

Engström, Per; Hagen, Johannes; Johansson, Edvard

Working Paper

Estimating Tax Noncompliance among the Self-Employed – Evidence from Pleasure Boat Registers

Discussion paper, No. 144

Provided in Cooperation with:

Aboa Centre for Economics (ACE), Turku

Suggested Citation: Engström, Per; Hagen, Johannes; Johansson, Edvard (2021) : Estimating Tax Noncompliance among the Self-Employed – Evidence from Pleasure Boat Registers, Discussion paper, No. 144, Aboa Centre for Economics (ACE), Turku

This Version is available at:

<https://hdl.handle.net/10419/233359>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

*Per Engström, Johannes Hagen, and
Edvard Johansson*

**Estimating Tax Noncompliance
among the Self-Employed –
Evidence from Pleasure Boat
Registers**

Aboa Centre for Economics

Discussion paper No. 144

Turku 2021

The Aboa Centre for Economics is a joint initiative of the economics departments of the University of Turku and Åbo Akademi University.



Copyright © Author(s)

ISSN 1796-3133

Printed in Uniprint
Turku 2021

*Per Engström, Johannes Hagen, and Edvard
Johansson*

**Estimating Tax Noncompliance among the
Self-Employed – Evidence from Pleasure Boat
Registers**

Aboa Centre for Economics

Discussion paper No. 144

March 2021

ABSTRACT

In this paper, we improve upon the Pissarides-Weber (PW) method for estimating tax evasion among the self-employed by utilizing unique register-based consumption measures from the Swedish and Finnish mandatory registers for pleasure boats. Our main contributions are twofold: i) The PW method crucially relies on the assumption that the intrinsic consumption preferences are the same for self-employed and wage earner households. Thanks to our rich data, we are able to test this assumption directly by contrasting the Engel curves between wage earner households and households that will become self-employed in the future. We find that the group that will become self-employed is only marginally more prone to own a boat before they actually become self-employed. This lends support to the assumption of similar intrinsic preferences between the two groups. ii) The large sample size that comes with register-based consumption measures allows us to challenge the standard functional form assumptions in previous PW applications. The estimated sizes of income underreporting in absolute monetary amounts are almost constant over reported income levels, whereas previous studies have assumed that the underreporting is proportional to income.

JEL Classification: D12, H24, H25, H26

Keywords: tax evasion, income underreporting, pleasure boats, Engel curves, permanent income, self-employment

Contact information

Per Engström
Uppsala University (Sweden)
Email: per.engstrom@nek.uu.se

Johannes Hagen
Jönköping University (Sweden)
E-mail: johannes.hagen@ju.se

Edvard Johansson
Åbo Akademi University (Finland)
E-mail: edvard.johansson@abo.fi

Acknowledgements

We are grateful to Spencer Bastani, Janya Hamaaki, Julie Brun Bjørkheim and seminar participants at IIPF 2020, Jönköping University, University of Turku, the Finnish Tax Administration and the Swedish Tax Agency for helpful comments. We acknowledge financial support from the Nordic Tax Research Council and Stiftelsen Inger, Arne och Astrid Oscarssons Donationsfond.

1. Introduction

The consumption-based method for estimating underreporting among the self-employed (SE), introduced by Pissarides and Weber (1989) is widely used in the empirical literature on tax evasion/avoidance. The Pissarides-Weber (PW) method is based on using “excess consumption” (consumption conditional on reported income) among the self-employed as indirect evidence of income underreporting. This method typically relies on survey data on consumption, which greatly limits its applicability as consumption surveys tend to be small, irregular and suffer from high non-response rates as well as potential recall bias. Using register-based proxies for household consumption is thus a possibility in terms of improving this research. However, such proxies are generally very difficult to find, as most types of consumption are not registered, in addition to the fact that the data would also have to fit the PW methodology. For example, using data from car registers is sub-optimal since many of the self-employed use their car as an asset in their firm.

In this project, we use the mandatory pleasure boat registers in Finland and Sweden. These data fit the PW framework well for a number of reasons. First, the owner has no clear tax incentives to register the pleasure boat as an asset in the firm except when the boat is actually instrumental for operations (such as fishing and shipping, which we exclude). Second, pleasure boats are an important expenditure category, and next to Canada, the Nordic countries are the most pleasure-boat-dense countries in the world.² Third, bias resulting from selection and attrition should be small since it is currently (Finland) and used to be (Sweden) mandatory to register any pleasure boat satisfying certain length and engine power criteria. Furthermore, the registers are/were not used for tax purposes, in which case one could expect underreporting of boats to be correlated with underreporting of income.

We have access to the Finnish boat register from 2016, while the Swedish boat register was in place during 1990–1992. These data are matched with population-wide, administrative databases from each country to get information on the employment status and other demographic characteristics of the boat owners. Having access to similar registers in two countries means that we can compare and validate the results. As far as we know, this is the first study in the literature on income underreporting among the self-employed to use a population-wide register-based measure of consumption.

By now, there is a host of applications of the PW method from different points in time, with different modifications of the basic PW model and using data from different countries. Apart from the UK (Lyssirotou et al., 2004; Pissarides and Weber, 1989), the method has been used in, for instance, Sweden (Apel, 1994; Engström and Hagen, 2017; Engström and Holmlund, 2009), Finland (Johansson, 2005), Canada (Schuetze, 2002), South Korea (Kim et al., 2017), Estonia (Kukk and Staehr, 2014), Spain (Martinez-Lopez, 2013), the US (Hurst et al., 2014) and Norway (Nygård et al., 2019). These studies typically find evidence of substantial underreporting of income among the self-employed. Estimates typically range from about 20 to 50 percent of the total income being hidden. The most common consumption measure is food expenditure since it is

² Population size/number of pleasure boats (SweBoat, 2019).

sufficiently mundane for individuals not to be afraid of truthfully reporting it and because – with some exceptions – it is unlikely to be registered as a business expense.

The method has also spawned a number of innovative PW-related methods using a register-based traces-of-income approach in order to estimate underreporting. For example, Feldman and Slemrod (2007) use charitable contributions instead of consumption to infer the true income of the self-employed. Braguinsky et al. (2014) use the market value of cars as an alternative consumption measure in a modified PW setting, allowing for hidden incomes among many different types of workers (i.e., not only the self-employed). Finally, Artavanis et al. (2016) use microdata on household credit from a Greek bank and replicate the bank underwriting model to infer the banks' estimates of individuals' true income.

Compared to these previous studies, our main contributions are twofold. Both contributions are enabled by the fact that we have population-wide, longitudinal data on reported incomes, SE-status and a consumption measure (pleasure boats). To the best of our knowledge, this is a unique feature of our study. First, the estimated sizes of income underreporting are roughly in line with previous estimates (20–30 percent of average reported disposable income in previous Swedish and Finnish studies). However, we find that the estimated underreporting in absolute monetary amounts is roughly constant over reported income levels, while previous PW studies have assumed that underreporting is proportional to income. What enables us to make a much more informed choice regarding functional form is the statistical power that comes from measuring N in millions instead of thousands.

Second, access to longitudinal data on income and employment status allows us to address a central critique against the PW method, namely that potential “excess consumption” among the self-employed can be the result of heterogeneous preferences rather than income underreporting. We do this by comparing Engel curves between ordinary wage earner (WE) households and WE households that will switch to become SE households in the near future. If SE households had an intrinsically higher preference for pleasure boats, we would estimate an excess consumption for the soon-to-be SE households compared to stable WE households. Our findings indicate that the future SE group is actually slightly more prone to own a boat compared to the ordinary WE group. The excess boat ownership is roughly 5 times lower (in both countries) for the future SE group compared to the current SE group. One interpretation of this analysis is thus that a conservative, *preference heterogeneity-adjusted* PW estimate of underreporting is around 80 percent of the standard estimate. Performing this type of analysis has not been possible in survey-based applications of the PW method since it requires a much larger dataset than what is offered in a typical survey. A well-powered PW study is based on around 5,000 responding households. Typically, around 10 percent of these are coded as SE households (i.e., around 500 households). The number of WE households in the survey switching to SE the following year is typically lower than 50 households per year. When this is combined with a rather noisy survey-based consumption measure, there is simply not enough statistical power to distinguish between the consumption patterns of the two WE groups. Furthermore, this exercise relies on a panel dimension in the SE measure, which typically requires a register-based indicator of SE status (few consumption surveys are balanced panels).

The rest of the paper is structured as follows. In Section 2, we describe the basic PW method and how we modify it so that we can use pleasure boats instead of food consumption (the standard consumption measure in PW studies). We also discuss and justify our choice of functional form. In Section 3, we describe the pleasure boat and income data from each country, as well as sample restrictions and key variables. Section 4 presents and compares PW estimates of income underreporting. Section 5 presents various robustness tests, including the preference-adjusted PW estimates. Section 6 includes a heterogeneity analysis for different sectors and Section 7 concludes the paper.

2. Method

2.1 The basic PW method

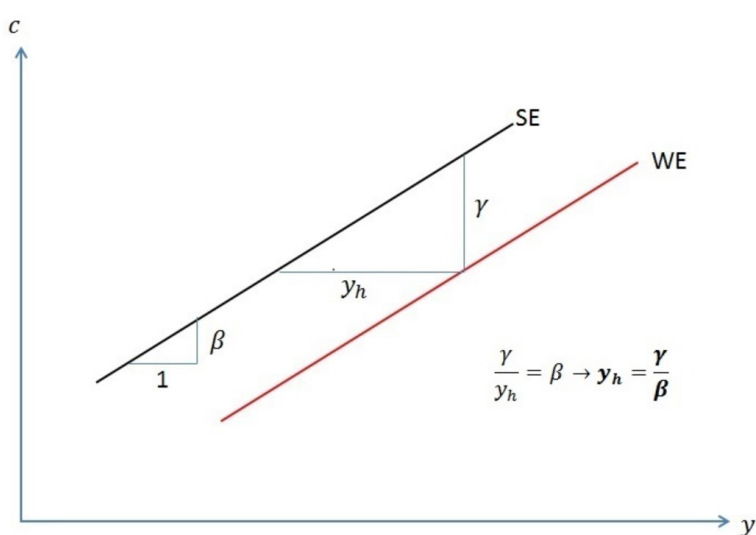
The consumption-based method for estimating underreporting among the self-employed introduced by Pissarides and Weber (1989) is frequently used in the empirical literature on tax evasion/avoidance. The PW method is based on using excess consumption among the self-employed as indirect evidence of income underreporting. The analysis is typically conducted at the household level and uses different survey measures of food consumption as the outcome.

The amount of underreporting is retrieved by estimating the following standard Engel curve:

$$c_{it} = \alpha \bar{X}_i + \beta y_{it} + \gamma SE_{it} + \varepsilon_{it} \quad (1)$$

where c_{it} is log consumption of household i at time t , y_{it} log of disposable income, \bar{X}_i represents factors affecting consumption and SE_{it} is a dummy for self-employed households. Assuming that preferences for consumption, conditional on disposable income and covariates, are equal for the self-employed and wage earners, and that the self-employed systematically underreport their income by a constant factor, the amount of underreporting (in logged form) is given by $y_h = \frac{\gamma}{\beta}$ (see Figure 1 below). This implies that the underreported income as a share of the true income is given by $1 - \kappa$ where we can estimate $\hat{\kappa} = \exp(-\frac{\hat{\gamma}}{\hat{\beta}})$. The log-log specification of Equation 1 above implies that SE households underreport a constant share of the disposable income.

Figure 1. Illustration of the basic PW method.



2.2 Modified method based on pleasure boats instead of food consumption

Instead of food consumption as the dependent variable, we use an indicator for boat ownership on the left-hand side of Equation (1), which clearly invalidates the use of a standard log-log specification. However, the choice of whether or not to log the income measure still remains. In Section 4.1, we show that using nominal income, as opposed to logged income, clearly fits the data better. When it comes to both Sweden and Finland, we get almost linear Engel curves for both groups (SE and WE) when using the nominal income measure (in EUR or SEK depending on country). Thus, the modified version of Equation (1) that we estimate is:

$$boat_{it} = \alpha \bar{X}_i + \beta y_{it} + \gamma SE_{it} + \varepsilon_{it} \quad (2)$$

where $boat_{it}$ is boat ownership of household i at time t and y_{it} is a nominal income measure.

This choice of functional form of the estimated Engel curves has important implications for the estimated underreporting. The standard log-log specification implies that the *share* of unreported income is assumed to be independent of income. The validity of this assumption is often tested by adding an interaction between the SE indicator and the (log) income measure in Equation 1 above (see, for instance, Engström and Hagen, 2017 and Hurst et al., 2015). If the estimated interaction is negative (positive), it implies that the share of underreported income decreases (increases) in income.

Typically, the formal test does not reject the null of equal slopes. However, the standard PW applications are based on relatively small samples, which gives the above test relatively low power. Both Engström and Hagen (2017) and Hurst et al. (2014) estimate (insignificant) negative interaction terms. Furthermore, Kukk et al. (2020) examined the corresponding relationship in a large number of European countries and found that the slopes of the Engel curves (log-log specification) are usually lower for the SE group than for the WE group. There is thus suggestive evidence that the standard assumption, that the share of underreporting is independent of income, is invalid. Our choice of functional form (i.e., using nominal income measures on the X-axis) instead implies that the *nominal underreporting* is independent of income, which is clearly consistent with a decreasing share of hidden income as income increases.³

By visually inspecting the graphs in the results section below, we see that the slopes of the Engel curves based on nominal income are almost identical for the SE and WE groups. In the results section, we also perform the corresponding formal tests and discuss the choice of functional form in more detail. Since we do not rely on small sample survey

³As noted by Braguinsky et al. (2014), independence of nominal underreporting and income is also consistent with a simple theoretical model of tax evasion. The argument is as follows: let the objective function of a taxpayer be $txY - aC(xY)$ where t represents the tax rate, x is the share of hidden income (choice variable), Y is the (exogenous) income, a is a parameter capturing the risk of getting caught and $C(\cdot)$ is a convex cost of getting caught. The first order condition directly gives that $xY = C'^{inv}(\frac{t}{a})$, which implies that the choice of nominal underreporting will be independent of income.

data, the statistical power of these tests is substantially higher compared to previous studies.

Apart from this modification of the functional form, our application of the PW method involves the usual interpretation. This means that we attribute any excess ownership of pleasure boats to income underreporting, conditional on reported disposable income (and a number of covariates). We match the data with register data on demographics and incomes in the respective country. This provides us with extensive panel data, thus enabling us to use measures of permanent household income in the analysis along the lines of Engström and Hagen (2017), which limits the need to find instruments for current income in the modified Engel curves.

3. Data

3.1 Pleasure boat register data

We use administrative data on pleasure boat ownership in two countries: Sweden and Finland. In Finland, pleasure boat owners are required to submit information about their boat(s) to the Finnish Communications Agency (TRAFI). The purpose of this register is to improve the safety of water traffic and facilitate control and rescue operations.⁴ It is also used for planning the use of Finnish water areas. One could argue that it is beneficial that the register is not used for taxation purposes. If it were to be used for these purposes, there would be a risk that underreporting income would be highly correlated with underreporting boat ownership. We have access to the Finnish boat register from 2016. In this year, there were around 204,000 registered pleasure boats in Finland.

The Swedish boat register was in place during 1988–1992 and we have access to the register from 1991. The register was implemented primarily for maritime safety and control reasons. The abolishment of the register was disputed, but critics argued that it infringed on personal integrity, was difficult and expensive to administer and even facilitated boat thefts (Motion 1989/90:T633). There were around 300,000 registered pleasure boats in Sweden during this period.⁵

The boat registers cover all pleasure boats satisfying certain length and engine power criteria.⁶ Hence, these registers include different types of boats, such as sailboats, powerboats and jet skis.

Apart from size (length and width) and engine power, the registers also provide boat-level information on production year, date when the boat was purchased by the current owner and location of the boat (municipality). There is no information on the estimated value of the boat nor the purchase amount. Instead, we use engine power and size measures as proxies for value in the robustness section below (see section 5.3 below). Boat owners are identified via a unique identification number that can be linked to other administrative databases (see 3.2).⁷

⁴ <https://www.traficom.fi/en/transport/boaters/watercraft-register>

⁵ The boat register is preserved by the Swedish National Archives (Riksarkivet): <https://riksarkivet.se/>

⁶ In Sweden, boats that were either 5–6 meters or powered by an engine of 10–15 kW had to be reported. The requirements in Finland are about the same (5.5 meters or 15 kW).

⁷ For households owning more than one boat, we use the boat with the most recent production year.

3.2 Income data

To calculate household incomes, we use register-based longitudinal databases from each country. Nordic register data on income are of very high quality due to the fact that they are automatically reported by third parties (for wage earners) and are reported separately for different types of income. Since the longitudinal income data are at the individual level, we aggregate income for the members of a given household to get the household income.

3.3 Key variables and sample restrictions

The three key variables are pleasure boat ownership, annual disposable income and self-employment status.

Boat ownership is a dummy variable taking the value of one if at least one household member is registered as a pleasure boat owner in year t .

In the Swedish case, current disposable income is defined as the household's disposable income in Statistics Sweden's IOT database.⁸ In the case of Finland, the corresponding data come from Statistics Finland's FOLK database.⁹ Disposable income is based on all types of (register-based) income, including transfers, income from labor and self-employment and capital income.

We use past and future income records to create multiple-year average measures of income. This approach has been used in the literature to account for the fact that transitory income fluctuations may attenuate the estimate of the income elasticity of consumption, which, in turn, leads to overestimating the degree of income underreporting (see, for instance, Engström and Hagen, 2017). Specifically, for each household in year t , we compute income measures that average income between $t - 3$ and $t + 3$. In the case of Finland, this concerns only the years $t - 3$ and $t + 2$ since the income data end in 2018.

The self-employment status of the household is based on information in the income register data. We define self-employed households as households where at least one of the adult members either report a positive income from self-employment or is considered linked to a closely held corporation (Johansson, 2007; Engström and Hagen, 2017).

From the full population of households in each country, we make four sample restrictions. First, we restrict the sample in each country to households where the oldest individual is between 18 and 64 years of age. In most cases, the oldest individual is also the registered boat owner. The individual variables such as age, gender and education pertain to the oldest member of the household in the subsequent analysis, whereas all income measures pertain to the entire household. Note that we do not drop single-person households. Second, we restrict the sample to households where the composition of the adult members does not change over the relevant time ($t - 3$ to $t + 3$ for Swedish households and $t - 3$ to $t + 2$ for Finnish households). As shown in the next sub-section, between one-third and half of the households are defined as not stable and are thus

⁸ <https://www.scb.se/vara-tjanster/bestalla-mikrodata/vilka-mikrodata-finns/individregister/inkomst--och-taxeringsregistret-iot/>

⁹ https://www.stat.fi/tup/mikroaineistot/aineistot_en.html

dropped from the sample.¹⁰ The main reason for restricting our analysis to what we henceforth refer to as “stable households” is for our multiple-year income measure to be comparable across households. Third, we drop households where at least one adult member is employed or self-employed in a boat-related sector, such as sea transport, ship dealing and ship renovation. Fourth, we keep households with incomes between the 5th and 95th percentiles (based on the unrestricted population).

3.4 Descriptive statistics

The first part of Table 1 reports descriptive statistics for wage earners and the self-employed in Sweden in 1991. We note that on average, the self-employed are older, more likely to be male and married while also having higher incomes than wage earners. Interestingly, the self-employed individuals are almost twice as likely to own a boat – 11 percent of self-employed households owned a boat in 1991 compared to 6 percent of employed households.

The second part of Table 1 reports corresponding statistics for Finnish households. The demographic patterns are very similar to those in Sweden – self-employed are on average older, have higher incomes and are more likely to be male and married compared to wage earners. The boat ownership rates are slightly higher in Finland but just as in Sweden, the SE group is more prone to owning a boat. The differences in boat ownership rates between Sweden and Finland get smaller as we include the non-stable households (see Table A 1 in the Appendix). The reason for this is that the stability criterion removes relatively more households in Sweden since it applies to more years in Sweden compared to Finland.

¹⁰ Our results are robust to including non-stable households.

Table 1. Descriptive statistics for Swedish and Finnish households

VARIABLE		Sweden (1991)		Finland (2016)	
		WE	SE	WE	SE
Boat owner	Mean	0.06	0.11	0.08	0.14
	SD	0.24	0.31	0.27	0.34
Current income	Mean	170,004	189,824	46,048	51,246
	SD	75,831	79,651	20,340	20,687
Permanent income	Mean	155,889	178,753	45,812	52,293
	SD	67,448	80,721	20,575	21,308
Age	Mean	41	45	47	49
	SD	13	11	11	9
Female	Mean	0.30	0.17	0.36	0.17
	SD	0.46	0.38	0.48	0.38
Elementary school	Mean	0.31	0.39	0.11	0.17
	SD	0.46	0.49	0.31	0.38
Upper secondary school	Mean	0.49	0.44	0.45	0.53
	SD	0.50	0.50	0.50	0.50
College 2 years	Mean	0.12	0.09	0.15	0.14
	SD	0.32	0.29	0.36	0.35
College 3 years or more	Mean	0.09	0.09	0.29	0.16
	SD	0.29	0.29	0.45	0.36
Household size	Mean	1.77	2.38	2.29	2.85
	SD	1.21	1.36	1.45	1.50
Married	Mean	0.37	0.64	0.44	0.63
	SD	0.48	0.48	0.50	0.48
<i>Number of households</i>		1,747,040		594,388	

Note: This table reports summary statistics for the self-employed and wage earners in Sweden and Finland, respectively. SD = standard deviation. The sample is restricted to stable households where the oldest member is between 18 and 64 years of age and with a combined income between the 5th and 95th percentiles (see 3.3). The income measures for Sweden and Finland are denoted in SEK (EUR 1 \approx SEK 7.75 in 1991) and EUR (current prices), respectively.

4. Results

4.1 Graphical results and choice of functional form

Figures 2 and 3 plot the relationship between permanent disposable income and boat ownership in Sweden (1991 data) and Finland (2016 data), respectively. Specifically, we plot boat ownership for equally spaced bins between the 5th and 95th percentiles.

These strikingly similar Engel curves reflect two interesting patterns. First, we see that the self-employed are more likely to own a boat at all income levels, thus indicating substantial underreporting.¹¹ We estimate the degree of underreporting in more detail in the next section (4.2).

Second, the linearity of the Engel curves, as well as the similar slopes, suggest that the self-employed underreport a certain amount of money, rather than a certain share of their income (note that we do not have logged values on the axes, as opposed to most previous studies). We thus provide new evidence on the functional form of underreporting, which has been a difficult task for previous studies using survey data including a quite small number of observations.

We conduct two additional analyses to validate our choice of a nominal income measure. First, when we plot the Engel curves based on logged income, we clearly see that the standard assumption of linear Engel curves is violated in both countries (Figures A2 and A3 in the Appendix). Second, we formally test the equality of slopes assumption in Tables A2 and A3. For Sweden, we find that the slopes are almost identical: 1.2 percent higher for SE based on the current income measure (statistically significant) and only 0.2 percent higher for SE based on the permanent income measure (statistically insignificant). The differences are somewhat greater in Finland: 6.4 percent lower for SE based on current income and 7.0 percent lower for SE based on the permanent income measure (both significant at conventional levels).

Recall that the standard log-log Engel curves presume that SE households underreport a constant share of true income. In unlogged form, such a relationship would imply a steeper slope for the SE group compared to the WE group – the two curves need to diverge in order for the relative income underreporting to be constant. The Finnish pattern, with a slightly flatter Engel curve for the SE group, instead implies that high-income households actually underreport a lower *absolute amount* compared to the low-income self-employed. We thus proceed under the assumption that there are no economically significant differences in the slopes of the Engel curves between the two groups. This is certainly true for the Swedish data, and based on the Finnish pattern, this is a conservative, rather than bold, assumption.

¹¹ The corresponding Engel curves using current disposable income instead of permanent income are presented in Figures A3 and A4 in the Appendix.

Figure 2 Boat ownership share and permanent disposable income (SEK) among Swedish households in 1991

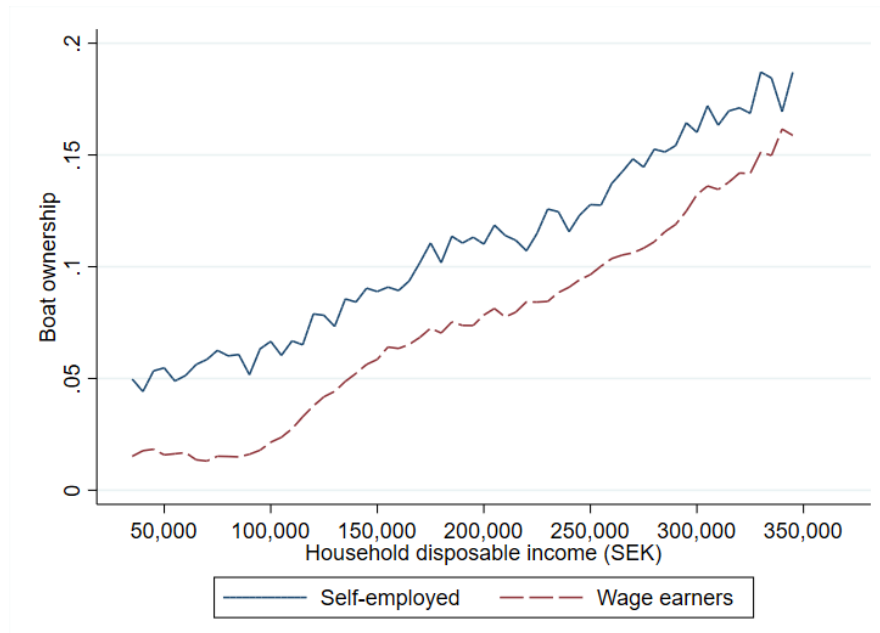
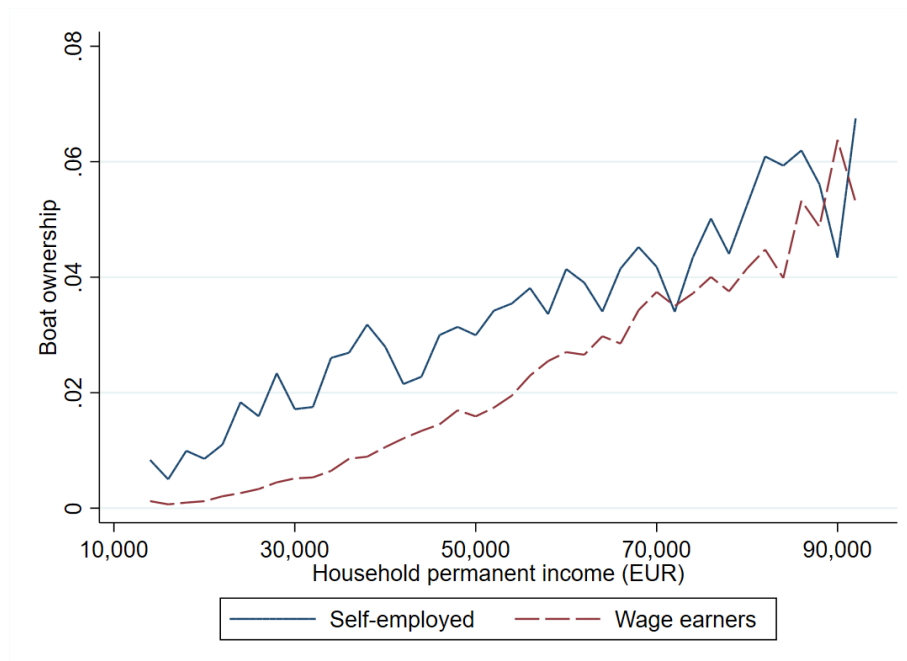


Figure 3. Boat ownership share and permanent disposable income (EUR) among Finnish households in 2016



4.2 PW estimates of boat ownership and income underreporting

In this section, we present the results from estimating Equation (1). That is, we regress a dummy for boat ownership on annual disposable income and a set of control variables. The control variables include age, gender, level of education, sector affiliation of the oldest household member, number of household members, marital status and

municipality of residence. Tables 2 and 3 report regression estimates for Sweden and Finland, respectively. We report results with/without controls and with current/permanent measures of household disposable income. The income for Sweden variables in Table 3 is expressed in 1000s of SEK in order to reduce the number of decimal points.

Recall that we do not regress the log of household income on the log of food expenditure as in the standard PW specification. Instead, we regress current (or permanent) household income on boat ownership (0/1). The implied underreporting is simply given by the estimate of the SE dummy (γ) divided by the estimated slope of the Engel curve (β).

In line with the previous literature, we find that for Sweden, the self-employment dummy as well as the measure for income are both positive and statistically significant at the 1% level. The implied estimates of underreported income are in the range of SEK 70,000–80,000 for the current income measure. This corresponds to approximately 24–27 percent of mean disposable income in Sweden in 1991.¹² For the permanent income measure, the implied underreporting is, as expected, somewhat lower. The implied underreporting is in the range of SEK 60,000–71,000 for the permanent income measure, or 21–24 percent of disposable income. As argued in Engström and Hagen (2017), the permanent income measure is preferred since the current measure may suffer from an attenuation bias that risks overestimating the hidden incomes.

For Finland, we find estimates of underreporting also in the range of 20–30 percent of disposable income. For our preferred estimate based on permanent income in column (4), self-employed households on average underreported their income by about EUR 14,700, which is roughly 21% of the average household disposable income.

¹² We get this by dividing the estimated amount of underreporting by the sum of mean reported disposable income from Table 1 and the amount of underreporting. Specifically, for the estimates in column (2) we have $70,603 / (224,639 + 70,603) = 0.239$.

Table 2. Estimation results on Swedish sample, boats in 1991 register

VARIABLES	(1)	(2)	(3)	(4)
Self-employed	0.0361*** (0.000614)	0.0331*** (0.000735)	0.0330*** (0.000622)	0.0303*** (0.000733)
Current income	0.000486*** (2.72e-06)	0.000420*** (5.28e-06)		
Permanent income			0.000554*** (6.04e-06)	0.000515*** (1.70e-05)
Observations	1,747,040	1,607,542	1,747,040	1,607,542
R-squared	0.026	0.071	0.028	0.073
Controls	No	Yes	No	Yes
Underreporting (SEK)	74,145	78,843	59,543	58,845
Underreporting, std. err. (SEK)	1,367	1,971	1,404	2,396
% of true income	28.09	29.35	24.99	24.77

Note: This table shows the OLS estimates of the Engel curve from Equation (2) and the corresponding standard errors using the Swedish sample in 1991. The sample is restricted to stable (see 3.3) households where the oldest member is between 18 and 64 years of age and with a combined income between the 5th and 95th percentiles. The implied underreporting is given by the estimate of the SE dummy (γ) divided by the estimated slope of the Engel curve (β). We use the delta method to calculate the standard errors for the estimated underreporting. The list of controls includes age, gender, level of education level, sector affiliation of the oldest household member, number of household members, marital status and municipality of residence. The income measures are expressed in 1000s of SEK. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3. Estimation results on Finnish sample, boats in 2016 register

VARIABLES	(1)	(2)	(3)	(4)
Self-employment	0.0416*** (0.00136)	0.0370*** (0.00162)	0.0381*** (0.00136)	0.0346*** (0.00162)
Current income	2.41e-06*** (1.90e-08)	2.19e-06*** (3.21e-08)		
Permanent income			2.47e-06*** (1.89e-08)	2.45e-06*** (3.32e-08)
Observations	585,853	584,731	585,853	584,731
R-squared	0.036	0.086	0.039	0.088
Controls	No	Yes	No	Yes
Underreporting (EUR)	17,299	16,902	15,466	14,093
Underreporting, std. err. (EUR)	596	782	578	690
% of true income	25.24	24.81	22.83	21.23

Note: This table shows the OLS estimates of the Engel curve from Equation (2) and the corresponding standard errors using the Finnish sample in 2016. Sample restrictions, list of controls and estimation procedures are otherwise similar to Table 2.

5. Robustness – Evaluating the key assumptions

5.1 Preferences for pleasure boats

How can we be sure that these ownership differences reflect income underreporting among the self-employed? One of the main critiques against the PW method is that excess consumption among the self-employed could be due to differences in preferences. Perhaps self-employed individuals simply have a stronger preference for boats than wage earners?¹³ We address this issue by comparing boat ownership among households defined as wage earners in year t but which transition to self-employment in the near future to households that are wage earners in all observed time periods. The presumption is that the future self-employed resemble the current self-employed in terms of preferences for boats but do not (yet) have the opportunity to hide income.

Figure 4 (for Sweden) shows the Engel curves of employed households and future self-employed households. We define future self-employed as households with a member who was employed in 1991 but became self-employed within 4 years. Figure 5 shows the corresponding Engel curves for Finnish households. Since we only observe self-employment status up to 2018 (i.e., two years after the boat records), we define future self-employed households as those with a member who was a wage earner in 2016 but then transitioned to self-employment in 2017–2018. The income measure on the horizontal axis is permanent household disposable income.¹⁴

Both figures show a similar pattern: future self-employed households are roughly as likely to own a boat in the year of interest as wage earners across all income groups. The only exception is a slightly higher share of boat owners among the Swedish SE with the lowest income. We interpret these results as evidence that the differences in boat ownership observed in Figures 1 and 2 are unlikely accounted for by significant preference differences between SE and WE households.

We have also estimated these Engel curves in the PW regression framework used in 5.1. Reassuringly, the dummy indicating future self-employment is insignificant in the Finnish setting. For Sweden, the dummy coefficient is statistically significant but economically much less significant compared to the baseline results: around 0.005 compared to around 0.03 for the current SE group (see Table 2 versus Table 4).

¹³ Feldman and Slemrod (2007) and Glazer and Konrad (1996) explore the notion that charitable giving is a signal of wealth or income. Pleasure boats are a highly “visible” good that may signal status, which might be more important to the self-employed individuals.

¹⁴ The corresponding graphs for current income are presented in Figures A5 and A6.

Figure 4. Boat ownership and permanent disposable income among Swedish households in 1991. Self-employed refer to households defined as wage earners in 1991 but which became self-employed within 4 years



Figure 5. Boat ownership and disposable income among Finnish households in 2016. Self-employed refers to households defined as households that were wage earners in 2016 but become self-employed within 2 years



Table 4. Boat ownership and disposable income for future self-employed households in Sweden

VARIABLES	(1)	(2)	(3)	(4)
Future self-employed	0.00471*** (0.000978)	0.00546*** (0.000981)	0.00449*** (0.000983)	0.00514*** (0.000983)
Current income	0.000491*** (3.00e-06)	0.000448*** (5.97e-06)		
Permanent income			0.000567** * (6.55e-06)	0.000568*** (2.42e-05)
Observations	1,393,039	1,339,609	1,393,039	1,339,609
R-squared	0.024	0.066	0.025	0.068
Controls	No	Yes	No	Yes
Underreporting (SEK)	9,573	12,201	7,918	9,042
Underreporting, std. err. (SEK)	1,995	2,200	1,745	1,796
% of true income	4.801	6.040	4.366	4.955

Note: This table shows the OLS estimates of the Engel curve from Equation (2) and the corresponding standard errors using the Swedish sample in 1991. Future self-employed refers to households employed in 1991 but which became self-employed within 4 years. Sample restrictions, list of controls and estimation procedures are otherwise similar to Table 2.

Table 5 Boat ownership and disposable income for future self-employed households in Finland

VARIABLES	(1)	(2)	(3)	(4)
Future self-employment	0.00798 (0.00494)	0.00411 (0.00484)	0.00661 (0.00492)	0.00271 (0.00483)
Current income	2.46e-06*** (1.89e-08)	2.20e-06*** (3.21e-08)		
Permanent income			2.52e-06*** (1.88e-08)	2.48e-06*** (3.32e-08)
Observations	585,853	584,731	585,853	584,731
R-squared	0.034	0.084	0.037	0.087
Controls	No	Yes	No	Yes
Underreporting (EUR)	3,248	1,866	2,619	1,093
Underreporting, std. err. (EUR)	2,011	2,202	1,953	1,950
% of true income	6.359	3.755	5.149	2.215

Note: This table shows the OLS estimates of the Engel curve from Equation (2) and the corresponding standard errors using the Finnish sample in 2016. Future self-employed refers to households employed in 1991 but which became self-employed within 4 years. Sample restrictions, list of controls and estimation procedures are otherwise similar to Table 2

5.2 The firm as a saving vehicle

In this subsection, we address another important difference between SE and WE households that constitutes a challenge for the PW method. The SE households may have substantial savings within their firms. The Swedish tax rules for incorporated firms allow for accumulating profits within the firm that may be used for future dividends or salary to the owner (see, for instance, Alstadsæter and Jacob, 2017). Furthermore, the firm may be sold in the future, generating a large one-time spike in income for that year. These possibilities to legally delay incomes for SE households represent a potential threat to the PW method since the current income measure does not include these legally delayed incomes (see, for instance, Hurst et al., 2014). Simply put, the PW method cannot separate legally hidden (delayed) incomes from illegally hidden incomes, while we only want the method to pick up the latter.

The permanent income measure addresses this problem to some extent since it includes future years in the income measure. As seen in Table 2 and Table 3, the permanent income measure also renders lower estimates of hidden income compared to the current income measure. This is consistent with the firm working as a saving vehicle for the SE group, but it may also only indicate that the permanent income measure works as intended. As discussed at length in Engström and Hagen (2017), the main reason why the permanent income measure presents lower estimates of underreporting is that current income is a rather noisy measure. Noise (i.e., classic measurement errors) in the income measure will attenuate the slope of the Engel curve, which, in turn, leads to an upward bias in the estimated underreporting.

In this subsection, we thus create an alternative income measure that is as forward-looking as possible, thus directly addressing the problem of legally hidden (delayed) incomes among SE households. Data access confines us to only use the Swedish data in this analysis since we only have two years of future data for Finland. We define “future income” as the average income between 1991 (the boat register year) and 1995 (the last year in our data). The idea is to capture future dividends and salaries saved by SE households within their firm (or the whole firm being sold off). We proceed by estimating the same type of modified Engel curves as in Table 2 and 3 above.

The results are presented in Table 6 below. When switching from the symmetric permanent income measure to the forward-looking income measure, the estimated income underreporting *increases* rather than *decreases*. This evidence speaks against SE households building up substantial savings within their companies. The estimated underreporting is almost SEK 100,000 compared to around SEK 60,000 in the preferred baseline specification (column 4). We thus find no evidence that this asymmetry in saving techniques between SE households and WE households challenges the interpretation that the excess boat consumption among SE households is primarily due to non-compliance.

Table 6. Future income for Swedish households

VARIABLES	(1)	(2)
Self-employment	0.0379*** (0.000639)	0.0298*** (0.000734)
Future income	0.000450*** (1.25e-05)	0.000298*** (1.63e-05)
Observations	1,747,040	1,607,542
R-squared	0.021	0.070
Controls	No	Yes
Underreporting (SEK)	84,106	99,977
Underreporting, std. err. (SEK)	3,053	6,034
% of true income	33.20	37.14

Note: This table shows the OLS estimates of the Engel curve from Equation (2) and the corresponding standard errors using the Swedish sample in 1991. The income measure is an average of household income between 1991 and 1995. Sample restrictions, list of controls and estimation procedures are otherwise similar to Table 2.

5.3 Proxies for boat value

So far, our analysis has disregarded which type of boat the household owns. Boats that cost several million EUR are thus lumped together with boats that only cost a few hundred EUR. Ideally, we would have access to exact prices and operating costs for all boats in the registers. In practice, however, such prices and costs are extremely difficult to estimate due to the massive number of different types of boats in the registers.¹⁵

Instead, we use very crude proxies of the boats' prices and operating costs, including length, width, area (width*length) and engine power. These proxies for a boat's price and operating costs will replace the simple indicator variable on the left-hand side of the Engel curves estimated in this section. The interpretation is thus that a household that does not own a boat technically owns a boat with a length of zero meters and so on. When analyzing engine power, we drop all households owning a sailing boat from the analysis since engine power is a bad proxy for the cost of a sailing boat. Table 7 (Sweden) and 8 (Finland) below report the results using the four cost/price proxies. In order to save space, we limit the analysis to the preferred specification using the full set of controls and the permanent income measure.

The estimates of underreporting based on the three size-based proxies for boat price (length, width and width*length) are consistent with the corresponding estimates in Table 2 and 3 above. However, when it comes to engine power in Sweden (and to a lesser extent in Finland), we get much higher estimates than before. This may be because SE households have less leisure compared to WE households (see, for instance, Hyytinen

¹⁵ The boat model information in the data consists of a text field filled in by the owners themselves. The number of different entries is vast (>1,000). Furthermore, the production year is missing for a substantial share of the boats. This makes the task of estimating the prices and operating costs of the boats practically infeasible.

and Ruuskanen, 2007). Furthermore, sailing boats are arguably very time-consuming compared to motorboats.

Table 7. Proxies for boat value and permanent disposable income among Swedish households in 1991.

VARIABLES	(1) Engine power (kW)	(2) Width	(3) Length	(4) Area
Self-employed	2.572*** (0.0517)	0.0687*** (0.00167)	0.211*** (0.00492)	0.521*** (0.0127)
Permanent income	0.0227*** (0.000895)	0.00115*** (3.90e-05)	0.00344*** (0.000118)	0.00830*** (0.000295)
Observations	1,583,283	1,607,542	1,607,542	1,607,542
R-squared	0.034	0.070	0.069	0.060
Controls	Yes	Yes	Yes	Yes
Underreporting (SEK)	113,543	59,501	61,373	62,851
Underreporting, std err (SEK)	4,805	2,458	2,528	2,679
% of true income	38.85	24.97	25.56	26.01

Note: This table shows the OLS estimates of the Engel curve from Equation (2) and the corresponding standard errors using the Swedish sample in 1991. The previous outcome variable (boat ownership) is replaced by the boat-specific variable indicated in each column heading. Sailing boats are dropped in column (1). Sample restrictions, list of controls and estimation procedures are otherwise similar to Table 2

Table 8. Proxies for boat value and permanent disposable income among Finnish households in 2016.

VARIABLES	(1) Engine power (kW)	(2) Width	(3) Length	(4) Area
Self-employed	3.091*** (0.133)	0.0534*** (0.00358)	0.147*** (0.00995)	0.404*** (0.0273)
Permanent income	0.000172*** (2.82e-06)	5.62e-06*** (7.69e-08)	1.51e-05*** (2.14e-07)	3.62e-05*** (6.04e-07)
Observations	589,406	589,412	589,410	589,413
R-squared	0.068	0.089	0.086	0.051
Controls	Yes	Yes	Yes	Yes
Underreporting (EUR)	18,459	9,928	10,201	11,654
Underreporting, std. err. (EUR)	816	654	678	788
% of true income	26.70	16.38	16.76	18.70

Note: This table shows the OLS estimates of the Engel curve from Equation (2) and the corresponding standard errors using the Finnish sample in 2016. The previous outcome variable (boat ownership) is replaced by the boat-specific variable indicated in each column heading. Sailing boats are dropped in column (1). Sample restrictions, list of controls and estimation procedures are otherwise similar to Table 2

6. Heterogeneity – sectoral analysis

An important benefit of the large sample sizes resulting from population-wide registers is that the results can be broken down into sector-specific estimates of income underreporting. Our data contain sectoral codes for firms in both Sweden and Finland.

The Finnish classification follows the General Industrial Classification of Economic Activities within the European Communities (NACE). The classification of Swedish firms in 1991 follows an older version of NACE (NACE 1970). For each country, we group firms into nine sectors based on one-digit sectoral codes.¹⁶ We estimate Equation (1) after adding treatment effects between the self-employment dummy and each sector dummy and removing the constant. This means that the coefficient associated with each interaction term can be interpreted as the excess boat ownership among the self-employed in a certain sector compared to wage earners in the same sector. Tables 9 and 10 report the estimated coefficients of the interaction terms for Sweden and Finland, respectively.

For Sweden, we see that most interaction terms have positive and significant coefficients. Hence, the self-employed are more likely to own a boat compared to wage earners in the same sector across a wide spectrum of the economy. The coefficients reveal that the highest excess boat consumption among the self-employed is found in the sectors of *real estate, renting and business activities; construction; retail, hotels and restaurants* while the lowest excess consumption is found in the sectors *electricity, gas and water supply; mining; agriculture*. These patterns pass a “sanity check” since they are roughly consistent with evidence from the Swedish Tax Agency (Skatteverket, 2008) based on more direct approaches for measuring tax evasion.

For Finland, we see that the highest excess boat consumption among the self-employed is observed in the sectors of *construction; wholesale and retail trade; industry; professional industries*. The coefficients associated with the remaining sectors – except *agriculture* which is negative – are also positive and significant, thus also indicating income underreporting. In the case of Finland, these estimates also pass the same type of “sanity check” as in the Swedish case as they are roughly consistent with information from the Audit Committee of the Finnish Parliament (2010).

¹⁶ The Swedish sectoral data readily contains nine sectors at the one-digit level. In the Finnish data, we merge several sectors with few self-employed individuals so that we get 9 sectors in total (see Tables 9 and 10).

Table 9. Boat ownership and sector affiliation, Sweden

VARIABLES	(1)	(2)
Agriculture x SE	0.0142*** (0.00282)	0.0125*** (0.00274)
Mining x SE	-0.00924 (0.00989)	-0.00383 (0.00978)
Manufacturing x SE	0.0242*** (0.00172)	0.0292*** (0.00167)
Electricity, gas and water supply x SE	-0.0184** (0.00842)	-0.00542 (0.00816)
Construction x SE	0.0475*** (0.00269)	0.0498*** (0.00262)
Retail, hotels and restaurants x SE	0.0521*** (0.00194)	0.0461*** (0.00188)
Transport, storage and communication x SE	0.0243*** (0.00249)	0.0239*** (0.00243)
Real estate, renting and business activities x SE	0.0403*** (0.00228)	0.0363*** (0.00223)
Public administration x SE	0.0236*** (0.00125)	0.0190*** (0.00121)
Observations	1,601,986	1,601,486
R-squared	0.096	0.135
Controls	No	Yes

Note: This table shows the OLS estimates of the Engel curve from Equation (2) after adding treatment effects between the self-employment dummy and each sector dummy and removing the constant, using the Swedish sample in 1991. Sample restrictions, list of controls and estimation procedures are otherwise similar to Table 2.

Table 10. Boat ownership and sector affiliation, Finland

VARIABLES	(1)	(2)
A. Agriculture x SE	-0.0272*** (0.00409)	-0.0241*** (0.00397)
B+C+D+E Industrial sector x SE	0.0601*** (0.00598)	0.0602*** (0.00584)
F. Construction x SE	0.0404*** (0.00408)	0.0374*** (0.00399)
G. Wholesale and retail trade x SE	0.0666*** (0.00456)	0.0552*** (0.00448)
H. Transport and storage x SE	0.0384*** (0.00491)	0.0345*** (0.00480)
I. Accommodation and food service activities x SE	0.0266*** (0.00635)	0.00886 (0.00623)
J+K+L+M+N+O+P+S Professional services x SE	0.0484*** (0.00280)	0.0412*** (0.00275)
Q. Human health and social work activities x SE	0.0317*** (0.00542)	0.0191*** (0.00532)
R. Arts, entertainment and recreation x SE	0.0311*** (0.00889)	0.0253*** (0.00874)
Observations	565,025	565,025
R-squared	0.115	0.151
Controls	No	Yes

Note: This table shows the OLS estimates of the Engel curve from Equation (2) after adding treatment effects between the self-employment dummy and each sector dummy and removing the constant, using the Finnish sample in 2016. Sample restrictions, list of controls and estimation procedures are otherwise similar to Table 2.

7. Conclusion

In this paper, we have proposed the notion that pleasure boats may be used as consumption measures in the Pissarides and Weber (1989) framework for estimating hidden incomes among the self-employed. The novelty of this paper is that we have register-based data on consumption (i.e., data from the pleasure boat registers of Sweden and Finland) that we link to high-quality panel data on income from registers for the entire populations of Sweden and Finland. This makes our estimates more reliable. Most importantly, however, our sample size is also much larger, by a factor of around 100–200, compared to what has been the norm in previous studies using the PW methodology. Furthermore, the use of panel data on income also limits the need to instrument for current income when using the PW methodology (Engström and Hagen, 2017).

The rich data allow us to present two main contributions to the existing literature. First, we provide evidence in favor of self-employed and wage earners having the same intrinsic preferences for consumption – one of the main assumptions behind the Pissarides-Weber model. This assumption has been notoriously hard to evaluate

empirically. We analyze this by exploiting the panel dimension in SE status among the households, in combination with the large sample sizes. Specifically, we compare WE households that we know will become self-employed in the near future with ordinary WE households and find very small differences in boat ownership. This suggests that the excess boat consumption among SE households mainly manifests itself after the households have become self-employed. We calculate a *preference heterogeneity adjusted* PW estimate of underreporting that is around 80 percent of the standard estimate. Thus, we argue that about 20% of the underreporting found in the ordinary specification is due to differences in preferences between the self-employed and wage earners.

Second, owing to the large sample sizes, we are able to challenge the functional form assumptions made in traditional PW studies. The choice of functional form of the estimated Engel curves has important implications for the estimated underreporting. The standard log-log specification implies that the *share* of unreported income is assumed to be independent of income. Instead, we find that a specification with a boat ownership indicator as dependent variable and household disposable income level as explanatory variable fits our data better. Our specification instead presumes that the absolute amount of hidden income is independent of income. This implies that the proportion of income that is hidden is much higher for the households reporting the lowest income, a finding consistent with, for instance, Brewer et al. (2017) and Braguinsky et al. (2014).

Finally, our large sample size also allows us to provide information on income underreporting at a sectoral level. Our qualitative results from Sweden and Finland are in line with what other studies have found using more direct measures of the underground economy but also show that underreporting can be found in a variety of sectors.

References

- Alstadsæter, A. and Jacob, M. (2012). *Income Shifting in Sweden—An Empirical Evaluation of the 3:12 Rules. Report to the Expert Group for Public Economics.* Ministry of Finance, Stockholm.
- Apel, M. (1994). An Expenditure-Based Estimate of Tax Evasion in Sweden. In: *Essays on Taxation and Economic Behavior. Economic Studies 18.* Department of Economics, Uppsala University.
- Artavanis, N., Morse, A., and Tsoutsoura, M. (2016). Measuring income tax evasion using bank credit: Evidence from Greece. *The Quarterly Journal of Economics*, 131(2), 739–798.
- Braguinsky, S., Mityakov, S., and Liscovich, A. (2014). Direct estimation of hidden earnings: Evidence from Russian administrative data. *The Journal of Law and Economics*, 57(2), 281–319.
- Braguinsky, S. and Mityakov, S. (2015). Foreign corporations and the culture of transparency: Evidence from Russian administrative data. *Journal of Financial Economics*, 117(1), 139–164.
- Brewer, M., Etheridge, B., and O’Dea, C. (2017). Why are Households that Report the Lowest Incomes So Well-Off? *The Economic Journal*, 127(605), 24–49.
- Engström, P. and Hagen, J. (2017). Income underreporting among the self-employed: A permanent income approach. *European Economic Review*, 92, 92–109.
- Engström, P. and Holmlund, B. (2009). Tax evasion and self-employment in a high-tax country: Evidence from Sweden. *Applied Economics*, 41(19), 2419–2430.
- Feldman, N. E. and Slemrod, J. (2007). Estimating tax noncompliance with evidence from unaudited tax returns. *The Economic Journal*, 117(518), 327–352.
- Glazer, A. and Konrad, K. A. (1996). A signaling explanation for charity. *The American Economic Review*, 86(4), 1019–1028.
- Hirvonen, M., Lith, P., and Walden, R. (2010). Suomen kansainvälistyvä talous. Eduskunnan tarkastusvaliokunnan Julkaisu 1/10.
https://www.eduskunta.fi/FI/naineduskuntatoimii/julkaisut/Documents/trvj_1+2010.pdf
- Hurst, E., Li, G., and Pugsley, B. (2014). Are household surveys like tax forms? Evidence from income underreporting of the self-employed. *Review of Economics and Statistics*, 96(1), 19–33.
- Hyytinen, A. and Ruuskanen, O.-P. (2007). Time Use of the Self-Employed. *KYKLOS*, 60(1), 105–122.
- Johansson, E. (2005). An estimate of self-employment income underreporting in Finland. *Nordic Journal of Political Economy*, 31(1), 99–109.
- Kim, B., Gibson, J., and Chung, C. (2017). Using Panel Data to Estimate Income Under-Reporting by the Self-Employed. *The Manchester School*, 85(1), 41–64.
- Kukk, M. and Staehr, K. (2014). Income underreporting by households with business income: Evidence from Estonia. *Post-Communist Economies*, 26(2), 257–276.
- Lyssiottou, P., Pashardes, P., and Stengos, T. (2004). Estimates of the black economy based on consumer demand approaches. *The Economic Journal*, 114(497), 622–640.
- Martinez-Lopez, D. (2013). The underreporting of income by self-employed workers in Spain. *SERIEs*, 4(4), 353–371.
- Motion 1989/90:T633. Avskaffande av båtregistret.
https://www.riksdagen.se/sv/dokument-lagar/dokument/motion/avskaffande-av-batregistret_GD02T633

- Nygård, O. E., Slemrod, J., and Thoresen, T. O. (2019). Distributional implications of joint tax evasion. *The Economic Journal*, 129(620), 1894–1923.
- Pissarides, C. A. and Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, 39(1), 17–32.
- Schuetze, H. J. (2002). Profiles of tax non-compliance among the self-employed in Canada: 1969 to 1992. *Canadian Public Policy/Analyse de Politiques*, 219–238.
- Skatteverket (2008), *Skattefelskarta för Sverige*, Rapport 2008:1, <https://skatteverket.se/download/18.6704c7931254eefbe718000349/1359706120429/rapport200801skattefelskarta.pdf>
- SweBoat (2019) Fakta om båtlivet i Sverige 2019, <http://service.sweboat.se/file.aspx?afile=2385aaa5-fc30-49d6-b024-d981b8a3efa0>

Appendix

Figures

Figure A 1. Boat ownership and log disposable income for Swedish households, 1991



Figure A 2. Boat ownership and log disposable income for Finnish households, 2017

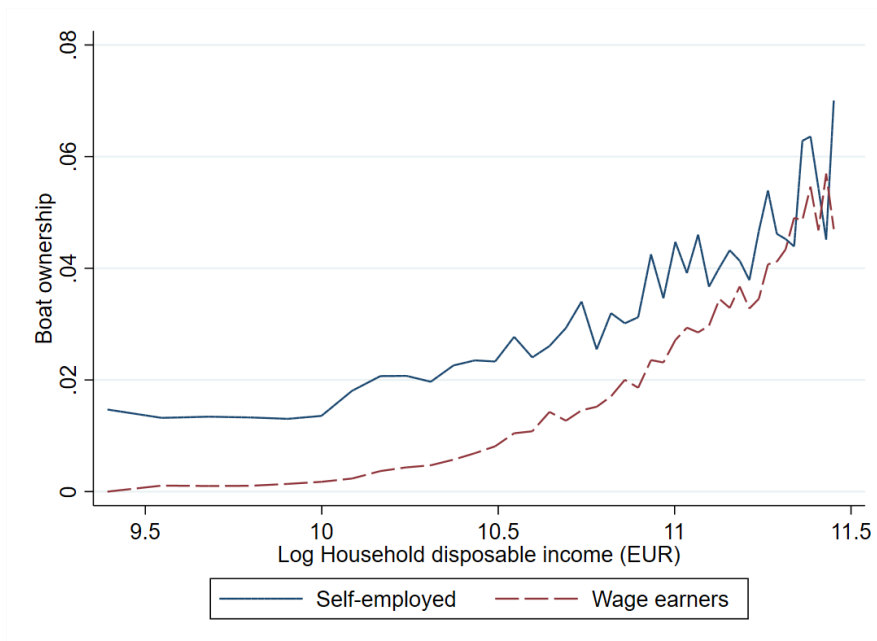


Figure A 3. Current income, Sweden

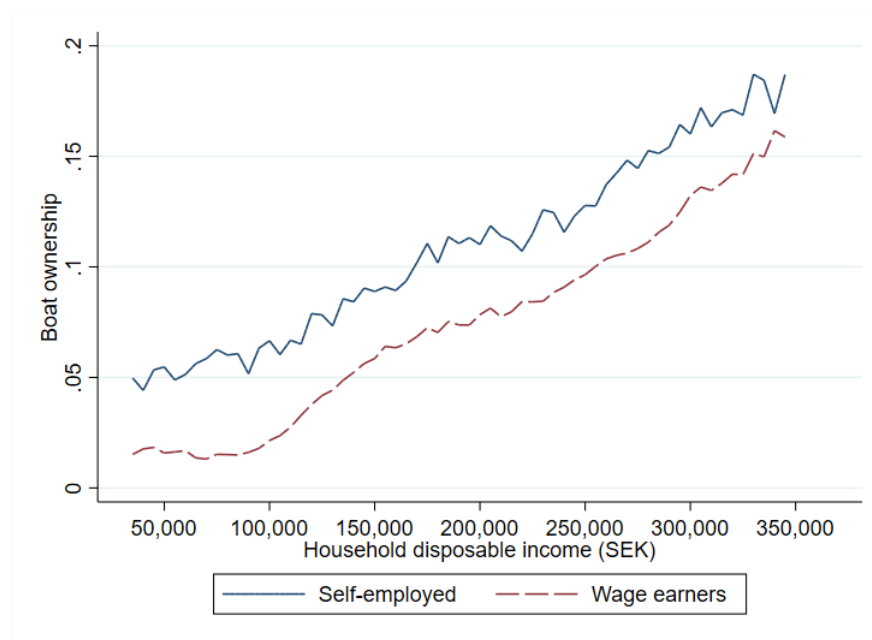


Figure A 4. Current income, Finland

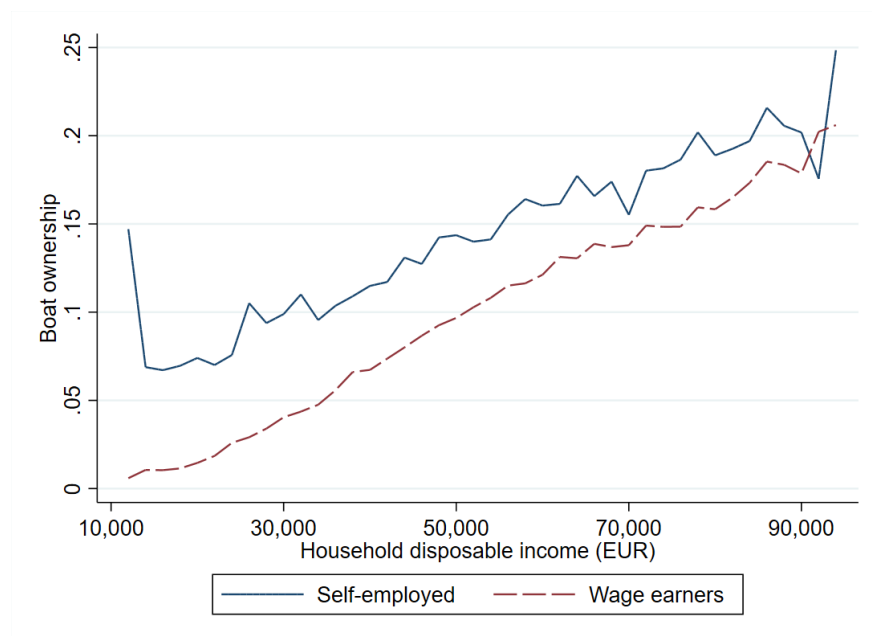


Figure A 5. Future SE current income, Sweden



Figure A 6. Future SE current income, Finland



Tables

Table A 1. Descriptive statistics for households, including non-stable households

VARIABLE		Sweden		Finland	
		WE	SE	WE	SE
Boat owner	Mean	0.05	0.10	0.06	0.12
	SD	0.22	0.30	0.25	0.33
Current income	Mean	159,688	186,566	43,195	47,578
	SD	75,527	79,313	19,230	20,180
Permanent income	Mean	148,072	176,564	43,018	48,667
	SD	67,857	81,803	19,346	20,658
Age	Mean	37	44	42	46
	SD	13	11	11	10
Female	Mean	0.39	0.23	0.36	0.17
	SD	0.49	0.42	0.48	0.38
Elementary school	Mean	0.28	0.36	0.12	0.18
	SD	0.45	0.48	0.32	0.38
Upper secondary school	Mean	0.52	0.45	0.47	0.52
	SD	0.50	0.50	0.50	0.50
College 2 years	Mean	0.11	0.10	0.12	0.13
	SD	0.32	0.30	0.32	0.34
College 3 years or more	Mean	0.10	0.10	0.30	0.18
	SD	0.29	0.30	0.46	0.38
Household size	Mean	1.77	2.32	2.25	2.67
	SD	1.14	1.31	1.34	1.45
Married	Mean	0.31	0.59	0.37	0.54
	SD	0.46	0.49	0.48	0.50
Stable household	Mean	0.55	0.66	0.48	0.55
	SD	0.50	0.47	0.50	0.50

Note: This table reports summary statistics for the self-employed and wage earners in Sweden and Finland, respectively. SD = standard deviation. The sample is restricted to (see 3.3) households where the oldest member is between 18 and 64 years of age and with a combined income between the 5th and 95th percentiles. The income measures for Sweden and Finland are denoted in SEK (EUR 1 \approx SEK 7.75 in 1991) and EUR (current prices), respectively.

Table A 2. Test equality of slopes, Sweden

VARIABLES	(1)	(2)
Self-employed	0.0227*** (0.00177)	0.0280*** (0.00403)
Current income	0.000409*** (5.41e-06)	
SE x current income	5.17e-06*** (9.12e-07)	
Permanent income		0.000511*** (1.74e-05)
SE x permanent income		1.22e-06 (2.17e-06)
Observations	1,607,542	1,607,542
R-squared	0.071	0.073
Controls	Yes	Yes
Underreporting (SEK)	55,406	54,792
Underreporting, std. err. (SEK)	4,209	7,643
% of true income	22.59	23.46

Note: This table shows the OLS estimates of the Engel curve from Equation (2) after adding an interaction term between the self-employment dummy and the income measure, using the Swedish sample in 1991. Sample restrictions, list of controls and estimation procedures are otherwise similar to Table 2.

Table A 3. Test equality of slopes Finland

VARIABLES	(1)	(2)
Self-employment	0.0310*** (0.00343)	0.0303*** (0.00341)
Current income	2.27e-06*** (3.35e-08)	
Current income x SE	-1.47e-07** (6.76e-08)	
Permanent income		2.55e-06*** (3.45e-08)
Permanent income x SE		-1.78e-07*** (6.67e-08)
Observations	585,848	585,848
R-squared	0.087	0.090
Controls	Yes	Yes
Underreporting (EUR)	13,660	11,900
Underreporting, std. err. (EUR)	1,474	1,305
% of true income	21.05	18.54

Note: This table shows the OLS estimates of the Engel curve from Equation (2) after adding an interaction term between the self-employment dummy and the income measure, using the Finnish sample in 2016. Sample restrictions, list of controls and estimation procedures are otherwise similar to Table 2.

The **Aboa Centre for Economics (ACE)** is a joint initiative of the economics departments of the Turku School of Economics at the University of Turku and the School of Business and Economics at Åbo Akademi University. ACE was founded in 1998. The aim of the Centre is to coordinate research and education related to economics.

Contact information: Aboa Centre for Economics, Department of Economics, Rehtorinpellonkatu 3, FI-20500 Turku, Finland.

www.ace-economics.fi

ISSN 1796-3133