EUROSÜSTEEM

Income Underreporting by Households with Business Income. Evidence from Estonia

Merike Kukk, Karsten Staehr

Working Paper Series
6/2013

The Working Paper is available on the Eesti Pank web site at:
http://www.eestipank.ee/en/publications/series/working-papers

For information about subscription call: +372 668 0998; Fax: +372 668 0954
e-mail: publications@eestipank.ee

ISBN 978-9949-493-21-0
Eesti Pank. Working Paper Series,
ISSN 1406-7161; 6/2013

Income Underreporting by Households with Business Income. Evidence from Estonia

Merike Kukk and Karsten Staehr*

Abstract

This paper estimates the extent of income underreporting by households with business income relative to households of wage earners in Estonia. The paper uses a modified version of the methodology pioneered by Pissarides and Weber (1989). The extent of income underreporting is estimated by comparing food Engel curves for households with and without business income. The baseline result is that the reported income of households with business income above 20% of total income must be multiplied by 2.6 in order to attain the same propensity of food consumption as households of wage earners. Households with business income above 0 but below 20% also underreport income, but to a lesser extent. The estimates are higher than those found for developed countries, but consistent with other studies of the shadow economy in transition countries. The analysis also shows that the presence of business income is a better indicator of income underreporting than a reported status of self-employment.

JEL Code: H26, E21, E26, H24

Keywords: tax evasion, business income, income underreporting, Engel curve, transition country

Corresponding author's e-mail address: merike.kukk@ttu.ee .

The opinions expressed in this paper are solely those of the authors and do not necessarily reflect the views of Eesti Pank.

* The authors would like to thank Arvo Valtin from Statistics Estonia for providing the underlying data, Dmitry Kulikov for arranging the database, and Jens Hölscher, Jaanika Meriküll, Tairi Rõõm and participants of the ECEE4 conference in Tallinn for valuable comments to earlier versions of the paper. Support from Estonian Base Financing grant No. B617A and Estonian Target Financing grant No. SF0140059s12 is acknowledged.

Non-technical summary

Many households earn substantial income from self-employment or other business activities, and reporting of such income to the authorities is largely undertaken by the individual households. This provides substantial scope for underreporting. Most studies of underreporting and tax evasion focus on "envelope wages" and unreported employment, while relatively few consider underreporting by individuals with business income. This paper seeks to fill this gap in the case Estonia, a fast developing transition country in Central and Eastern Europe.

Pissarides and Weber (1989) introduced an innovative methodology for providing estimates of income underreporting by the self-employed, using data from household budget surveys. Presuming that food consumption and income are closely related for otherwise comparable households, different food consumption behaviour between the self-employed and wage earners may stem from underreported income. Moreover, if individuals provide consistent data about their income to all data collectors, the underreporting results based on data from household budget surveys may also be used as rough proxies of income underreporting to tax authorities.

This paper estimates the extent of income underreporting to the Estonian Household Budget Survey (HBS) in 2002–2007 by the households with business income relative to the households of wage earners. The analysis uses the methodology by Pissarides and Weber (1989) but modifies it to take into account the availability of a self-reported measure of regular or permanent income in the Estonian HBS.

The baseline estimation considers income underreporting by households for which business-related income comprises 20% or more of total reported income. The baseline result is that the reported income should be multiplied by 2.6 to attain the same propensity of food consumption as households of wage earners. In other words, households with business income over 20% have left unreported 62% of their "true" income, i.e. the sum of reported income and unreported income. The "true" income in this context is relative to the income reported by wage earners. The estimated income underreporting can be in the form of underreporting of gross income or over-reporting of business expenses.

The underreporting of income is somewhat lower when the share of business-related income is less than 20%, but it is still substantial; the income should be multiplied by 2.0 for households where business-related income is 10–20% or by 1.6 for household where it is 5–10%. The sample period 2002–2007 was characterised by rapid economic development, EU membership in

2004 and rapid institutional changes, but no trend in the underreporting is apparent when the sample is split into three time subsamples.

The underreporting results for households with business income are somewhat higher for Estonia than those in most other studies using the methodology of Pissarides and Weber (1989). This applies in particular when Estonia is compared with developed economies, but less so when Estonia is compared with Russia, another transition economy. Studies of the extent of envelope wages and the overall size of the shadow economy show a similar pattern, i.e. that the extent of unreported activity is much larger for transition economies than for developed economies.

It is noticeable that two different ways of identifying households involved in business activities, i.e. the self-reported employment status and the share of business income in total reported income, provide different results when data from the Estonian HBS are used. Since the self-reported employment status refers only to the household head and since many households where the household head is reported to be a wage earner obtain business income, the presence of business-related income is seen as the more reliable indicator of unreported income.

Contents

1. Introduction

The shadow economy depicts income or production not reported to the authorities. The main aim of such underreporting is arguably to evade taxation, but it may also occur because the activity is illegal or because reporting imposes an administrative burden. The volume of different income components can be underreported, e.g. wage payments, income from self-employment and corporate income (Black et al. (2012)). Reliable estimates of the size of the shadow economy are important for analysis of the incidence and welfare of taxation policies.

Many studies focus on estimates of the *aggregate* size of the shadow economy; see Schneider et al. (2010) for a review of the methods used to provide such estimates and some results. Estimates of the extent of specific forms of underreporting are also important as detailed information on e.g. revenue effects, the excess burden and the distributional impact of different tax measures are crucial for informed policy-making.

Underreporting by individuals who earn their income from self-employment or other business-related activities has received limited attention in the empirical literature. This is unfortunate since evasion of the tax on business income appears to be widespread. Slemrod (2007) uses data in a report from the US Internal Revenue Service and estimates that while only 1% of wage and salary income and 12% of capital gains were left unreported in 2001, as much as 43% of the business income of individuals was not reported to the tax authorities in the USA. There are no similar analyses of the contribution of different forms of tax evasion in other countries, but it is reasonable to assume a similar pattern elsewhere. As business income of individuals is largely the subject of individual reporting, there is generally ample scope to evade taxes on business income. Some studies even argue that the chance to engage in tax evasion could be a reason for people to choose self-employment or engagement in other business-related activities (Bruce (2000)).

It is challenging to estimate the prevalence of underreporting of business income, as the main reason for underreporting is to avoid data on the true income becoming available to the authorities. Unlike most employed individuals, individuals with business-related income have substantial discretion about the information made available to the tax authorities. If individuals underreport their income to the tax authorities, they may also underreport their income to other collectors of data. This point was illustrated by the 2007 Eurobarometer survey on undeclared work in which the data on underreporting by the self-employed were deemed unreliable because the data in many cases were inconsistent with other studies (Eurobarometer (2007)).

Pissarides and Weber (1989) introduced an innovative methodology for providing estimates of income underreporting by the self-employed, using data from household budget surveys on, inter alia, income and consumption. The starting point is the Engel curve of food consumption, which posits that food consumption and income are closely related for otherwise comparable households. Different food consumption behaviour between the self-employed and wage earners may consequently stem from underreported income. The "true" income comprising both reported and unreported income, which would match the level of food consumption, can then be computed.

Although individuals do not have any direct incentive to underreport their business-related income to a household budget survey, individuals may still feel compelled to provide consistent data about their income to all data collectors, particularly if they have any suspicion that the information will be shared with the tax authorities. Individuals will not have a similar incentive to underreport consumption levels as there is no direct tax evasion involved. Moreover, household budget surveys typically require the respondents to provide detailed information on their purchases of individual items and this can lead to a fairly precise estimation of their consumption levels. In short, it is reasonable to assume that household budget surveys contain relatively reliable consumption data, while data on business income may be subject to underreporting (Hurst et al. (2010)).

This paper uses a modified version of the methodology pioneered by Pissarides and Weber (1989) to provide estimates of income under-reporting by households with business income relative to households of wage earners in Estonia. The analysis is based on data from the Estonian Household Budget Survey (HBS) for the period 2002–2007, a period when Estonia experienced rapid economic growth. The method has been used in a handful of studies of tax evasion by the self-employed, but mainly in cases of developed countries (see the literature survey in Section 2).

This paper contributes to the literature in four ways. First, we focus on Estonia, a transition country from Central and Eastern Europe. Like other transition countries, Estonia differs from developed countries by having a lower income level and evolving institutional and administrative systems. The shadow economy is generally larger in the transition countries in Central and Eastern Europe than in West European countries (Tafenau et al. (2010), Schneider et al. (2010)).

Studies of tax evasion in transition economies have hitherto focused mainly on informal employment and envelope wages, i.e. the use of cash payments of salaries (see e.g. Schneider (2011), Feld and Schneider (2010), Meriküll and Staehr (2010), Williams (2009)). Other forms of tax evasion, including income underreporting by individuals with business income, have

received less attention. Kim et al. (2009) use the methodology of Pissarides and Weber (1989) to assess income underreporting by self-employed households in Russia in 1990s, but the methodology has not previously been use to assess underreporting by self-employed households for any of the EU countries from Central and Eastern Europe.[1]

Second, Estonia joined the European Union in 2004 along with a number of other European transition countries. In the run-up to accession many reforms were passed in order to achieve compliance with the *acquis communautaire*. The result was numerous changes in company law, taxation rules, industrial policy and government institutions. Meanwhile the Estonian economy grew very rapidly due in part to the improved confidence stemming from the accession process. The paper seeks to analyse whether these changes have affected the income underreporting by households with business income.

Third, the methodology of Pissarides and Weber (1989) requires data on the permanent income of households. The permanent income is usually not observed and additional assumptions are required to compute a proxy, and this usually gives rise to a range of estimates of the underreporting by self-employed households, with a lower and an upper bound. The Estonian Household Budget Survey makes it possible to disentangle transitory and permanent income using self-reported information, and this implies that we are able to provide a point estimate of the extent of underreporting. Kim et al. (2009) also provide a point estimate but use statistical filtering to extract a measure of the permanent income.

Fourth, the paper contributes to the literature by analysing in detail the importance of different ways of identifying a self-employed household. We consider two different definitions; the self-reported employment status of the household head and a definition based on the share of total household income coming from business-related income. The latter definition assumes a given threshold and ascribes a household as self-employed if the share of business-related income in total income exceeds this threshold. We also investigate the importance of the choice of threshold value.

The rest of the paper is organised as follows: Section 2 provides a review of the empirical literature that uses food Engel curves to estimate income underreporting of the self-employed or households with business income. Section 3 discusses the methodology of Pissarides and Weber (1989) and

[1] Estimates of income underreporting by self-employed households or households with business income are particularly important since many other countries, including Estonia, seek to encourage entrepreneurship and self-employment. One policy measure is support for the unemployed to assist them in establishing businesses and becoming self-employed (Leetmaa and Nurmela (2010)).

develops the methodology used in the paper. Section 4 introduces the data from the Estonian Household Budget Survey used in the empirical analysis. Section 5 examines the main properties of the consumption and income data. Section 6 provides the results of the estimations. Finally, Section 7 summarises the empirical findings.

2. Literature estimating income underreporting

Pissarides and Weber (1989), henceforth P&W, introduced an innovative methodology for providing estimates of income underreporting by self-employed households. They consider a household to be self-employed when business-related income exceeds a given share of its reported income. They calculate the unreported taxable income of the self-employed in the UK using income and expenditure data from the Family Expenditure Survey (FES). A detailed description of the methodology is provided in Section 3. They find that to find the "true" income that includes both reported and unreported income, the reported income must be multiplied by a factor of 1.51–1.64 for the blue-collar self-employed and by 1.28–1.54 for the white-collar self-employed. In other words, the blue-collar self-employed leave 34–39% of their "true" income unreported, while white-collar self-employed leave 23–35% unreported.

Estimations for the USA using the P&W methodology have been done by Hurst et al. (2010). Using data from the Consumer Expenditure Survey (1980–2003) and the Panel Study of Income Dynamics (1980–1997), they find that the self-employed underreport their income by about 30% of their "true" income to the household surveys used in the analysis. They observe greater underreporting of income in the early part of the sample and relate it to higher tax rates in that period. They also find evidence that the self-employed with higher education misreport their income to a lesser extent.

Mirus and Smith (1996) use the Canadian Family Expenditure Survey from 1990 and estimate that self-employed households conceal 12.5% of their "true" income. The research on income tax non-compliance by self-employed households in Canada was continued by Schuetze (2002). He investigated a longer period from 1969 to 1992 and estimated the share of underreporting for different years, demographic characteristics and occupations. His findings suggest that the degree of non-compliance by self-employed households varies significantly with occupation, age and the number of household members that are self-employed. His estimations suggest that households which obtained 30% or more of reported income from business concealed on average between 11% and 23% of the total household income.

There are two studies that cover the Nordic countries. Johansson (2005) estimates income underreporting by the self-employed in Finland for the years 1994–1996. He finds that in households where only the head of the household was self-employed, on average 16.5% of the "true" income was not reported. In households in which at least two adults were self-employed, income was underreported by 42% on average. Engström and Holmlund (2006) set up a hypothesis that the incentives for underreporting should be stronger in countries with high tax rates and examine the connection between food expenditure and reported income in households in Sweden. They estimate that households with at least one self-employed member underreport their income by around 30%. They also distinguish between the self-employed with unincorporated and incorporated businesses, the latter of which must follow more stringent regulation. Engström and Holmlund (2006) conclude that underreporting is twice as prevalent among the self-employed who are unincorporated as among the self-employed with incorporated businesses.

Kim et al. (2009) refine the P&W method to eliminate transitory income fluctuations and obtain the permanent income component for estimating consumption propensities. They use panel data from the Korea Labour Income Panel Survey from 2000–2005 and the Russian Longitudinal Monitoring Survey from 1994–2000. Their approach leads to a point estimate of underreported income instead of the interval provided in other studies using the P&W method. They find that in Korea 38% and in Russia 47% of the "true" income of self-employed households is not reported. For comparison, the list of studies using the P&W methodology is given in Table 1.

The P&W method has led to the development of alternative methods for estimating income underreporting. Lyssiotou et al. (2004) examine the use of non-parametric methods and propose a consumer demand system approach. They use the 1993 UK FES data and reach a larger share of underreporting than Pissarides and Weber (1989); the blue-collar self-employed leave 54% of their "true" income unreported and the white-collar self-employed leave 39% of their "true" income unreported.

Wangen (2005) develops two additional methods based on the P&W methodology to estimate income underreporting. His second method gives much wider intervals than the P&W method and indicates that actual business-related income is about 3.5 times higher than reported income for the UK. The large gap is in part the result of the model specification but it also suggests that refinements of the P&W methodology may lead to larger estimates of income underreporting by self-employed households.

Table 1: Studies using the methodology of Pissarides and Weber (1989) to estimate underreported income of the self-employed

	Country	Database and time period	Definition of self-employment	Unreported income as share of "true" income
Pissarides and Weber (1989)	UK	Family Expenditure Survey 1982	Share of business income over 25%	White collar: 23–35%, blue collar 34–39%
Hurst et al. (2010)	USA	Consumer Expenditure Survey 1980–2003, Panel Study of Income Dynamics 1980–1997	Reported self-employment	From CEX: 31%, from PSID: 29%
Schuetze (2002)	Canada	Family Expenditure Survey 1969–1992 (of these 6 years)	Share of business income over 30%	11–23%
Johansson (2005)	Finland	Household expenditure survey 1994–1996	Reported self-employment	One person self-employed: 10–24%, two people self-employed: 37–47%
Engström and Holmlund (2006)	Sweden	Household Budget Survey 1999–2004	Reported self-employment	With incorporated business: 15–20%, with unincorporated business: 40–50%
Kim et al. (2009)	Russia, South Korea	Longitudinal Monitoring Survey 1994–2000 (Russia), Labour Income Panel Survey 2000–2005 (South Korea)	Reported self-employment	Russia: 47%, Korea: 38%

3. Methodology

This paper uses a modified version of the methodology pioneered by Pissarides and Weber (1989) to provide estimates of income not reported by households with business income.[2] The idea is to estimate a food Engel curve

[2] We also considered using the consumer demand system approach by Lyssiotou et al. (2004). Their method assumes that wage income in all households is correctly reported and only self-employment income is underreported, which is a stricter assumption than the corresponding one in P&W. Moreover, the use of the P&W methodology facilitates comparison with the results for other countries.

for all households but allow for a shift dummy for households with business income. After controlling for different household characteristics and wealth proxies that can induce differences in consumption behaviour, the propensity of food consumption is expected to be the same for both groups of taxpayers and the estimated shift dummy or gap will therefore reveal the share of earnings unreported by households with business income.

The crucial identifying assumption of the model is the attribution of the estimated gaps in food consumption to income underreporting. There are other potential explanations for an expenditure gap, such as heterogeneity in preferences, which may bias estimates of underreporting. Using food consumption with additional control variables to capture household heterogeneity is meant to address this problem. The risk of confusing heterogeneity with underreporting would be higher in a comparison of spending on durable goods, which have a wider variety of brands with larger differences in prices. Additionally, food typically cannot be classed as a business expense, which cars or telecommunication costs can, and when this happens, it is easier to report personal expenses as business expenses, possibly leading to the underreporting of consumption along with the underreporting of income.

According to the Permanent Income Hypothesis, consumption is smoother than income as it is not affected by transitory income (Friedman (1957)). This means that consumption depends on the permanent component of income and – in principle – not on current income. As the saving or dissaving of transitory income can be mistaken for misreporting, transitory income should be treated separately from permanent income.

The food Engel curve can be estimated using the following specification, where subscript i is the indices for the households:

$$\log c_i = \alpha + \beta \log y_i^{\text{perm}} + X_i'\phi + \varepsilon_i \,. \tag{1}$$

The term $\log c_i$ denotes the logarithm of food consumption, $\log y_i$ is the logarithm of the permanent component of income, X_i is a column vector of control variables affecting consumption, and ε_i is an error term. The coefficient β is the income elasticity of food consumption.

The income process of a wage earner can be expressed as:

$$\log y_i = \log y_i^{\text{perm}} + \log y_i^{\text{trans}} \,. \tag{2}$$

Usually only current income is reported in survey data, and it is difficult to distinguish between the permanent and transitory components of current income. P&W therefore use the assumption that the log of current income is the sum of the logarithms of the permanent and the transitory income components and that the transitory income has a log-normal distribution across

households. These assumptions allow them to derive the permanent income component from current income. Realistically, they also assume that short-term fluctuations in current income are different for households of wage earners and households with business income as the latter group experiences more volatile income and this difference has to be taken into account in the estimations.

The Estonian HBS makes it possible to exclude the transitory component of the income directly through the use of self-reported information.[3] We can therefore, when we derive the final model used to calculate the share of underreporting by the self-employed, leave aside further assumptions regarding the distribution of the permanent and transitory components of the reported income.

It is assumed that the households with business income misreport their earnings by a factor $\kappa_i \geq 1$. When income is underreported, the "true" permanent component of income is thus:

$$\log y_i^{\text{perm}} = \log \kappa_i + \log y_i^{\text{rep\_perm}} . \tag{3}$$

The term $\log \kappa_i$ captures the unreported log permanent income that must be added to the log reported income, $\log y_i^{\text{rep\_perm}}$, to attain the "true" log permanent income, $\log y_i^{\text{perm}}$. The P&W model assumes that employed households provide unbiased reporting of their income to the household budget survey, i.e. $\kappa_i = 1$ for the wage earners. In Estonia envelope wages contribute considerably to the total income of a small fraction of wage earners (Meriküll and Staehr (2008), EKI (2011)). Fortunately the assumption of P&W is not too restrictive in this respect: if the households of wage earners also systematically misreport their income to household surveys, κ_i would be an estimate of the relative difference in underreporting by households with business income and households of wage earners.

The random variable κ_i is assumed to have a log-normal distribution, and $\log \kappa_i$ can be expressed as the deviation from its mean:

$$\log \kappa_i = \mu_\kappa + \nu_i . \tag{4}$$

The variable μ_κ is the mean of the logarithm of κ_i and $\mu_\kappa > 0$ would consequently entail underreporting by households with business income. The random variable ν_i has zero mean and constant variance $\sigma_\nu^2 > 0$ within each group. The variance for the group of households with business income is la-

[3] The survey includes questions on current income and regular income. The difference can be considered as transitory income. Further investigation of the two income components is provided in Kukk et al. (2012).

belled $\sigma^2_{\nu|\mathrm{R}}$, while the variance for the group of households of wage earners is labelled $\sigma^2_{\nu|\mathrm{W}}$.

By combining equations (1), (3) and (4), the following consumption function is obtained:

$$\log c_i = \alpha + \beta \log y_i^{\text{rep\_perm}} + \beta\mu_\kappa + X_i'\phi + \xi_i . \tag{5}$$

The term $\beta\mu_\kappa$ denotes the shift of the intercept of the Engel curve for households with business income compared to households of wage earners and $\xi_i = \varepsilon_i + \beta\nu_i$ is an error term. Since $\mathrm{E}[\log y_i^{\text{rep\_perm}} \xi_i] \neq 0$, reported permanent income has to be instrumented. The first stage estimation takes the form:

$$\log y_i^{\text{rep\_perm}} = Z_i'\delta + \omega_i , \tag{6}$$

where Z_i is a column vector of identifying instruments and ω_i is the error term of the first stage income regression. The instrumented or predicted log reported permanent income is labelled $\log \hat{y}_i^{\text{rep\_perm}}$.

The true or total regular income of households with business income is found as the reported regular income multiplied by κ. The underreporting factor κ is estimated as:

$$\kappa = \exp(\mu_\kappa + \tfrac{1}{2}\sigma^2_{\nu|\mathrm{B}}) . \tag{7}$$

The variance $\sigma^2_{\nu|\mathrm{B}}$ is not known, so additional assumptions must be applied. In eq. (6) the residual ω_i contains unexplained variation in permanent income and the deviation of actual from reported income ν_i. Assuming that unexplained variation in permanent income has the same variance for both groups, the only difference in the variance of the error term ω_i in eq. (6) between the two groups stems from the variance in the log underreporting factor, and $\sigma^2_{\nu|\mathrm{B}}$ can therefore be estimated as $\sigma^2_{\nu|\mathrm{B}} = \sigma^2_{\omega|\mathrm{B}} - \sigma^2_{\omega|\mathrm{W}}$.

Unlike estimations undertaken using the P&W methodology in unaltered form, we obtain a point estimate of κ. There is no need to impose additional assumptions on the distribution of permanent income as the Estonian HBS makes possible the use of a self-reported measure of permanent income which omits transitory income fluctuations. This eliminates the need to report a range of estimates as the residual would include an additional error component (deviation of actual from permanent income) and the estimation of $\sigma^2_{\nu|\mathrm{B}}$ would need an additional assumption regarding the covariance between the random variable of eq. (4) and the random variable of the permanent income process. Kim et al. (2009) also provide a point estimate, but their method is based on statistical filtering of the current income variable.

4. Dataset and sample restrictions

The paper uses data from the Estonian Household Budget Survey (HBS), which is conducted on an annual basis by Statistics Estonia among a representative cross-section of Estonian households. This paper uses the data from the years 2002 to 2007.[4] The dataset has previously been used by Kulikov et al. (2009) to investigate the saving behaviour of Estonian households and by Kukk et al. (2012) to estimate consumption sensitivities to shocks in income processes of different persistence. These studies contain a more detailed description of the dataset.

The total consumption is the sum of twelve consumption categories. For the analysis we use the consumption of food which includes eating outside the home. Wider consumption measures, such as non-durable and total consumption, include expenditure items like telecommunication services, mobile telephones and computers that may be reported as business expenses instead and these measures are therefore not suitable for use in the estimation of income underreporting by households with business income.

The Estonian HBS makes it possible to extract two separate monthly household income figures. The *current* after-tax household income contains five income categories, which are wage income, business-related income, property income, transfers and other income. In the questionnaire, business-related income is any type of earnings from self-employment and also earnings from activities that take place outside a regular wage contract, such as income from a start-up business or individual consultancy services. It includes income from the following sources: 1) registered self-employment activities, 2) provision of services, 3) self-production, 4) intermediation of products and services, and 5) other business activities. It also contains dividends or any other kind of payment from a self-managed company – only dividends from investment with no active involvement in the company are regarded as capital gains.

Additionally, the dataset includes the *regular* after-tax income, which is the household's assessment of its average monthly income when transitory income fluctuations are omitted. Kukk et al. (2012) show that the variable denoting the difference between current and regular income is indeed transitory income. The regular income cannot be considered to be fully permanent, but as pointed out by Pissarides and Weber (1989), it is important to exclude short-term fluctuations in the income, while permanent income does not need to correspond one-to-one to the income in the Permanent Income Hypothesis.

[4] The data collection was discontinued in 2008 due to budgetary constraints at Statistics Estonia.

The regular income is in all likelihood a good proxy for permanent income in the P&W model.

The Estonian Household Budget Survey includes a question on the occupational status of the household head. The household head is taken to be self-employed if the occupational status of the *household head* is "entrepreneur with employees, inc. farmer", "self-employed" or "employed in family enterprise, inc. farmer".

Different household characteristics that act as control variables and as instruments for income figures are nationality, gender, region of residence (Statistics Estonia's classification of 5 regions), the age of the household head, the number of children under 16, and a categorical variable indicating whether one or two adults are working. The dataset contains only a limited number of proxies of wealth, of which we use two: a dummy for renting housing rather than owning housing, and a dummy for owning a second real estate property.

We also use the variable of temporary income which is defined as the difference between current and regular income reported in the Estonian HBS. Although theory hypothesises that consumption is insensitive to temporary income, Kukk et al. (2012) show that consumption does react to temporary income shocks, although by substantially less than it does to regular income shocks. The variable will therefore be included in the consumption model to avoid omitted variable bias.

All income and consumption variables are in real terms. The nominal income consumption series has been deflated by the HICP index (with the index = 1 in 2005), while nominal food consumption has been deflated by the HICP index for the food category.

Observations with missing income and food expenditure have been excluded. Moreover, all observations where the economic activity status of the household head is classified as "inactive" have been omitted from the sample. Finally, the sample has been restricted to households with two adults, with or without children, since more precise estimates of the parameters can be obtained by focusing on a fairly homogeneous group. In the final sample there are 6016 cross-sectional observations.

The P&W methodology prescribes the division of the sample into a part deemed unlikely to underreport income and a part deemed likely to underreport. There are basically two methods for dividing the sample. One method considers the share of business-related income in total income and defines a household as self-employed if the share exceeds a given threshold. This is the method used in Pissarides and Weber (1989) and later in Schuetze (2002). The second method uses the household's self-reported employment status.

This is the method used in Hurst et al. (2010), Johansson (2005) and Engström and Holmlund (2006).

The Estonian HBS makes it possible to use either of the two methods, as data on the occupational status of the household head and the share of business income in total income are available. The first method has, however, some drawbacks. First, the self-reported employment status refers only to the household head, but the household head will typically not be the only income earner and the employment status of the other adult household member is not known. Second, even if the household head is the only or the main income earner, the household head might have both business-related income and wage income. Third, experience shows that households that combine employment and self-employment tend to report their status as employed.

Table 2 shows that the two identification methods divide the sample into different groups: 1/3 of the households where the household head is reported to be self-employed do not report any business-related income. Moreover, almost half of the households where the household head is reported to be a wage earner earn some business-related income. The number of these households (2565 households) exceeds by far the number of households where the household head is self-employed and the share of business income in total income household is positive (366 households).

Table 2: Split of the sample using different definitions of self-employment

		Self-reported occupational status		
		Wage earner	Self-employed	Total
Share of business income in total income	0%	2923	162	3085
	> 0%	2565	366	2931
	Total	5488	528	6016

Source: Estonian HBS 2002–2007.

This shows that many households earned both wage income and business income; households that report the household to be self-employed comprise only a small part of the households who are involved in self-employment activities. In this paper we therefore focus on underreporting by households with business income, but for robustness checks we also consider those who state that the household head is self-employed.

5. The properties of the main variables

The P&W methodology assumes a linear relationship between log food consumption and log regular income. To assess the validity of the assumption in this sample we undertake additional estimations and for this purpose we use the sample of wage earners, i.e. households with no business income, to ensure that underreporting of business income does not distort the results. Figure 1 shows a scatter plot between the regular income and food consumption of wage earners together with a lowess curve of bandwidth 0.25. The figure reveals a linear relationship between the two variables. We also estimated the Engel curve for different income quartiles of wage earners and a quintile regression (not reported). The coefficients are very stable across different income groups, suggesting that the linear Engel curve can be used for further estimations.

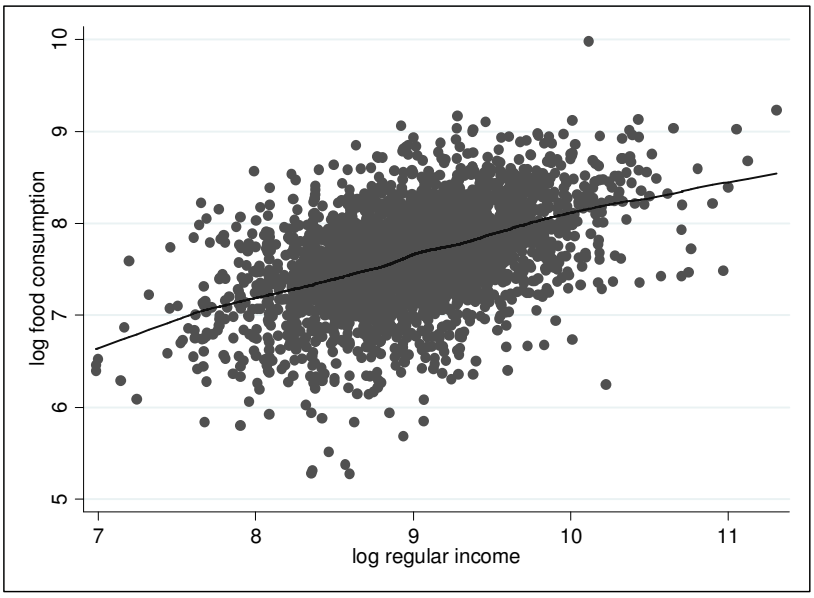

Figure 1: Lowess curve for wage earners

Note: Bandwidth = 0.25.

Table 3 shows the distribution of the share of business-related income to total after-tax income across households. While almost half of the households in the sample (2931 households) report some positive business-related income, about 2/5 of this group report business-related income of less than 5%, while about 1/5 report business-related income of over 20% of their total after-tax income. Very large shares of households combine employment and self-employment and the balance between the two ways of generating income varies greatly between households. One adult may for instance be employed

while the other is self-employed or one person may combine paid employment and self-employment.

Table 3: Split of the sample into different groups based on the share of business-related income in total after-tax income

Share of business-related income in total after-tax income	Number of households	Share of total households within the sample
0%	3085	51.3%
0–5%	1067	17.7%
5–10%	662	11.0%
10–20%	544	9.0%
20–50%	380	6.3%
≥50%	278	4.6%
Total	6016	100.0%

Source: Estonian HBS 2002–2007.

The P&W methodology provides estimates of underreporting with respect to *total* income and therefore the underreporting factor should be different for households with different shares of business-related income. In our estimations we will focus on households whose reported business-related income is over 20% of reported total income, but we also provide estimations for households where the share of business-related income is between 0 and 20%. We compare these households with business income with households of wage earners, defined as households with *no* business income.

The estimation of the Engle curve may lead to erroneous results if the share of business-related income is correlated with the regular income variable. If, for instance, households with higher regular income were to have a higher share of business-related income, the estimated results would be different for different income groups as the importance of the business income would vary across income groups. However, we do not find any evidence of this form of correlation. There is a weak negative correlation between regular income and the share of business-related income among households who receive business income (–0.18). If only households with more than 20% business-related income (the baseline sample) are considered, the correlation coefficient is 0.04. Figure 2 shows a scatter plot of the share of business-related income to log regular income and no systematic pattern is apparent.

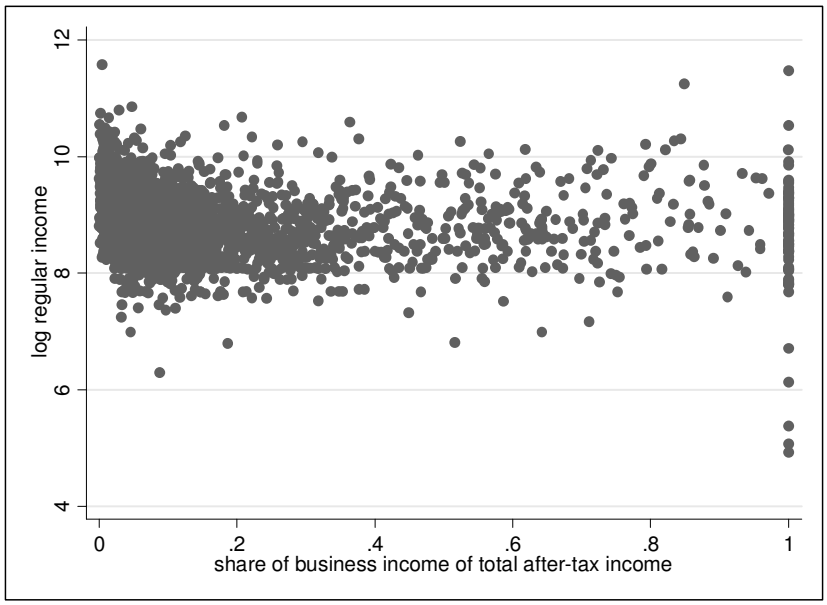

Figure 2: Scatter plot of share of business-related income and log regular income

Table 4 gives the main descriptive statistics of the income and consumption measures for households of wage earners (with no business income) and households with business income above 20% of total reported income. Households of wage earners have a lower mean of food consumption than households with business income, but report a higher level of regular income. Given the same propensity of food consumption for the two groups and that food consumption is correctly reported, the gap can be explained by income underreporting by the households with business income.

Table 4: Comparison of the main variables for the two household groups

	Wage earners		**Business income ≥ 20% of total reported income**	
Variable	**Mean**	**St. dev.**	**Mean**	**St. dev.**
Log food consumption[a]	7.638	0.533	8.044	0.537
Log regular income[a]	8.991	0.548	8.747	0.653
No. of obs.	3085		658	

Source: Estonian HBS 2002–2007.

Note: [a] The variables are expressed in 2005 prices. During the sample period, the kroon (isocode EEK) was the currency in Estonia; the exchange rate was fixed at 15.65 EEK for 1 EUR.

The kernel density functions in Figure 3 show that the distributions of food consumption and regular income for wage earners and for households with business income are very similar while the means of the distributions are different. As the distributions are so similar, it is reasonable to use the P&W method to estimate the gap between the two groups.

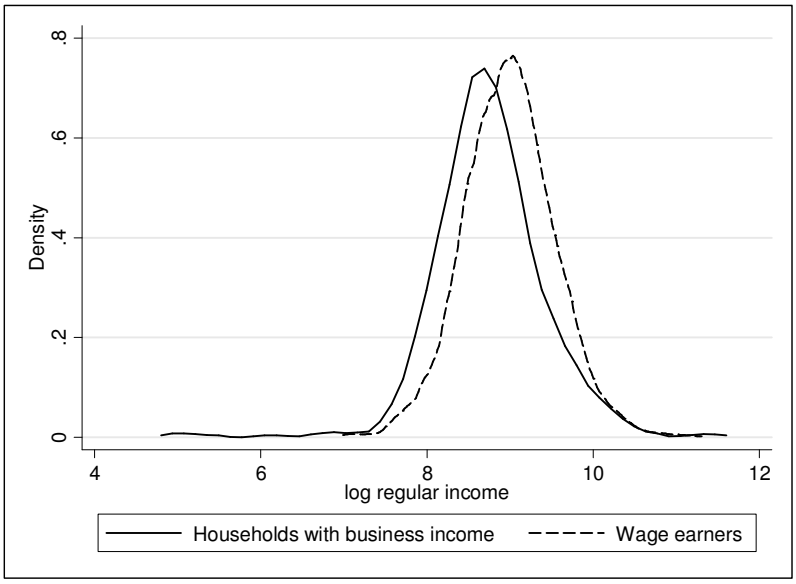

Figure 3: Kernel density functions of total after-tax income for households of wage earners and households with business-related income over 20%
Notes: Kernel = epanechnikov, bandwidth = 0.1315.

6. The empirical model and the estimations

We are estimating the following Engel curve, which is the empirical equivalent of eq. (5):

$$\log c_i = \alpha + \beta \log \hat{y}_i^{\text{rep\_reg}} + \gamma D_i + X_i'\phi + \tau_t + \xi_i \,. \qquad (8)$$

where $\log c_i$ is the log of household food consumption, $\log \hat{y}_i^{\text{rep\_reg}}$ is the instrumented log reported regular income and D_i is a dummy variable which in the baseline estimation takes the value one for households with business-related income of over 20% of reported after-tax income. The control variables consist of a number of household characteristics captured in a column vector X_i as well as time fixed effects τ_t for the 71 months in the sample. Household characteristics include the log temporary income, age and age squared of the household head, the number of children under 16, a dummy

for there being two income earners, a dummy for the ownership of a second real estate property and a dummy for renting of the main residence. The dataset lacks wealth variables that may affect consumption along with the income variable (Attanasio (1999)). Nevertheless, inclusion of dummies for real estate should capture the wealth effect on consumption as real estate presents the main share of the wealth of households (ECB (2013)).[5]

We use the household head's education level, gender, nationality and regional dummies as additional instruments for regular income. According to the standard earnings model of Mincer, education level is a main determinant of income level (Mincer (1976)). Significant income gaps are also observed between genders and nationalities in Estonia (Leping and Toomet (2008), Anspal and Rõõm (2011)).

It is assumed that the estimated coefficients of the control variables and of the instruments for the income variable $\log \hat{y}_i^{\text{rep\_perm}}$ do not differ between households of wage-earners and households with business income.

The dummy variable D_i in eq. (8) replaces $\beta\mu_k$ in eq. (5), i.e. the estimated coefficient γ is equivalent to $\beta\mu_\kappa$ in the theory model.[6] The underreporting factor κ indicates how much the regular income of the households with business income must be scaled up to attain the "true" regular income that would be comparable to that of wage earners:

$$\kappa = \exp\left(\frac{\gamma}{\beta} + \tfrac{1}{2}(\sigma^2_{\omega|\mathrm{B}} - \sigma^2_{\omega|\mathrm{W}})\right). \tag{9}$$

The underreported income is correspondingly computed as $(\kappa - 1)/\kappa$ and gives the share of unreported income out of total after-tax income.

As it is not possible to compute confidence intervals for $\hat{\kappa}$ in eq. (9), we also compute the simplified measure κ_S:

$$\kappa_S = \exp(\gamma / \beta). \tag{10}$$

[5] Additionally, instrumentation should reduce or fully eliminate the problem of biased estimates of the income coefficient due to the omission of relevant wealth variables (Wooldridge (2002)).

[6] If the business related income is taken as endogenous, the dummy variable D_i would need to be instrumented. Experimentation with different instruments from the dataset did not produce reasonable results, arguably due to weak instruments, and we therefore decided not to instrument D_i.

The advantage of using κ_S is that we can calculate the standard error of the measure. It makes it possible to test whether the estimated coefficients are significantly different for different sub-samples.

The estimation of the food Engel curve in eq. (7) is given in Table 5. Here we present the variables of main interest: the coefficients for regular income and the business income dummy. The estimations for the full baseline model and for the first stage regression are given in Table A.1 in Appendix A.

Column (1) in Table 5 shows the results of the baseline model where a dummy for households with business income over 20% is included. The households for which the share of business-related income is between 0 and 20% are excluded from the sample altogether, hence households with no business-related income behave as a comparison group. The estimated coefficients presented in the table are statistically significant at the 1% level. The propensity of consumption of food, β, is 0.61, while the coefficient of the business income dummy, γ, is 0.55. For the first stage regression the *F*-statistic for the overall goodness of fit is 40.49 and the coefficient of determination R^2 is 0.45 and this suggests that the chosen variables are appropriate instruments of the regular income variable.

As a robustness check we run the estimations with different sets of control variables. The results are shown in Table B.1 in Appendix B. The estimations are very stable regardless of the number and choice of control variables. The estimations confirm that the propensity to consume food is markedly higher for the households with business-related income than for the wage earners.

The estimated underreporting factor refers to underreporting of total income of the household, not only of business income. Different shares of business-related income in total income may therefore lead to different estimates of the underreporting factor. This issue is examined in more detail in the estimations in columns (2) and (3) in Table 5. Column (2) reports the coefficients of two different dummies, one for which the business income is in the interval 20–50% and one for which the business income is over 50%, but it follows that the estimated coefficients, γ_1 and γ_2, are very similar. Column (3) shows the results when dummies for households with lower shares of business income are included. The lower the share of business income is, the lower the estimated coefficients of the dummy variables γ_3, γ_4 and γ_5 are, which is unsurprising given that the estimated lower dummy coefficient indicates lower underreporting of *total* income. The results in columns (2) and (3) made us choose the 20% cut-off of business income in reported total income for our baseline scenario.

Table 5: Food consumption estimations

	(1)	(2)	(3)
	Baseline regression (business income ≥ 20%)	**Split of households with different share of business income**	**Including households with lower share of business-related income**
β (consumption propensity)	0.612*** (0.039)	0.611*** (0.039)	0.625*** (0.031)
γ (business income ≥ 20%)	0.546*** (0.025)	..	..
γ_1 (business income ≥ 50%)	..	0.545*** (0.028)	0.546*** (0.028)
γ_2 (business income 20–50%	..	0.548*** (0.039)	0.581*** (0.039)
γ_3 (business income 10–20%	..	..	0.444*** (0.019)
γ_4 (business income 5–10%)	..	..	0.290*** (0.017)
γ_5 (business income < 5%)	..	..	0.131*** (0.016)
R^2	0.297	0.298	0.331
Endogeneity test [p-value]	36.23 [0.000]	36.66 [0.000]	50.94 [0.000]
Hansen *J*-test [p-value]	13.44 [0.062]	13.45 [0.062]	18.69 [0.001]
No of obs.	3743	3743	6016
R^2 of the first regression	0.453	0.454	0.449
F-stat of the first regression [p-value]	40.49 [0.000]	40.16 [0.000]	57.54 [0.000]

*Notes: IV estimations with GMM estimator. Education level, nationality, gender and five regional dummies are used as instruments. Log temporary income, age and age squared of the household head, the number of children, two income earners, ownership of a second real estate property, renting the main residence, and 71 monthly time fixed effects are included in the estimations, but the results are not shown in the table. Robust standard errors are reported in round parentheses below the coefficient estimates. Superscripts ***, ** and * indicate that the coefficient is statistically different from 0 at the 1%, 5% and 10% level, respectively. F-stat is the test statistic for the Wald test of overall goodness of the fit of the first regression.*

Table 6 shows the underreporting factors κ and κ_S computed from the estimated coefficients in Table 5 and the residual variances from the first stage of the IV regression, cf. eqs. (9) and (10). Column (1) shows that the simple underreporting factor, κ_S, is 2.4 and the standard errors suggest that this result is attained with substantial precision. The standard underreporting factor taking into account the first stage variances, κ, is 2.6. Focusing on the latter result, the reported income for households with business income over 20% should be multiplied by 2.6 to attain the same propensity of food consumption as households of wage earners. Put differently, households with business income over 20% have left unreported $1.631/2.631 \approx 62\%$ of their "true" income.

Table 6: Estimates of income underreporting by households with business income

	(1)	**(2)**	**(3)**
	Business income ≥ 20%	**Business income 10–20%**	**Business income 5–10%**
Underreporting factor κ_S	2.442	2.035	1.591
(with standard errors)	(0.137)	(0.080)	(0.053)
Underreporting factor κ	2.631	2.005	1.561
Share of "true" income unreported	62.0%	50.1%	35.9%

Notes: Authors' calculations based on the results in Table 5.

It is important to underscore that the "true" income in this context is relative to the income reported by wage earners. As discussed in Sections 1 and 3, it is conceivable that households of wage earners receive envelope wages and misreport their income to the HBS. We can only calculate the gap of underreporting between households with business income and households of wage earners and the computed underreporting of the "true" income should therefore be seen as relative to the underreporting by households of wage earners.

For the households that have business-related income of less than 20% (columns (2) and (3)), the income variance does not differ from the income variance of wage earners and hence the two underreporting factors κ_S and κ are very similar. In order to obtain the same propensity of food consumption as households of wage earners, the income should be multiplied by 2.0 for households where business-related income is 10–20% or by 1.6 for household where it is 5–10%. These results indicate considerable underreporting of

total income in households in which the share of business-related income is relatively small.[7]

The results in Table 6 suggest that the extent of underreporting by households with business income is very substantial. The baseline model entails that 62% of the income is left unreported, but the share of unreported income would be somewhat smaller if a lower cut-off than 20% is chosen. The corresponding results from other countries discussed in Section 2 generally show lower shares.

Among developed countries, the upper limit of the share of underreporting is estimated to be around 30–40%. Estonia is a transition country with particular economic and institutional structures, and the substantially higher estimates for Estonia than for developed countries are consistent with the estimates of the shadow economy in Estonia and other European transition economies. This applies to the overall size of the shadow economy where the estimates for the transition countries typically are two or three times the estimates for developed economies (Tafenau (2010), Schneider at al. (2010)). It also applies to the payment of envelope wages and the extent of unreported employment (Williams (2008), (2009)).

Among the group of transition countries, the only methodology study is from Russia where the result is underreporting of 47% of "true" income, i.e. a slightly smaller share than was found in this study (Kim et al. (2009)). The Russian study, however, uses a different definition of self-employed households, so the results are not directly comparable. Moreover, the underreporting factor for Russia is estimated with large standard errors.

Previous studies using the P&W methodology have not emphasised the importance of the definition of the self-employed. Further investigation shows, however, that the underreporting results depend on how we distinguish between different types of household. As shown in Table 2 there are a significant number of households that report themselves as self-employed but who do not have business-related income, and also a large number of households who have a substantial share of business-related income but are reported as employed. Our estimations in Table 7 show that the share of business-related income in total reported income is the main defining identification scheme. First, separate estimations for the households with business in-

[7] *If* it is assumed that business income is the main source of the underreporting, the estimations would imply very substantial underreporting of business income. Households with business-related income of less than 20% would need to multiply their *business income* by a factor of at least by 6 to attain the given underreporting factor in total income. We cannot, however, assume that the source of the underreporting is only business income; the estimations imply that even small shares of business-related income are a good indicator of income underreporting.

come where the household head is self-employed (column (1)) and where the household head is a wage earner (column (2)) provide a very similar result, i.e. substantial underreporting of income. Second, the estimation in column (3) shows that there is no evidence of income underreporting for households where the household head is reported to be self-employed but where the household reports having business income.

The upshot is that the presence of business-related income is the most reliable indicator of income underreporting. This finding is consistent with the findings in a study by Nastav and Bojnec (2008). They search for a relationship between small business and the shadow economy in ten New Member States of the European Union. The size of small businesses is proxied by the number of self-employed persons, but they do not find any significant relationship between the two variables.

Table 7: Estimates of income underreporting across different sub-samples

	(1)	**(2)**	**(3)**
	Self-employed, share of business income ≥ 20%	**Wage earners, share of business income ≥ 20%**	**Self-employed, share of business income = 0**
Underreporting factor κ_S	2.202	2.484	0.947
(with standard errors)	(0.158)	(0.145)	(0.06)
Underreporting factor κ	2.438	2.640	1.007
Share of "true" income unreported	59.0%	62.1%	0.0%

Notes: Authors' calculations.

The results for Estonia are not *fully* comparable with those of other studies that use the P&W methodology. Previous studies have typically defined some households as wage earners even if they report earning some business-related income. This applies when the studies use self-reported employment status but also when the definition of self-employment is based on the share of business income when households with business income below the threshold value are considered to be wage earners. We find that households with business income underreport their income even if the share of business income in total income is small (Table 6, column (3)).

Based on the findings in Table 6 we define a household as a wage earner if the household reports *no* business income; the comparison group in previous studies is thus more mixed than it is in our estimations. When we use a similar distinction to that in previous studies, we obtain somewhat lower but still

qualitatively similar results. If we use the same cut-off point of 25% business-related income as used in Pissarides and Weber (1989) for instance, and include households with business income of lower than 25% in the sample of wage-earners, the estimated underreporting parameter for Estonia would be 2.37, which entails a share of unreported income equal to 57.8%. Nevertheless, we prefer to define households of wage earners as those with no business income precisely because we find evidence of underreporting among households with a share of business income between 0 and 20%.

The underreporting by households with business income found in this study can in fact originate from two sources. One possibility is for households to leave some business-related income unreported. Another possibility is to over-report the expenses related to business activities. This is because the Estonian tax code allows households to deduct their business-related expenses from business income before reporting the income to the tax authorities. This lets households report some personal expenses like car maintenance expenses, telecommunication bills and even housing costs (if the office is at home) as business expenses, thus lowering the reported business income. This type of tax behaviour is considered to be used quite extensively in Estonia. The two sources of underreporting can explain the high share of underreporting found with the P&W methodology.

The gap between the consumption propensities for households with business income and households of wage earners could be due to factors other than underreporting of income. If households involved in business activities have preferences that lead to a different consumption pattern than households of wage earners, then the estimations could also capture this effect. However, this is considerably less likely for food consumption than for non-durables or durables, for which consumption can vary more across different goods and brands. Moreover, we employ a large number of control variables to account for heterogeneity in tastes across households.

We also investigate the dynamics of the underreporting factor over time by estimating it separately for the years 2002–2003, 2004–2005 and 2006–2007. This exercise is particularly pertinent as the period was characterised by rapid economic and institutional change (Meriküll and Staehr (2010)). Estonia joined the EU in 2004 and many reforms were passed in the years before to achieve compliance with the *acquis communautaire*. The result was numerous changes in company law, taxation rules, industrial policy and government institutions. The country saw rapid development with annual rates of economic growth above 6 percent. The flat tax rate on personal and corporate income was 26% until 2004, but was then gradually lowered to 22% in 2007.

The estimations are run for households that have business-related income over 20%. The results in Table 8 provide some evidence for parameter

changes across time as the underreporting factor is lower for the period after 2004, but the difference is not statistically significant. In any case, the share of "true" income not reported is very large in all three subsamples; it is relatively constant across the subsamples and with no clear direction of change. This suggests that EU accession, institutional development and rapid economic change did not materially affect the extent of underreporting of households with business income.

Table 8: Estimates of income underreporting across different time periods

	(1)	(2)	(3)
	2002–2003	**2004–2005**	**2006–2007**
Underreporting factor κ_S	2.755	2.241	2.483
(with standard errors)	(0.275)	(0.186)	(0.327)
Underreporting factor κ	3.130	2.337	2.569
Share of "true" income unreported	68.1%	57.2%	61.1%

Notes: Authors' calculations.

We also produced estimations for different subsamples but again the wide confidence intervals did not allow us to draw any conclusions (not reported). We did not find any statistically significant difference between the coefficient estimates for households of different educational levels, nor for households living in different regions, but the results evidently hinge on the small number of observations in the different subsamples.

7. Final comments

This paper estimates the extent of income underreporting to the Estonian Household Budget Survey in 2002–2007 by the households with business-related income relative to the households of wage earners. The analysis uses the methodology pioneered by Pissarides and Weber (1989) but modifies it to take into account the availability of a self-reported measure of regular or permanent income in the Estonian HBS. If individuals provide consistent data about their income to all data collectors, the underreporting results based on data from the HBS may also be used as rough proxies of income underreporting to tax authorities.

The baseline estimation considers income underreporting by households for which business-related income comprises 20% or more of total reported

income. The baseline result is that the reported income should be multiplied by 2.6 to attain the same propensity of food consumption as households of wage earners. In other words, households with business income over 20% have left unreported 62% of their "true" income, i.e. the sum of reported and unreported income. The "true" income in this context is relative to the income reported by wage earners. The estimated income underreporting can be in the form of underreporting of gross income or over-reporting of business expenses.

The underreporting of income is somewhat lower when the share of business-related income is less than 20%, but it is still substantial; the income should be multiplied by 2.0 for households where business-related income is 10–20% or by 1.6 for household where it is 5–10%. The sample period 2002–2007 was characterised by rapid economic development, EU membership in 2004 and rapid institutional changes, but no trend in the underreporting is apparent when the sample is split into three time subsamples.

The underreporting results for households with business income are somewhat higher for Estonia than those in most other studies using the methodology of Pissarides and Weber (1989). This applies in particular when Estonia is compared with developed economies, but less so when Estonia is compared with Russia, another transition economy. Studies of the extent of envelope wages and the overall size of the shadow economy show a similar pattern, i.e. that the extent of unreported activity is much larger for transition economies than for developed economies.

It is noticeable that the two different ways of identifying households involved in business activities, i.e. the self-reported employment status and the share of business income in total reported income, provide different results when data from the Estonian HBS are used. Since the self-reported employment status refers only to the household head and since many households where the household head is reported to be a wage earner obtain business income, the presence of business-related income is seen as the more reliable indicator of unreported income.

The estimates may represent an upper bound for the underreporting of income, given that a fraction of the food consumption difference may originate from sources other than underreporting. The upshot is, nevertheless, that the underreporting by households with business income is very pronounced. If households are consistent in their reporting to household survey and tax authorities, the estimated income underreporting is a possible indicator of the level of tax evasion or tax avoidance by Estonian households with business income. The present analysis suggests that detailed studies using tax and register data would indeed be a worthwhile undertaking.

References

ANSPAL, S. AND RÕÕM, T. (2011): "Gender pay gap in Estonia: empirical analysis". Research Report for the Ministry of Social Affairs of Estonia, http://www.sm.ee/fileadmin/meedia/Dokumendid/V2ljaanded/Publikatsioonid/2011/Gender_pay_gap_Estonia_analysis.pdf.

ATTANASIO, O. (1999): "Consumption". In J. Taylor & M. Woodford (eds.). *Handbook of Macroeconomics*, Rotterdam: Elsevier, pp. 741–812.

BLACK T., BLOOMQUIST, K., EMBLOM, E., JOHNS, A., PLUMLEY, A. AND STUK, E. (2012): "Federal tax compliance research: tax year 2006 tax gap estimation". U.S. Department of the Treasury, Internal Revenue Service Research, Analysis & Statistics Working Paper, March 2012.

BRUCE, D. (2000): "Effects of the United States tax system on transitions into self-employment". *Labour Economics*, Vol. 7, No. 5, pp. 545–574.

ECB (2013): "The Eurosystem Household Finance and Consumption Survey. Results from the first wave". European Central Bank, Statistics Paper Series, No. 2.

ENGSTRÖM, P. AND HOLMLUND, B. (2006): "Tax evasion and self-employment in a high-tax country: evidence from Sweden". CESifo Working Paper, No. 1736.

EUROBAROMETER (2007): "Undeclared work in the European Union". Special Eurobarometer, No. 284, European Commission, http://ec.europa.eu/public_opinion/archives/ebs/ebs_284_en.pdf.

FELD, L. P. AND SCHNEIDER, F. (2010): "Survey on the shadow economy and undeclared earnings in OECD countries". *German Economic Review*, Vol. 11, No. 2, pp. 109–149.

FRIEDMAN, M. (1957): *A Theory of the Consumption Function*. Princeton and Oxford: Princeton University Press.

HURST, E., LI, G., AND PUGSLEY, B. (2010): "Are household surveys like tax forms: evidence from income underreporting of the self employed". *NBER Working Paper*, No. 16527.

JOHANSSON, E. (2005): "An estimate of self employment income underreporting in Finland". *Nordic Journal of Political Economy*, Vol. 31, No. 1, pp. 99–109.

KIM, B., GIBSON, J., AND CHUNG, C. (2009): "Using panel data to exactly estimate income under-reporting of the self employed". Labor

Economics Working Papers, No. 22996, East Asian Bureau of Economic Research.

KUKK, M., KULIKOV, D, AND STAEHR, K. (2012): "Consumption sensitivities in Estonia: income shocks of different persistence". Bank of Estonia Working Papers, No. 3/2012.

KULIKOV, D., PAABUT, A. AND STAEHR, K. (2009): "A micro-econometric analysis of household saving in Estonia: income, wealth and financial exposure". In D. Mayes (ed.). *Microfoundations of Economic Success: Lessons from Estonia*, Cheltenham: Edward Elgar, pp. 190–241.

LEETMAA, R. AND NURMELA, K. (2010): "EEO Review: Self-employment, 2010. Estonia". European Employment Observatory. http://www.eu-employment-observatory.net/resources/reviews/Estonia-SERvw2010.pdf.

LEPING, K. O. AND TOOMET, O. (2008): "Emerging ethnic wage gap: Estonia during political and economic transition". *Journal of Comparative Economics*, Vol. 36, No. 4, pp. 599–619.

LYSSIOTOU, P., PASHARDES, P. AND STENGOS, T. (2004): "Estimates of the black economy based on consumer demand approaches", *Economic Journal,* Vol. 114, No. 497, pp. 622–640.

MERIKÜLL J. AND STAEHR, K. (2010): "Unreported employment and envelope wages in mid-transition: comparing developments and causes in the Baltic countries". *Comparative Economic Studies*, Vol. 52, No. 4, pp. 637–670.

MINCER, J. (1974): *Schooling, Experience, and Earnings*, New York: NBER Press.

MIRUS, R. AND SMITH, R. (1996): "Self-employment, tax evasion, and the underground economy: micro based estimates for Canada". Working Paper, No. 1002, International tax program, Harvard Law School.

PISSARIDES, C. AND WEBER, G. (1989): "An expenditure based estimate of Britain's black economy". *Journal of Public Economics*, Vol. 39, No. 1, pp. 17–32.

SCHNEIDER, F. (2011): "The shadow economy and shadow economy labor force: what do we (not) know?". Discussion Paper, No. 5769, Institute for the Study of Labor (IZA).

SCHNEIDER, F., BUEHN, A. AND MONTENEGRO, C.E. (2010): "New estimates for the shadow economies all over the world". *International Economic Journal*, Vol. 24, No. 4, pp. 443–461.

SCHUETZE, H. (2002): "Profiles of tax noncompliance among the self employed in Canada: 1969–1992". *Canadian Public Policy*, Vol. 28, No. 2, pp. 219–237.

SLEMROD, J. (2007): "Cheating ourselves: the economics of tax evasion". *Journal of Economic Perspectives*, Vol. 21, No. 1, pp. 25–48.

TAFENAU, E., HERWARTZ, H. AND SCHNEIDER, F. (2010): "Regional estimates of the shadow economy in Europe". *International Economic Journal*, Vol. 24, No. 4, pp. 629–636.

WANGEN, K. R. (2005): "An expenditure based estimate of Britain's black economy revisited". Discussion Papers, No. 414, Research Department of Statistics Norway.

WILLIAMS, C. C. (2008): "Envelope wages in Central and Eastern Europe and the EU". *Post-Communist Economies*, Vol. 20, No. 3, pp. 363–376.

WILLIAMS, C. C. (2009): "The commonality of envelope wages in Eastern European economies". *Eastern European Economics*, Vol. 47, No. 2, pp. 37–52.

WOOLDRIDGE, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*, Cambridge MA: MIT Press.

Appendix A

Table A.1: Full estimation of IV regression model

	First stage IV regression	**Second stage IV regression**
Dependent variable:	$\log y_i^{\text{regular}}$	$\log c_i$
β (regular income)		0. 612*** (0.039)
Household education level 2	0.116*** (0.032)	..
Household education level 3	0.325*** (0.033)	..
Non-Estonian	–0.226*** (0.019)	..
Female	–0.160*** 0.015	..
Region 2	–0.251*** (0.021)	..
Region 3	–0.183*** (0.026)	..
Region 4	–0.219*** (0.026)	..
Region 5	0.208*** (0.022)	..
γ (business income dummy)	–0.212*** (0.025)	0.546*** (0.025)
Log temporary income	–0.076*** (0.015)	0.170*** (0.016)
Age	–0.001** (0.001)	0.002* (0.001)
Age^2	0.0002*** (0.000)	–0.0002*** (0.000)
Number of children	0.086*** (0.010)	0.087*** (0.010)
Two income earners	0.333*** (0.021)	–0.057*** (0.022)
Owning second real estate	–0.107*** (0.025)	–0.050** (0.025)

	First stage IV regression	Second stage IV regression
Renting the main residence	–0.077***	–0.051*
	(0.026)	(0.029)
Constant	8.447***	2.262***
	(0.082)	(0.331)
R^2	0.453	0.297
F-stat	40.49	..
[p-value]	[0.000]	
No. of obs.	3743	3743

*Notes: Full estimations of IV regression. 71 monthly dummies are included in the estimations but are not shown in the table. Robust standard errors are reported in round parentheses below the coefficient estimates. Superscripts ***, ** and * indicate that the coefficient is statistically different from 0 at the 1%, 5% and 10% level, respectively. F-stat is the test statistic for the Wald test of overall goodness of the fit of the first regression.*

Appendix B

Table B.1: Robustness test of the regressioon to different sets of control variables. Dependent variable: $\log c_i$

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
β (regular income)	0.602*** (0.032)	0.618*** (0.037)	0.603*** (0.038)	0.604*** (0.038)	0.604*** (0.038)
γ (dummy for business income)	0.555*** (0.026)	0.583*** (0.025)	0.551*** (0.025)	0.547*** (0.025)	0.547*** (0.025)
Temporary income and two income earners	No	Yes	Yes	Yes	Yes
Age and number of children	No	No	Yes	Yes	Yes
Renting the main residence and owning second real estate	No	No	No	Yes	Yes
Time fixed effects	No	No	No	No	Yes
R^2	0.213	0.244	0.278	0.278	0.278
No. of obs.	3743	3743	3743	3743	3743

*Notes: IV estimations. Education level, nationality, gender and regional dummies are used as instruments. Robust standard errors are reported in parentheses below the coefficient estimates. Superscripts ***, ** and * indicate that the coefficient is statistically different from 0 at the 1%, 5% and 10% level, respectively*

Working Papers of Eesti Pank 2013

No 1
Ardo Hansson, Martti Randveer
Economic Adjustment in the Baltic Countries

No 2
Philip Du Caju, Theodora Kosma, Martina Lawless, Julián Messina, Tairi Rõõm
Why Firms Avoid Cutting Wages: Survey Evidence from European Firms

No 3
Michael Ehrmann, Chiara Osbat, Jan Strasky, Lenno Uusküla
The Euro Exchange Rate During the European Sovereign Debt Crisis – Dancing to its Own Tune?

No 4
Liina Malk
Relaxation of Employment Protection and its Effects on Labour Reallocation

No 5
Guido Baldi, Karsten Staehr
The European Debt Crisis and Fiscal Reaction Functions in Europe 2000–2012