TARC

Tax Administration Research Centre

Self-Employment Underreporting in Great Britain: Who and How Much?

Ana Cinta Gonzalez Cabral
University of Exeter

Gareth Myles
University of Exeter and Institute of Fiscal Studies

Christos Kotsogiannis
University of Exeter

Discussion Paper: 010-14

HM Revenue & Customs

HM TREASURY

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Self-Employment Underreporting in Great Britain: Who and How Much?

Ana Cinta G. Cabral
University of Exeter

Christos Kotsogiannis
University of Exeter

Gareth Myles
University of Exeter

October 6, 2015

Abstract

The present paper estimates the degree of income underreporting by the self-employed in Great Britain using the Pissarides and Weber (1989) approach. Firstly, it critically assesses whether the estimate obtained can be related to underreporting. Secondly, the paper extends the model to incorporate the analysis of the determinants of noncompliance based on characteristics found on the tax returns and also on other socioeconomic factors. The research finds income reported by the self-employed must be multiplied by a factor of 1.24 to obtain their true income, which translates into a lower bound estimate of the black economy of 1.6% of GDP. We find that self-employment underreporting varies by sex, age, type of occupation, region, education and characteristics based on the self-employment activity. These findings can provide guidelines to the tax administration about the profile of individuals that need to be addressed with their compliance activitites.

[0]Acknowledgements: The authors would like to thank the University of Exeter Business School faculty, attendants to the LAGV 2015, IIPF Annual Conference in Dublin 2015, IFS and HMRC for their comments on earlier versions of the paper. Funded by ESRC grant no. ES/K005944/1.

Contact details: anacintagcabral@gmail.com

1 Introduction

The self-employed have been found to markedly underreport their incomes to the tax authorities. In the UK, business taxpayers (self-employed and partnerships) have been found to be the largest contributors to the self-assessment tax gap which amounts to £4 billion in 2012.[1] In the US, around 57% of nonfarm proprietor income was found to be underreported in 2001, and around 30% in tax year 1992.

After the economic crisis, we have observed a positive trend in the number of self-employed individuals in all EU countries. This trend is especially pronounced in the UK: since the first quarter of 2008, the number of self-employed individuals has increased by 15% while the number of employees has increased only by 1%. Considering only the year from the first quarter of 2013, the UK had a growth rate of 8% in the number of self-employed which was the highest in the EU-15. Only Slovenia, Cyprus, Bulgaria, and Lithuania exhibited higher growth rates. The self-employed now represent 14% of the labour force in the UK and this share has seen steady growth since the beginning of the economic downturn.[2] Since it is becoming an increasingly common labour market choice, concern over the distortions introduced into the economy by the self-employed engaging in non-compliance with taxation has grown and with it the need to reassess the extent and causes of underreporting of income by the self-employed.

Our aims in the present paper are threefold. First, we aim to measure the extent of income underreporting by the self-employed in Great Britain using a generalisation of the methodology of Pissarides and Weber (1989). We make use of the Secure Access version of the Living Costs and Food Survey for the period 2010-2012 to provide alternative estimates using food expenditure, utilities expenditure, and expenditure on a basket of non-durables. The results from the three product groups are entirely consistent. We also use three different meaures of self-employment income and again find results in the expected direction. To date the use of this method has relied on a theoretical presumption that the outcome of the estimation is explained by underreporting. Our second aim is to assess whether the discrepancy observed could be due to other alternative reasons that have more to do with the possible heterogeneous behaviour of occupational groups and differences in preferences than with underreporting. We test three different possibilities: differing preferences for consumption, differing financial situations and funding of current spending and possible measurement errors in the survey. Thirdly, we extend the method to be able to explore the characteristics mostly associated with noncompliance.

The paper has two major innovations. Firstly, the analysis of the profile of the non-compliant self-employed is, to our knowledge, unprecedented which makes this paper the first of its kind. Surprisingly, the exploration of the characteristics of the noncompliant has been highly studied in the criminology literature while it has not aroused much interest in the economics literature. We should bear in mind however that the theory of crime cannot be directly applied to tax compliance due to the complexity of the latter as Graetz and Wilde (1985) explain. However, some of the findings from both literatures are common when it comes to the characteristics of the individuals that commit crime. Despite having to take into account some reservations from the direct application of crime theory, we observe commonalities that are worth exploring and therefore can help build a theory of the determinants of tax non-compliance. Besides, the data used corresponds to survey data and not to the

[1] Source: "Measuring the Tax Gaps 2013-2014 edition. Tax Gap estimates for 2012-13."

[2] See Figure 2 and Figure 3 for a graph about the evolution of self-employed workers. Statistics are obtained using Eurostat Labour Force Survey.

administrative data which are typically used for this purpose. Using survey data instead of administrative data provides us with a wider range of observable characteristics for the individuals than those available from tax returns which enables a deeper investigation of the traits related to noncompliance.

The use of the Secure Access version of the Living Costs and Food Survey[3][4] provides us with a wide range of variables. The availability of this data permits the second innovation of this paper which is to critically assess the results of the expenditure-based method and to analyse how confident we can be the discrepancy observed is due to underreporting. There are several other explanations that challenge the assumptions of the model and which could be potentially biasing the results and leading to a falsely attributing to underreporting a simple failure of one of the assumptions of the theoretical model.

The main results of the paper can be quickly summarised. We find that, on average, the self-employed underreport their incomes by 19.3% which translates into a lower-bound estimate of the black economy (defined as unreported taxable income) of 1.6% of GDP. We find that this result is not caused by differing preferences between the self-employed and the employed, nor it is caused through savings funding current consumption. We do find that using documents used to report to inland revenue to fill in the income section of the survey leads to a higher level of underreporting being detected. Regarding the characteristics of the noncompliant individuals, we find the classical results that males underreport more than females and that individuals become more compliant as they age. We also find that the self-employed operating in partnerships underreport more than own-account self-employed. The number of self-employed in the household also matters: households with two self-employed underreport more than households of one self-employed. Also, those who regard themselves as having a job underreport less than those who consider themselves as owning a business.

The structure of the paper is as follows. Section 2 contains a brief literature survey. A description of the methodology is offered on section 3. Section 4 provides an outline of the data used and the estimation procedure. Section 5 provides the results followed by section 6 which concludes.

2 Measuring the black economy

Administrative and survey data are the sources of microdata available for the measurement of the black economy, defined as unreported taxable income. The US probably counts with the most complete and comprehensive dataset regarding noncompliance collected firstly under the Tax Compliance Measurement Program (TCMP) and then under the National Research Program (NRP). Bennet (2005) reports that using data from the National Research Program (NRP) for tax year 2001, gross receipts from Schedule C which corresponds to the one that the self-employed need to file, was the most misreported item with $48.6 million underreported. It was also ranked number one by the 1988 Tax Compliance Measurement Program (TCMP) with $20.6 million underreported.

Using audit data to investigate tax evasion has mainly two difficulties despite it constituting a more complete, direct, and detailed data source than surveys. The first is that some audits are risk-based and not completely random so care should be taken when grossing-up the results to be representative of the population. For example, for NRP 2001 the IRS

[3] Office for National Statistics,Department for Environment, Food and Rural Affairs. (2014). Living Costs and Food Survey, 2006-2012: Secure Access. [data collection]. 4th Edition. UK Data Service. SN: 7047

[4] The original data creators, depositors or copyright holders, the funders of the Data Collection (if different) and the UK Data Archive bear no responsibility for the analysis hereby presented.

oversampled high-income returns and individuals filing a Schedule C. The second difficulty relates to the fact that the amount of underreporting discovered in audits depends on the ability of the examiner (this is the *detection rate*). However, to ameliorate the latter a detection controlled methodology has been developed by Erard and Feinstein (2012).

The second kind of micro methods uses surveys. This kind of data source also has its risks as people are normally wary of answering questions regarding their involvement in evasion activities. Responses to surveys can then be modified even if the individuals are assured of confidentiality and that the data will not be linked without their consent to administrative data. However, what we know for sure is that expenditure rather than income can constitute a more reliable source to produce an estimate of income underreporting. Individuals are more likely to modify their income received but are less likely to modify at the same time their expenditure.

One of the first papers exploiting the discrepancy between income and expenditure to measure the size of the black economy in the UK was that of Dilnot and Morris (1981). They computed excess expenditures for all 7200 households in the survey and characterised as "black-economy households" all those where expenditure exceeded income by 20% and by at least £3. They reported an upper bound and a lower bound distinguishing whether the discrepancy could be explained by the circumstances of the household or whether they included pensioner and unemployed households. They found that the black economy was between £3.2 and £4.2 billion representing around 2.3-3% of 1977 GNP. Most of the black economy households were headed by an individual working in a skilled or semiskilled occupation. 22% of the black economy households were headed by a self-employed individual. Dilnot and Morris calculated participation ratios for different observed characteristics. The self-employed were much more likely than other groups to be part of the black economy sample than other employees. Individuals in skilled or semiskilled manual occupations were more likely to participate in black economy activities than unskilled.

Pissarides and Weber (1989)[5] built on this idea and provided an improved method based on expenditure equations to measure the underreporting of the self-employed in Great Britain. The method tries to obtain a measure of income underreporting by the self-employed through a comparison of the relationship between food expenditure and income from this group to that of the employees who are assumed to be honest reporters. They recorded that true self-employment income is on average 1.55 times the income reported by the self-employed in Great Britain using the 1982 Family Expenditure Survey, with the uplift factor being higher in blue-collar households than in white-collar households (1.5 versus 1.6). Using this estimate of underreporting they obtained that the size of the black economy was 5.5% of GDP. Lyssiotou *et al.* (2004) using the same approach for 1992 estimated the coefficient of self-employment underreporting to be on average 1.28,[6] and again higher for blue-collar households (1.39) than white-collar (1.18). Besides, they also introduced a new methodology using complete demand systems for which they found that self-employment income should be multiplied by a factor of 2.18 for the case of blue-collar households and 1.64 for white-collar households. They estimated the size of the black economy to be 10.6% of GDP in 1993.

Similar studies have also been conducted in other countries. In Canada, Schuetze (2000) using the equivalent to the Family Expenditure Survey for 1969-1992 found an average underreporting factor of 1.2.[7] Hurst *et a*l (2014) for the US found that self-employment

[5] Pissarides and Weber will be shortened PW for the sake of simplicity

[6] This result was obtained averaging the values of blue collar and white collar workers.

[7] The term uplift factor, coefficient and underreporting are going to be used interchangeably. They both refer to the factor by which reported self-employment income needs to be multiplied by to obtain true income.

income is underreported by 30%.Enström and Holmlund (2009) find the need to scale self-employment income by a factor of 1.3 for Sweden. Johansson (2005) found that the self-employed underreport their incomes in Finland by 16.5% for households of one self-employed (42% for households of two self-employed) representing 1.3% (3.2%) of GDP. The results of these studies are summarized in Table 2. It should be noted that these estimates are not comparable between different countries or even for the same country for different years as macroeconomic conditions and samples as well as definitions of the variables used for the analyses vary.

3 Methodology

As outlined in the introduction, this paper is composed of three separate parts. The first part of this paper which aims at measuring the extent of underreporting of the self-employed in Great Britain employs the expenditure method outlined by Pissarides and Weber (1989). We also provide the estimation using an alternative specification described by Hurst *et al.* (2014). A summary of both methodologies will be provided in the present section. The second part critically assesses if there are other alternative explanations rather than underreporting that can be driving our results. The third part of the paper extends the PW model in order to provide an analysis of the characteristics that drive noncompliance. The underlying intuition of the model is kept, this is, we will still be using an expenditure based method but this time the interaction between the characteristics to be analysed and the self-employed dummy variable is included in order to capture the varying underreporting behaviours associated with them.

3.1 The Pissarides and Weber method

The model has two types of households: self-employed and employed households. All households, i, are assumed to report their expenditure on items, j, this is, C_{ij} correctly. Income, however, is assumed to be reported correctly by employed households, hence their true income Y_i equals their reported income $Y_i^{'}$, $Y_i = Y_i^{'}$. Self-employment income is suppose to be misreported. Thus, for self-employed households

$$Y_i = k_i Y_i^{'} \quad \text{where } k_i \geq 1. \tag{1}$$

where k_i is a random variable that captures the factor by which self-employed income has to be scaled to arrive to their true income. For the employed it follows that $k_i = 1$.

The expenditure function for household i for each item of expenditure j can be written as

$$\ln C_{ij} = \beta_j \ln Y_i^p + X_i \alpha_j + \varepsilon_{ij}, \tag{2}$$

where β_j is the elasticity of income for good j, Y_i^p represents permanent income, α_j is a vector of parameters, and ε_{ij} is a white noise error. It is assumed that permanent income, the measure of income that affects consumption, is related to measured income as

$$Y_i = p_i Y_i^p \tag{3}$$

where p_i is a random variable. It is assumed that both k_i and p_i are log-normally distributed

Now permanent income is

$$\ln Y_i^p = \ln Y_i^{'} - \ln p_i - \ln k_i. \tag{4}$$

It is assumed that the mean of p_i is the same for both households as their measured income should be affected by chance events in the same manner. However income from self-employment is more variable than income from employment. Thus,

$$\bar{p}_S = \bar{p}_E, \tag{5}$$

and

$$\sigma^2_{u_S} > \sigma^2_{u_E}. \tag{6}$$

As k_i and p_i are log-normally distributed we can write them as deviations from their means

$$\ln p_i = \mu_p + u_i, \tag{7}$$

$$\ln k_i = \mu_k + \nu_i, \tag{8}$$

where u_i and ν_i are random variables with zero mean and constant variances σ^2_u and σ^2_ν respectively.

Using the log-normality of p_i , the mean of p_i and the mean of its log μ_p are related as follows

$$\ln \bar{p} = \mu_p + \frac{1}{2}\sigma^2_u, \tag{9}$$

so knowing that the mean of p_i, $\bar{p}_i$, is the same for both groups and that the variance of u_i is larger for the self-employed than for employed households we find that

$$\mu_{p_S} - \mu_{p_E} = -\frac{1}{2}(\sigma^2_{u_S} - \sigma^2_{u_E}). \tag{10}$$

Substituting (4), (7) and (8) into (2) we get

$$\ln C_{ij} = X_i\alpha_j + \beta_j \ln Y_i^{'} - \beta_j(\mu_p - \mu_k) - \beta_j(u_i - v_i) + \epsilon_{ij}. \tag{11}$$

In order to estimate (11) income, $Y_i^{'}$, is treated as endogenous and instrumented. This provides an independent estimate of income variance for the two groups

$$\ln C_{ij} = X_i\alpha_j + \beta_j \ln Y_i^{'} + \gamma_j SE_i + \eta_i \tag{12}$$

where SE_i is a dummy variable that takes the value 1 if the individual is self-employed and zero otherwise.[8]

From the first-stage regressions of income on a Z_i set of identifying instruments, we can obtain an independent estimate of the variance of income for both occupational groups:

$$\ln Y_i^{'} = X_i\delta_1 + Z_i\delta_2 + \varsigma_i, \tag{13}$$

The residual ς_i is a composite of three errors: unexplained variations in permanent income, deviations from actual to permanent income, u_i, which is assumed to be the same for both groups, and deviations from actual to reported income, ν_i.

Assume that the unexplained variations of permanent income have the same variance for both groups. The residual income variance for the self-employed, $var\varsigma_{i_S}$, should then be larger than the residual income variance, $var\varsigma_{i_E}$, for the employed. This is because $var\varsigma_{i_S}$

[8] A definition of a self-employed household is provided in Section 4.

contains the variance of v_i, the underreporting component, and the variance of u_i is larger for the self-employed. Hence,

$$\begin{aligned} var\varsigma_{i_S} - var\varsigma_{i_E} &= \sigma^2_{Y_S} - \sigma^2_{Y_E} = var(u-v)_S - var(u)_E = \\ &= \sigma^2_{u_S} + \sigma^2_{v_S} - 2cov(uv)_S - \sigma^2_{u_E}. \end{aligned} \tag{14}$$

From (7), (8) and (10), the estimated coefficient γ_j in (12) is given by

$$\gamma_j = -\beta_j((\mu_{p_S} - \mu_{p_E}) - \mu_k) = \beta_j \left[\mu_k + \frac{1}{2}(\sigma^2_{u_S} - \sigma^2_{u_E})\right]. \tag{15}$$

The estimates that are of interest here are the mean value of k_i which informs about the factor reported income from those self-employed households need to be multiplied by in order to have an estimate of true income. By the log-normality of k_i,

$$\ln \bar{k} = \mu_k + \frac{1}{2}\sigma^2_{v_S}. \tag{16}$$

Substituting μ_k from (15) into (16) we obtain,

$$\ln \bar{k} = \frac{\gamma_j}{\beta_j} + \frac{1}{2}\left(\sigma^2_{v_S} - \sigma^2_{u_S} - \sigma^2_{u_E}\right). \tag{17}$$

Assuming that the covariance between u and v is zero, then we can just use (14) in (17) to estimate the level of $\bar{k}$. We cannot provide a certain value for $\bar{k}$ but we can provide a feasible range over which these values can lie, for which we are going to consider variations of $\sigma^2_{v_S}$ and $\sigma^2_{u_S}$. From (17) we can see that $\sigma^2_{v_S}$ and $\sigma^2_{u_S}$ are negatively related, therefore $\bar{k}$ will reach a lower bound when $\sigma^2_{v_S}$ takes its lowest value of 0. Then we can use (14) into (17) to obtain this lower bound, for which we will use the subindex l

$$\ln \bar{k}_l = \frac{\gamma_j}{\beta_j} - \frac{1}{2}(\sigma^2_{Y_S} - \sigma^2_{Y_E}). \tag{18}$$

We can also provide an upper bound for $\bar{k}$. As was noted earlier in (6), the variance of u for the self-employed is larger than for the employed. Therefore, the minimum feasible value of $\sigma^2_{u_S}$ is $\sigma^2_{u_E}$ which will provide the upper bound for $\bar{k}$, marked with subindex h

$$\ln \bar{k}_h = \frac{\gamma_j}{\beta_j} + \frac{1}{2}(\sigma^2_{Y_S} - \sigma^2_{Y_E}). \tag{19}$$

A covariance of zero implies that whatever the income of the self-employed is in different years they will underreport by the same proportion. It is also worth investigating what would happen if the covariance was positive.[9] Consider the upper bound. If the coefficient of correlation $\rho > 0$, (14) and (17) imply,

$$\ln \bar{k}_h = \frac{\gamma_j}{\beta_j} + \frac{1}{2}(\sigma^2_{Y_S} - \sigma^2_{Y_E}) + cov(uv)_S \tag{20}$$

[9] This would be the case if for example, if the second source of income becomes too high then concealment might be more difficult and some declaration will be necessary.

As usual, we can write the covariance as

$$cov(uv)_S = \rho\sigma_{u_S}\sigma_v \tag{21}$$

An approximation for σ_{u_S} is available. The variance of income from the employed is just composed of the variance of u for this group and unexplained variations in permanent income. Then, the variance of u for the employed cannot exceed the residual variance of income for that group, $\sigma^2_{Y_E}$. Since we are in the upper bound case, we assume that $\sigma^2_{u_S} = \sigma^2_{u_E}$ and thus an upper bound for σ_{u_S} would be $\sigma_{Y_{PE}}$. Then (14) can be written as

$$\sigma^2_{Y_S} - \sigma^2_{Y_E} = \sigma^2_{v_S} - 2\rho\sigma_{u_S}\sigma_v = \sigma^2_{v_S} - 2\rho\sigma_{Y_E}\sigma_v. \tag{22}$$

We can then solve for σ_v in (22) for different values of ρ, estimate the covariances and use them in (20) to appreciate the possible effect of $\rho > 0$ in our estimates of income underreporting by the self-employed households.

3.1.1 Hurst: A comparison

Hurst et al.(2014) provide a reduced version of the methodology in which they focus on obtaining the degree of income underreporting κ instead of the adjustment factor k. They assume that a fraction κ of the income of the self-employed is misreported. Thus, if again Y_i^p is permanent income, v_{ij} is transitory income by group k and X_{ij} is the vector of household characteristics. Income for the self-employed can be written as:

$$Y_i = \kappa Y_i^p \exp(X_i\alpha + v_i) \tag{23}$$

where for the employed $\kappa = 1$.

The Engel curve for the employed could then be written as:

$$\ln C_{ij_E} = X_{i_E}\alpha_{j_E} + \beta_{j_E}\ln Y_{i_E} + \varepsilon_{ij_E} \tag{24}$$

and for the self-employed,

$$\ln C_{ij_S} = X_{i_S}\alpha_{j_S} + \beta_{j_S}\ln Y_{i_S} + \varepsilon_{ij_S} \tag{25}$$

Thus, the coefficient κ of underreporting can be estimated from

$$\ln C_{ij} = X_i\alpha_j + \beta_j \ln Y_i + \gamma_j SE_i + \varepsilon_{ij} \tag{26}$$

Thus we can estimate the coefficient of underreporting using γ_j and β_j from (26) to be:

$$1 - \hat{\kappa} = 1 - \exp(-\hat{\gamma}_j/\hat{\beta}_j) \tag{27}$$

Pissarides and Weber (1989) focus on estimating the scaling factor k which relates to κ as $\kappa = 1 - (1/k)$. Hurst *et al.*(2014) use a reduced form of the method that does not adjust for the distinct variance of income of the self-employed and employed and possibly biasing the results of underreporting upwards. We provide this estimation for comparison purposes as some literature has built estimates based on this method see Engström and Holmlund (2009).

3.2 Determining the non-compliance characteristics

The PW model is now extended to account for the characteristics of the non-compliant. Therefore, in order to do so we interact the self-employment dummy variable with the socioeconomic and demographic characteristics of the individuals. We will be working with characteristic N which has n categories, say Gender (N) which has two categories (n), male and female and we will be interacting the dummy variable SE_i with these n categories.

The Engel curves will then be given by,

$$\ln C_{ij} = X_i\alpha_j + \beta_j \ln Y_i^{'} + \sum \gamma_{j_n} SE_i * N + \eta_i$$

In fact, what γ_{jn} is capturing is the difference in the intercept between the employees and the self-employed in that category. The model itself works in the same manner but we allow for different characteristics of the individuals to exhibit different underreporting behaviours. This enables us to obtain an average underreporting behaviour for each category within the characteristic explored using the employees as a benchmark. In the model, this is allowed by including the interaction of the self-employed dummy variable with the categories of the characteristics c.

Therefore, we can write that the mean value of k for each category n within the characteristic N can be obtained from,

$$\ln \bar{k}_n = \frac{\gamma_{j_n}}{\beta_j} + \frac{1}{2}(\sigma^2_{v_{Sn}} - \sigma^2_{u_{Sn}} - \sigma^2_{u_E}) \tag{28}$$

Equation (14) for this model would be rewritten as:

$$\begin{aligned} \sigma^2_{Y_{Sn}} - \sigma^2_{Y_E} &= var(u-v)_{S_n} - var(u)_E = \\ &= \sigma^2_{u_{Sn}} + \sigma^2_{v_{Sn}} - 2cov(uv)_{S_n} - \sigma^2_{u_E} \end{aligned} \tag{29}$$

If we assume the covariance between u and v for all categories is 0, then we can provide an upper and lower bound for the average value of k. Then a lower bound would be obtained for each category when $\sigma^2_{v_{Sn}}$ equals zero:

$$\ln \bar{k}_{n,l} = \frac{\gamma_{j_n}}{\beta_j} - \frac{1}{2}(\sigma^2_{Y_{Sn}} - \sigma^2_{Y_E}). \tag{30}$$

We can also provide an upper bound for $\bar{k}_n$. The minimum feasible value of $\sigma^2_{u_{Sn}}$ is $\sigma^2_{u_E}$ which will provide the upper bound for $\bar{k}_n$, denoted by subindex h.[10]

$$\ln \bar{k}_{n,h} = \frac{\gamma_{j_n}}{\beta_j} + \frac{1}{2}(\sigma^2_{Y_{S_n}} - \sigma^2_{Y_E}). \tag{31}$$

Intuitively, what the method does is to estimate the Engel curve for food for the employed and for the self-employed. If we assume that the food function is the same for both groups, any deviations of the Engel curve for food for the self-employed from that of the employed are indicative of income underreporting. This is, if for a certain level of expenditure the employed report a certain level of income which is higher than that reported by the self-employed. The difference between the income of the honest group, this is the employees, and that (mis)reported by the self-employed gives an indication of the amount of income

[10] The effect of a positive covariance can be assessed for all estimators as was introduced before.

underreporting by the self-employed. In the extension provided to fit our second aim of analysing noncompliance what we would be looking at is different Engel curves for self-employed in different categories under the same characteristic. This is we would be looking at a self-employed male curve and a self-employed female curve which will be evaluated with respect to the employee engel curve.The difference between the income reported by a self-employed under a certain category and the income reported by the employees for a given level of consumption provides an estimation of income underreporting by the self-employed in that category. This can be represented as:

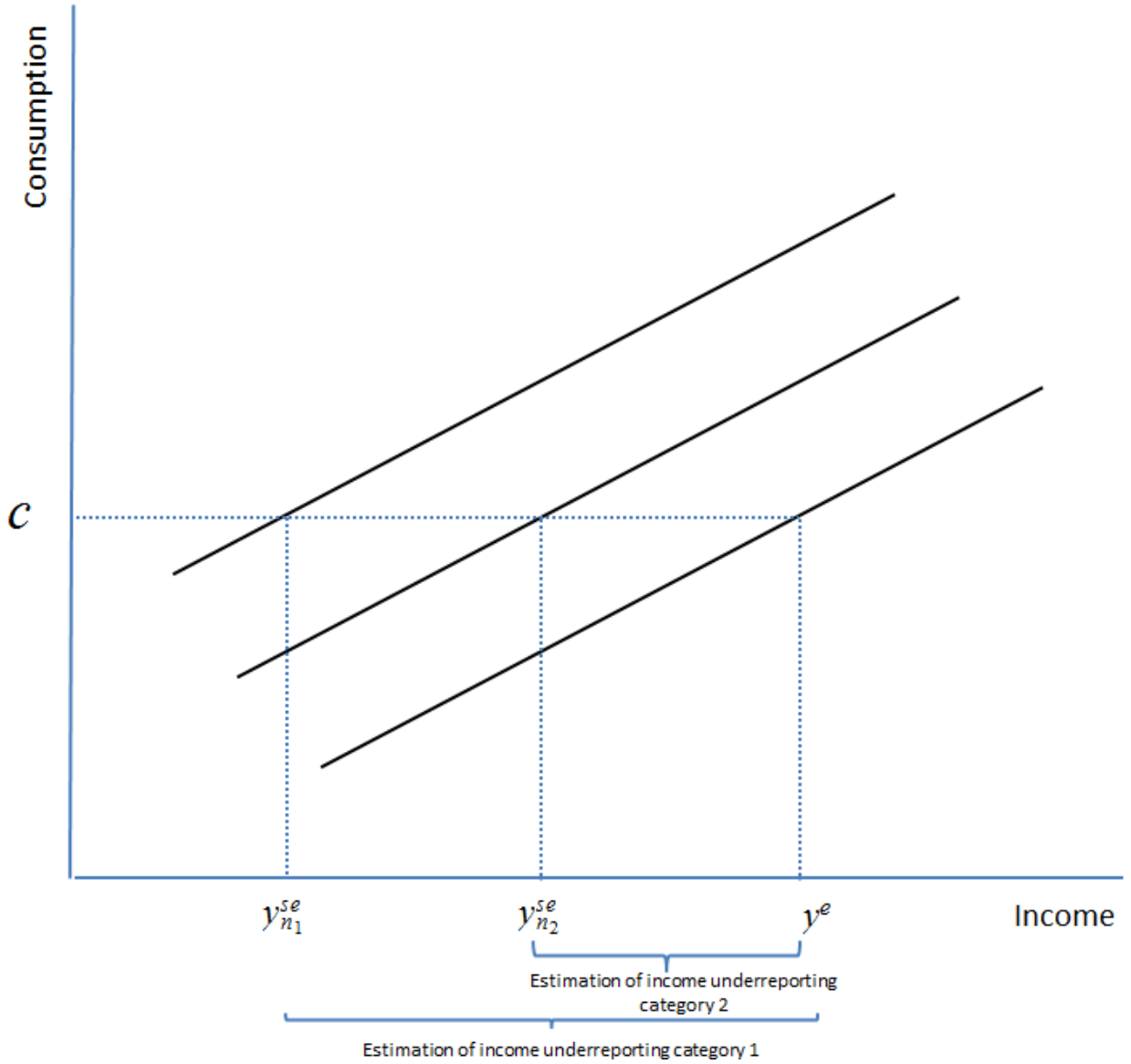

Figure 1: Extent of underreporting

It could be the case that self-employed in different categories under the same characteristic underreport in the same manner. Then the curves would overlap meaning that the characteristic yields no different underreporting behaviour and is not determinant to the analysis of noncompliance.

The key to this method lies in the assumption that the employed report their income correctly whereas the self-employed tend to underreport their income. This assumption is

reasonable since the self-employed are responsible for reporting taxable income but the employed operate under a withholding tax. This provides the self-employed with an opportunity for evasion that the employed do not have access to. We accept that although the rate of compliance among the employed is very high it does not represent full compliance and thus the result of the amount of underreporting by the self-employed would be higher were this fact taken into account. Therefore, our estimate will represent a lower bound.

4 Data and estimation

The data to be used comes from the Secure Access version of the Living Costs and Food Survey (LCF) produced by the Office of National Statistics for the years 2010-12. The LCF uses as a unit of survey the household and it captures expenditure decisions and income earned from all the individuals within a household. Data for 2010, 2011 and 2012 are going to be pooled in order to obtain a reasonable sample size that can allow a reliable analysis. The data was deflated using the consumer price index base 2010.

The sample selected in the analysis aimed at choosing a homogenous kind of household in order to be able to compare the different reporting of households leaving out the concern that differences in non-compliance could be due to the composition of the household. Therefore, the sample was restricted to households of two adults, either cohabitees, married or civil partners who live in Great Britain and the HRP is either employed or self-employed. We also restrict the age of the HRP to be less than 60 in order to leave out different expenditure behaviour after retirement as evidenced by Aguiar and Hurst (2005).

As has been done in previous research, income reported had to be also adjusted in order to account for the fact that income reported by the self-employed dates back to the last available record which could well have been obtained a year before and consequently, not to the date of the interview. Failing to correct this time lag could lead to spurious results. Self-employment income reported was updated to the time of the interview using the monthly rate of inflation calculated from the amount of self-employment income per self-employed worker drawn from the Blue Book for the corresponding years.

The measure of income to be used is disposable labour income. Underreporting using total household disposable income are not significantly different, although lower underreporting is found. The use of labour income alone partials out the incidence of underreporting in other sources of income. In the survey, self-employed individuals were asked about the profit from their activity and about how much their drawings amount to both for business and non-business purposes or personal use. In the event that the individuals were not able to respond to any of these questions they were asked for an estimation of how much their income was once expenses were deducted.

We aim at capturing all income the self-employed take into consideration when taking their consumption decision. We assume that the self-employed make their consumption decision based on their total earnings both labour earnings and those reinvested back into the business. We can think about reinvesting back into the business as a way for the self-employed to save. If this is the case, taking drawings only into account will lead to a much higher underreporting that it is actually true. In order to test these hypothesis we are going to use three different measures of self-employment income: a comprehensive measure; self-employment income as profits only and self-employment income only from drawings.

The comprehensive measure of self-employment income variable was computed as follows. The profit figure (transformed into a weekly amount) was taken if reported. If the individual

reported a loss (or zero profit) or was not able to report a profit figure, then the estimation of the weekly drawings was taken. If none of the former were available, then the weekly equivalent of the estimation of income minus expenses was taken. This measure of self-employment income we are going to call it a comprehensive measure because it takes into account as much data on income as possible in order to avoid missingness in the answers. The second measure considers only profits as the measure of self-employment income aims at reinforcing the validity of the comprehensive measure. We provide and discuss the results of these three alternative measures in the results section.

We are going to define a self-employed household as those which draw more than 25% of their income from labour income from self-employment. This is done in order to avoid households that have a substantial amount of self-employment income coming from a subsidiary source to classify themselves as employees. Alternative specifications of the self-employment dummy variable are taken into account for robustness.

Regarding expenditure, we are going to consider three different measures of expenditure: food, utilities and a basket of nondurable goods. We are going to argue that food expenditure is reported accurately for several reasons. Food being a necessity is not an expenditure that can be altered by transitory shocks, its consumption cannot be postponed to future periods and zeros for infrequency of purchase are not present. It cannot generally be claimed as a business expense. None of the individuals in the sample actually mention food as a business expense. It is expected to be correctly reported as there is no social stigma associated to it. We know other items of expenditure such as tobacco or alcohol are usually misreported on surveys. Food does not represent an item that can be suspected of showing a certain lifestyle as opposed to expenditure on holidays or newly bought cars, and therefore the interest in misreporting is minimal.Lastly, food expenditures are recorded using a diary.[11] Utilities was chosen as an item of expenditure that could be easily recalled due to the regular payments. However, it raises concerns as it represents an item typically claimed as a business expense by the self-employed. The higher level of consumption could be masquerading the differing nature of the self-employment activity and its particular fiscal treatment. Therefore, estimates although presented for comparison should, in principle, be taken with caution. A composite measure of expenditure is also presented and is formed of a basket of nondurable goods.[12]

The descriptive statistics of the data are presented below in Table 1. We see that consistently for each item of expenditure listed the self-employed exhibit higher levels of expenditure whereas their reported income is lower than that of the employed.

Table 1: Summary Statistics

	Employed Household			Self-Employed Household		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Ln(Food Expenditure)	4026	4.35	0.558	738	4.432	0.579
Ln(Expenditure on Utilities)	4021	3.222	0.565	736	3.314	0.585
Ln(Expenditure on Non-Durables Basket)	4036	5.734	0.524	739	5.81	0.552
Ln(Disposable Labour Income)	4034	6.528	0.504	738	6.473	0.695

[11] Diaries are kept along two weeks for frequent items of expenditure as is the case of food by all individuals over 16 years old for a fortnight and children between 7 and 15 are asked to keep a simplified version. Other items of regular expenditure, such as the mortgage, are captured in the household questionnaire.

[12] The basket of nondurable goods comprises expenditure on food, alcohol and tobacco, clothing, utilities, non-durable expenditure on recreation, non-durable expenditure on transportation and communication, helath, education and other miscellaneous nondurable expenditures.

From the summary statistics we can observe the self-employed report a lower level of income when reporting a higher level of expenditure for all income and expenditure measures which we attribute to underreporting of income by the self-employed in order to sustain that level of expenditure.

Regarding the estimation method, two-stage least squares is going to be applied. Instead of using permanent income to estimate the Engel curve for food, we use income reported to the survey. In order to reduce measurement error, we instrument for income. The instruments to be used are: education variables (whether the household reference person (HRP)[13] and spouse had higher education) and whether the HRP person of the household was in a white-collar or blue-collar occupation.

Pissarides and Weber (1989) also instrument for the self-employment dummy variable. The reason behind this is that a self-employed household is defined from the proportion that self-employment income represents to total household income, and if the first is underreported then proportion may vary and consequently whether the household is considered self-employed or not. However, instrumenting can bring its own problems. The bias that can result in the estimates due to weak instruments could outweigh the effect of failing to instrument and accepting the bias introduced by considering that some households that behave as self-employed are classified in our sample as employed.

In this paper we opt for using the more conservative approach of only instrumenting what is truly necessary, income, and we accept any biased introduced by the possible misclassification of some cases.

5 Results

5.1 Self-Employment Underreporting

In order to provide an estimation of income underreporting by the self-employed, four key parameters are required: the estimation of the residual income variance for the dishonest and the honest group (self-employed and employed) from the first-stage regressions (σ^2_Y); alongside the marginal propensity to consume (β) and the coefficient of the self-employment dummy variable (δ) from the second-stage regressions.

We report the estimations of income underreporting using the Pissarides and Weber method and the variant by Hurst *et al.*(2014)[14]. Using food expenditure, we find that the self-employed underreport their income by 19.3% which translates into a midpoint multiplier of 1.24 as can be seen in Table 3. This is, income reported by the self-employed should be multiplied by a factor of 1.24 to obtain true income.[15] This value is similar to that obtained

[13] From 2001/02 the concept of household reference person (HRP) was adopted on all government-sponsored surveys in place of head of household. The household reference person is the householder who: owns the household accommodation, or 1) is legally responsible for the rent of the accommodation, or 2) has the household accommodation as an emolument or perquisite, or 3) has the household accommodation by virtue of some relationship to the owner who is not a member of the household. If there are joint householders the household reference person will be the one with the higher income. If the income is the same, then the eldest householder is taken

[14] For ease of speech, we are going to report in the text the midpoint underreporting from the Pissarides and Weber method, the result using Hurst et al. (2014) are reported in the corresponding tables.

[15] The estimates presented are obtained assuming the covariance between u_i and v_i to be zero. We cannot empirically assess the correlation between the two but assuming the highest correlation of 1, our estimates would need to be scaled up by 35%. On the other hand if the correlation was 0, the estimates would be correct and needed to be adjusted upwards between 0-35% if the correlation lies in between.

for countries such as Sweden or Canada (Engstrom & Holmlund, 2009; Schuetze, 2002). We provide the same estimation using two alternative measures of expenditure for comparison purposes: expenditure in utilities and a nondurable basket. The concern with using expenditure in utilities, despite being an item of expenditure one may expect to be easily recalled, is that it can be claimed as a business expense biasing therefore the results of underreporting upwards. This would lead to a misattribution of the discrepancy to underreporting when it is only a feature of the self-employed activity and fiscal treatment. The degree of underreporting revealed using utilities is slightly larger: 21.6%. We use also a basket of all non-durable goods in order to give a composite measure of expenditure. We find a very similar degree of underreporting as using food expenditure, 19.2%.[16]

For robustness we have used alternative measures of self-employment income. On the basis that the self-employed make their consumption decision based on their available and retained earnings, we observe that the level of underreporting estimated using the comprehensive measure of self-employment income and profit is remarkably close, 19.4% and 19.3%. Using only drawings from the business results in a substantial decrease on the sample size. Not many individuals report taking drawings from their business or report missing answers. Using drawings, Table **4** shows underreporting doubles what has been found for the previous measures of income. We think this results overestimate the amount of underreporting as it may not consider the influence of retained earnings in shaping consumption decisions.

Table 6 shows that the definition of the self-employed used based on income does not significantly alter the results when using self-reports.We provide two alternative definitions. In the estimation of the first column, we make use of self-reported employment status to classify a household as self-employed. We would mark a household as self-employed if the HRP is self-employed in either their main job or any of the subsidiaries. In the second column, we use a more comprehensive definition. We want to classify as self-employed a household if they have any source of self-employment income.We do not take into account the importance or amount of self-employment income, only the opportunity. We proceed to classify a household as self-employed if either the HRP or the spouse is self-employed in any of their main or subsidiary jobs.

5.2 Is it underreporting?

As we have noted in the literature survey, there has been a recent wave of research using the expenditure-based method. However, the method relies on several assumptions that have remained untested or unchallenged. One may wonder whether the estimate we obtain is truly attributable to underreporting or if it could be a feature of the assumptions posed to obtain the estimate. In this section we aim at evaluating whether the estimation results are caused by heterogeneous preferences for food consumption, heterogeneous spending behaviour, or measurement error in the survey.

5.2.1 Preference Heterogeneity

One of the key assumptions in the set up of the model is that the self-employed and the employed have the same elasticity of consumption with respect to income so that both

[16]The authors have also used alternative measures of expenditure that include utilities and housing and a basket of goods similar to the one presented in the paper including housing expenditures. The decision to not include it is that we would need to use estimations of imputed rent for individuals own the houses which will blurr the estimation. Therefore, the authors have decided to leave this item of expenditure out of the analysis.

occupational groups have the same relationship between income and expenditure. This assumption could easily be false if the groups have different preferences towards food. We have tested this assumption by estimating Engel curves separately for the two occupational groups.

We obtain the result that the income elasticities for both occupational groups using the IV specification are statistically indistinguishable from each other. This supports one of the main assumptions of the model for estimating underreporting. Also, our data on food expenditure distinguishes between food eaten in the house and eaten out of the house. Since food eaten out is typically more expensive a difference in the pattern of consumption between the two groups could explain the estimation results. We have tested whether there is a different preference between food eaten in and out between the occupational groups. Estimating the share of food eaten in with respect to total food expenditure, we find that there is no significant difference between the two groups. Therefore, the higher level of expenditure observed for the self-employed cannot be justified by a higher expenditure on food outside from home.[17]

5.2.2 Heterogeneous Spending behaviour: Financial Constraints

Another reason why we observe a discrepancy between expenditure and income may be a different level of financial stability for the self-employed. What we know, and is controlled for in the model, is that income for the self-employed is more volatile than income for the employed. However, it could still be the case that in the years observed, 2010-12, it that the level of expenditure of the self-employed is not funded by current income only but also by past savings. Therefore, the model will be misleadingly attributing to underreporting the fact that some current consumption is not funded by current income but by past income. Using variables from the Secure Access data, we are able to filter out those individuals who claim to be financially constrained. By this is meant individuals who find their current income is not sufficient to fund their current consumption and need recourse to other alternative sources such as loans, savings, or money from relatives. We have created a smaller dataset of individuals not financially constrained and we rerun the analysis to obtain their level of underreporting to be 22.9%.

The main category of expenditure analysed in this paper is food expenditure for the reasons discussed above. However, we have also presented estimates using utilities as an expenditure category. One of the main concerns about using utilities is that it is one of the main categories claimed by the self-employed as a business expense. Therefore, a higher level of expenditure on utilities relative to income may not be a sign of underreporting but the result of a differential fiscal treatment for the self-employed. We have sieved out of the estimation all those self-employed that claim any utilities bill as a business expense. Using utilities as the expenditure category on those self-employed individuals who do not claim any business expenses, we still observe a very similar level of underreporting (20.8%).

Therefore, there is evidence that neither the role of savings in funding current consumption nor the possibility of claiming of business expenses from some expenditure categories are the driver of the results obtained.

[17] These results are available on request.

5.2.3 Measurement Error

One of the downsides of using a survey to reveal underreporting is the extent to which we believe the information reported to the survey is accurate. The nature of the survey makes it likely that expenditure is more accurately reported than income since the survey makes available instruments for making expenditure as accurate as possible For example, in the case of food a diary is filled by each member of the family. In the case of utilities, we trust on it being a recurrent expense and therefore easy to recall. However, there is extensive literature on misreporting of expenditure in surveys.[18]

On the other hand, in the case of income, underreporting may not be the only reason behind the discrepancy found. Individuals are asked about their current income. Employees are usually paid on a monthly basis, so their claim of income earned should not deviate in excess of what they actually earn. That is, it should be free from recall error. For the self-employed, this is not the case. Recalling their last profit of their businesses may not be such an easy task. Therefore, it may be true that the discrepancy found is due to measurement error, it is due to the self-employed supplying approximate figures for their income that have little to do with their true levels.

In order to check whether the discrepancy stems from simple measurement error, we have carried out two checks. In the first check, we have selected those individuals who consulted any documents in order to give their profit figures. These documents could be either a Notice of Tax Assessment (form 300), their Annual Accounts (or the summary), their Tax Return or any other documents. We find that the level of underreporting is only larger than in the case when no document is consulted (22.8% versus 19.3%).

The second check is to look at how the estimate looks when we have administrative data. One of the differences between filling out a survey and reporting to the revenue service is the incentive to report the true income. In the survey, individuals may report their true income more accurately than they do to the revenue service, mainly because from the survey there is no win and no loss. However, one may also argue that some of the individuals who are afraid of their data being linked despite reassurance of the anonymity of their responses, will give the same answer to the survey as they did to the revenue service for the sake of consistency and avoiding self-incrimination.

Our aim here is to disentangle whether these effects are biasing the results. If we observed administrative data on incomes rather than survey incomes, how would the estimate of underreporting vary? Is there any measurement error in incomes due to failure to recall;do individuals tend to report closer to their true income or on the other hand do individuals tend to consistency? In order to do so, we repeat the estimation only for individuals who have made use of documents that have already been used with the revenue service or that are from the revenue service: the Notice of Tax Assessment and the Tax Return. Using only individuals who have looked at these documents we find a larger value of underreporting of 30.7%. However, we need to take these results with care as the sample size of self-employed using these documents is very low. Even so, this highlights the potential advantages of using administrative and survey data together.

5.3 Determinants of noncompliance

This sections constructs a profile of the noncompliant self-employed. We have analysed the characteristics of the HRP in the household and also the characteristics of the self-employed.

[18] See Bee et al. (2015)

We will start exploring noncompliance using characteristics observable in the tax returns: age, gender and region.

In order to assess the impact of age on self-employment underreporting, we have stopped controlling for age and introduced the interaction of the self-employment dummy variable and different age categories. We have done this analysis taking into account the age of the household reference person and also the age of the self-employed individual. The results from Table 9 and 10 reveal that as individuals age, they become more compliant. This is, in households were the HRP is less than 35 years old, income is underreported by on average 27.3%; if between 35 and 45, 18.9% and between 45 and 60 by 14.2%. We can observe that the coefficient of the dummy variable loses significance as the age bracket increases. The same applies when looking at the age of the self-employed[19], with underreporting coefficients of 26.4%, 17.5% and 14.9% for the respective age brackets. The finding that income underreporting decreases as as age increases is consistent with the findings in the literature on tax audits in the US using TCMP (Clotfelter, 1983; Feinstein, 1991) and also in experiments (Kleven, Knudsen, Kreiner, Pedersen, & Saez, 2011). Criminological studies provide evidence of this same finding for crimes in general. It seems that for all types of criminals and crimes there exists a common distribution for age which shows to be invariant to social and cultural conditions. so that crime is negatively correlated with age (Hirschi & Gottfredson, 1983). There are two theories to explain this occurrence: the age theory and the desistance theory. The age theory relates to the fact that as the organism ages, the tendency to commit a crime goes down. The desistance theory argues that this change in the behaviour of the criminal is due to factors related to age: social position, social status, material possession which make the cost of criminality increase with age (Tittle, 1980) (Torgler & Valev, 2006).[20][21]

Gender is another characteristic observed generally by the revenue service. We find that men underreport significantly more than women. We consider two specifications for the regressions. In one, we distinguished the gender of the individual only if the household had only one self-employed individual and we created a variable to capture underreporting by households where both partners were self-employed. From Table 15 we can observe that households of two self-employed underreport more than households of one self-employed. In households of one self-employed, the self-employment dummy variable is not significant when the self-employed is the wife. In the second specification contained in Table 16, we take the sex of the of the main earner of self-employment income and the result is invariable. The criminology literature also supports the fact that women are more compliant than men. Our result is also in line with what has been found in studies using surveys and experiments analysing tax compliance.[22] Vogel (1974) found using a survey in Sweden that men think of themselves as having better illegal opportunities than women. Torgler and Valev (2006)

[19] In households where there are two self-employed individuals, we take the age of the main earner of self-employment income.

[20] Quetelet was one of the first to tackle this issue and estates that the declining is a natural process brought by "the enfeeblement of physical vitality and the passions" (Brown Miller, 1988,p.13) Maturational reform constitutes one of the most important theories within desistance theories.

[21] Hirschi and Gottfredson (1983) however support the difficulty in assessing and understanding the relationship between age and crime. The aging-out effect has been proved to be invariant with respect to social and cultural conditions, such as race and gender, which are correlated with crime and therefore these variables cannot be used to explain the effect of age. They even assert that there is no existent variable in criminology that is useful to provide a sound explanation of this effect.

[22] See Tittle (1980), Torgler and Schneider (2006); for experiments, Spicer and Becker (1980), Spicer and Hero (1985), Baldry (1987), Kleven et al. (2011), Friedland, Maital and Rutenberg (1978).

using data from eight Western European countries from the World Values Survey and the European Values Survey that span the period from 1981 to 1999 also find that in the case of bribes men are more frequently asked for bribes by government officials than women. They find in general women are more willing to comply.

Gender differences in delinquency have been largely investigated in criminology and many theories have been proposed. The fact that this difference in behaviour was due to the inequitable role in society of women and men was discarded since entrance in the labour market of women did not affect crime rates. The most accepted theory to support this difference points at self-control and opportunities to commit crimes as the drivers. Delinquent men are found to be more exposed to delinquent companions that can influence their behaviour through imitation (Mears, Ploeger, & Warr, 1998).

Opposite results can also be found in the literature such as Baldini, Bosi & Lalla (2009) who use the discrepancy between the income reported to the fiscal authorities and the income reported to the Survey of Household Income and Wealth in Italy for the region of Modena as a measure of income tax evasion. Using a probit model they find females are more likely to underreport than men. However their study refers only to a particular region of Italy and the results of the study cannot be generalized to the whole population. Schuetze (2000), however, finds that the sex of the individual yields no difference in compliance behaviour.

Knowing how noncompliance is distributed geographically can provide first evidence of which regions are being less compliant and therefore may want to be more targeted with the audits. Besides, this is one of the aspects that are incorporated in the self-assessment tax return filled by the self-employed. We have tried considering all the regions of England as well as Scotland and Wales.[23] From Table 17, we observe that the most noncompliant region in the UK is Greater London followed by the East of England.

One of the advantages of using survey to audit data is the richness of information about socioeconomic and demographic characteristics of the individuals interviewed and of the household in general. The information we fill in the tax return is merely for identification purposes more than for the sake of providing a detailed profile of the individual. We are going to analyse the role of the occupation, number of self-employed individuals in the household, the position in the household, the legal form of the activity and their consideration of the self-employed activity.

Pissarides and Weber's original analysis was divided by occupations: white and blue-collar. The definition of what constitutes each category is not explicit in the paper. We define as white-collar those individuals who are employers in small or large organizations, higher managerial, higher professional, lower professional and higher technical, lower managerial, higher supervisory and intermediate positions. Blue-collar are those who work on their own account, or in lower technical, lower supervisory and semi-routine or routine activities.[24] The results in Table 5 suggest that self-employed in blue-collar occupations tend to underreport more than those in white-collar. This result is consistent with what has been found previously in the UK by Lyssiotou et al. (2004) and Pissarides and Weber (1989).

The number of self-employed individuals present in the household is a variable that should also be taken into account. We observe that households with two self-employed individuals underreport more than households with only one self-employed individual with 29% and 13% respectively as can be seen in Table 12. This result is in line with intuition. A self-employed

[23] Restrictions on the sample size due to the Secure Access nature of the data used led to the pooling of certain regions together in order to achieve a sufficient sample size that ensured anonymity.

[24] For illustration purposes only, the manager of a restaurant will be considered white-collar while the waiters will be considered blue-collar.

individual has higher chances of underreporting their income than an employee due to the absence of third party reporting. A household that counts with two self-employed individuals then has more opportunity to underreport than that with a single self-employed individual in the household. This results are in line with those found by Johansson (2005) for Finland but run against what Schuetze (2000) found for Canada.

We would also like to check what pattern we observe when we take into account which role in the household the self-employed individual has. The position of the self-employed is going to be identified for households with only one self-employed individual and we would consider the underreporting of households with two self-employed individuals separately as shown in Table 13. We find that households where the self-employed individual is the HRP underreport more than households where the self-employed is the spouse. This however relates to the findings about women being more compliant due to the large amount of male HRP in the sample.

The legal form of the activity may also be a parameter that affects noncompliance. Self-employed can work as a sole trader or form a partnership. We find that individuals working in a partnership underreport more than individuals acting as sole traders (27.2% and 17.9% correspondingly, see Table 14). Having a business with other partners can allow more concealment of self-employment income as the amount of income is shared among different individuals. Therefore, larger figures of income and expenses than in the case of a sole trader will not arouse the curiosity of the revenue agency. Some may think that avoiding in a partnership should be more difficult than as a sole trader. However, the only part in which the individuals need to reach an agreement about is in how much income to conceal and that choice is already made when presenting their partnership tax return to the tax authority. Then the only formality required is to maintain consistency between the individuals' partnership pages in their personal tax return and the partnership tax return prepared by the nominated taxpayer.

The survey explicitly includes a question about whether the self-employed consider themselves as having a job or a business. In fact, there are cases in which the self-employed will operate in a rented facility. The case of hairdressers serves as a good example. It is common to find self-employed hairdressers working in a salon where the owner, say another self-employed hairdresser, simply hires the chair to them. The first one, the one that hires the chair may regard herself as having a job and the second one, the owner of the salon, as having a business. Similar cases are found with insurance companies' selling agents.

This question, although it might seem trivial, offers a psychological insight into noncompliance. If an individual has a self-employed job they may not feel as free to underreport as when they have their own business. A fear that information about their noncompliance may leak from the person who owns the company may exist. Another explanation may be that concealing income is not socially acceptable and therefore can serve as a deterrent for those who work in someone else's premises. In fact, Table 11 reveals this choice exhibiting that the income from self-employed who own a business is underreported by 21.2% which is higher than for those who have a self-employed job, 17.4%.

6 Conclusions

In this paper we use the Pissarides and Weber (1989) expenditure-based method to measure income underreporting by the self-employed in Great Britain using data from 2010-12. The aim has been threefold: firstly, to provide an estimate of income underreporting by the

self-employed; secondly, to provide support and a critical analysis of the linkage of the discrepancy obtained to underreporting and thirdly, to elaborate on the characteristics of the noncompliant individuals.

We have provided estimates of income underreporting by the self-employed under alternative specifications of the expenditure and income variables. It has been shown from our results that self-employment income should be multiplied by an average factor of 1.24 to obtain true income which amounts to 19.3% underreporting. Taking into account that for the period between 2010-12 self-employment income represented 5.5% of GDP, this yields an estimate of the size of the black economy of 1.6% of GDP. This estimate of the size of the black economy is a lower bound as we acknowledge that all tax evasion does not occur only within the sector of the self-employed. However, due to the lack of quality data for measuring this phenomenon we require a plurality of methods to triangulate the size of the black economy.

There have been many studies blindly applying the expenditure-based method to estimate underreporting. The fact that the method relies on a set of assumptions raises doubts about the source of this discrepancy. Our second aim in this paper has been to enrich the literature providing support that the discrepancy obtained through this method can be truly linked to underreporting. We have found that differing preferences and differing way of funding current consumption do not lie behind the primary reasons for the discrepancy found. This only raises more confidence in the estimates obtained. However, the use of income figures by the survey respondents in documents reported to the revenue agency result in detecting higher levels of underreporting. This highlights the potential of linking administrative and survey data.

Regarding the characteristics most related to noncompliance we find that:

- Men are less compliant than women
- Age is inversely related to compliance. This is, as individuals get older, they turn more compliant
- Self-employed operating in partnerships underreport more than own-account self-employed.
- Blue-collar households underreport more than white-collar.
- Self-employed who consider themselves having a business underreport more than those who consider themselves as having a job.
- Households with two individuals underreport more than households of one individual.
- Households where the self-employed is the head underreport more than when the self-employed is the spouse which is driven by the fact that in most households the spouse is the wife
- We find that self-employment underreporting varies for the different regions.

When the tax revenue agency gets involved into a process of tracing noncompliance they aim at, in words of Louise XIV's Finance Minister, "plucking the goose as to obtain the largest possible amount of feathers with the smallest possible amount of hissing". Audits which are one of the strategies used by the revenue agency to uncover evasion are very costly and resources available are limited. The estimation results provide a framework for directing

the target of the audits to certain profiles in the population in order to minimize the cost while producing the higher revenue possible.

Other compliance activities orientated towards educating and encouraging compliance in the taxpayers need to be as well correctly orientated to the most noncompliant population in order to achieve the highest efficiency possible. Therefore, according to the results of our analysis, young male self-employed individuals in blue-collar occupations should be the focus of audits and compliance activities in general, especially those working in partnerships.

We consider our findings to be strong and consistent, firstly as the method to obtain black economy estimates is through micro data avoiding the difficulties of macro methods; secondly because we have tested how our estimate relates to underreporting by being able to single out other potential explanations of the discrepancy; and lastly because our findings are consistent with the evidence found in studies using audit data and also to the findings in compliance experiments. Evidence about the profile of the noncompliant has never been provided, to our knowledge, for the UK case analysing survey data making the findings of this study unique. It contributes to a literature which is still at its infancy.

There is clearly room for further research on this area. The literature should evolve and move one step forwards from only the measurement to more profound grounds, analysing the patterns of noncompliance in order to be able to inform tax policy choices. We could also check if the pattern of noncompliance found on this paper is consistent for employees and if the characteristics of the individuals who underreport more on their income tax are consistent for other types of taxes. All in all and back to the subject of our study, according to ONS, self-employed individuals tend to be older and the number of women in self-employment is increasing at a faster rate than the number of men (although in levels men are still predominant in self-employment). These changes emphasise the importance of understanding the determinants of income underreporting by the self-employed.

7 Appendix

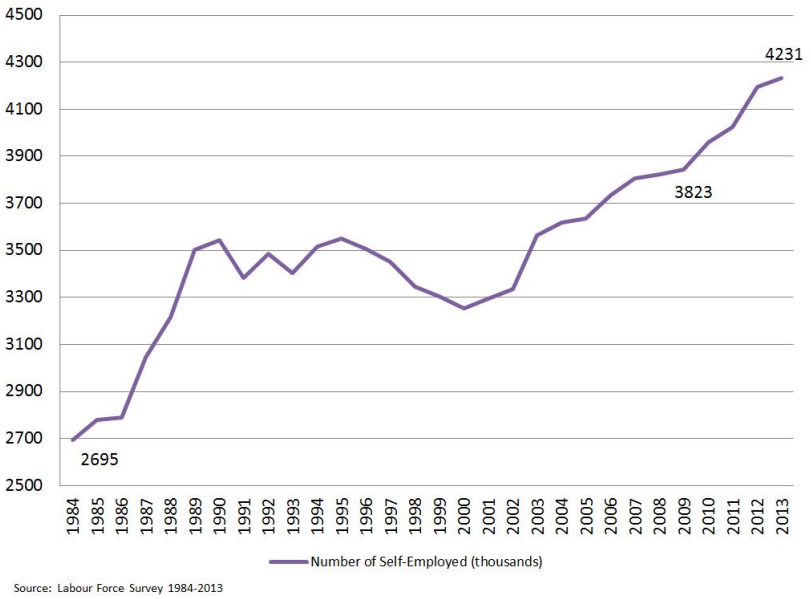

Figure 2: Historical Evolution of the Number of Self-Employed in the UK

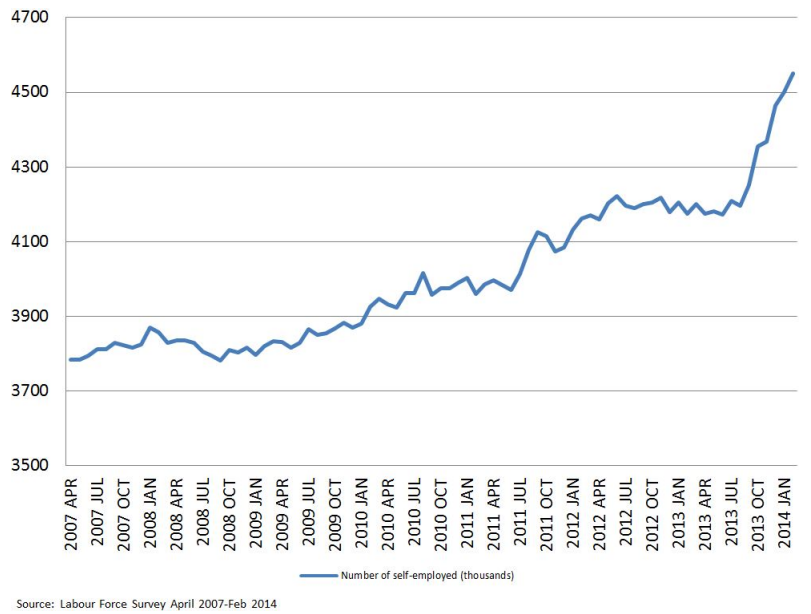

Figure 3: Evolution of the Number of Self-Employed since the Economic Downturn

Table 2: Summary table of previous results.

Author	**Region**	**Period considered**	**Methodology**	**Estimation of Income Underreporting**
Pissarides and Weber (1989)	Great Britain	1982	OLS, IV SF	White-Collar: (1.54, 1.28) ; Midpoint: 1.55 Blue-Collar: (1.64 , 1.51) ; Midpoint: 1.28
Lyssiotou et al. (2004)	UK	1993	OLS, IV SF	White-Collar: (1.26, 1.09) ; Midpoint: 1.18 Blue-Collar: (1.41 , 1.37) ; Midpoint: 1.39
Lyssiotou et al. (2004)	UK	1993	IV-GMM SF	White-Collar: 1.64 Blue-Collar: 2.18
Schuetze (2000)	Canada	1990	OLS, IV SF	(1.23, 1.17) Midpoint: 1.2
Johansson (2005)	Finland	1994-1996	OLS, IV SF	Head is SE: (23.5%,10%); Midpoint: 17% Both spouses SE: (46.9%, 37.3%); Midpoint: 42%
Besim and Jenkins (2005)	North Cyprus	1998-1999	OLS CoU	11-14% for the self-employed 9.7-14.2% for private employees
Engstrom and Holmlund (2010)	Sweden	1999-2001 and 2003-2004	OLS, IV SF	1.34 at least two (self) employed 1.31 at least one (self) employed
Kim et al. (2009)	Korea	2000-2005 1994-2000	OLS; Panel Data, Fixed Effects SF, CoU	1.614 (0.380) 1.880 (0.468)
Tedds (2010)	Canada	1982-1986 1992-1996	Nonparametric estimation	$ 2.612 billions $ 5.796 billions
Hurst et al. (2014)	US	1980-2003	CoU	CoU: 30%
Kukk and Staehr (2014)	Estonia	2002-2007	OLS, IV SF	Reported Self-Employment status: 1.386 Self-Employed (if self-employment income>0): 2.277 Self-Employed (if SE income>20%): 2.631

Notes:

1. CoU stands for coefficient of underreporting (κ). SF stands for scaling factor (k). The relationship between is 2. SE stands for self-employed

Table 3: Estimation of Income Underreporting using different measures of Expenditure

	Food Expenditure		Utilities		Non-Durables Basket	
	IV	OLS	IV	OLS	IV	OLS
Dummy SE	0.089***	0.065***	0.097***	0.060***	0.118***	0.086***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Elasticity of Consumption	0.438***	0.250***	0.418***	0.139***	0.581***	0.337***
	(0.04)	(0.01)	(0.04)	(0.02)	(0.03)	(0.01)
Variance SE	0.735		0.739		0.738	
Variance E	0.495		0.496		0.496	
Spread	0.240		0.243		0.242	
HLP						
	0.184***	0.230***	0.207***	0.353***	0.183***	0.225***
	(0.043)	(0.068)	(0.045)	(0.101)	(0.031)	(0.046)
PW						
Upper Bound						
Multiplier	1.422***	1.505***	1.465***	1.796***	1.422***	1.497***
	(0.075)	(0.133)	(0.084)	(0.279)	(0.054)	(0.088)
Underreporting	0.297***	0.336***	0.317***	0.443***	0.297***	0.332***
	(0.037)	(0.059)	(0.039)	(0.087)	(0.027)	(0.039)
Lower Bound						
Multiplier	1.058***	1.120***	1.085***	1.331***	1.055***	1.111***
	(0.055)	(0.099)	(0.062)	(0.207)	(0.04)	(0.066)
Underreporting	0.054	0.107	0.079	0.249**	0.052	0.100*
	(0.05)	(0.079)	(0.053)	(0.117)	(0.036)	(0.053)
Midpoint						
Multiplier	1.240***	1.312***	1.275***	1.564***	1.238***	1.304***
	(0.065)	(0.116)	(0.073)	(0.243)	(0.047)	(0.077)
Underreporting	0.193***	0.238***	0.216***	0.360***	0.192***	0.233***
	(0.042)	(0.067)	(0.045)	(0.099)	(0.031)	(0.045)
Observations						
Self-Employed	734		732		735	
Employed	4018		4012		4028	
Total	4752		4744		4763	

Notes:

1. Robust standard errors in parenthesis

2. The stars signal significance at the following levels:

* p<0.10 ** p<0.05 *** p<0.01

Table 4: Estimation of Underreporting for different measures of Income.

	Profit		Drawings		Proxy	
	IV	OLS	IV	OLS	IV	OLS
Dummy SE	0.079**	0.040	0.177***	0.089	0.089***	0.065***
	(0.03)	(0.03)	(0.06)	(0.05)	(0.02)	(0.02)
Elasticity of Consumption	0.405***	0.186***	0.376***	0.161***	0.438***	0.250***
	(0.04)	(0.01)	(0.04)	(0.01)	(0.04)	(0.01)
Variance SE	0.856		.978		0.735	
Variance E	0.566		.616		0.495	
Spread	0.290		0.362		0.240	
HLP						
	0.177***	0.196	0.376***	0.423**	0.184***	0.230***
	(0.067)	(0.132)	(0.096)	(0.197)	(0.043)	(0.068)
PW						
Upper Bound						
Multiplier	1.494***	1.528***	2.137***	2.311***	1.422***	1.505***
	(0.122)	(0.25)	(0.329)	(0.788)	(0.075)	(0.133)
Underreporting	0.331***	0.346***	0.532***	0.567***	0.297***	0.336***
	(0.055)	(0.107)	(0.072)	(0.147)	(0.037)	(0.059)
Lower Bound						
Multiplier	0.989***	1.011***	1.200***	1.298***	1.058***	1.120***
	(0.081)	(0.165)	(0.185)	(0.442)	(0.055)	(0.099)
Underreporting	-0.011	0.011	0.167	0.23	0.054	0.107
	(0.083)	(0.162)	(0.128)	(0.263)	(0.05)	(0.079)
Midpoint						
Multiplier	1.241***	1.270***	1.669***	1.804***	1.240***	1.312***
	(0.102)	(0.208)	(0.257)	(0.615)	(0.065)	(0.116)
Underreporting	0.194***	0.212*	0.401***	0.446**	0.193***	0.238***
	(0.066)	(0.129)	(0.092)	(0.189)	(0.042)	(0.067)
Observations						
Self-Employed	311		131		734	
Employed	4230		4416		4018	
Total	4541		4547		4752	

Notes:

1. Robust standard errors in parenthesis

2. The stars signal significance at the following levels:

* p<0.10 ** p<0.05 *** p<0.01

Table 5: Estimation of Underreporting by type of Occupation

	IV		OLS	
	SE White-Collar	SE Blue-Collar	SE White-Collar	SE Blue-Collar
Dummy SE	0.079**	0.094***	0.088***	0.019
	(0.03)	(0.04)	(0.03)	(0.03)
Elasticity of Consumption	0.443***		0.235***	
	(0.04)		(0.01)	
Variance SE	.737	.719	.737	.719
Variance E		.500		
Spread	0.241	0.223	0.241	0.223
HLP				
	0.163***	0.190***	0.312***	0.077
	(0.062)	(0.06)	(0.081)	(0.111)
PW				
Upper Bound				
Multiplier	1.386***	1.415***	1.687***	1.241***
	(0.103)	(0.105)	(0.199)	(0.149)
Underreporting	0.279***	0.293***	0.407***	0.194**
	(0.054)	(0.053)	(0.07)	(0.097)
Lower Bound				
Multiplier	1.030***	1.078***	0.202**	0.946***
	(0.076)	(0.08)	(0.148)	(0.113)
Underreporting	0.029	0.073	1.470***	-0.057
	(0.072)	(0.069)	(0.094)	(0.127)
Midpoint				
Multiplier	1.208***	1.246***	1.470***	1.094***
	(0.09)	(0.093)	(0.173)	(0.131)
Underreporting	0.172***	0.198***	0.320***	0.086
	(0.061)	(0.06)	(0.08)	(0.11)
Observations				
Self-Employed	377	347	377	347
Employed	4028		4028	
Total	4405	4375	4405	4375

Notes:

1. Robust standard errors in parenthesis

2. The stars signal significance at the following levels:

* p<0.10 ** p<0.05 *** p<0.01

Table 6: Alternative definition of Self-Employed Household

	Self-Employed Head		Any Self-Employed	
	IV	OLS	IV	OLS
Dummy SE	0.100***	0.073***	0.099***	0.077***
	(0.02)	(0.02)	(0.02)	(0.02)
Elasticity of Consumption	0.437***	0.252***	0.434***	0.251***
	(0.04)	(0.01)	(0.04)	(0.01)
Variance SE	0.697		0.693	
Variance E	0.494		0.491	
Spread	0.203		0.202	
HLP				
	0.205***	0.252***	0.204***	0.263***
	(0.038)	(0.06)	(0.037)	(0.057)
PW				
Upper Bound				
Multiplier	1.419***	1.509***	1.415***	1.530***
	(0.068)	(0.121)	(0.066)	(0.119)
Underreporting	0.295***	0.337***	0.293***	0.346***
	(0.034)	(0.053)	(0.033)	(0.051)
Lower Bound				
Multiplier	1.114***	1.185***	1.114***	1.205***
	(0.053)	(0.095)	(0.052)	(0.094)
Underreporting	0.102**	0.156**	0.103**	0.170***
	(0.043)	(0.068)	(0.042)	(0.065)
Midpoint				
Multiplier	1.266***	1.347***	1.265***	1.367***
	(0.061)	(0.108)	(0.059)	(0.106)
Underreporting	0.210***	0.258***	0.209***	0.269***
	(0.038)	(0.06)	(0.037)	(0.057)
Observations				
Self-Employed	897		974	
Employed	3855		3778	
Total	4752		4752	

Notes:

1. Robust standard errors in parenthesis
2. The stars signal significance at the following levels:
   * p<0.10 ** p<0.05 *** p<0.01

Table 7: Use of Documents

	Any Document consulted		HMRC documents	
	IV	OLS	IV	OLS
Dummy SE	0.104**	0.087**	0.152*	0.147**
	(0.04)	(0.04)	(0.08)	(0.07)
Elasticity of Consumption	0.433***	0.254***	0.433***	0.266***
	(0.04)	(0.02)	(0.04)	(0.02)
Variance SE	0.793		0.770	
Variance E	0.496		0.496	
Spread	0.298		0.275	
HLP				
	0.214***	0.290**	0.297**	0.424***
	(0.081)	(0.114)	(0.13)	(0.16)
PW				
Upper Bound				
Multiplier	1.542***	1.707***	1.692***	2.064***
	(0.158)	(0.273)	(0.313)	(0.573)
Underreporting	0.351***	0.414***	0.409***	0.516***
	(0.067)	(0.094)	(0.109)	(0.134)
Lower Bound				
Multiplier	1.051***	1.163***	1.195***	1.458***
	(0.108)	(0.186)	(0.221)	(0.405)
Underreporting	0.048	0.14	0.163	0.314*
	(0.098)	(0.138)	(0.155)	(0.19)
Midpoint				
Multiplier	1.296***	1.435***	1.444***	1.761***
	(0.133)	(0.23)	(0.267)	(0.489)
Underreporting	0.228***	0.303***	0.307**	0.432***
	(0.079)	(0.112)	(0.128)	(0.158)
Observations				
Self-Employed	177		57	
Employed	4028		4028	
Total	4205		4085	

Notes:

1. Robust standard errors in parenthesis
2. The stars signal significance at the following levels:
   * p<0.10 ** p<0.05 *** p<0.01

Table 8: Other Alternative Explanations for the Discrepancy

	Not Financially Constrained		No claim of Business Expenses (1)	
	IV	OLS	IV	OLS
Dummy SE	0.100**	0.087**	0.099***	0.072***
	(0.04)	(0.04)	(0.03)	(0.02)
Elasticity of Consumption	0.406***	0.307***	0.441***	0.147***
	(0.08)	(0.03)	(0.04)	(0.02)
Variance SE		.750		.709
Variance E		.480		.495
Spread		0.270		0.214
HLP				
	0.218***	0.247**	0.202***	0.390***
	(0.083)	(0.105)	(0.05)	(0.102)
PW				
Upper Bound				
Multiplier	1.509***	1.568***	1.425***	1.864***
	(0.161)	(0.218)	(0.09)	(0.311)
Underreporting	0.338***	0.362***	0.298***	0.464***
	(0.071)	(0.089)	(0.044)	(0.089)
Lower Bound				
Multiplier	1.083***	1.126***	1.101***	1.440***
	(0.116)	(0.157)	(0.069)	(0.24)
Underreporting	0.077	0.112	0.092	0.306***
	(0.098)	(0.124)	(0.057)	(0.116)
Midpoint				
Multiplier	1.296***	1.347***	1.263***	1.652***
	(0.138)	(0.187)	(0.079)	(0.275)
Underreporting	0.229***	0.258**	0.208***	0.395***
	(0.082)	(0.103)	(0.05)	(0.101)
Observations				
Self-Employed		152		534
Employed		977		3995
Total		1129		4529

Notes:

1. Robust standard errors in parenthesis
2. For the estimation regarding business expenses, the expenditure variable is utilities.
3. The stars signal significance at the following levels:

* p<0.10 ** p<0.05 *** p<0.01

Table 9: Estimation of Underreporting by Age of the HRP

	IV			OLS		
	<35	35-45	>45	<35	35-45	>45
Dummy SE	0.140***	0.086**	0.060*	0.094**	0.066**	0.034
	(0.05)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
Elasticity of Consumption		0.439***			0.236***	
		(0.04)			(0.01)	
Variance SE	0.572		0.783		0.764	
Variance E			0.496			
Spread	0.077		0.288		0.268	
HLP						
	0.273***	0.177**	0.127*	0.327***	0.244**	0.132
	(0.075)	(0.074)	(0.067)	(0.119)	(0.106)	(0.11)
PW						
Upper Bound						
Multiplier	1.433***	1.439***	1.377***	1.549***	1.565***	1.386***
	(0.147)	(0.129)	(0.106)	(0.273)	(0.219)	(0.175)
Underreporting	0.302***	0.305***	0.274***	0.354***	0.361***	0.278***
	(0.072)	(0.062)	(0.056)	(0.114)	(0.089)	(0.091)
Lower Bound						
Multiplier	1.320***	1.026***	0.953***	1.427***	1.117***	0.959***
	(0.135)	(0.092)	(0.073)	(0.252)	(0.156)	(0.121)
Underreporting	0.242***	0.026	-0.05	0.299**	0.105	-0.043
	(0.078)	(0.087)	(0.081)	(0.124)	(0.125)	(0.132)
Midpoint						
Multiplier	1.376***	1.233***	1.165***	1.488***	1.341***	1.172***
	(0.141)	(0.111)	(0.09)	(0.262)	(0.188)	(0.148)
Underreporting	0.273***	0.189***	0.142**	0.328***	0.254**	0.147
	(0.074)	(0.073)	(0.066)	(0.119)	(0.104)	(0.108)
Observations						
Self-Employed	156	251	328	156	251	328
Employed			3619			
Total		4752			4752	

Notes:

1. Robust standard errors in parenthesis
2. The stars signal significance at the following levels:
   * p<0.10 ** p<0.05 *** p<0.01

Table 10: Estimation of Underreporting by Age Band of the Self-Employed

	IV			OLS		
	<35	35-45	>45	<35	35-45	>45
Dummy SE	0.134***	0.079**	0.064*	0.080*	0.063*	0.039
	(0.05)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
Elasticity of Consumption		0.439***			0.236***	
		(0.04)			(0.01)	
Variance SE	0.576		0.779		0.751	
Variance E			0.496			
Spread	0.081		0.283		0.255	
HLP						
	0.263***	0.164**	0.135**	0.288**	0.235**	0.151
	(0.075)	(0.076)	(0.065)	(0.128)	(0.107)	(0.106)
PW						
Upper Bound						
Multiplier	1.418***	1.403***	1.385***	1.466***	1.532***	1.411***
	(0.145)	(0.128)	(0.104)	(0.264)	(0.214)	(0.176)
Underreporting	0.295***	0.287***	0.278***	0.318***	0.347***	0.291***
	(0.072)	(0.065)	(0.054)	(0.123)	(0.091)	(0.089)
Lower Bound						
Multiplier	1.300***	1.021***	0.965***	1.345***	1.115***	0.984***
	(0.133)	(0.093)	(0.072)	(0.242)	(0.155)	(0.123)
Underreporting	0.231***	0.02	-0.036	0.256*	0.103	-0.017
	(0.078)	(0.089)	(0.078)	(0.134)	(0.125)	(0.127)
Midpoint						
Multiplier	1.359***	1.212***	1.175***	1.406***	1.323***	1.197***
	(0.139)	(0.111)	(0.088)	(0.253)	(0.185)	(0.15)
Underreporting	0.264***	0.175**	0.149**	0.289**	0.244**	0.165
	(0.075)	(0.075)	(0.064)	(0.128)	(0.105)	(0.104)
Observations						
Self-Employed	148	251	334	148	251	334
Employed			3619			
Total		4752			4752	

Notes:

1. Robust standard errors in parenthesis
2. The stars signal significance at the following levels:
   * p<0.10 ** p<0.05 *** p<0.01

Table 11: Estimation of Underreporting by their consideration of the self-employed activity.

	IV		OLS	
	Job	Business	Job	Business
Dummy SE	0.080**	0.099***	0.033	0.074***
	(0.04)	(0.03)	(0.03)	(0.03)
Elasticity of Consumption	0.439***		0.236***	
	(0.04)		(0.01)	
Variance SE	0.716		0.716	
Variance E	0.496			
Spread	0.221		0.221	
HLP				
	0.167**	0.203***	0.132	0.268***
	(0.065)	(0.054)	(0.108)	(0.082)
PW				
Upper Bound				
Multiplier	1.372***	1.462***	1.317***	1.593***
	(0.107)	(0.1)	(0.164)	(0.178)
Underreporting	0.271***	0.316***	0.241**	0.372***
	(0.057)	(0.047)	(0.095)	(0.07)
Lower Bound				
Multiplier	1.050***	1.075***	0.008	1.172***
	(0.082)	(0.073)	(0.126)	(0.131)
Underreporting	0.048	0.07	1.162***	0.147
	(0.074)	(0.064)	(0.124)	(0.095)
Midpoint				
Multiplier	1.211***	1.269***	1.162***	1.383***
	(0.095)	(0.087)	(0.145)	(0.154)
Underreporting	0.174***	0.212***	0.14	0.277***
	(0.065)	(0.054)	(0.107)	(0.081)
Observations				
Self-Employed	309		309	
Employed	4028		4028	
Total	4738		4738	

Notes:

1. Robust standard errors in parenthesis
2. The stars signal significance at the following levels:
   * p<0.10 ** p<0.05 *** p<0.01

Table 12: Estimation of Underreporting by Number of Self-Employed in the Household.

	IV		OLS	
	One	Both	One	Both
Dummy SE	0.059**	0.138**	0.048**	0.076*
	(0.02)	(0.06)	(0.02)	(0.04)
Elasticity of Consumption	0.434***		0.235***	
	(0.04)		(0.01)	
Variance SE	0.657		0.657	
Variance E	0.496			
Spread	0.161		0.161	
HLP				
	0.127**	0.273***	0.183**	0.276**
	(0.05)	(0.097)	(0.078)	(0.137)
PW				
Upper Bound				
Multiplier	1.256***	1.711***	1.343***	1.716***
	(0.071)	(0.227)	(0.128)	(0.324)
Underreporting	0.204***	0.415***	0.256***	0.417***
	(0.045)	(0.078)	(0.071)	(0.11)
Lower Bound				
Multiplier	1.044***	1.107***	0.104	1.110***
	(0.059)	(0.147)	(0.106)	(0.209)
Underreporting	0.042	0.096	1.230***	0.099
	(0.054)	(0.12)	(0.086)	(0.17)
Midpoint				
Multiplier	1.150***	1.409***	1.230***	1.413***
	(0.065)	(0.187)	(0.117)	(0.266)
Underreporting	0.130***	0.290***	0.187**	0.292**
	(0.049)	(0.094)	(0.078)	(0.133)
Observations				
Self-Employed	567		567	
Employed	4028		4028	
Total	4595		4595	

Notes:

1. Robust standard errors in parenthesis
2. The stars signal significance at the following levels:
   * p<0.10 ** p<0.05 *** p<0.01

Table 13: Estimation of Underreporting by Position in the Household of the Self-Employed.

	IV			OLS		
	Spouse SE	HRP SE	Both SE	Spouse SE	HRP SE	Both SE
Dummy SE	0.032	0.094***	0.143**	0.040	0.056**	0.077*
	(0.04)	(0.03)	(0.06)	(0.04)	(0.03)	(0.04)
Elasticity of Consumption		0.443***			0.236***	
		(0.04)			(0.01)	
Variance SE	0.427		0.825		0.794	
Variance E			0.496			
Spread	-0.069		0.330		0.298	
HLP						
	0.07	0.173***	0.152*	0.511***	0.205**	0.148
	(0.083)	(0.053)	(0.095)	(0.138)	(0.085)	(0.135)
PW						
Upper Bound						
Multiplier	1.042***	1.498***	1.716***	1.148***	1.536***	1.721***
	(0.093)	(0.099)	(0.224)	(0.187)	(0.165)	(0.323)
Underreporting	0.04	0.332***	0.417***	0.129***	0.349***	0.419***
	(0.086)	(0.044)	(0.076)	(0.142)	(0.07)	(0.109)
Lower Bound						
Multiplier	1.110***	1.019***	1.110***	1.223***	1.046***	1.114***
	(0.099)	(0.067)	(0.145)	(0.2)	(0.112)	(0.209)
Underreporting	0.099	0.019	0.099	0.182	0.044	0.102
	(0.08)	(0.065)	(0.117)	(0.133)	(0.103)	(0.168)
Midpoint						
Multiplier	1.076***	1.259***	1.413***	1.185***	1.291***	1.418***
	(0.096)	(0.083)	(0.184)	(0.193)	(0.139)	(0.266)
Underreporting	0.07	0.205***	0.292***	0.156	0.225***	0.295**
	(0.083)	(0.052)	(0.092)	(0.138)	(0.083)	(0.132)
Observations						
Self-Employed	174	428	131	174	428	131
Employed			3619			
Total		4752			4752	

Notes:

1. Robust standard errors in parenthesis
2. The stars signal significance at the following levels:
   * p<0.10 ** p<0.05 *** p<0.01

Table 14: Estimation of Underreporting based on Association Status.

	IV		OLS	
	Own Account	Partnership	Own Account	Partnership
Dummy SE	0.083***	0.119**	0.050**	0.096**
	(0.03)	(0.06)	(0.02)	(0.04)
Elasticity of Consumption	0.437***		0.236***	
	(0.04)		(0.01)	
Variance SE	0.695		0.695	
Variance E		0.496		
Spread	0.200		0.200	
HLP				
	0.173***	0.238**	0.191**	0.335***
	(0.046)	(0.099)	(0.074)	(0.127)
PW				
Upper Bound				
Multiplier	1.363***	1.782***	1.393***	2.041***
	(0.076)	(0.231)	(0.128)	(0.391)
Underreporting	0.266***	0.439***	0.282***	0.510***
	(0.041)	(0.073)	(0.066)	(0.094)
Lower Bound				
Multiplier	1.074***	0.966***	0.089	1.106***
	(0.06)	(0.125)	(0.101)	(0.212)
Underreporting	0.069	-0.035	1.245***	0.096
	(0.052)	(0.134)	(0.084)	(0.173)
Midpoint				
Multiplier	1.218***	1.374***	1.245***	1.574***
	(0.068)	(0.178)	(0.115)	(0.301)
Underreporting	0.179***	0.272***	0.197***	0.365***
	(0.046)	(0.094)	(0.074)	(0.122)
Observations				
Self-Employed	607		607	
Employed	4028		4028	
Total	4635		4635	

Notes:

1. Robust standard errors in parenthesis
2. The stars signal significance at the following levels:
   * p<0.10 ** p<0.05 *** p<0.01

Table 15: Estimation of Income Underreporting by Sex of the Self-Employed

	IV			OLS		
	Wife SE	Husband SE	Both SE	Wife SE	Husband SE	Both SE
Dummy SE	-0.013	0.090***	0.156***	-0.016	0.063***	0.087**
	(0.05)	(0.03)	(0.06)	(0.06)	(0.02)	(0.04)
Elasticity of Consumption		0.439***			0.237***	
		(0.04)			(0.01)	
Variance SE	0.672		0.864		0.726	
Variance E			0.496			
Spread	0.177		0.369		0.230	
HLP						
	-0.03	0.186***	0.299***	-0.068	0.232***	0.307**
	(0.127)	(0.049)	(0.092)	(0.25)	(0.075)	(0.129)
PW						
Upper Bound						
Multiplier	1.076***	1.414***	1.833***	1.038***	1.499***	1.854***
	(0.133)	(0.085)	(0.24)	(0.243)	(0.147)	(0.344)
Underreporting	0.071	0.293***	0.454***	0.037	0.333***	0.461***
	(0.114)	(0.042)	(0.071)	(0.225)	(0.066)	(0.1)
Lower Bound						
Multiplier	0.875***	1.068***	1.110***	0.845***	1.132***	1.123***
	(0.108)	(0.064)	(0.145)	(0.197)	(0.111)	(0.208)
Underreporting	-0.142	0.063	0.099	-0.184	0.116	0.109
	(0.141)	(0.056)	(0.118)	(0.277)	(0.087)	(0.165)
Midpoint						
Multiplier	0.976***	1.241***	1.471***	0.941***	1.315***	1.488***
	(0.12)	(0.074)	(0.193)	(0.22)	(0.129)	(0.276)
Underreporting	-0.025	0.194***	0.320***	-0.062	0.240***	0.328***
	(0.126)	(0.048)	(0.089)	(0.248)	(0.075)	(0.125)
Observations						
Self-Employed	83	519	133	83	519	133
Employed			3619			
Total		4752			4752	

Notes:

1. Robust standard errors in parenthesis
2. The stars signal significance at the following levels:
   * p<0.10 ** p<0.05 *** p<0.01

Table 16: Estimation of Underrepoting by Sex of the main earner of Self-Employment Income.

	IV		OLS	
	SE Wife	SE Husband	SE Wife	SE Husband
Dummy SE	0.055	0.098***	0.007	0.070***
	(0.05)	(0.03)	(0.04)	(0.02)
Elasticity of Consumption	0.439***		0.236***	
	(0.04)		(0.01)	
Variance SE	0.879		0.879	
Variance E	0.496			
Spread	0.384		0.384	
HLP				
	0.118	0.199***	0.029	0.255***
	(0.107)	(0.045)	(0.178)	(0.069)
PW				
Upper Bound				
Multiplier	1.477***	1.424***	1.341***	1.532***
	(0.18)	(0.08)	(0.246)	(0.142)
Underreporting	0.323***	0.298***	0.254*	0.347***
	(0.082)	(0.04)	(0.137)	(0.061)
Lower Bound				
Multiplier	0.871***	1.095***	-0.264	1.178***
	(0.106)	(0.062)	(0.145)	(0.109)
Underreporting	-0.148	0.087*	1.066***	0.151*
	(0.14)	(0.052)	(0.232)	(0.079)
Midpoint				
Multiplier	1.174***	1.259***	1.066***	1.355***
	(0.143)	(0.071)	(0.196)	(0.126)
Underreporting	0.148	0.206***	0.062	0.262***
	(0.104)	(0.045)	(0.172)	(0.069)
Observations				
Self-Employed	137		137	
Employed	4028		4028	
Total	4165		4165	

Notes:

1. Robust standard errors in parenthesis
2. The stars signal significance at the following levels:

* p<0.10 ** p<0.05 *** p<0.01

Table 17: Estimates of Self-Employment underreporting by region, 2010-2012. Pooled data

	IV							OLS						
	North	Yorkshire & The Humber	Midlands	East	Greater London	South	Wales & Scotland	North	Yorkshire & The Humber	Midlands	East	Greater London	South	Wales & Scotland
Dummy SE	0.082	-0.093	0.064	0.160***	0.235***	0.061	0.031	0.019	-0.157**	0.025	0.151***	0.244***	0.046	0.016
	(0.06)	(0.09)	(0.06)	(0.05)	(0.06)	(0.04)	(0.07)	(0.05)	(0.07)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)
Elasticity of Consumption				0.427***							0.231***			
				(0.04)							(0.01)			
Variance SE	0.714	0.932	0.844	0.601	0.811	0.713	0.617	0.714	0.932	0.844	0.601	0.811	0.713	0.617
Variance E				0.496							0.496			
Spread	0.418	0.636	0.548	0.305	0.515	0.417	0.321	0.418	0.636	0.548	0.305	0.515	0.417	0.321
HLP														
	0.175*	-0.245	0.139	0.312***	0.423***	0.133*	0.071	0.078	-0.972*	0.103	0.479***	0.651***	0.18	0.065
	(0.106)	(0.266)	(0.118)	(0.079)	(0.09)	(0.077)	(0.147)	(0.206)	(0.574)	(0.191)	(0.12)	(0.083)	(0.125)	(0.213)
PW														
Upper Bound														
Multiplier	1.383***	1.096***	1.467***	1.541***	2.132***	1.316***	1.151***	1.238***	0.692***	1.408***	2.036***	3.525***	1.391***	1.145***
	(0.178)	(0.234)	(0.201)	(0.177)	(0.333)	(0.117)	(0.182)	(0.277)	(0.201)	(0.301)	(0.469)	(0.838)	(0.213)	(0.261)
Underreporting	0.277***	0.088	0.318***	0.351***	0.531***	0.240***	0.131	0.192	-0.445	0.290*	0.509***	0.716***	0.281**	0.126
	(0.093)	(0.195)	(0.093)	(0.075)	(0.073)	(0.067)	(0.137)	(0.181)	(0.42)	(0.152)	(0.113)	(0.067)	(0.11)	(0.2)
Lower Bound														
Multiplier	1.062***	0.588***	0.920***	1.373***	1.411***	1.012***	1.006***	0.951***	0.372***	0.883***	1.813***	2.334***	1.070***	1.000***
	(0.137)	(0.126)	(0.126)	(0.158)	(0.22)	(0.09)	(0.159)	(0.213)	(0.108)	(0.189)	(0.418)	(0.555)	(0.163)	(0.228)
Underreporting	0.059	-0.699*	-0.087	0.271***	0.291***	0.012	0.006	-0.052	-1.691**	-0.132	0.448***	0.571***	0.065	0
	(0.121)	(0.363)	(0.149)	(0.084)	(0.111)	(0.088)	(0.157)	(0.235)	(0.783)	(0.242)	(0.127)	(0.102)	(0.143)	(0.228)
Midpoint														
Multiplier	1.223***	0.842***	1.193***	1.457***	1.771***	1.164***	1.078***	1.094***	0.532***	1.146***	1.924***	2.929***	1.231***	1.072***
	(0.157)	(0.18)	(0.164)	(0.168)	(0.277)	(0.103)	(0.17)	(0.245)	(0.155)	(0.245)	(0.443)	(0.697)	(0.188)	(0.245)
Underreporting	0.182*	-0.187	0.162	0.314***	0.435***	0.141*	0.073	0.086	-0.88	0.127	0.480***	0.659***	0.187	0.067
	(0.105)	(0.254)	(0.115)	(0.079)	(0.088)	(0.076)	(0.147)	(0.204)	(0.547)	(0.186)	(0.12)	(0.081)	(0.124)	(0.213)
Observations														
Self-Employed	95	57	104	90	90	209	90	95	57	104	90	90	209	90
Employed				4028							4028			
Total	4123	4085	4132	4118	4118	4237	4118	4123	4085	4132	4118	4118	4237	4118

Notes:
1. Robust standard errors in parenthesis
2. The stars signal significance at the following levels:

* p<0.10 ** p<0.05 *** p<0.01

8 References

Aguiar, M., & Hurst, E. (2004). Consumption vs. expenditure (No. w10307). National Bureau of Economic Research.

Baldini, M., Bosi, P., & Lalla, M. (2009). Tax Evasion and Misreporting in Income Tax Returns and Household Income Surveys. Politica Economica, 3, 333-348.

Baldry, J. (1987). Income Tax Evasion and the Tax Schedule: Some Experimental Results. Public Finance, 42, 357-383.

Bee, A., Meyer, B.D. & Sullivan, J.X. The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?. No. w18308. National Bureau of Economic Research, 2012.}

Bennet, C. (2005). Preliminary Results of the National Research Program's reporting . In I. R. Services, IRS Research Bulletin (Publication 1500). Washington D.C.

Besim, M., and Jenkins, G. P. (2005). Tax compliance: when do employees behave like the self-employed?. Applied Economics, 37 (10), 1201–1208.

Brown, W., & Miller, T. (1988). Following-up Previously Adjudicated Delinquents: A Method. In R. Jenkins, & W. Brown, The Abandonment of Delinquent Behaviour: Promoting the Turnaround. New York: Praeger.

Bruce, D. (2000). Effects of the United States tax system on transitions into self-employment. Labour Economics, 7, 545-574.

Cowell, F. (1990). Cheating the government. Cambridge, Massachusetts: The MIT Press.

Dilnot, A., & Morris, C. (1981). What do we know abobut the black economy? Fiscal Studies, 2(1), 58-73.

Engstrom, P., & Holmlund, B. (2009). Tax evasion and self-employment in a high-tax country: evidence from Sweden. Applied Economics, 41(19), 2419-2430.

Erard, B., & Feinstein, J. (2007). Econometric Models for Multi-Stage Audit Processes: An Application to the IRS National Research Program. International Center for Public Policy Working Paper Series, GSU paper0723.

Erard, B., & Feinstein, J. (2012). The Individual Income Reporting Gap: What We See and What We Do't. IRS Research Bulletin, Publication 1500.

Feinstein, J. (1991). An Econometric Analysis of Income Tax Evasion and Its Detection. The RAND Journal of Economics, 22(1), 14-35.

Foster, K. (1996). A comparison of the Census characteristics of respondents and non-respondents to the 1991 FES. ONS Survey Methodology Bulletin No. 38.

Friedland, N., Maital, S., & Rutenberg, A. (1978). A simulation study of income tax evasion. Journal of Public Economics, 10, 107-116.

Graetz, M., & Wilde, L. (1985). The Economics of Tax Compliance: Fact and Fantasy. National Tax Journal, 38(3), 355-363.

Hirschi, T., & Gottfredson, M. (1983). Age and the Explanation of Crime. American Journal of Sociology, 89(3), 552-584.

HMRC (2014). Levelling the tax playing field. Retrieved from:

http://www.hmrc.gov.uk/budget2013/level-tax-playing-field.pdf

Hurst, E., Li, G., & Pugsley, B. (2014). Are Household Surveys Like Tax Forms: Evidence from Income Underreporting of the Self Employed. The Review of Economics and Statistics, 96(1), 19-33.

ISWGNA. (2006). Estimates of the unrecorded economy and national accounts. Declaration of ISWGNA. Retrieved from OECD:

http://www.oecd.org/tax/administration/36333066.pdf

James, H., & Rassekh, F. (2000). Smith, Friedman, and Self-Interest in Ethical Society. Business Ethics Quarterly, 10(3), 659-674.

Johansson, E. (2005). An estimate of self-employment income underreporting in Finland. Nordic Journal of Political Economy, 31, 99-109.

Kim, B., Gibson, J., and Chung, C. (2009). "Using panel data to exactly estimate income under-reporting by the self-employed." Working Paper 09-02, Korea Institute for International Economic Policy.

Kleven, H., Knudsen, M., Kreiner, C., Pedersen, S., & Saez, E. (2011). Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. Econometrica, 79(3), 651-692.

Kukk, M., Staehr, K. (2014). Income underreporting by households with business income: evidence from Estonia. Post-Communist Economies, Taylor & Francis Journals, 26(2), 257-276.

Lyssiotou, P., Pashardes, P., & Stengos, T. (2004). Estimates of the Black Economy based on Consumer Demand Approaches. The Economic Journal, 114, 622-640.

Mears, D., Ploeger, M., & Warr, M. (1998). Explaining the Gender Gap in Delinquency: Peer Influence and Moral Evaluations of Behavior. Journal of Research in Crime and Delinquency, 31, 251-266.

Newey, W. (1985). Generalized Method of Moments Specification Testing. Journal of Econometrics, 29, 229-256.

Parente, P., & Santos Silva, J. (2012). A cautionary note on tests of overidentifying restrictions. Economic Letters, 115(2), 314-317.

Pissarides, C., & Weber, G. (1989). An expenditure-based estiamte of Britain's black economy. Journal of Public Economics, 39(1), 17-32.

Schneider, F. E. (2000). Shadow economies: size, causes and consequences. Journal of Economic Literature, 38, 77-114.

Schneider, F., & Williams, C. (2013). The Shadow Economy. Institute of Economic Affairs.

Schuetze, H. (2000). Profiles of Tax Non-compliance Among the Self-Employed in Canada: 1969 to 1992. Canadian Public Policy, 28(2), 219-237.

Smith, S. (1986). Britain's Shadow Economy. London: Claredon Press. Oxford.

Spicer, M., & Becker, L. (1980). Fiscal Inequity and Tax Evasion: An Experimental Approach. National Tax Journal, 171-175.

Spicer, M., & Hero, R. (1985). Tax Evasion and Heuristics. A Research Note. Journal of Public Economics, 26, 263-267.

Tanzi, V. (1982). The Underground Economy in the United States and Abroad. Lexington: Lexington Books.

Tedds, L. M. (2010). Estimating the income reporting function for the selfemployed.Empirical Economics, 38 (3), 669 – 687.

Tittle, C. (1980). Sanctions and Social Deviance. New York: Praeger.

Torgler, B., & Valev, N. (2006). Women and Illegal Activities: Gender Differences and Women's Willingness to Comply over Time. Working Paper No. 2006-15. Center for Research in Economics, Management and the Arts.

Torgler, B., & Valev, N. (2010). Gender and public attitude toward corruption and tax evasion. Contemporary Economic Policy, 28, 554-568.

Torgler, J., & Schneider, F. (2006). What shapes attitudes toward paying taxes? Working Paper, Department of Economics, Johannes Kepler University of Linz, No. 0608.

Van Eck, R. (1987). Secondary Activities and the National Accounts. Working Paper. Central Bureau of Statistics.

Vogel, J. (1974). Taxation and Public Opinion in Sweden: an Interpretation of Recent Survey Data. National Tax Journal, 27, 499-513.