AN EXPENDITURE-BASED ESTIMATION OF SELF-EMPLOYMENT INCOME UNDERREPORTING IN FINLAND

NOVEMBER 2000
Key words: Tax Evasion, Self-Employment, Food Consumption

JEL Classification: H26, J23, D12

© Swedish School of Economics and Business Administration & Edvard Johansson

Edvard Johansson
Department of Economics and Political Science
Swedish School of Economics and Business Administration
P.O.Box 479
00101 Helsinki, Finland

Distributor:

Library
Swedish School of Economics and Business Administration
P.O.Box 479
00101 Helsinki
Finland

Phone: +358-9-431 33 376, +358-9-431 33 265
Fax: +358-9-431 33 425
E-mail: publ@shh.fi
http://www.shh.fi/link/bib/publications.htm

SHS intressebyrå IB (Oy Casa Security Ab), Helsingfors 2000

ISBN 951-555-656-2
ISSN 0357-4598
An Expenditure-Based Estimation of Self-Employment Income

Underreporting in Finland

Edvard Johansson*

October 2000

Abstract

This paper estimates the extent of income underreporting by the self-employed in Finland using the expenditure based approach developed by Pissarides & Weber (1989). Household spending data are for the years 1994 to 1996. The results suggest that self-employment income in Finland is underreported by some 27% on average. Since income for the self-employed is about 8% of all incomes in Finland, the size of this part of the black economy in Finland is estimated to be about 2.3% of GDP.

JEL Classifications H26, J23, D12

* Swedish School of Economics and Business Administration, P.O. Box 479, 00101 Helsinki, Finland. Phone: +358-9-43133406. Fax: +358-9-431334382. E-mail: edvard.johansson@shh.fi. Helpful comments by Tom Berglund and Henry Ohlsson and financial support from the Nordisk skattevitenskaplig forskningsråd and the Jenny ja Antti Wihurin rahasto is gratefully acknowledged.
1. Introduction

There has recently been an increase in the interest in self-employment. This is true for academic economic research as well as for society in general. One major explanation for this increased interest is that self-employment is regarded by many, including governments in many countries, to be essential for the well-being of the economy. Thus, the self-employed are frequently considered to be entrepreneurs, who do not only create jobs for themselves but also for others.

Consequently, there has been quite a few papers written on self-employment. One of the major research questions in this literature has been the question of why some individuals quit their wage or salary job to become self-employed. Well-known papers in this literature include Evans & Leighton (1989), Evans & Jovanovic (1989), Holtz-Eakin et al. (1994) and Blanchflower & Oswald (1998). There are also two papers examining transitions from wage or salary work to self-employment using data from Finland, namely Johansson (2000) and Uusitalo (1999). All these papers document factors that can explain why some individuals quit their wage or salary jobs to become self-employed.

Tax evasion is an interesting, but not yet very well understood or documented, explanation of why someone would want to give up his or her salary job to become self-employed. It is probably much easier to transform private consumption to business expenses and avoid tax for a person that is self-employed than for a person that works as an employee. In time series studies of the self-employment rate, e.g. Blau (1987) and Parker (1996), it has also been found that when the marginal tax rate increases, the number of self-employed individuals in the economy increases. There are also two papers, Bruce (2000) and Schuetze (2000), that uses micro-data to
investigate whether individuals enter self-employment because they want to avoid tax. However, this strand of the self-employment literature, even though important, is still in its early stages.

In this context it is, therefore, natural to think about the magnitude of income underreporting by the self-employed. If it is possible to underreport self-employment income substantially, we should be more concerned about the results obtained in the above mentioned studies.

However, measuring income underreporting is difficult, since the whole motivation behind underreporting of income is to avoid the true income to be registered in data available to others. In Pissarides & Weber (1989), (henceforth P&W), however, an interesting expenditure-based estimation technique is presented. In short, this method assumes that income underreporting occurs among the self-employed but not the employees in employment. Food expenditure, on the other hand, is assumed to be correctly reported both by employees in employment and the self-employed. Food expenditure equations are then estimated, using the assumption that the “marginal propensity to consume food” is the same both for the self-employed and the employees, and inverted to arrive at a true self-employment income. Pissarides & Weber estimate that the true income of the self-employed on average is about 55% higher than the reported income.

A small number of other studies also exist that calculate self-employment income underreporting by the P&W method. Using data from Sweden, Apel (1994) estimates the true income of the self-employed in Sweden to be about 25% higher than the reported income. For the UK, there exist two more studies apart from the P&W study, namely Baker (1993) and Cullinan (1997).
Cullinan (1997) investigate the extent of underreporting in the UK for the years 1987 and 1992. He finds that underreporting amounts to between 19% and 23% of self-employment income in 1987 and to between 23% and 37% in 1992. Baker (1993) uses data from the same source as Cullinan (1997) and Pissarides and Weber (1989), but investigate the issue for every year from 1978 to 1991. His estimates suggest that actual self-employment incomes may be between 1.3 to 1.5 times reported self-employment incomes. He finds no clear time series pattern in the underreporting measures, but finds considerable variation between occupations and industry groups. Schuetze (2000), and Mirus and Smith (1996) estimate underreporting of self-employment in Canada. Mirus and Smith (1996) firstly, report that self-employment income on average is underreported by some 12.5%. In the study by Schuetze (1999), underreporting is disaggregated across years, demographic characteristics and occupation. It is found that the degree of underreporting of self-employment income varies significantly by occupation, age and the number of households members self-employed. However, no significant time trend is found in this study.

By using this expenditure based estimation method it is also possible to measure the size of the underground, or the “black” economy for a country. Remember that one of the assumptions in this line of research is that underreporting of income occurs exclusively in the self-employed sector of the economy. If the estimated underreporting of income by the self-employed is then multiplied by the size of the self-employed sector of the economy, one arrives at an estimate of the “black” sector of an economy. Pissarides & Weber (1989) report that the size of Britain's black economy is 5.5% of GDP, since the estimated underreporting is 55% and the size of the
self-employed sector of the economy is 10%. The corresponding figure in Apel (1994) is about 1% of GDP.

In this paper we estimate underreporting of self-employment income in Finland using the P&W method. The data we use is data from a survey of household expenditures compiled by Statistics Finland during the years 1994 to 1996. The result of the estimation is that self-employment income on average is underreported by 27% Since the share of total income in Finland that comes from self-employment is about 8% of GDP\(^1\), we estimate that the “black” economy has a size that is about 2.3% of GDP.

The paper is organised as follows. In section 2 we review the Pissarides & Weber (1989) methodology of estimating underreporting of self-employment income. In section 3 we present the data and the results. Section 4 concludes the paper.


The expenditure-based estimation approach, originally developed by P&W (1989) relies on three major assumptions. 1) the reporting of expenditure by all groups of the population is accurate. 2) the reporting of income by some groups in the population is accurate. 3) the expenditure function on some items are the same for all groups in the population.
Income underreporting is then modelled in the following way: $C_{ij}$ is the consumption of household $i$ on item $j$ on after tax income $Y'_i$. $Z_i$ is a vector of household characteristics. It is assumed that $C_{ij}$ for food is correctly reported by all households, $Y'_i$ is correctly reported by employees in employment and $Z_i$ is correctly recorded for all households. Let then $Y_i$ be the “true” income for household $i$. Then $Y_i = Y'_i$ for employees in employment but for the self-employed we have the following:

$$Y_i = k_i Y'_i, \quad k_i \geq 1.$$  

where $k_i$ is a random variable that shows the extent of under-reporting of income by self-employed household $i$. A bigger $k_i$ indicates more under-reporting by household $i$.

For each item of expenditure $j$ there is an expenditure function,

$$\ln C_{ij} = Z_i \alpha_j + \beta_j \ln Y_i^p + \varepsilon_{ij}$$

where $\alpha_j$ is a vector of parameters, $\beta_j$ is a scalar, the “marginal propensity to consume” good $j$ and $\varepsilon_{ij}$ is white noise. $Y_i^p$ is the measure of income that influences consumption decisions. This measure of income is likely to be less volatile than observed income $Y'_i$. Pissarides & Weber
refer to this measure of income as permanent income, without necessarily requiring that that the expenditure function conforms exactly to the permanent income hypothesis.

The distinction between permanent income and measured income is needed in this context because for a given level of permanent income, the measured income of the self-employed may be more variable than the measured income of employees in employment. If this is true, then the consequence is that the measure of the income under-reporting by the self-employed will have to be adjusted accordingly. In general it is assumed that permanent and measured income are related by

\[ Y_i = p_i Y_{p}' \]

where \( p_i \) is a random variable. The expected value of \( p_i \) depends on random events. In a “good” year, \( p_i \) will have a mean that is bigger than 1. It is assumed that the mean of \( p_i \) is the same both for employees and the self-employed. However, the variance of \( p_i \) is expected to be bigger for the self-employed than for the employees.

(1) and (3) will then imply that the log of permanent income is

\[ \ln Y_i'' = \ln Y_i' - \ln p_i + \ln k_i \]
Thus, if permanent income is used instead of observed income in (2) there are two additional random regressors – $\ln p_i$ and $k_i$, each entering with coefficient $\beta_j$.

Since no data on $p_i$ or $k_i$ do not exist, Pissarides & Weber make the assumption that $p_i$ and $k_i$ are both log-normally distributed random variables to make estimation tractable. They then write them as deviations from their means:

\[
\ln p_i = \mu_p + u_i \\
\ln k_i = \mu_k + v_i
\]

The random variables $u_i$ and $v_i$ have zero means and constant variances $\sigma_u^2$ and $\sigma_v^2$ within each occupational group. No assumptions about any covariation between $u_i$, $v_i$ and $\varepsilon_y$ is made at this stage.

So far the arguments are as follows: a) the employees in employment report their incomes correctly, so for them $k_i = 1$ and $\sigma_v^2 = 0$, whereas for the self-employed we should generally expect to find $\mu_k > 0$ and $\sigma_v^2 > 0$. b) the mean of $p_i$, denoted by $\bar{p}$ is the same for each occupational group, but the variance $\sigma_u^2$ is bigger for the self-employed than for employees in employment.

By the log-normality of $p_i$, the mean of $p_i$ and the mean of its log $\mu_p$, are related by
(7) \[ \ln \bar{p} = \mu_p + \frac{1}{2}\sigma^2_u \]

and so, if the subscript SE is used to denote the self-employed and the subscript EE to denote employees in employment, we have that:

(8) \[ \mu_{p_{SE}} - \mu_{p_{EE}} = - \frac{1}{2}(\sigma^2_{u_{SE}} - \sigma^2_{u_{EE}}) \leq 0 \]

By the convexity of logs, equality between the means of \( p_i \) implies that the mean of the log of \( p_i \) of the self-employed is less than the mean of the log of \( p_i \) of the employees.

Substituting now from (4), (5) and (6) into (2) we get:

(9) \[ \ln C_{ij} = Z_i \alpha_j + \beta_j \ln Y_i - \beta_j (\mu_p - \mu_k) - \beta_j (u_i - v_i) + \epsilon_{ij} \]

Suppose then that equation (9) is estimated separately for the self-employed and for the employees but \( \alpha_j \) and \( \beta_j \) are restricted to be common. Then the intercepts of the equations should differ because \( \mu_p - \mu_k \) is not the same in each group. The variance of the errors of each equation should also differ, with the self-employed generally having bigger variance. These differences in the estimates can be used to obtain an estimate of income under-reporting for the self-employed.
Thus, the residual consumption variance for the self-employed is hypothesised to be higher than the residual variance for the employees. This variance is a composite of the variances and covariances of three types of errors, as can be seen from (9). In order to get an independent estimate of the variance of errors of income, we need to calculate the residual variance from a reduced –form income regression of the form

\[
\ln Y'_i = Z_i \delta_i + X \delta_z + \zeta_i
\]

where \( X_i \) is a set of identifying instruments. The residual \( \zeta_i \) is a composite of three errors; unexplained variations in permanent income, deviations of actual from permanent income, \( u_i \) and deviations of actual income from reported income \( v_i \). By the previous argument concerning the properties of \( u_i \) and \( v_i \) for the self-employed and the employees, the residual income variance of the self-employed should exceed the residual income variance of the employees.

From the discussion of equations (5), (6) and (8) we can now conclude that the following equation holds for \( \gamma_j \):

\[
\gamma_j = \beta_j \left[ \mu + \frac{1}{2} \left( \sigma_{uE}^2 - \sigma_{vE}^2 \right) \right]
\]

Now, what we really are interested in is the mean value of \( k_j \), the number by which average reported self-employment income should be multiplied to give average true income. If we let this number be \( \overline{k} \) we have:
Thus, to go from the estimated coefficient on the self-employment dummy to an estimate of the mean of income under-reporting we need to know the variances of $u_i$ and $v_i$. Generally these are not known, but inferences about them can be made from the estimated residual income variance.

Suppose first that the unexplained variations in permanent income in (10) have the same variance for both the employees and the self-employed. This should not be a too unrealistic assumption, given that these variations are due to omitted variables and that the self-employment dummy is one of the regressors in (11). Remembering that $\sigma^2_{vEE}=0$ we have that:

\[
\text{var} \zeta_{SE} - \text{var} \zeta_{EE} = \text{var}(u - v)_{SE} - \text{var} u_{EE}
\]

and if expanding the right hand side:

\[
\sigma^2_{\zeta_{SE}} - \sigma^2_{\zeta_{EE}} = \sigma^2_{uSE} + \sigma^2_{vSE} - 2\text{cov}(uv)_{SE} - \sigma^2_{uEE} = \sigma^2_{vSE} + \rho \sigma_{uSE} \sigma_{vSE} - \sigma^2_{uEE}
\]

where $\rho$ is the partial correlation coefficient between $u_{SE}$ and $v_{SE}$. From (11) and (12) we obtain:

\[
\ln \bar{k} = \mu_k + \frac{1}{2} \sigma^2_{vSE} = \frac{\gamma_j}{\beta_j} + \frac{1}{2} \left( \sigma^2_{vSE} - \sigma^2_{uSE} + \sigma^2_{uEE} \right)
\]
Comparing (14) and (15) we can see that the residual income variances do not contain enough information to solve for $\ln k$. Nevertheless, it is possible to calculate a range for the mean of under-reporting by using (14).

The question now is how the estimate the mean of under-reporting is affected when $\sigma_{vSE}^2$ and $\sigma_{uSE}^2$ vary over their feasible range that satisfies condition (14), for given values of the other variances (which are treated as parametric) and given the partial correlation coefficient $\rho$ between $u_{SE}$ and $v_{SE}$. If $\rho = 0$ then $\sigma_{vSE}^2$ and $\sigma_{uSE}^2$ are negatively related, so (15) gives a lower bound when $\sigma_{uSE}^2$ takes its lowest value. The lowest value $\sigma_{vSE}^2$ can take is 0 which implies, using (14) and (15), that the lower bound is:

$$\ln k_l = \mu_k + \frac{1}{2} \sigma_{vSE}^2 = \frac{\gamma_j}{\beta_j} - \frac{1}{2} \left( \sigma_{vSE}^2 - \sigma_{\xi SE}^2 \right)$$

Since the income of the self-employment have at least as much variance as the income from the employees the lowest values that $\sigma_{uSE}^2$ can take is $\sigma_{uEE}^2$. The upper bound of underreporting is therefore:

$$\ln k_u = \mu_k + \frac{1}{2} \sigma_{vSE}^2 = \frac{\gamma_j}{\beta_j} + \frac{1}{2} \left( \sigma_{vSE}^2 - \sigma_{\xi EE}^2 \right)$$

If $\rho \neq 0$, i.e., if the covariance between $u_{SE}$ and $v_{SE}$ is not zero it is not possible to calculate a range of underreporting without further information on $u$ and $v$. A zero covariance would imply
that whatever the income of the self-employed household turned out to be in a particular year, the household would tend to report the same percentage of it to the tax authorities. P & W however argue that a self-employed household would like to keep his or her income approximately constant over years as to not raise the suspicions of the tax authorities. Thus, if a self-employed person has had a particularly high income in one year he or she might be less inclined to declare all of it than if the income was more “normal” Similarly, if the income is particularly low he or she might be more inclined to declare it.

Such a behaviour implies a positive covariance between $u_{SE}$ and $v_{SE}$. If so, P & W argue that that the upper bound of the mean of under-reporting, $\bar{k}_u$, needs to be adjusted upwards. Thus for a positive $\rho$, (14) and (15) imply that

$$\ln \bar{k}'_u = \mu_k + \frac{1}{2} \sigma^2_{vSE} - \frac{1}{2} \sigma^2_{\xi SE} + \sigma^2_{\xi EE} + \text{cov}(uv)_{SE}$$

remembering that $\sigma^2_{uSE} = \sigma^2_{uEE}$ for the upper bound. In order to be able to calculate the covariance we need information of the partial correlation coefficient $\rho$ and the standard deviations $\sigma_{uSE}$ and $\sigma_{vSE}$. An estimate of the upper bound on $\sigma_{uSE}$ is available. Since $\sigma^2_{uEE}$ and therefore $\sigma^2_{vSE}$, cannot exceed the residual variance of the employees in employment, because the latter is the is the sum of the variance of $u$ and the residual variance of permanent income, it is possible to use $\sigma^2_{\xi EE}$ as an upper limit of $\sigma^2_{uSE}$. From (14) we obtain that:
Equation (19) can be solved for $\sigma_{vSE}$, which can be combined with the assumed upper bound for $\sigma_{uSE}$, i.e. $\sigma_{\xi SE}$, to get an estimate of the upper bound on $\text{cov}(\mu
u)_{SE}$ for a given $\rho$.

### 3. An estimate of income under-reporting for the self-employed in Finland

In this section we estimate the extent of self-employment income underreporting in Finland using the Pissarides & Weber (1989) approach.

The data set used in this study is called Kulutustutkimus and is a household expenditure survey compiled by Statistics Finland. This survey gives information on expenditures for different groups of goods and services for different types of households. Data are collected by a combination of personal interviews, diaries, and register information, the income data being taken from register information. The data in this study are from the years 1994-1996.

In this study households with two adults are studied. A household is considered to be self-employed if the socio-economic status of the head of the household is self-employment, which means that the head of the household has been self-employed for more than 6 months during the year. Following earlier research, I also assign a small part of households to the self-employed that have not reported self-employment as the main occupation but yet have a sizeable part, more than 25% of total income from self-employment. Also following earlier research, I exclude households where the head of the household is older than 64 years old and household whose
main occupation is farming. The reason why I do not include farmers in the study is that they are likely to have an expenditure function for food that is not the same as the one for other households, since they can be expected to produce large parts of their food themselves. Finally, I also exclude households where the head is not working all of the 12 months of the year. This because it is necessary to reduce the effect of hours worked on the results.

Table 1 gives some descriptive statistics for the sample considered in this study. There seems to be some interesting differences between the self-employed households and the households headed by an employee. Firstly, it looks like the (reported) disposable income is slightly higher for the employees, but the number for the log of the food expenditure looks higher for the self-employed. Secondly, the variance of the income of the self-employed is much higher than for the employees. Thirdly, although not a particular topic for this paper, it is interesting to notice that the self-employed seem to live in owner-occupied housing to a greater extent than do the employees.

To get an estimate of the underreporting of income by the self-employed we estimate equation (9) for food expenditure by running a regression of the form:

\[
\ln C_i = Z_i \alpha + \beta \ln Y_i + \gamma SE_i + \eta_i
\]

where \(SE_i\) is a dummy variable taking the value 1 if the household is headed by a self-employed and 0 if it is headed by an employee in employment.
The working assumption is this paper is that the expenditure function for food for the self-employed and the employees are the same. I tested for differences in the coefficients of the self-employed and the employees by introducing an interactive term where the self-employment dummy was multiplied with income. The t-value obtained was –0.139 and the null hypothesis was thereby not rejected. Similarly, we tested for non-linearity in the effect of income on expenditure by introducing the square of the log of income in the expenditure function. The t-value of that coefficient was -0.430, and therefore we reject the hypothesis of non-linearity in the effect of income on food expenditure.

The results from the food expenditure regression is presented in table 3. The coefficient of income is positive and significant as expected. The coefficient on the self-employment dummy is also positive and significant. Reported income $Y_i'$ was treated as endogenous and instrumented. This enables an independent estimate of the residual variance of reported income for each group, which can then be used in the calculation of income underreporting.

As elaborated in section 2, we need the marginal propensity to consume, $\beta_j$, the coefficient of the self-employment dummy, $\gamma_j$, the residual variance of reported income for the self-employed and the employees $\sigma_{\eta SE}^2$, and $\sigma_{\eta EE}^2$, and the residual variances of reported income $\sigma_{\zeta SE}^2$ and $\sigma_{\zeta EE}^2$, to estimate the extent of income underreporting. These estimates are given in table 3. We can also see that the residual variance from the income regression is, as expected, bigger for the self-employed than for the employees in employment, i.e. $\sigma_{\eta SE}^2 > \sigma_{\eta EE}^2$. 
According to (16) the estimate of the lower bound of income underreporting is

\[
\ln \bar{k}_l = \mu_k + \frac{1}{2} \sigma_{\text{vSE}}^2 = \frac{\gamma_j}{\beta_j} - \frac{1}{2} \left( \sigma_{\text{SE}}^2 - \sigma_{\text{EE}}^2 \right)
\]

Using the estimates in table 2 we can calculate the lower bound to be:

\[
\ln \bar{k}_l = 0.132
\]

and taking the antilog we arrive at:

\[
\bar{k}_l = 1.141
\]

Similarly we can compute an upper bound of the underreporting using (17) and assuming that the covariance between \( u_{SE} \) and \( v_E \) is zero. Then:

\[
\ln \bar{k}_u = \mu_k + \frac{1}{2} \sigma_{\text{vSE}}^2 = \frac{\gamma_j}{\beta_j} + \frac{1}{2} \left( \sigma_{\text{SE}}^2 - \sigma_{\text{EE}}^2 \right)
\]

This implies that

\[
\ln \bar{k}_u = 0.287
\]
and accordingly

\[ \tilde{k}_i = 1.332 \]

As argued in section 2, the upper bound of the estimate needs to be adjusted upwards if the correlation between \( u_{SE} \) and \( v_{E} \) is positive. For a positive \( \rho \) the upper bound can, according to (18) be written as:

(18)

\[
\ln \tilde{k}_u' = \mu_k + \frac{1}{2} \sigma_{vSE}^2 = \frac{1}{2} (\sigma_{\xi SE}^2 - \sigma_{\xi EE}^2) + \text{cov}(uv)_{SE}
\]

For this we also need an estimate of \( \text{cov}(uv)_{SE} \). This can be obtained by using Eq. (19):

(19)

\[
\sigma_{\xi SE}^2 - \sigma_{\xi EE}^2 = \sigma_{vSE}^2 - 2 \rho \sigma_{\xi SE} \sigma_{vSE}
\]

and the estimates in Table 2 to solve for \( \sigma_{vSE} \). We obtain that

\[ \sigma_{vSE} = 0.523 \]

for \( \rho = 1 \). Combined with the upper bound of \( \sigma_{uSE}^2 \), i.e \( \sigma_{\xi SE}^2 \), this implies an estimate of \( \text{cov}(uv)_{SE} \) for \( \rho = 1 \):

\[ \text{cov}(uv)_{SE} = 0.110 \]
This implies then that:

\[ \bar{F}_u = 1.488 \]

using a smaller correlation coefficient, for example \( \rho = 0.5 \), the upper bound is

\[ \bar{F}_u = 1.397 \]

Taken together, it seems like a reasonable midpoint for the underreporting estimate is somewhere between 25 and 30\%, maybe 27\%.

4. Concluding remarks

In this paper the extent of income underreporting by the self-employed is estimated by the expenditure based approach developed by Pissarides & Weber (1989). This approach relies on two major assumptions. Firstly, it is assumed that income underreporting for reasons of tax evasion occurs in the self-employed sector of the economy only. Secondly, it is assumed that both the self-employed and the employees in employment report their food consumption correctly. A food consumption regression was then estimated for individual households. This regression was then inverted and we arrived to the conclusion that self-employment income on average must be multiplied by 1.27 to arrive at the true income for the self-employed. Since income from self-employment is about 8\% of all income in Finland, the estimate of the black economy in Finland is about 2.3\% of GDP.
It is likely the this estimate is an underestimate of total tax-evasion in the economy. This because it is relatively unlikely that tax evasion is occurring in the self-employed sectors of the economy only. It is conceivable that e.g. farmers, students, and retired individuals receive income from work that do not show up in any official records. However, since it is likely that the food consumption functions for these individuals or households are different from households analysed in thes study, tax underreporting by those households will naturally not show up in a study like this. Hence, the measure of the underground economy obtained in an analysis like this should be taken as a lower bound of the true “black” sector of an economy.

There is clearly room for more research in this area. One aim for the future would be to disaggregate underreporting according to industries or time periods and relate these disaggregated measures to the opportunity to and incentive for tax evasion. Thus, do we see more tax evasion in a years with higher taxation? and do we see more tax evasion in industries or sectors where the opportunity to cheat is greater? Schuetze (2000) provides some evidence on these matters, but the literature is still very much in its infancy.
References


### Table 1: Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>self-employed</th>
<th>employees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log (disposable income)</td>
<td>12.020</td>
<td>12.063</td>
</tr>
<tr>
<td></td>
<td>(0.52)</td>
<td>(0.29)</td>
</tr>
<tr>
<td>Log (food expenditure)</td>
<td>10.075</td>
<td>10.003</td>
</tr>
<tr>
<td></td>
<td>(0.47)</td>
<td>(0.45)</td>
</tr>
<tr>
<td>Own house</td>
<td>0.883</td>
<td>0.773</td>
</tr>
<tr>
<td>Number of children</td>
<td>1.197</td>
<td>1.161</td>
</tr>
<tr>
<td></td>
<td>(1.27)</td>
<td>(1.18)</td>
</tr>
<tr>
<td>Age</td>
<td>42.676</td>
<td>40.792</td>
</tr>
<tr>
<td></td>
<td>(9.66)</td>
<td>(9.35)</td>
</tr>
<tr>
<td>N</td>
<td>299</td>
<td>1755</td>
</tr>
</tbody>
</table>

Note: Standard deviations in parentheses

### Table 2: Estimates needed for calculation of income under-reporting

<table>
<thead>
<tr>
<th>$\beta_j$</th>
<th>$\gamma_j$</th>
<th>$\sigma_{\eta SE}^2$</th>
<th>$\sigma_{\eta EE}^2$</th>
<th>$\sigma_{\zeta SE}^2$</th>
<th>$\sigma_{\zeta EE}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.315</td>
<td>0.066</td>
<td>0.153</td>
<td>0.140</td>
<td>0.203</td>
<td>0.044</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses are t-values
Table 3: Results from 2SLS regression. Dependent variable: log of food expenditure

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t-value</th>
<th>SE</th>
<th>Coefficient</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE</td>
<td>0.066</td>
<td>(2.75)</td>
<td>Q294</td>
<td>0.037</td>
<td>(0.96)</td>
</tr>
<tr>
<td>LNINC</td>
<td>0.315</td>
<td>(4.95)</td>
<td>Q394</td>
<td>0.141</td>
<td>(3.69)</td>
</tr>
<tr>
<td>AGE</td>
<td>0.017</td>
<td>(0.34)</td>
<td>Q494</td>
<td>-0.039</td>
<td>(-1.01)</td>
</tr>
<tr>
<td>AGESQ*10</td>
<td>-0.001</td>
<td>(-0.20)</td>
<td>Q195</td>
<td>-0.074</td>
<td>(-2.05)</td>
</tr>
<tr>
<td>CH</td>
<td>0.232</td>
<td>(6.58)</td>
<td>Q295</td>
<td>0.007</td>
<td>(0.20)</td>
</tr>
<tr>
<td>CHSQ</td>
<td>-0.010</td>
<td>(-2.82)</td>
<td>Q395</td>
<td>-0.012</td>
<td>(-0.33)</td>
</tr>
<tr>
<td>CH07</td>
<td>-0.054</td>
<td>(-1.84)</td>
<td>Q495</td>
<td>-0.038</td>
<td>(-0.99)</td>
</tr>
<tr>
<td>OHO</td>
<td>-0.013</td>
<td>(-0.46)</td>
<td>Q196</td>
<td>-0.080</td>
<td>(-2.22)</td>
</tr>
<tr>
<td>REG2</td>
<td>0.008</td>
<td>(-0.36)</td>
<td>Q296</td>
<td>-0.025</td>
<td>(-0.68)</td>
</tr>
<tr>
<td>REG3</td>
<td>0.0001</td>
<td>(0.03)</td>
<td>Q396</td>
<td>-0.058</td>
<td>(-1.73)</td>
</tr>
<tr>
<td>REG4</td>
<td>-0.0001</td>
<td>(-0.005)</td>
<td>Q496</td>
<td>-0.012</td>
<td>(-1.07)</td>
</tr>
<tr>
<td>REG5</td>
<td>0.068</td>
<td>(1.74)</td>
<td>Constant</td>
<td>5.491</td>
<td>(1.134)</td>
</tr>
<tr>
<td>REG6</td>
<td>0.070</td>
<td>(0.96)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N 2053
Adjusted R^2 0.311
χ^2 24.30

Notes: Numbers in parentheses are t-values. Log of disposable income was treated as endogenous and instrumented.
Additional instruments: ROOM, WFT, WHT, WSE, WUE, and the product of SE with AGE, AGESQ, CH, CHSQ, CH07, OHO, REG2, REG3, REG4, REG5, REG6, ROOM, WFT, WHT, WSE, WUE. χ^2 is a test for overidentifying restrictions.
Appendix

Variables used in the regression

SE  dummy for self-employed household
LNFOOD logarithm of the household’s expenditure on food
LNINC logarithm of the household’s disposable income
AGE  age of the head of the household
AGESQ  age of the head of the household squared
CH  number of children in the household
CHSQ  number of children in the household squared
CH07  number of children between 0 and 7 in the household
OHO  dummy for owner occupied house
REG1  regional dummy for Etelä-Suomen Lääni
REG2  regional dummy for Länsi-Suomen Lääni
REG3  regional dummy for Itä-Suomen Lääni
REG4  regional dummy for Oulun Lääni
REG5  regional dummy for Lapin Lääni
REG6  regional dummy for Ahvenanmaa
Q194-Q496  dummy for different time periods of information collection on food consumption where Q194 is the first quarter of 1994 and Q496 is the fourth quarter of 1996
ROOM  number of rooms in the house
WFE  number of months spouse has been working during year
WSE  number of months spouse has been self-employed during the year
WUE  number of months spouse has been unemployed during the year
WUE  number of weeks spouse has been working part-time during the year

Notes:

1 Source: The Income Distribution Statistics of Statistics Finland, various issues, and author’s own calculations.

2 This feature of the administratively collected self-employment income has earlier been documented in Johansson (2000).