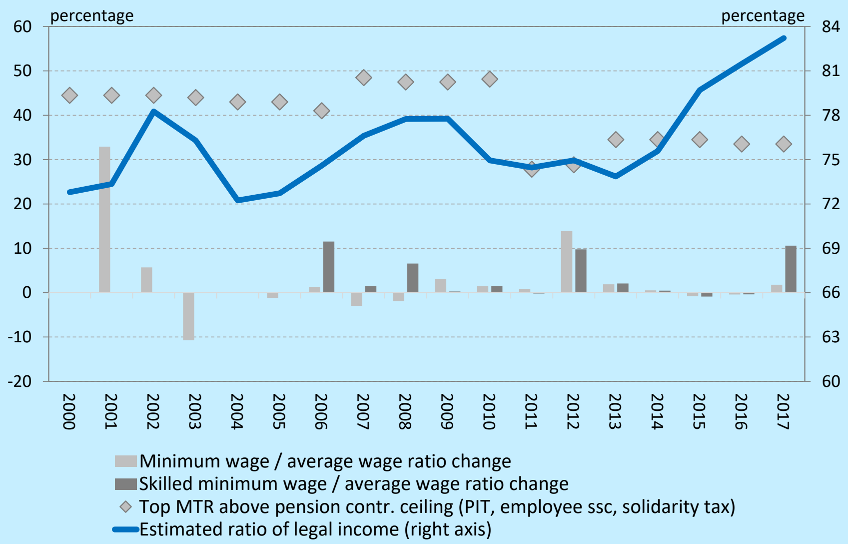


Figure 2
Estimated share of declared labor income



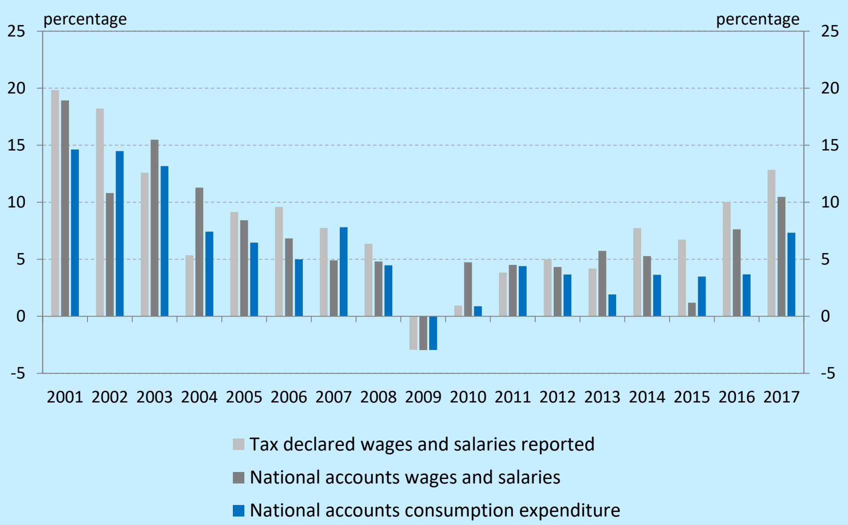
Notes: *Estimated ratio of legal income is calculated as a fraction of the declared wages and salaries reported to the Hungarian Tax Authorities and the national account wages and salaries (D.11). Source: HCSO, MNB.*

Figure 3
Economic measures and the estimated share of declared labor income



Notes: *Estimated ratio of legal income is calculated as a fraction of the declared wages and salaries reported to the Hungarian Tax Authorities and the national account wages and salaries (D.11). Top marginal tax rate for top earners above the pension contribution ceiling includes PIT, employee social security contribution and the 4% solidarity tax in effect between 2007-2009. Source: HCSO, MNB.*

Figure 4
Labor income and consumption expenditure growth rate



Notes: *Estimated ratio of legal income is calculated as a fraction of the declared wages and salaries reported to the Hungarian Tax Authorities and the national account wages and salaries (D.11). Consumption expenditure series include the household consumption expenditure (P.3) in the national accounts, aggregated labor income series include the national account wages and salaries (D.11) and declared labor income series include the employees labor income declared under the PIT. Source: HCSO, MNB.*

3 Estimation of the undeclared workforce

As we discussed earlier, we make a distinction between *illegal employment* (often referred to as ”black employment”) and *partly legal employment* (often referred to as ”grey employment”), and study these two phenomena separately. Undeclared workforce belongs to illegal employment, as people in this category work without any declared work contract, and they do not pay taxes and contributions after their earned income. As a consequence, they will not be covered by health insurance, and they do not accumulate pension rights either. Their activity is completely hidden from the authorities.

In turn, people who are employed partly legally will have some kind of official work contract, and they will declare *some* of their incomes to the authorities. But this form of employment is only partly legal as in this case only some part of their total income is officially declared (”legal” and taxed), while there is another part (the ”illegal” part) which is paid unofficially and directly to the employee. This latter part remains hidden from the authorities and hence will not be subject of any taxes and social security contributions. Partly legally employed people will be covered by the social security system, and they will also accumulate (some) pension rights, based on the legal part of their income. However, by not paying taxes and contributions after at least some part of their income, these partly illegally employed people effectively pay a reduced personal income tax and social security contribution.

In this section we estimate the number of entirely ”illegal” employees, i.e. the number of those people who are working without any type of legal work contract. Of course, for the employers, this kind of employment is much cheaper, as – besides the employee’s taxes and contributions – they also avoid paying the employer’s social security contribution (17-27% of gross income in the 2010s in Hungary). So there is a clear incentive from the employers’ side for this kind of illegal employment. Authorities can fight against this by frequent employment inspections, and with fines if they discover illegal employment.[7]

3.1 METHOD TO ESTIMATE THE NUMBER OF UNDECLARED WORKFORCE

When estimating the number of those employed illegally, we follow the method of Augusztinovics and Köllő (2007) and Elek et al. (2009), who use the discrepancy method and compare the number of employees in two different data sets. The first data set is the Labor Force Survey (LFS) of the Hungarian Central Statistical Office (KSH), which contains all employed individuals irrespective of whether they have a work contract or not.[8] The second data set is the administrative data set of the Hungarian Pension Authority (ONYF), which contains all *declared* work contracts after which the employed individual paid pension contribution.[9]

One crucial assumption behind this method is that the LFS data set indeed contains all employed individuals, and most importantly those who work illegally. If some of the individuals who work illegally fail to report their employment status (maybe because they think that some kind of authority might fine them for this), then we will underestimate the number of those employed illegally. Elek et al. (2009) share this concern and compare four different data sets that try to estimate total employment. The result of this comparison is that the number of employed individuals is the highest in the Labor Force Survey in almost all age categories (the only exception is the 15-24 age category). While this does not exclude the possibility that we still underestimate the total employment – and especially the illegal part of it – in the LFS, this downward bias is probably small.

[7] We will report the results of these inspections in the next subsection.

[8] The Labor Force Survey is an internationally harmonized methodology to estimate the number of employed. In this survey, anybody is regarded to be employed if she worked at least one hour for payment in the preceding week of the survey – irrespectively of whether this was legal or illegal. The survey contains socio-demographic information (such as age, gender, education, settlement), as well as some information about the employer (e.g. location of work). For a more detailed description of the LFS data, see Bak and Szabó (2016).

[9] From now on, we will refer to the first data set as ”LFS”, and or the second data set as ”ONYF”. Since LFS is based on a survey, it contains sampling error: according to the Central Statistical Office, the true employment’s 95% confidence interval contains an uncertainty of around $\pm$18 thousand individuals. In the ONYF data we observe the whole population, so in that case we do not face estimation uncertainty.

Besides this, there are several other caveats that make the direct comparison of the two data sets difficult. In the following paragraphs we list these differences and explain the ways with which we make the two data sets comparable.

1. **Age differences** The LFS surveys all individuals between 15 and 74 years of age, while the administrative data set of the ONYF contains all individuals who pay pension contributions. To make the two data sources comparable, we drop all individuals from the ONYF data who are older than 74 years or younger than 15 years.[10]

2. **Foreign employees** In the administrative data set of ONYF, we have data about foreigners who do not actually live in Hungary and work (in some form of legal employment status) only temporarily in Hungary. As these individuals do not have a permanent address in Hungary, they will never be found by the Labor Force Survey (which only considers individuals with a permanent address in Hungary). Therefore we drop all individuals of the ONYF data who do not have a permanent Hungarian address.[11]

3. **Hungarians working abroad** This is the opposite case relative to the previous one: some Hungarians work in other (neighboring) countries, undertaking daily cross-border commuting. As these individuals live in a domestic household, they will be counted by the LFS; but as they do not have a legal employment contract in Hungary, they will not be present in the administrative data set of ONYF. Luckily, the LFS has a question about the location (county) of employment, and one possible answer is foreign location. Therefore we can drop these individuals who work abroad from the LFS data.

4. **Working without contribution payment** In some periods of 2001-2017, pensioners could work legally with no obligation to pay a pension contribution, which means that they are included in the LFS data, but not in the ONYF admin data on pension contributions. Unlike Elek et al. (2009), who had access to admin data on individuals covered by health insurance and could back out the number of pensioners in legal employment status, we could not precisely identify the number of working pensioners in those periods when they did not have to pay pension contributions. Therefore we decided to drop all pensioners from both data sets. For the LFS, we have an indicator on whether the respondent is beneficiary of old-age pension or not, and we used this variable to drop all old-age pensioners. For the admin data of ONYF, we used a separate data on new pensioners from previous years (which has the same person identifiers as the pension contribution data itself) to select the individuals who were already retired at the time of pension contribution payment.

5. **Definition of employment** The admin data of the ONYF reports the type of employment, based on which the individual is obliged to pay pension contribution and accumulates pension rights. There are around 100 different categories of this, not all of which corresponds to employment in the sense of the LFS. For example, recipients of unemployment or childcare benefits are present in the ONYF admin data (as these beneficiaries accumulate pension rights), but they are not considered as employed in the LFS employment definition. In turn, those who are on sick leave, are also present in the ONYF admin data set, but – as they are only temporarily away from their existing regular job – are considered as employed by the LFS. So from the admin data set, we have to drop unemployment or childcare benefit recipients, but not sick leave recipients – and we have to decide for all possible employment statuses whether we should count them in or not. To harmonize the two different data sets for all possible employment statuses, we used Table A of Appendix A in Elek et al. (2009), who already did this kind of matching.

There were also some discrepancies between the two data sources that we could not correct for. These are mainly related to those individuals who are employed legally (and hence are present in the ONYF admin data set on contribution payments), but remain hidden from the LFS as they do not have a permanent address. These people might live in prisons or other specialized institutions. We think, however, that the relatively low number of these people might not distort significantly our calculations on the proportion of illegally employed individuals. We also think that their number is probably quite stable, so any dynamics in the number and proportion of illegally employed will not be driven by changes in the number of these individuals.

Due to the corrections listed above, we have to emphasize that our results will not be valid for the whole Hungarian economy, but only for the ”common denominator” of the two available data sources. This provides a very large, but not perfect coverage for the national economy.

[10] In the ONYF data, we have information about the individuals' birth years, based on which we can always select those who are in the 15-74 age category.

[11] In the ONYF data, we have the zip code of the permanent address of each individual. Experts of the pension authority have informed us that missing data here indicates that the individual does not have a permanent address in Hungary, and that most of these individuals are foreigners (as opposed to Hungarians without any official address).

We compare these ”harmonized” data sets by calculating the simplest possible statistics from them: *how many* individuals were employed on average in a particular point in time. The LFS contains quarterly data about the number of employed (either ”illegal” or ”legal”) on an average week of the particular quarter. For the admin data, we calculated a very similar number: the number of people with an employment contract (in any legal form, with the LFS-compatible definition of employment) on the 15th day of each month, and then took the average of the monthly numbers to generate quarterly data. We then compared these two quarterly time series, and calculated the number of illegally employed individuals as the difference of the two; while the proportion of illegal employment was calculated as the number of illegally employed, divided by total (LFS-based) employment.

Among the previous studies that used a similar approach to ours, Augusztinovics and Köllő (2007) estimated that out of a total employment of around 3.72 million individuals in 2001, 368 thousand (or slightly less than 10%) were employed illegally. Elek et al. (2009) used updated data sources for the years 2001-2005, and found that total employment was relatively stable at 3.87-3.92 million, of which 630-670 thousand (or 16.3-17.1%) was illegal.

In this paper, we have access to these two data sources for all years between 2001-2017, and we will calculate the number and proportion of those employed illegally for these 17 consecutive years. We note that although some of our data refers to the same period as Elek et al. (2009), our results will be slightly different, for three reasons. The first reason is the different treatment of pensioners in our data sources: as we drop all pensioners that were present in Elek et al. (2009), our absolute numbers can be expected to be lower (by a magnitude of around 100 thousands). The second reason is that pension contribution data is subject to some revision even after many years, as some employment contracts are sometimes declared with a delay of as many as 4-5 years. Because of this, our numbers for legal employment might be larger by around 10-20 thousand individuals especially in 2004-2005. (For this very same reason, our calculations for the years of 2015-2017 might also be less reliable.) And finally, while Elek et al. (2009) used a roughly 5% representative sample of the ONYF’s admin data set, we have access to the full data set.

In the following two subsections we present our results for illegal employment. First we will calculate illegal employment for the whole economy, and analyse its changes during the 17 years of observation period. Then in the second subsection, we show the evolution of illegal employment by gender, age categories and regions; and we will also report the proportion of illegal employment among those employed at firms (e.g. excluding self-entrepreneurs).

3.2 RESULTS: ILLEGAL EMPLOYMENT AT THE NATIONAL LEVEL

Figure 5 shows the quarterly time series of total employment, calculated from the two different data sources, between 2001Q1 and 2017Q4. We report the total employment after the corrections (the steps of which are detailed in the previous subsection) of the two data sets, so the levels of the series can be compared and the difference is a good approximation of the number of those who are employed illegally. The first data set – depicted by the solid blue line – is the Labor Force Survey (”LFS”), and it contains all employed individuals, irrespective of whether their employment is legal or illegal. The second data set – depicted by the dashed orange line – is the administrative data set of the Hungarian pension authority (”ONYF”), and it only contains those persons whose employment form is legal.

If our corrections are correct, then we should always see smaller total numbers in the admin ONYF data set than in the LFS data. Therefore, as a useful first check, we prepared the same figures by genders, age categories and regions,[12] and found in general that the ONYF data employment is almost always smaller in these subgroups than the LFS-based employment. The only exception is the age category of 55-64 old women, for which category we see higher employment in the admin data in some periods in the first half of the 2010s, and for the 65-74 age category we also see slightly higher (a few hundred persons, only occasionally exceeding 1000) ONYF-data-based employment. This probably reflects that we cannot perfectly identify all individuals in the admin data set who are working besides being pensioners,[13] but this correction still eliminates much larger fluctuations in the two data sets, so we decided to keep it.

[12] The most interesting of these will be reported in the next subsection.

[13] We think that this might be the case as women in the 55-64 age category can retire with 40 accumulated service years from 2010, so probably our new pension data is not perfect for those women who used this opportunity for early retirement.

Using these calculations, on Figure 6 we show the yearly average number of illegally employed individuals (bars, left scale), as well as the proportion of illegal employment (solid line, right scale). To convert our quarterly data into yearly frequency, we simply averaged the quarterly absolute numbers and proportions for each calendar year.

Based on Figures 5 and 6, we divide the 2001-2017 period into three sub-periods: the pre-crisis period of 2001-2006, the crisis period of 2007-2012, and the recovery period of 2013-2017.

1. **In June 2006**, as a response to high budget and current account deficits, Hungary implemented an austerity package, so the economy started to stagnate already in 2007 (and events in the autumn of 2008 further deteriorated the economic situation). Therefore, in contrast to other countries, in Hungary the **pre-crisis period** is in 2001-2006. During these years, LFS-based total (legal and illegal) employment was relatively stable and fluctuated between 3.72-3.76 million, while the ONYF admin data-based (legal) employment stood between 3.30-3.33 million. Note that in previous studies, Augusztinovics and Köllő (2007) found 3.72 million employed individuals for 2001, of which they estimated that 3.35 was legally employed; while for the period of 2001-2005 Elek et al. (2009) reported 3.87-3.90 million of total employment and 3.21-3.26 million legally employed persons. But as the initial corrections were not the same in these two previous studies, one should not directly compare the absolute numbers.[14] The resulting number of illegally employed fluctuated between 397-447 thousand, while the proportion of illegally employed in this period was also relatively stable and stood at 10.7-11.9%. These latter figures are a bit larger than in Augusztinovics and Köllő (2007), but lower than in Elek et al. (2009).
2. **During the 2007-2012 crisis** we find that the total number of employed decreases by around 200 thousand persons, while the number of legally employed individuals only drops by around 80 thousands. Therefore the absolute number and the proportion of illegally employed individuals both decreased, and bottomed out in 2010 at around 266 thousands and 7.5%. While the exact reasons behind this decline require further investigations, we assume that one important factor could be that it is probably much cheaper and quicker to lay off illegal employees than legal ones. Another factor could be some kind of composition effect across industries: for example, the crisis hit especially hard the construction sector, where we hypothesise – and labour market inspections always find – a larger-than-average proportion of illegal employees.
3. **The period of 2013-2017** can be labeled as a recovery period, when total employment grew by around 630 thousands (from 3.55 million in 2010 to 4.18 million in 2017), but legal employment only increased by around 430 thousands (from 3.25 million in 2011 to 3.68 million in 2017). We note that the relatively moderate increase in the number of legally employed might partly be due to lags in registering legal employment (which might severely affect the figures of 2015-2017). As a consequence, the number and proportion of illegally employed individuals both increased steadily, and they reached record levels of 497 thousands and 11.9% by 2017.

The main conclusion from these three sub-periods is that *illegal employment is procyclical in Hungary*: it decreases in recessions, and increases during the recovery period. Of course, whether this is a general phenomena for all recessions, or it is just specific to the recession following the financial crisis of 2008-2009, needs further investigation. Nevertheless, we hypothesise that the relative costs of laying off and re-employing illegal and legal employees might be part of the explanation, and industry composition effects might also have played a role in it. More data, and region- and industry-specific analysis might shed light on these questions.

We do two different ”reality checks” for our results. First we compare them with the findings of the official labor market inspections undertaken by the relevant Hungarian labor market authority (ITM (2021)). In this report, the Hungarian Ministry of Innovation and Technology (which is responsible for labor market developments) summarizes the findings of the 2020 inspections related to the labor market, and compares them with the previous years results. Most importantly, they report the number of inspected employees each year between 2011-2020, and the number whom they found to be employed illegally.[15] The dashed green line of Figure 7 shows their results between 2011-2018, which is the relevant period for us.[16] The solid red line on the figure shows our estimated proportions on illegal employment.

[14] Elek et al. (2009) did not drop those who worked while being pension beneficiaries, which can contribute to their higher reported LFS-based employment. On the other hand, their admin data was relatively recent and hence subject of ex post revisions (especially so in 2004-2005), which might explain why we find a larger number of legally employed individuals. Relative to the earlier study of Augusztinovics and Köllő (2007), we have even more methodological differences, so the fact that our figure is similar to theirs is probably due to that the effects of these differences on the reported number of employees happen to cancel each other out.

[15] The relevant table can be found on page 3 of the document. Unfortunately, the document is only available in Hungarian.

[16] According to the report, the proportion of illegal employment further increased in 2019 to 17.5%, and then it dropped back to 15% in 2020. These changes are in line with our earlier hypothesis about the procyclicality of illegal employment.

The two different estimates show a similar picture on the development of illegal employment in the 2010s: during the continuous recovery from the financial crisis, the proportion of those who were employed without official work contracts grew steadily.

As a second check, we compare our findings with the results of the cross-country report on illegal employment of the International Labor Office (ILO (2018)). In this report, the ILO uses a harmonized methodology across countries to estimate the size of informal economy in each country.[17] For European countries, the ILO uses the EU-SILC data set. Unfortunately, we only have an estimate for Hungary for 2016, when the estimated proportion of those employed in an illegal way was 12.2% – a slightly higher number than our estimate for the same year (10.9%), but slightly lower than the estimate of the Hungarian labour market authority (12.7%). For some European countries (e.g. Serbia, Russia), however, there is comparable data over time (ILO (2018), Figure 6 on page 15). These data indicate that – similarly to the Hungarian case – in the aftermath of the financial crisis the proportion of illegally employed was growing steadily (from 12 to 16% in Russia, and from 8 to 16% in Serbia). All in all, these comparisons suggest that our baseline results are close to what other institutions estimate with different methodologies.

At the end of this subsection, we estimate the magnitude of forgone tax revenues due to illegal employment for the most recent year 2017. For this year we estimated a daily average of 497.7 thousand illegal employees. This figure also includes those who claim in the Labour Force Survey that they were employed at least one day on the previous week. Also part of this ”employment” might have been very unofficial (e.g. helping out the neighbour with gardening, or cleaning the house for somebody). Unfortunately, we have no information on their average hidden income, but probably the monthly wage from these activities did not reach the minimum wage. We calculate the forgone tax revenues with the assumption that the average undeclared wage was the minimum wage, and consider this as an upper bound for the true figure. In 2017 the monthly minimum wage was 127,500 forints, after which employees had to pay 33.5% personal income tax and social security contribution, and employers also paid a 22% social security contribution. Therefore the total forgone tax revenues could have equaled 422.6 billion forints, or 1.1% of the GDP in 2017. But as mentioned earlier, this figure is probably an upper bound of the tax avoidance due to illegal employment, and is subject of a significant amount of estimation uncertainty.

3.3 RESULTS: ILLEGAL EMPLOYMENT AT DIFFERENT SUB-GROUPS

Now we turn to the analysis of gender-, age- and region-specific differences in the evolution of illegal employment.

Figures 8 and 9 show the estimated **number and proportion of illegal employment among males and females** separately. According to the figures, illegal employment is more widespread among males: for 2017 we estimate that 13.9% of males and 9.6% of females were employed without work contract. The difference, however, is decreasing over time: in 2001, the proportion of illegally employed males and females were 15.4% and 6.4%, respectively. In terms of absolute numbers, while in 2001 there were 316 and 108 thousands of illegally employed males and females, respectively, by 2017 these numbers stood at 315 and 182 thousands. So all of the increase in absolute numbers can be attributed to females.

Without further investigations, it is hard to determine what causes the relative increase in the illegal employment of females. According to the LFS, in 2017 and relative to 2001, male employment increased by 225 thousands and female employment grew by 207 thousands, i.e. the number of ”extra” employment is around the same for both genders. But according to the ONYF data on declared employment statuses, we see 225 thousand more employed males in 2017 than in 2001, while female employment only increased by 132 thousands.[18] So the absolute number of new employment statuses is around the same for both genders, but the number of new legal employment statuses is much lower for females – and that is while their illegal employment increased in relative terms.

[17] The ILO processes micro data for more than 100 countries that represent more than 90 percent of the world’s employed population aged 15 years old and over. According to the ILO definition, ”for a job held by an employee to be considered as informal, the employment relationship should not be, in law or in practice, subject to national labour legislation, income taxation, social protection or entitlement to certain employment benefits (advance notice of dismissal, severance pay, paid annual or sick leave, etc.)”. (ILO (2018), page 10.)

[18] One concern might be that this relatively moderate increase in female employment is due to the pension regulations implemented from 2011, according to which females with 40 service years can retire irrespective of their age. Around 150 thousand females in the age category of 55-64 have used this opportunity for early retirement between 2011 and 2017. However, according to the cohort distribution of illegally employed females, all the increases in female illegal employment between 2001-2017 came from the 25-54 age categories. Therefore the observed increase in female illegal employment is apparently not a consequence of the newly implemented Female-40 preferential pension rule.

In terms of the distribution of **illegal employment across different age cohorts**, we find the highest proportion among the very young (15-24 years), but this high proportion is at the same time very volatile (Figure 10). At the beginning of the investigated period, in 2002-2003 22-23% of young adults were employed illegally, but this proportion decreased to 7% by 2011. After the recovery period of 2012-2017, the illegal employment proportion of young adults again approached 20%. The volatility is probably due to the relatively low employment at this age; but the especially strong procyclicality also indicates that these young adults are probably among the first employees who are dismissed in a recession, especially if they are employed illegally.

In contrast, we see the smallest proportion of illegal employment among the middle aged (45-54 years, see Figure 11). For this age cohort, the proportion of illegal employment stood between 4-8% until 2014, and only later increased to more than 8%.[19] We note that illegal employment proportion is moderate and relatively less volatile for the age cohorts in between (25-34 and 35-44 age categories): these proportions are between 8-14% for both cohorts and all years, with similar procyclicality that we have also seen in aggregate figures.

As a robustness test, we also did the calculations for the 15-54 age category (Figure 12) , which should not be affected by those old-age pensioners whose contribution payment obligations changed several times between 2001-2017. As noted earlier, in our baseline calculations these pensioners were dropped; therefore our results for the 15-54 age category should be similar to our baseline calculations. This is indeed the case, we estimate the proportion of illegally employed in this category to be around 11-12% in 2001-2005, which then drops to 8.5% in 2008. This proportion then increases gradually in the recovery period of the 2010s to 13.1% in 2017 – a very similar pattern to what we reported earlier for the 15-74 age category.

Next we investigate the **number and proportion of illegal employees when we exclude those who are self-employed**. We hypothesise that illegal employment is more wide-spread among self-employed, so our expectation is that in this sub-group we will find lower illegal employment. According to Figure 13, we indeed see lower illegal employment if we exclude self-employed from the sample. In this subgroup, the proportion of illegal employment fluctuated in the range of 7-9% before the financial crisis, and dropped to as low as 4% by 2011. Then in the recovery period of 2012-2017, it started to increase again gradually, and reached 10% by 2017.[20]

Finally, we investigate **regional patterns in illegal employment**, where the analysis is based on the permanent address of employees.[21] Figure 14 shows the absolute number of illegal employees by regions between 2001-2017. In Hungary, the biggest region is Central Hungary, which contains Budapest and its metropolitan area (see white bars on Figure 14). During the period observed, we see a strong reallocation of illegal employment towards this central region, at the expense of less developed regions in the North East (Northern Hungary, Northern Great Plain). Interestingly, the other two relatively developed region, Western- and Central-Transdanubia did also increase their share in illegal employment. But the increase was most dramatic in the central region: while Budapest and its metropolitan area accounted for less than one third of all illegal employment in 2001, by 2017 it had more than half of all illegal employees. Whether this is just a statistical artifact due to poor data quality on the actual location of working, or is related to the procyclicality of illegal employment, remains to be seen in future waves of the data sets, where better information will be available. Nevertheless, we note that in terms of economic activity there is a similar reallocation after the financial crisis to the one that we see in the proportion of illegal workforce, and this positive correlation between economic activity and illegal employment is consistent with the hypothesis of the procyclicality of illegal employment.

As a further remark, we note that although the proportion of illegal employment is similar in 2001-2005 and 2017, in 2017 the same proportion is achieved at a much higher employment rate. If this increase in employment rate significantly affected the composition of those employed illegally, then the proportion of forgone tax revenue might still be different. In particular, if there are relatively more unskilled workers within those employed, and unskilled workers with lower wages have a higher

[19] Note that in absolute numbers, we are talking about 20-30 thousand ”new” illegal employees, some part of which might disappear with ex post data revisions which might influence exactly these years.

[20] We note again that this last figure might be revised downward, as more employers might register their 2015-2017 employees in the coming years.

[21] The regional estimation of illegal employment might be biased if individuals work at different regions than their official permanent address is. In Hungary, this is of special concern as many people fail to change their permanent address immediately after moving. Also, anecdotal evidence suggests that many people are commuting each week to other regions and only return to their families for the weekend. If there are changes in the extent of these wide-spread phenomena over time, then even the estimated dynamics of illegal employment will be biased. Ideally, we should use information for the actual place of working (as opposed to permanent address), but this information is only available in LFS. Future updates of the ONYF admin data, however, will already contain this information.

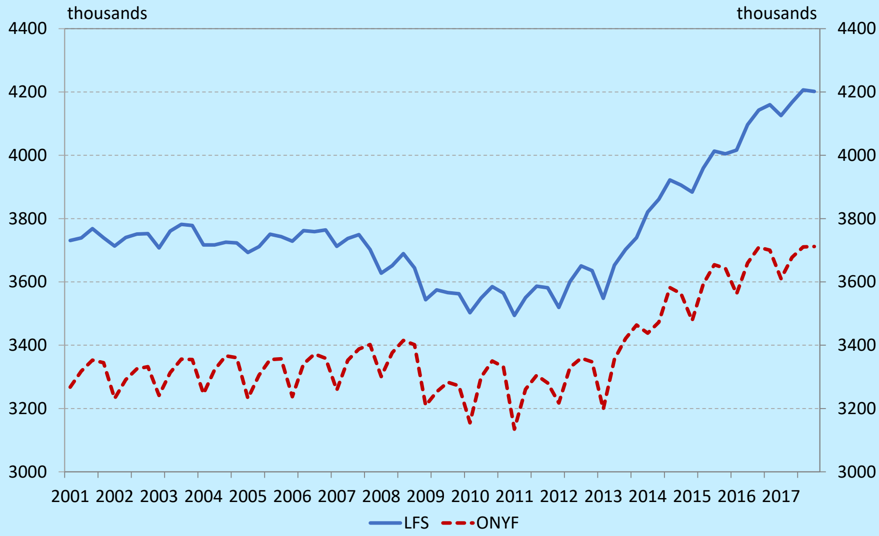
incidence for illegal employment, then the proportion of forgone tax revenue is smaller. Of course, we should have detailed information on the characteristics of illegal employees if we wanted to decide whether this is indeed the case.

We note that our result in this section are similar to those reported by Elek and Köllő (2019), who also find that between 2001-2006 illegal employment was more widespread among the young, among the self-employed, among males, and among those in the central region.

Finally, similarly to Elek and Köllő (2019), it would be very interesting to report illegal employment at the sectoral level, as there are some sectors (e.g. construction, services) where we might expect higher incidence of illegal employment. However, as our data refers to individuals (as opposed to firms or employers), we do not directly observe the sector of employer. From 2011, however, we do observe the occupations of non-self-employed individuals, from which – in certain cases – we can infer the sector in which these individuals work (e.g. waiters and waitresses mostly work in the service sector, and bricklayers mostly work in the construction sector).[22] Based on this sectoral allocation, we find that between 2011-2017, the proportion of illegal employees is between 20-30% in the construction sector, and between 14.5-20.5% in the services sector – both ranges are much higher than the overall range of 5-10% among non-self-employed employees (Figure 13).

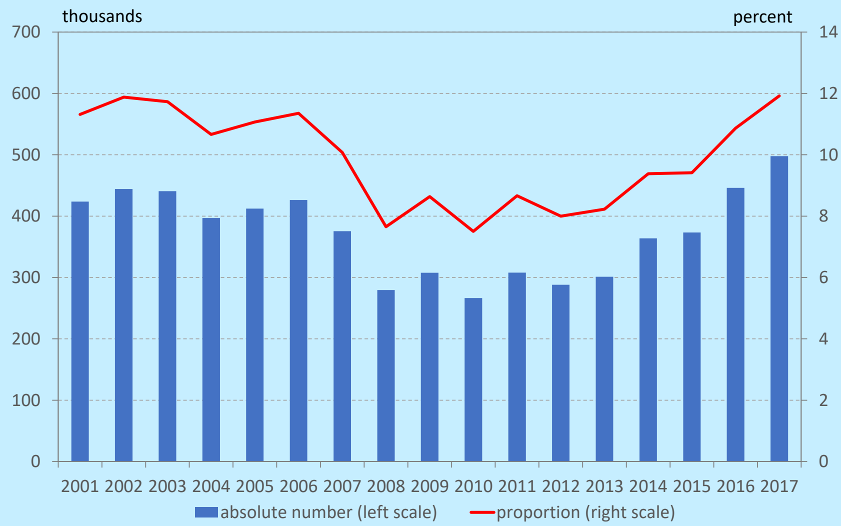
[22] Of course, this sectoral allocation is not perfect: there might be waiters and waitresses working outside the service sector, and also bricklayers working outside the construction sector. Also, in this way we do not observe all employees of the construction sector: we are unable to identify the bookkeeper of a construction firm.

Figure 5
Total employment in the two different data sets



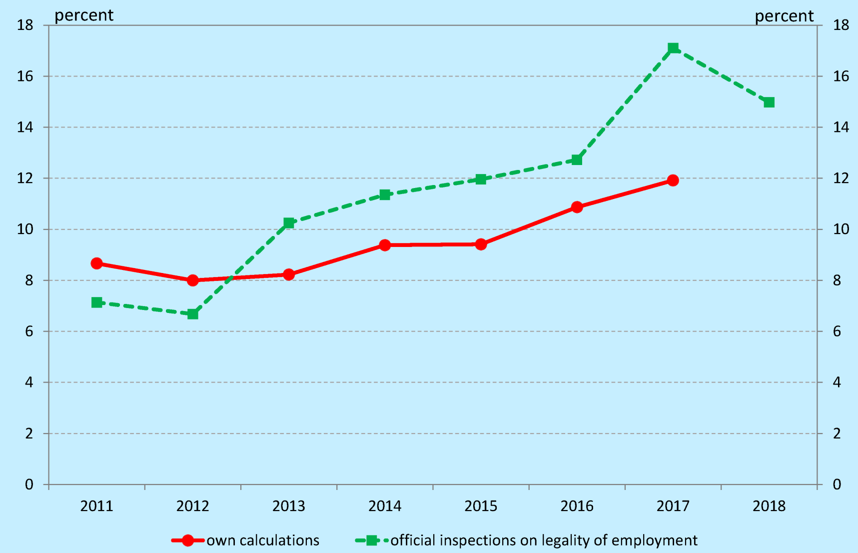
Notes: *The figure shows the quarterly total employment, calculated from the two different harmonized data sets, between 2001-2017. The difference between the number of employed individuals in the two data sets can be regarded as the number of those who are employed illegally.*

Figure 6
Illegal employment between 2001-2017, absolute number and proportion



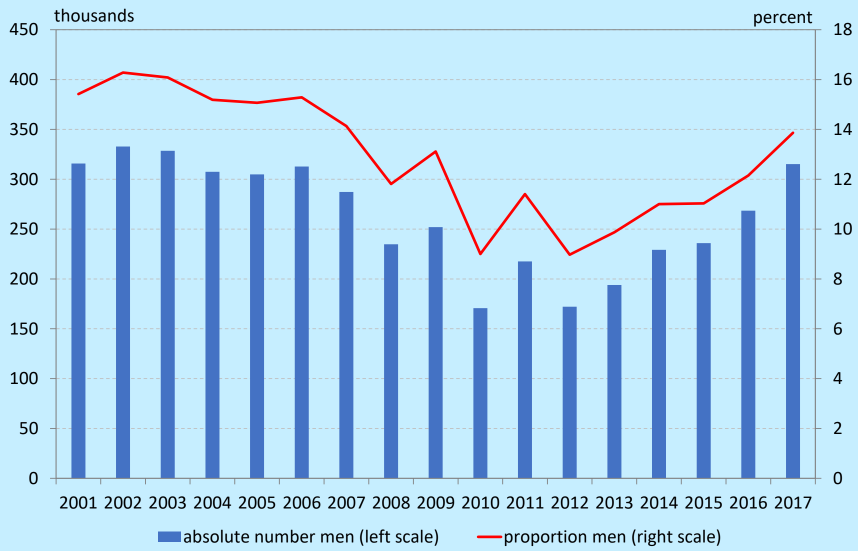
Notes: *The figure shows the yearly average number and proportion of those who are employed illegally, calculated from the two different harmonized data sets between 2001-2017.*

Figure 7
Proportion of illegal employment, two different estimates



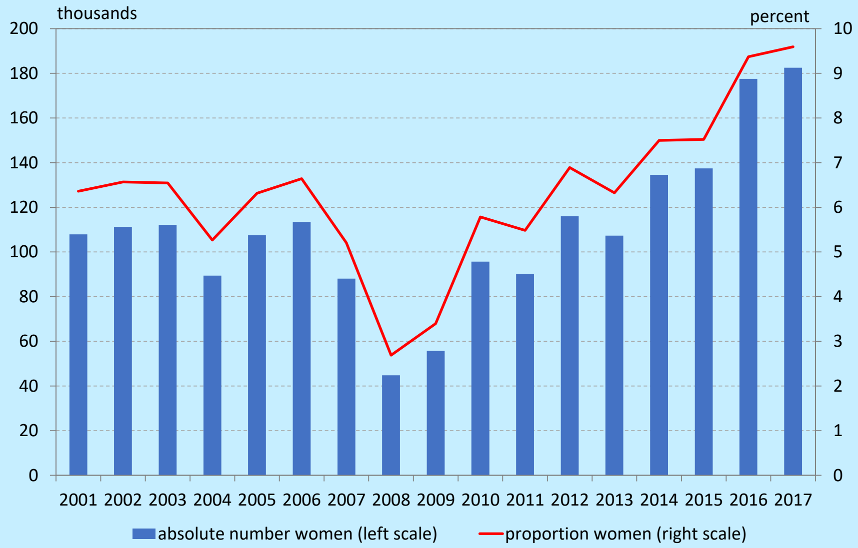
Notes: *The figure shows the yearly proportion of those who are employed illegally between 2011-2018. The solid line with circular markers shows our own calculations, and the dashed line with rectangular markers is taken from labor market inspections.*

Figure 8
Illegal employment of males, 2001-2017



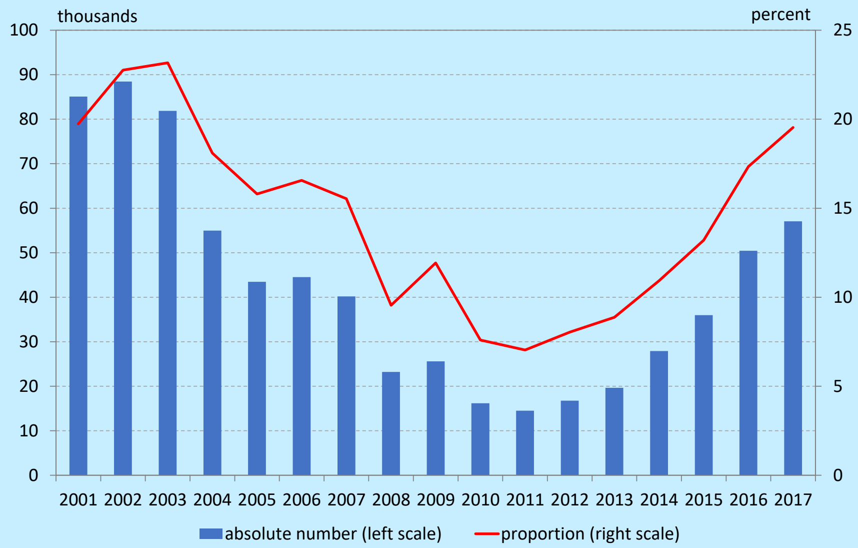
Notes: *The figure shows the yearly average number and proportion of males who are employed illegally, calculated from the two different harmonized data sets between 2001-2017.*

Figure 9
Illegal employment of females, 2001-2017



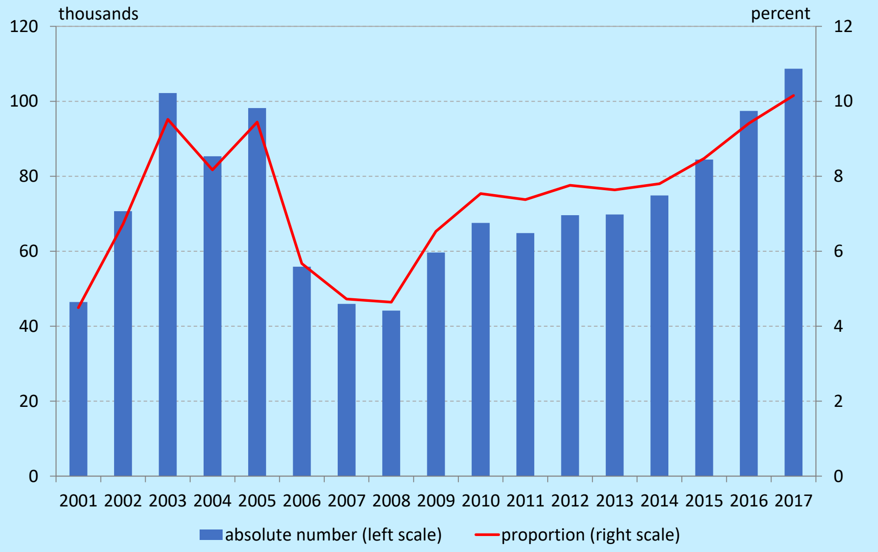
Notes: *The figure shows the yearly average number and proportion of females who are employed illegally, calculated from the two different harmonized data sets between 2001-2017.*

Figure 10
Illegal employment of young adults (15-24), 2001-2017



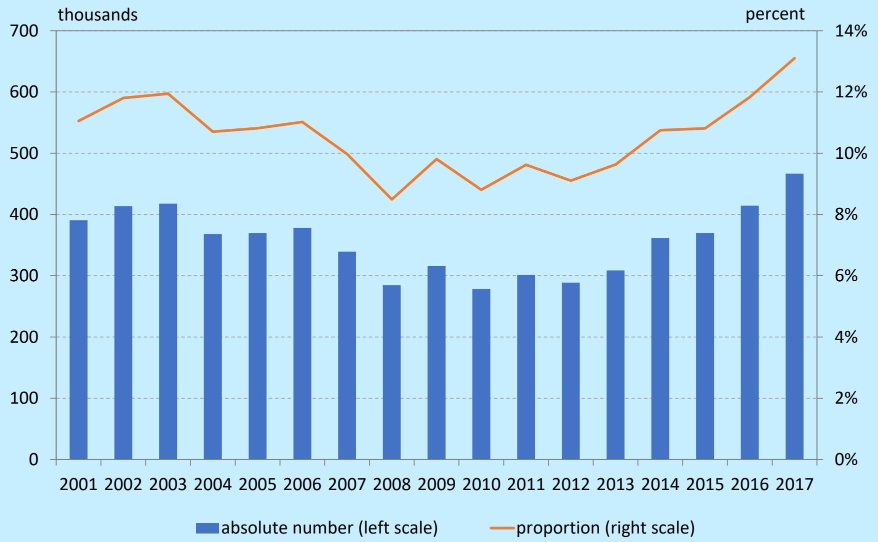
Notes: *The figure shows the yearly average number and proportion of young adults (15-24 years of age) who are employed illegally, calculated from the two different harmonized data sets between 2001-2017.*

Figure 11
Illegal employment of middle-aged adults (45-54), 2001-2017



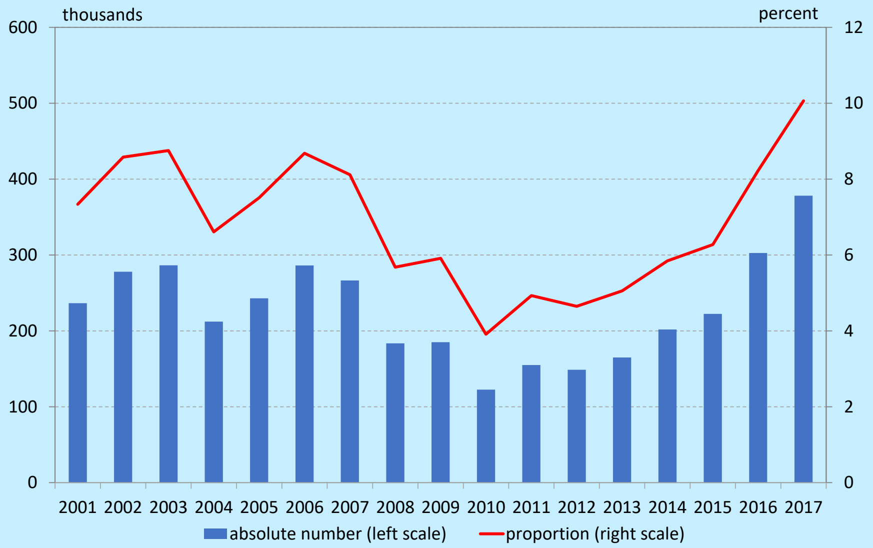
Notes: *The figure shows the yearly average number and proportion of middle aged adults (45-54 years of age) who are employed illegally, calculated from the two different harmonized data sets between 2001-2017.*

Figure 12
Illegal employment of not pensioners age group (15-54), 2001-2017



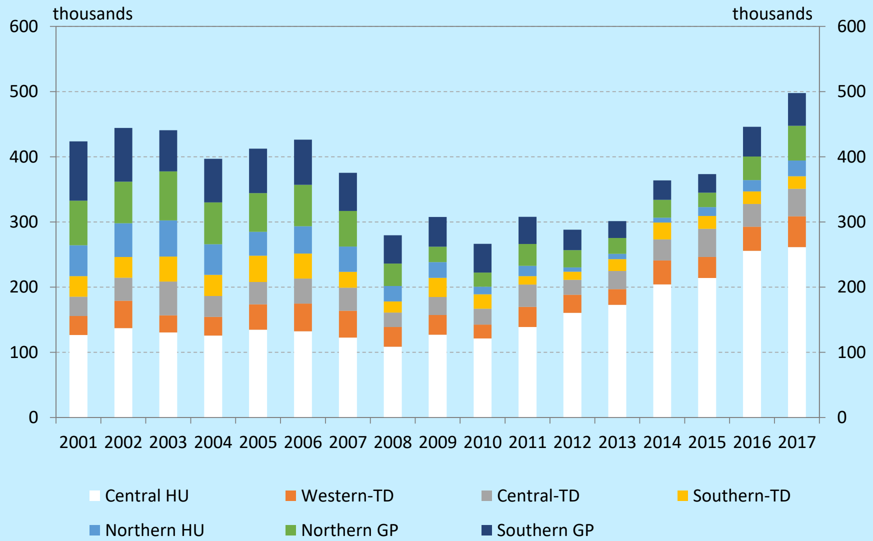
Notes: *The figure shows the yearly average number and proportion of middle aged adults (15-54 years of age) who are employed illegally, calculated from the two different harmonized data sets between 2001-2017.*

Figure 13
Illegal employment between 2001-2017, self-employed excluded



Notes: *The figure shows the yearly average number and proportion of those who are employed illegally, when self-employed are excluded, calculated from the two different harmonized data sets between 2001-2017.*

Figure 14
Illegal employment between 2001-2017 in Hungarian regions



Notes: *The figure shows the yearly average number of those who are employed illegally in the seven different regions of Hungary, calculated from the two different harmonized data sets between 2001-2017.*

4 Entrepreneurs' tax evasion estimation

This section of the paper focuses on estimating the income underreporting by comparing households' observed expenditure and reported income among groups potentially likely to underreport income and not likely. We estimate the income underreporting of entrepreneurs and employees compared to a baseline group, where income underreporting is not prevalent. We include in this baseline group those working in the public administration (administration of the state, defence and judical activities), museums, libraries, primary, secondary or tertiary level education. Their wage is determined by salary grids and have very limited opportunity to under-report income. Entrepreneurs have the opportunity to misreport part of their income due to lack of third-party reporting of their income, also they might claim back personal expenses as business expenses. Employees in Hungary might receive unreported wages as part of their compensation (i.e. "envelopes with cash"), hence we also estimate income underreporting among them. If for a given level of reported income the self-employed, and the employees report systematically larger consumption than the baseline group with no hidden income, then we hypothesise that this extra consumption is financed from unreported income.

For the analysis we use the Hungarian Household Budget Survey (HBS), that is yearly collected by the Hungarian Central Statistical Office (HCSO). The anonymized database includes information on demographic, health, labor market and housing characteristics, detailed household level expenditure data and both individual and household level income of about 7.500-10.000 households each year. The unit of observation for the estimation is the household as the available consumption data is aggregated at the household level.

4.1 EMPIRICAL STRATEGY TO ESTIMATE ENTREPRENEURS' TAX EVASION

The empirical strategy of the paper builds on Gorodnichenko et al. (2009) with slight modifications. According to the permanent income hypothesis consumption depends on expected future long term income (permanent income). If consumption persistently deviates from the reported income everything else held constant, then it might suggest part of the income is unreported. The final regression specification is based on the following equations estimating the relations between actual consumption (C_h), reported (Y_h^R), actual true (Y_h^*) and permanent (Y_h^P) income variables.

Households might report income different from actual income due to various reasons including personal and job characteristics ($\mu X_{1,h}$). $Entr_h$ is an indicator for households with entrepreneurs, while the baseline group contains household with employees. The coefficient of γ shows the average tax compliance difference between these two types of households.

$$lnY_h^R - lnY_h^* = -\gamma Entr_h + \mu X_{1,h} + error \quad (1)$$

Actual income (Y_h^*) might deviate from permanent income (Y_h^P) due to life cycle factors such as age, education, household composition, labor force participation and savings ($\eta X_{2,h}$). Unobservable transitory income shocks are included in the error term.

$$lnY_h^* - lnY_h^P = \eta X_{2,h} + error \quad (2)$$

Non-durable consumption is a fraction of permanent income, that is affected by taste shifters such as number of household members, number of small children, age, education, marital status ($\theta X_{3,h}$).

$$lnC_h - lnY_h^P = \theta X_{3,h} + error \tag{3}$$

While we have data on C_h and Y_h^R in the household surveys, Y_h^P and Y_h^* are not observable. Combining equations 1 - 3 we can still estimate γ and the income underreporting coefficient based on these two observable variables (equation 4). The dependent variable is the consumption-income gap, while γ shows the effect of being an entrepreneur on tax evasion and captures income underreporting. We use log specification as it is less sensitive to outliers.

$$lnC_i - lnY_h^R = \gamma Entr_h + \beta X_h + error \tag{4}$$

In Hungary even employees in the private sector might receive unreported wages as part of their compensation (i.e. "envelopes with cash"). Therefore we consider a baseline group, where income underreporting is not prevalent, which contains those working in the public administration (administration of the state, defence and judical activities), museums, libraries, primary, secondary or tertiary level education.[23] $Entr_h$ is an indicator for households with entrepreneurs, and $Privat_h$ for household with employees working in the private sector.[24] The coefficients of γ_1 and γ_2 show the average income underreporting in this two types of households compared to the baseline group. Interaction of the entrepreneur dummy and time trend, also private sector dummy and time trend are included to estimate possible evolvement in income underreporting.[25] The final regression specification is equation 5.

$$lnC_h - lnY_h^R = \alpha + \gamma_1 Entr_h + \gamma_2 Privat_h + \mu_1 Entr_h trend + \mu_2 Privat_h trend + \delta Year + \beta X_h + \varepsilon_h \tag{5}$$

This regression implicitly assumes that (1) all households report their correct level of consumption expenditure; (2) reported income in the survey is similar to income reported to tax authorities due to fear of anonymity; and (3) the consumption-income function is the same for both the evading and the honest households. The last assumption ensures that observed differences in consumption are not coming from differences in preferences.[26]

We assume that households report same income figures in the Statistical Office survey and in their tax report as they might suspect collusion of the two institutions. However if this assumption is not true and reported income in the survey is closer to real income, so it is higher than income on the tax report, then we underestimate income underreporting based on the survey data and real tax evasion is even greater than the estimated result. To check this assumption in their study Cabral and Gemmell (2018) use a unique database containing matched register incomes declared to the tax administration in New Zealand and income and expenditure data from household survey. They find that income reported in the household survey is indeed higher than income declared to the tax authorities, hence they conclude that survey-based estimates are likely to correspond to a lower bound of true income underreporting.

Data For the analysis we use the Hungarian Household Budget Survey (Háztartási költségvetési és életkörülmény adatfelvétel) collected by the Hungarian Central Statistical Office. The anonymized database provides detailed information on about 7.500-10.000 households each year, such as demographic, health, labor market and housing characteristics, and both individual and household level income. Moreover, it contains detailed household level consumption data as participants record their daily consumption and expenditure in a diary for two weeks (or for a month prior 2015) and also answer to a questionnaire including retrospective questions regarding consumption. The database contains household level representative weights.

[23] TEAOR code 84.1-84.3, 85.1-85.4, 91.

[24] See the definition of the three groups in Table 6 in the Appendix.

[25] Initially, we interacted the entrepreneur and private sector dummys with the year dummies, and obtained yearly estimates for the extent of underreporting. Motivated by the declining time pattern of these estimates, we decided to include these interaction variables with time trends, to check whether the decline in the extent of underreporting is statistically significant or not.

[26] It is not easy to test this last assumption empirically. As a quasi test, we calculated the average of consumption-income ratios by income deciles. These average ratios are similar and hover around 0.8 for the income deciles in the middle of the reported income distribution. As discussed in more details later, we drop the lowest and highest income deciles from our sample, where the consumption-income ratio is significantly higher and smaller, respectively.

The database is a rotational panel, where third of the sample contains new households, while the remaining households participated also in the previous year. Hence, households are present in the sample for three years.[27] Since 2015 it is also possible to fill out the questioners online, but according the Statistical Office less than 10 percent of the survey was submitted online.

4.2 RESULTS: ENTREPRENEURS' TAX EVASION

In this subsection, first we present a graph motivating our analysis, then descriptive statistics of the three groups to be compared, and finally the estimation results of the main regression specification. We motivate our main hypothesis with Figure 15 showing that there is on average a positive consumption-income surplus between the group of entrepreneurs and the baseline group suggesting income underreporting among the prior group. Figure 15 also shows that this consumption-income surplus difference decreased significantly after 2012 suggesting whitening in the labor market.[28]

The descriptive statistics for the three comparable groups is presented in Table 1. The unit of observation for the estimation is the household as the available consumption data is aggregated at the household level. The main estimation sample contains similar composition households and restricted to couples (either married or cohabiting) with children not older than 18 years-old, where the head of household works more than 30 hours per week, no one receives pension income, and the children are not working. We exclude households with zero (missing) income or zero reported consumption data. As in Engström and Holmlund (2009) and Nygård et al. (2019) households with income from agriculture are also excluded because their food purchases pattern might be different. Observations in 2012 are excluded as in the 2012 wave of the Household Budget Survey the net entrepreneur income is equal to the gross, which distorts the dependent variable of consumption - net income gap. We obtain a sample of 12.555 household-year observations for the period of 2008-2017.

The household structure (average number of members, children and small children) is very similar among the three types of households. The head of households employed in the public sector are the most educated, 41% of them have tertiary education, while 33% of the entrepreneur household heads and only 17% of the private sector employed household heads. It seems entrepreneur households have larger and more valuable real estates on average. In order to control for this, wealth proxies are added to the regressions, and also a robustness check is performed where only financially unconstrained households are included in the estimation sample. Consumption-income ratio is larger both in entrepreneur and private sector employee households compared to the baseline reference group of households. Considering similar consumption function, suspicion might arise of income underreporting among private sector employee and entrepreneur households.

The main estimation results of equation 5 are reported in Table 2. The dependent variable is the log consumption-income ratio, while the main explanatory variables are indicators that equal one for households with employees or with entrepreneurs. The baseline group includes households in the public sector, where income underreporting in not prevalent. The sample excludes the bottom and top 10% of per capita income yearly to exclude very poor and rich households who might have different consumption function. Controls in column 1 include the number of household members, number of children younger than 3 years, the education attainment, age, gender, occupational classification code of the head of household, county location and year dummies. In column 2 additional trend variables are added, which are interaction terms of the household type indicator and a count variable for the 10 year period. The main specification is presented in column 3, where additional proxies for wealth are added, such as square meter and value of the real estate, number of rooms, car ownership and real estate ownership type.[29] The Hungarian household budget survey does not contain actual wealth data, hence we could only control for available non-perfect wealth proxies.

The estimated coefficient of income underreporting is significant in all specifications both for households with employees in the private sector and entrepreneurs, while the time trend is only significant for households with entrepreneurs. The income

[27] But the data cannot be linked between years 2012 and 2013 due to technical changes in the anonymization process of the HCSO.

[28] The experts in the Central Statistical Office confirmed that there is no structural change in the HBS data during the analyzed period. Nonetheless, we present the descriptive statistics for the sample for the years before and after 2012 in Table 7 and Table 8 in the Appendix. The two samples are similar apart from that self-employees are slightly older and more educated on average after 2012, and the fraction of them living in Budapest is less, but it does not indicate structural break in the sample. The last line of the Table shows that consumption-income ratio of self-employees dropped by 14 percentage point on average suggesting possible income whitening.

[29] 1. Owner of the household accommodation, without mortgage, 2. owner of the household accommodation with mortgage repayment, 3. tenant who pays rent at a market price, 4. tenant who pays reduced rent, 5. resident of the accommodation because the owner is a relative (who does not live there), or it is part of a remuneration.

correction factor ($corr_Y = e^{\gamma} - 1$) and the average yearly trends are calculated as the exponent of the respective estimated coefficients minus one and are presented in the bottom panel of Table 2. This income correction factor ($corr_Y$) should be interpreted as unreported income relative to *reported* income. Then the share of unreported income relative to *total or true* income can be calculated as $corr_Y/(1 + corr_Y)$. Based on the baseline regression specification in column 3, we can estimate the true (or total) income of households with entrepreneurs in 2008 by multiplying their reported income with 1.284, and this multiplicative income correction factor decreased by 2.9 percentage points yearly on average between 2008-2017.[30] Note that the 28.4% income correction factor in 2008 ($corr_Y$) corresponds to 22% income underreporting ($corr_Y/(1 + corr_Y)$), while the 4.29% income correction factor in 2017 corresponds to 4.1% income underreporting. The estimated income correction factor for households with private sector employees is 4.29% on average in 2008 (this estimate is statistically significant) and this did not change during the analysed period of 2008 and 2017. These results suggest that labor market whitening only took place among entrepreneurs and not in the private sector.

Our results – while they cannot be directly compared to other papers on income underreporting in Hungary due to slightly different estimation methods, data and time frame – are in the same magnitude as alternative estimates. Keresztély and Madari (2021) estimated 15-25 percent income underreporting in 2015 among households in the private sector (including both employees and entrepreneurs) compared to those working in state-owned firms, and also a declining trend in the next years. Benedek and Lelkes (2011) compare the household budget survey and the administrative reported tax records in 2005 and finds that the average income underreporting is about 9-11 percent, with self-employees hiding more income than employees.

Possible explanation behind the income whitening among self-employee households could be the introduction of the small taxpayers' lump sum tax regime ("KATA") in 2013. Under the KATA taxation entrepreneurs only pay a fixed lump sum monthly tax that is independent from their actual income under the yearly income threshold of 6 million forint (12 million from 2017).[31] This income threshold is very generous as it is was about 4.4 times the skilled minimum wage during the period of 2013-2016, and 6 times larger in 2017. The monthly lump sum tax amount is 50.000 forint[32] for full time workers (and 25.000 forint for non full time workers), which substitutes the corporate income tax, personal income tax, social security tax, health insurance, pension, labour market and vocational training contribution liabilities. [33]

Marmoly (2019) presents a short analysis of KATA taxpayers based on the microdatabase of the Hungarian Ministry of Finance. The newly introduced simplified lump sum tax regime was very popular, the number of KATA taxpayers increased from 50 thousands in 2013 to about 200 thousands in 2017. In 2013 at the introduction of the new simplified tax 70 percent of them were previously active self-employee, 25 percent inactive or part time self-employee, 3 percent PIT paying employee, and only a negligible share had other not self-employed income or had no declared income. For 2017 the share of those with previously no declared income reached 10 percent. The author argues that as only 10 percent of them is younger then 24 years, therefore these taxpayers were likely to work illegally before, and due to the introduction of the simplified lump sum tax they started to declare their income to the tax authorities, which is the whitening effect of the KATA taxation regime. For 2017 the share of those previously active self-employees decreased to 45 percent, while those previously employees increased to 30 percent. It would be very interesting to analyse how the declared income of them changed after switching to the KATA taxation, as this channel could also have contributed to labor market whitening.

Back-of-the-envelope calculation: lost tax revenue This subsection contains a back-of-the-envelope calculation about the lost tax revenue due to self-employees income underreporting in year 2016. Because of the lack of administrative micro data on reported entrepreneurial income this estimation is based on many assumptions, hence the magnitude of the budget loss is more relevant than the actual figure. Based on the regression estimation results in Table 2 self-employed households underreported their total income on average by 4.95% in 2016 - assuming a linear time trend. However the total income of these households also included non-entrepreneurial income, specifically in 2016 only 60% of their income was entrepreneurial income. Derived from these two figures entrepreneurial income was underreported by 8.2% (=4.95%/60%) in 2016 on average. In Hungary during this year self-employed could be taxed under different tax regimes, in all of which the tax amount was determined proportionally to the tax base, expect under KATA taxation. The first column of Table 3 reports the annual budget tax revenue

[30] This means that for 2017, the estimated income correction factor – with this linear trend specification – decreased to around 2.3%.

[31] About 20.000 EUR per yearly income (40.000 EUR from 2017).

[32] About 170 EUR lump sum tax per month.

[33] The effective tax rate of households with self-employees (total tax ratio as a share of total income including also non taxable income) decreased remarkably from 32% to 20% after 2013 as presented in Table 7 and 8 in the Appendix.

of the different self-employed tax regimes in 2016. For the first three categories we simply calculate the lost tax revenue as 8.2% of the actual tax revenue from self-employees in 2016 (assuming no behavioral response) reported in column 3.

Under the KATA tax regime a lump sum tax is paid below a yearly income threshold as already described in the 4.2 subsection. Because of the lump sum tax there is no incentive for those much below this threshold to underreport income, the incentive is present only for those just below and above. In 2016 those reporting yearly income above the 6 million HUF threshold had to pay an additional 40% tax on their income exceeding this threshold. Marmoly (2019) presents graphs of reported income density, the 2016 graph shows bunching (excess density mass) at this threshold, which is evidence for income underreporting. We calculate the lost tax revenue as the bunching mass multiplied by the 8.2% estimated income underreporting fraction and the relevant tax rate (40%). All in all, we estimate 22 billion forints forgone tax revenue due to income underreporting for the different self-employment taxation regimes in 2016.

4.3 ROBUSTNESS CHECKS FOR INCOME UNDERREPORTING ESTIMATION

In this section we present different robustness check estimations, and find that in all of them the coefficients of interest is close to the baseline specification. Relative to the baseline 4.3 income correction factor for private sector employees the robustness estimates range between 0 and 7.7, and similarly to the baseline specification there is no significant whitening estimated in either of the robustness checks. While relative to the baseline estimate of 28.4 income correction factor for entrepreneur households, the robustness estimates range between 19.8 and 31.3, while the yearly average labor market whitening factor ranges between 1.9 and 3.3.

Most consumption-income gap estimation papers define non-durable goods expenditure as a proxy for consumption and exclude durable good consumption as it is very infrequent and has a high variance. Other papers such as Cabral and Gemmell (2018) and Tedds (2010) use expenditure on food as a proxy for consumption. They argue that it is more accurately reported as food is a necessity good that is not affected by transitory income shocks and very difficult to postpone its purchases in time, also it is not an expenditure that tax evader households might like to hide such as expensive holidays or car purchases, and finally taste for food is very likely to be similar among households and in time. The results of the estimation where consumption only includes food and beverage expenditure (also eating-out spending and excluding beverages with alcohol content) [34] is presented in column 1 of Table 4. The estimated income correction factor for self-employee households is large and very close to the baseline estimation with all non-durable expenditures, and also has a decreasing trend over time, while the income correction factor for employees in the private sector household is insignificant in this specification.

In column 2 of Table 4 the month of the consumption survey is also included in the regression to control for possible seasonal effects.[35] The estimates are very close to the baseline specification. We also run the regression for different household samples who might have different consumption function that could bias the estimation result. First, we restrict the sample to those households with male heads, then we expand the sample and also include households with only one parent and similarly as before children not older than 18 years. The result for the first sample is presented in column 3, and for the second in column 4, suggesting these sample selections do not distort the estimates.

We also run a regression specification with a stricter definition for the three compared household groups. We restrict the actual sample to only those households where all adults are either self-employees, or all are public sector employees or all are private sector employees. The sample size was decreased by about 1.800 households, but the estimated correction factor for self-employees income underreporting and its negative trend over time is the same as in the original sample, while the income corerction factor for private sector employees became insignificant as presented in column 5 of Table 4.

Cabral et al. (2019) argue that entrepreneurs might have more volatile income than employees and hence could be on a different level of financial stability. This way entrepreneurs might fund their current expenditure not only by current income, but also by past savings, which could lead to higher consumption-income gap estimates even in case of no income underreporting. To exclude this possibility we rerun the regression for a sample containing not financially constrained households.[36] If the

[34] Classification of Individual Consumption According to Purpose - COICOP 01, COICOP 11.1

[35] The month of the consumption survey ranges between March and July.

[36] They answered yes to the survey question whether they could pay for an unforseen expenditure.

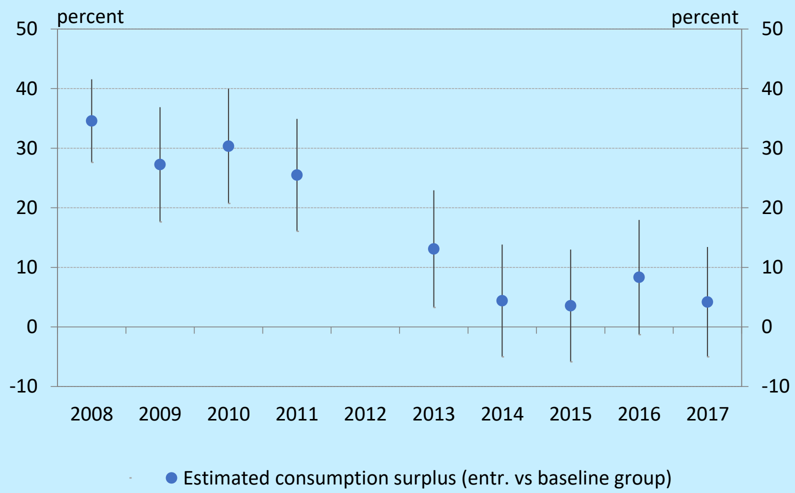
estimated income correction factors were much smaller in this restricted sample than in the main regression specification then it would suggest that entrepreneurs in the full sample do finance their expenditure from past savings, which biases the estimated income underreporting upward. The result of this restricted sample is presented in column 6, which are close to the main estimates indicating the discrepancy in the expenditure-income gap is due to income underreporting.

In the main estimation specification we assumed the same consumption function for all households independent of their income level (though we excluded the bottom and top 10 percent per capita income households). Table 5 contains different robustness estimates to check whether this assumption might bias the result. In column 1 top and bottom 20 income percentile are excluded, while in column 2 and 3 the regression is estimated separately for those below and above median income and finally in column 4 income decile dummies are included in the regression. The results presented are very similar to the main specification, which suggests that different consumption function by income level do not distort the estimates.

The consumption function might not be the same for people with different skills and occupations. Pissarides and Weber (1989) run the regression separately for two broad occupational groups, "white-collar" and "blue-collar" workers. We re-estimate the regression separately for highly skilled professionals and manual workers, and results presented in column 5 and 6 in Table 5 respectively.[37] Based on the estimation results income underreporting is slightly more prevalent among highly skilled professionals in Hungary.

[37] Highly skilled includes those with FEOR 1 (managers) and 2 (professionals) and manual workers includes FEOR 6 (agricultural and forestry occupations), 7 (industry and construction industry occupations), 8 (machine operators, assembly workers, drivers of vehicles), 9 (occupations not requiring qualifications).

Figure 15
Entrepreneurs' consumption-income surplus relative to the baseline group



Notes: *Dots represent the estimated consumption-income surplus of entrepreneurs relative to the baseline group of selected state employees. We estimate these surpluses from yearly cross-sectional regression, with a similar specification to equation 5. Bars represent standard errors. Observations in 2012 are excluded as in the 2012 wave of the Household Budget Survey the net entrepreneur income is equal to the gross, which distorts the dependent variable of consumption - net income gap.*

Table 1
Household descriptive statistics

	Baseline group	Employees	Entrepreneurs
Nbr. of HH members	3.1	3.1	3.0
Nbr. of children	1.1	1.1	1
Nbr. of children (<3 years)	0.15	0.19	0.14
Head of HH age	44.2	42.8	46.4
Primary education	7.4%	11.5%	1.8%
Secondary education	51.3%	71%	65.6%
Tertiary education	41.3%	17.4%	32.6%
Nbr. of cars	1.1	1.1	1.2
Nbr. of rooms	2.9	2.7	3.3
Real estate sqm	82.3	78.3	96.7
Real estate value (m HUF)	11.1	10.5	16.4
Food consumption (HUF)	796,390	728,207	816,658
Total consumption (HUF)	2,904,284	2,604,571	3,257,753
Net income (HUF)	3,989,505	3,445,467	3,912,221
Consumption-income ratio	72.8%	75.6%	83.3%
Number of observations	3,479	7,983	1,093

Notes: *Descriptive statistics represent mean values for the period 2008-2017 (excluding 2012). Monetary values are deflated to 2008 real forint. Sample contains households with two adults, and children younger than 19 years, where at least one adult is working more than 30 hours, none of the children are working and they have no pension, nor agriculture income. Total consumption includes non-durable goods.*

Table 2
Estimation of income underreporting

	(1)	(2)	(3)
D_empl	0.029***	0.036**	0.042***
	(0.008)	(0.015)	(0.015)
D_entr	0.139***	0.285***	0.250***
	(0.011)	(0.021)	(0.021)
Trend_empl		-0.001	-0.001
		(0.003)	(0.003)
Trend_entr		-0.031***	-0.029***
		(0.004)	(0.004)
Time trend	No	Yes	Yes
Wealth proxy	No	No	Yes
R-squared	0.078	0.087	0.126
Number of observations	9,923	9,923	8,667
Income correction factor (priv.)	2.94%	3.67%	4.29%
Income correction factor (entr.)	14.9%	33.0%	28.4%
Avg. yearly trend (entr.)		-3.1%	-2.9%

Notes: *Standard errors are shown in parentheses. *** = significant at 1-percent level; ** = significant at 5-percent level; * = significant at 10-percent level. This table contains the estimated average income gaps for households with employees or with entrepreneurs compared to the baseline group. Dependent variable is the log consumption-income ratio. The main explanatory variables are indicators that equal one for households with employees or with entrepreneurs. The baseline group includes households in the public administration (administration of the state, defence and judical activities), museums, libraries, primary, secondary or tertiary level education. The trend variable is an interaction term of the respective previous indicator and a count variable for the 10 year period. Controls in column 1 and 2 include the number of household members, number of children younger than 3, the education attainment, age, gender, occupational classification code of the head of household, and county location and year dummies. In column 3, additionally proxies for wealth are added, such as square meter and value of the real estate, number of rooms, car ownership and real estate ownership type. Top and bottom income deciles are excluded. The income correction factor and the avg. yearly trends are calculated as the exponent of the respective estimated coefficients minus one.*

Table 3
Lost tax revenue of self-employees in 2016 - back-of-the-envelope calculation

	Tax revenue (bn HUF)	Lost tax revenue (bn HUF)
SZJA	139	11
EVA	81	7
KIVA	14	1
KATA	70	5
Total	303	22

Data source: *MNB, own calculation.*

Table 4
Estimation of income underreporting by various income groups

	(1)	(2)	(3)	(4)	(5)	(6)
D_empl	0.027	0.045**	0.045***	0.036***	0.001	0.074***
	(0.023)	(0.019)	(0.016)	(0.014)	(0.022)	(0.027)
D_entr	0.220***	0.244***	0.252***	0.272***	0.251***	0.237***
	(0.032)	(0.027)	(0.022)	(0.019)	(0.031)	(0.033)
Trend_empl	0.002	-0.001	-0.000	0.002	0.003	-0.004
	(0.004)	(0.003)	(0.003)	(0.002)	(0.004)	(0.004)
Trend_entr	-0.034***	-0.029***	-0.029***	-0.028***	-0.029***	-0.023***
	(0.006)	(0.004)	(0.004)	(0.003)	(0.005)	(0.005)
Interview months	No	Yes	No	No	No	No
R-squared	0.108	0.124	0.123	0.183	0.137	0.109
Number of observations	8,667	7,639	7,986	12,491	6,846	3,418
Inc. corr. factor (priv.)	2.7%	4.6%	4.6%	3.67%	0%	7.68%
Inc. corr. factor (entr.)	24.6%	27.7%	28.7%	31.3%	28.5%	26.7%
Avg. yearly trend (entr.)	-3.3%	-2.9%	-2.9%	-2.8%	-2.9%	-2.3%

Notes: *Standard errors are shown in parentheses. *** = significant at 1-percent level; ** = significant at 5-percent level; * = significant at 10-percent level. This table contains the estimated income gaps between households in the state, private and entrepreneur sectors. Dependent variable is the log consumption-income ratio. The baseline group includes households in the public administration (administration of the state, defence and judical activities), museums, libraries, primary, secondary or tertiary level education. The trend variable is an interaction term of the respective previous indicator and a count variable for the 10 year period. Controls include the number of household members, number of children younger than 3, the education attainment, age, gender, occupational classification code of the head of household, and county location and year dummies, square meter and value of the real estate, number of rooms, car ownership and real estate ownership type. Top and bottom per capita income deciles are excluded. In column 1 the dependent variable is the log food expenditure-income gap. In column 2 the month of the consumption survey is included (data for 2008 and 2012 is not available). In column 3 only households with male head are included. In column 4 the sample also includes households with one adult, and children under 19. In column 5 the sample includes households with all adults working as self-employee, or all as a public sector employee or all in the private sector. In column 6 the sample contains not financially constrained households (they could pay for an unforseen expenditure). The income correction factor and the avg. yearly trends are calculated as the exponent of the respective estimated coefficients minus one.*

Table 5
Estimation of income underreporting in various samples

	(1)	(2)	(3)	(4)	(5)	(6)
D_empl	0.026	0.029	0.013	0.010	0.063*	0.043*
	(0.016)	(0.023)	(0.019)	(0.014)	(0.035)	(0.023)
D_entr	0.216***	0.210***	0.205***	0.181***	0.243***	0.217***
	(0.023)	(0.031)	(0.028)	(0.020)	(0.044)	(0.034)
Trend_empl	0.002	-0.002	0.004	0.003	-0.003	-0.003
	(0.003)	(0.004)	(0.003)	(0.002)	(0.006)	(0.004)
Trend_entr	-0.020***	-0.023***	-0.022***	-0.019***	-0.027***	-0.025***
	(0.004)	(0.005)	(0.005)	(0.004)	(0.008)	(0.006)
Income deciles	No	No	No	Yes	No	No
R-squared	0.186	0.204	0.157	0.242	0.168	0.123
Number of observations	6,578	4,419	4,248	8,667	1,697	4,491
Inc. corr. factor (priv.)	2.6%	2.9%	1.3%	1.0%	6.5%	4.4%
Inc. corr. factor (entr.)	24.1%	23.4%	22.8%	19.8%	27.5%	24.3%
Avg. yearly trend (entr.)	-2.0%	-2.3%	-2.2%	-1.9%	-2.7%	-2.5%

Notes: *Standard errors are shown in parentheses. *** = significant at 1-percent level; ** = significant at 5-percent level; * = significant at 10-percent level. This table contains the estimated income gaps between households in the state, private and entrepreneur sectors. Dependent variable is the log consumption-income ratio. The baseline group includes households in the public administration (administration of the state, defence and judical activities), museums, libraries, primary, secondary or tertiary level education. The trend variable is an interaction term of the respective previous indicator and a count variable for the 10 year period. Controls include the number of household members, number of children younger than 3, the education attainment, age, gender, occupational classification code of the head of household, and county location and year dummies, square meter and value of the real estate, number of rooms, car ownership and real estate ownership type. Top and bottom per capita income deciles are excluded.In column 1 top and bottom 20 income percentile are excluded. In column 2 and 3 the regression is estimated separately for those below and above median income. In column 4 income decile dummies are included. In column 5 and 6 the regression is estimated separately for highly skilled professionals and manual workers respectively. The income correction factor and the avg. yearly trends are calculated as the exponent of the respective estimated coefficients minus one.*

5 Conclusion

In this paper, we have investigated tax evasion in the labour market in the 2010s in Hungary, following major labour market tax reforms at the beginning of the decade. By implementing a flat personal income tax reform after 2012, the Hungarian government significantly decreased the marginal tax rate of high-income workers, while the average tax rate also declined. At the same time, the tax system became much simpler. All of these changes can contribute to a decline in income non-reporting or income under-reporting.

First we showed that the evolution of aggregate time series is broadly consistent with a whitening process. In the 2010s, the growth rate of declared income – taken from personal income tax declarations – is generally larger than the growth rate of incomes in the national accounts (which contains both declared and undeclared income). Moreover, the growth rate of the two alternative income series generally exceeds the growth rate of the consumption expenditures, which also indicates that the importance of hidden income, as a source of financing consumption, might have decreased.

While these results are all consistent with a decline in income under-reporting, they do not exclude the possibility of alternative explanations. Therefore, we use several micro data sets to investigate separately the evolution of income under-reporting at the intensive margin (when part of labour income remains unreported for those who have legal work contracts) and at the extensive margin (when employees remain completely unreported).

For the income underreporting of legally employed workers, we use the Hungarian Household Budget Survey (HBS) to compare the consumption patterns of self-employed entrepreneurs (whom we suspect might hide part of their income) and specific groups of public sector employees (who are unlikely to hide their income). We find that the extra consumption of entrepreneurs, which was quite significant in 2008-2010, gradually decreased over time. We interpret this as evidence that income under-reporting of entrepreneurs did decline gradually in recent years.

For the estimation of the number of illegally employed workers, we use another two micro data sets: the Hungarian Labour Force Survey (LFS), which contains both legal and illegal employees, and the admin data of the Hungarian Pension Authority (ONYF) that keeps track of all individuals who have any kind of legal employment status. From this comparison we find that although the proportion of illegal employment declined significantly by 2010-2012 (relative to its level until 2006), in the recovery period of 2012-2017 this proportion gradually increased again, and reached a level (around 12%) that is comparable to the pre-crisis period. As we have discussed, this pattern of illegal employment is consistent with the findings of the labour market inspections of the Hungarian authorities, and also with the patterns that some other European countries were experiencing during the recession in the aftermath of the global financial crisis. Therefore we hypothesise that the observed dynamics in the proportion of illegal employment is rather due to the procyclical nature of illegal employment, and is not a consequence of a whitening process at this margin. A thorough investigation of this hypothesis might be subject of future research.

References

Augusztinovics, M. and J. Köllő (2007), 'Munkapiaci pálya és nyugdíj, 1970-2020', *Közgazdasági Szemle*, vol. LIV no. 6, pp. 529–559

Bak, D. and L. Szabó (2016), 'Munkaerőpiac (Labor market)', *MNB Oktatási füzetek*, no. 6, pp. 1–37

Bakos, P., P. Benczúr and D. Benedek (2008), 'The elasticity of taxable income: Estimates and flat tax predictions using the Hungarian tax changes of 2005', *Magyar Nemzeti Bank Working Papers*, no. 7/2008, pp. 1–32

Benczúr, P., G. Kátay, A. Kiss and O. Rácz (2014), 'Income taxation, transfers, and labour supply at the extensive margin', *Banque de France Working Papers*, no. 487, pp. 1–28

Benedek, D. and O. Lelkes (2011), 'The Distributional Implications of Income Under-Reporting in Hungary', *Fiscal Studies*, vol. 32 no. 4, pp. 539–560

Bonnet, F., J. Vanek and M. A. Chen (2019), 'Women and Men in the Informal Economy: A Statistical Brief.' *Women in Informal Employment: Globalizing and Organizing (WIEGO) and International Labour Office (ILO)*,

Cabral, A. C. G. and N. Gemmell (2018), 'Estimating Self-Employment Income-Gaps from Register and Survey Data: Evidence for New Zealand', *Victoria University of Wellington Working Paper Series*, vol. 7 no. 7/2018, pp. 1–45

Cabral, A. C. G., C. Kotsogiannis and G. Myles (2019), 'Self-Employment Income Gap in Great Britain: How Much and Who?', *CESifo Economic Studies*, vol. 65 no. 1, pp. 84–107

Elek, P. and J. Köllő (2019), 'Eliciting permanent and transitory undeclared work from matched administrative and survey data', *Empirica*, vol. 46, pp. 547–576

Elek, P., A. Scharle, B. Szabó and P. A. Szabó (2009), 'A bérekhez kapcsolódó adóeltitkolás Magyarországon', *Közpénzügyi füzetek*, no. 23, pp. 1–60

Engström, P. and B. Holmlund (2009), 'Tax evasion and self-employment in a high-tax country: evidence from Sweden', *Applied Economics*, vol. 41 no. 19, pp. 2419–2430

Gorodnichenko, Y., J. Martinez-Vazquez and K. Sabirianova Peter (2009), 'Myth and Reality of Flat Tax Reform: Micro Estimates of Tax Evasion Response and Welfare Effects in Russia', *Journal of Political Economy*, vol. 117 no. 3, pp. 504–554

ILO, I. L. O. (2018), 'Women and Men in the Informal Economy: A Statistical Picture. 3rd Edition', *International Labour Office*,

ITM, I. T. M. (2021), 'Repont on the results of the 2020 labor market inspections', *Hungarian Ministry for Innovation and Technology*,

Keresztély, T. and Z. Madari (2021), 'A háztartások eltitkolt jövedelmének becslése inverz keresleti függvény segítségével', *Közgazdasági Szemle*, vol. LXVIII no. 7, pp. 736–752

Kiss, A. and P. Mosberger (2015), 'The elasticity of taxable income of high earners: evidence from Hungary', *Empirical Economics*, vol. 48 no. 2, pp. 883–908

Marmoly, N. (2019), 'Kinek jött be a kata?', *Számvitel, adó, könyvvizsgálat*, vol. 61 no. 11, pp. 11–13

Nygård, O. E., J. Slemrod and T. O. Thoresen (2019), 'Distributional Implications of Joint Tax Evasion', *Economic Journal*, vol. 129 no. 620, pp. 1894–1923

Pissarides, C. A. and G. Weber (1989), 'An expenditure-based estimate of Britain's black economy', *Journal of Public Economics*, vol. 39 no. 1, pp. 17–32

Szabó, P. and P. Pozsonyi (2011), 'GNI Inventory of Hungary Version 2.2', *Hungarian Central Statistical Office*,

Tedds, L. (2010), 'Estimating the income reporting function for the self-employed', *Empirical Economics*, vol. 38 no. 3, pp. 669–687

Tonin, M. (2011), 'Minimum wage and tax evasion: Theory and evidence', *Journal of Public Economics*, vol. 95 no. 11, pp. 1635–1651

Appendix A

Table 6
Variable definition

Household types	
Baseline group (public sector)	At least one adult works at public administration (administration of the state, defence and judical activities), education, library, museums based on NACE codes, and no one is entrepreneur.
Entrepreneur	At least one adult is an entrepreneur in the household (self-reported at the HBS), and no one works at the public sector.
Private sector	No one works at the public sector, and no one is an entrepreneur.
Dependent variables	
Consumption expenditure	Total non-durable purchased consumption.
Food consumption expenditure	Food and non-alcoholic beverage, also including eating-out expenditure. COICOP (Classification of Individual Consumption According to Purpose) 01, 11.1
Explanatory variables	
Net income	Disposable income including wage, entrepreneur and capital income, social transfers and maternity benefits net of taxes and negative transfers.
Industry sectors	One digit FEOR code (Hungarian Standard Classification of Occupation).

Table 7
Household descriptive statistics, 2007-2011

	Baseline group	Employees	Entrepreneurs
Nbr. of HH members	3.0	3.2	3.1
Nbr. of children	1.0	1.1	1.1
Nbr. of children (<3 years)	0.13	0.21	0.14
Head of HH age	44	42	45
Primary education	7%	12%	2%
Secondary education	51%	71%	67%
Tertiary education	42%	16%	32%
Nbr. of cars	1.1	1.1	1.2
Nbr. of rooms	2.8	2.7	3.3
Real estate sqm	80.9	77.4	97.9
Real estate value (m HUF)	12.4	11.5	18.7
Budapest	15%	16%	28%
Food consumption (HUF)	668,069	594,908	711,509
Total consumption (HUF)	2,511,607	2,199,334	2,908,976
Gross income (HUF)	4,886,456	4,002,999	4,689,810
Net income (HUF)	3,600,672	3,046,927	3,208,762
Effective tax rate	26%	24%	32%
Consumption-income ratio	70%	72%	91%
Number of observations	1,772	4,386	608

Notes: *Descriptive statistics represent mean values for the period 2008-2011. Monetary values are deflated to 2008 real forint. Sample contains households with two adults, and children younger than 19 years, where at least one adult is working more than 30 hours, none of the children are working and they have no pension, nor agriculture income. Total consumption includes non-durable goods.*

Table 8
Household descriptive statistics, 2013-2017

	Baseline group	Employees	Entrepreneurs
Nbr. of HH members	3.1	3.1	3.0
Nbr. of children	1.1	1.1	1.0
Nbr. of children (<3 years)	0.18	0.16	0.14
Head of HH age	45	44	48
Primary education	8%	11%	2%
Secondary education	51%	71%	64%
Tertiary education	40%	19%	34%
Nbr. of cars	1.1	1.1	1.2
Nbr. of rooms	3.0	2.8	3.2
Real estate sqm	83.8	79.3	95.1
Real estate value (m HUF)	9.9	9.2	13.5
Budapest	12%	14%	18%
Food consumption (HUF)	922,156	890,746	948,474
Total consumption (HUF)	3,289,138	3,098,697	3,694,982
Gross income (HUF)	6,039,869	5,355,376	5,959,851
Net income (HUF)	4,370,593	3,931,426	4,794,083
Effective tax rate	28%	27%	20%
Consumption-income ratio	75%	79%	77%
Number of observations	1,757	3,597	485

Notes: *Descriptive statistics represent mean values for the period 2013-2017. Monetary values are deflated to 2008 real forint. Sample contains households with two adults, and children younger than 19 years, where at least one adult is working more than 30 hours, none of the children are working and they have no pension, nor agriculture income. Total consumption includes non-durable goods.*

MNB Working Papers 2022/4

Income tax evasion estimation in Hungary

Budapest, May 2022

mnb.hu

1013 BUDAPEST, KRISZTINA KÖRÚT 55.