Nordic Journal of Political Economy

Volume 31 2005 Pages 99-109

An estimate of self-employment income underreporting in Finland

Edvard Johansson

Edvard Johansson*

An estimate of self-employment income underreporting in Finland

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. Depending on how a self-employed household is defined, the results suggest that self-employment income in Finland is underreported by some 16-40%. 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 1.3-3.2% of GDP.
JEL Classifications: H26, J23, D12

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. Interestingly, in a growing body on the determinants of self-employment, some studies, (see Blau, 1987, and Parker 1996 for time-series evidence, and Bruce 2000, and Schuetze 2000, for micro-econometric evidence), argue that individuals enter self-employment to avoid tax. 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 by a large extent, then we should be more concerned about the results obtained in the above mentioned studies.

It is also the case that in some countries, including Finland, the share of self-employed of the labour force has increased (see Blanchflower 2004, for an overview of recent trends of self-employment in an international perspective). In Finland, the share of the total non-agricultural employment that is self-employed has risen from under 7% in the mid 1980s to over 9% in the early years of the 21st century. This fact makes it more interesting to study income underreporting by the self-employed.

However, measuring income underreport-

* The Research Institute of the Finnish Economy, Lönnrotinkatu 4 B, 00120 Helsinki, Finland. Phone: +358-9-6099 0269. Fax: +358-9-601753. E-mail: edvard.johansson@etla.fi. Helpful comments by Henry Ohlsson and two anonymous referees have considerably improved the quality of the paper. Financial support from the Nordisk skattevitenskaplig forskningsråd is gratefully acknowledged.

ing 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, new approach based on food expenditure is used. P&W, using UK data estimate that self-employment income is underreported by some 55%. 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. The "marginal propensity to consume food" is assumed to be the same both for the self-employed and the employees. It is then possible to calculate how much income was needed for the self-employed to consume the amount of food actually consumed.

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 depending on which households are defined to be self-employed, self-employment income on average is underreported by some 16-40%. Since the share of total income in Finland that comes from self-employment is about 8% of GDP[1], this methods yield the result that the "black" economy has a size that is about 1.3 – 3.2% of GDP.

Previous research

Apart from the original P & W study which used UK data from the 1982 Family Expenditure Survey, there have been a limited number of studies applying the same methodology. Baker (1993), and Cullinan (1997) also have investigated income underreporting by the self-employed for the UK. Baker (1993) also used data from the Family Expenditure Survey, but for all the years from 1978 to 1991. The result in this study was that actual self-employment income was about 1.3 to 1.5 time larger than reported income. Cullinan (1997) again used the UK Family Expenditure Survey, but this time for the years 1987 and 1992. The result was that self-employment income is underreported by some 19-37%. Outside the UK, the issue has been investigated for Sweden by Apel (1994), and for Canada by Mirus and Smith (1996), and Schuetze (2002). Apel (1994) used data from the 1988 "Hushållens utgifter" data set, the Swedish equivalent to the UK Family Expenditure Survey, and found that reported self-employment income should be multiplied by a factor of 1.35 in order to arrive at the true income. Mirus and Smith (1996) used data from the 1990 Canadian Survey of Family Expenditures. Their result was that self-employment income was underreported by some 12.5%. Schuetze (2002) used the same data as Mirus and Smith (1996) but used six years of data instead of only one in an attempt to disaggregate underreporting by year and occupation. In summary, this study found that the self-employed have underreported earnings by some 11-23%.

This method is obviously not the only way to calculate the size of the black economy, and this paper should be seen as part of a much larger literature on the measurement of the "shadow" or "underground" economy. Indeed, Schneider and Enste (2000) divide the ways to calculate the size of the black economy into three broad categories, "direct"

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

approaches, "indirect" approaches, and "model-based" approaches. The "direct" approach makes use of direct evidence in the form of surveys or tax auditing evidence in order to estimate the size of the underground economy. The general idea of the "indirect" methods is that the researcher uses various economic or other indicators that indirectly may be used to extract information of the size of the underground economy. "Model-based" methods explicitly consider the multiple causes and consequences of the black economy by using dynamic multiple-indicators multiple causes (MIMIC) models (Scheider and Enste 2000:97-98).

Earlier research has produced some results regarding the size of the black economy for the Nordic countries using other methods than the one applied in this paper. Regarding "direct" approaches representative studies for Norway and Denmark are Goldstein et al. (2002) for Norway, and Mogensen et al. (1995) for Denmark. These studies, which are based on survey responses, indicate that the size of the black economy is some 2-5% for in Denmark, and 1-2% in Norway.

Schneider and Enste (2000) summarise most of the international research done using "indirect" methods. Lackó (1996, 1997, 1998, 1999) calculate the size of the black economy for the Nordic countries for the year 1990. In these studies, Finland's black economy is estimated at being 13.3% of GDP, Denmark's 16.9%, Sweden's 11.0%, and Norway's 9.3%. Schneider and Enste (2000, p. 104) also provide averages for 1994-1995 and 1996-1997 for the Scandinavian countries for the underground economy. Interestingly, they report that the size of the underground economy has increased in all three Scandinavian countries. In Denmark from 17.8% to 18.2%, in Norway from 18.2% to 19.4%, and in Sweden from 18.6% to 19.5%.

In summary, it is fair to say that the frequently used, more macroeconomically oriented indirect ways to measure the size of the underground economy has yielded estimates of the size of the underground economy that is far larger that the ones obtained by the method used in this paper. The most obvious explanation for this is that the present method assumes that underreporting of income takes place in the self-employed sector of the economy only. This assumption will bias the estimates of the underground economy down, if underreporting of income also to some extent occurs among employees. And as Goldstein et al. (2002) shows for Norway, this is probably occurring at least to some extent.

A review of the P&W (1989) model

The expenditure-based estimation approach, originally developed by P & W, 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_i is the food consumption of household *i* on after tax income Y_i'. Z_i is a vector of household characteristics. It is assumed that C_i 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:

$$(1) \qquad Y_i = k_i' Y_i', \ k_i \geq 1.$$

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 underreporting by household *i.*

For food expenditure there is an expenditure function,

(2) $$1n C_i = Z_i \alpha + \beta\, 1n Y_i^p + \varepsilon_i$$

where α is a parameter, β is a scalar, the "marginal propensity to consume" food and ε_i is white noise, and Z_i is a set of household characteristics. 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'. P & W 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 underreporting by the self-employed will have to be adjusted accordingly. In general it is assumed that permanent and measured income are related by

(3) $$Y_i = p_i Y_i^p$$

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

(4) $$1n Y_i^p = 1n Y_i' - 1n p_i + 1n k_i$$

Substituting this into 2 yields the following equation:

(5) $$1n C_i = Z_i \alpha + \beta\, 1n Y_i' - \beta\, 1n p + \beta\, 1n k_i + \varepsilon_i$$

It is now possible to run the regression:

(6) $$1n C_i = Z_i \alpha + \beta\, 1n Y_i' + \gamma\, SE_i + \eta_i$$

where SE_i is a dummy taking the value 1 if the household is self-employed and 0 if not. Remembering the above mentioned assumptions about p and k, i.e. that $1n\ p$ is the same for both groups, $k = 1$ for the employed whereas $k \geq 1$ for the self-employed, we can see that a rough estimate of income underreporting may be obtained by

(7) $$1n k = \gamma / \beta$$

However, when treating p_i and k_i as random variables, things become more complicated, and to make estimation possible P & W assume that they are log normal and write them as deviations from their means, $1n p_i = \mu_p + u_i$, and $1n k_i = \mu_k + v_i$, where u_i and v_i have zero means and constant variances σ_u^2 and σ_v^2 within each occupational group. Taking this into account, and substituting into (2) from (4) we get the following expression:

(8) $$1n C_i = Z_i \alpha + \beta\, 1n Y_i' - \beta(\mu_p - \mu_k) - \beta\,(u_i - v_i) + \varepsilon_i$$

Compared to equation (5) it is clear that not only the intercept, but also the variance of the errors will differ between the self-employed and the employees, with the self-employed generally having bigger variance.

What one would like to estimate is average the average of income underreporting (denoted by $\bar{k}$). Assuming log-normality, it can be shown that $\bar{k}$ is given by:

$$(9) \qquad \ln \bar{k} = \mu_k + \frac{1}{2}\sigma^2_{vSE}$$

where σ^2_{vSE} denotes the variance of k_i, and the subscript *SE* denotes the group of self-employed. Further, given that p_i and k_i are treated as random, it can the be shown that in (6) $\gamma = \beta\rfloor \mu_k + 1/2\ (\sigma^2_{vSE} - \sigma^2_{uEE})\rfloor$, where σ^2_u is the variance of p_i and the subscripts *SE* and *EE* refer to the self-employed and employed respectively. However, from this estimate it is not possible to isolate μ_k, because log p_i will differ between the self-employed and the employees. Then, if γ from equation is substituted into (8) we get the following expression for $\bar{k}$:

$$(10) \qquad \ln \bar{k} = \frac{\gamma}{\beta} + \frac{1}{2}(\sigma^2_{vSE} - \sigma^2_{uSE} + \sigma^2_{uEE})$$

However, no data on p_i and k_i of course exist, so underreporting cannot directly be estimated using (9). Instead, following P & W, an income equation is estimated in order to obtain estimates for the income variance of errors for the self-employed $(\sigma^2_{\zeta SE})$ and $(\sigma^2_{\zeta EE})$ for the employees separately. Under a set of assumptions (see the appendix for further explanations) it is then possible to calculate an lower bound of underreporting by the self-employed as

$$(11) \qquad \ln \bar{k}_l = \frac{\gamma}{\beta} - \frac{1}{2}(\sigma^2_{\zeta SE} - \sigma^2_{\zeta EE})$$

and an upper bound as

$$(12) \qquad \ln \bar{k}_u = \frac{\gamma}{\beta} + \frac{1}{2}(\sigma^2_{\zeta SE} - \sigma^2_{\zeta EE}).$$

Self-employment income underreporting in Finland

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. 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.

To start with, we consider a household 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. This is similar to other research. However, the definition of which household is to be considered self-employed is not entirely straightforward, and it is not unlikely that the way a self-employed household is defined also may affect the estimates of underreporting. For example, it may be the case that a household with both spouses self-employed have greater opportunities to underreport income than a household

Table 1: Descriptive Statistics

	Head of household self-employed		Both spouses self-employed	
	self-employed	employees	self-employed	employees
Log (food expenditure)	10.07	10.00	10.14	10.00
	(0.48)	(0.45)	(0.38)	(0.46)
Log (disposable income)	12.03	12.06	12.03	12.06
	(0.48)	(0.30)	(0.41)	(0.32)
Own house	0.89	0.77	0.89	0.78
	(0.32)	(0.42)	(0.32)	(0.41)
Number of children	1.22	1.16	1.21	1.16
	(1.24)	(1.17)	(1.10)	(1.19)
Age	42.82	40.79	43.11	40.91
	(9.37)	(9.33)	(7.94)	(9.44)
N	285	1769	148	1906

Note: Standard deviations in parentheses

with only the head of the household being self-employed and the other spouse doing something else. Because of this, we also consider an alternative definition, in which we only consider household where both spouses have worked at least 6 months each as self-employed during the year as a self-employed household.

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 employee households, for both definitions of a self-employed household. 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[2]. 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.

Then for the actual results. Before going to any regression results, a first crude approximation of the magnitude of underreporting by the self-employed can be obtained by considering the differences between the means of income and consumption for the employees and the self-employed, and assuming that the proportion of income that is spent on food is the same for both groups. From table 1 we find that this difference is 0.10 log points, i.e. about 10.5%, for households headed by a self-employed, and for the definition where both spouses are self-employed, the difference is some 0.17 log points, or around 18.5%. Thus, this crude measure indicates that if both spouses are self-employed, more underreporting occurs. The straightforward explanation for this is that households where both spouses are self-employed have better opportunities to underreport income as a larger part of the income comes from self-employment.

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

Then for regression-based results. In a first step, it is useful to examine the degree of underreporting by the self-employed treating p_i and k_i (see equation 4) as constants. This implies a straightforward estimation of (6), which we perform for both definitions of a self-employed household (see columns 1 and 3 of table 2)[3]. Indeed, the coefficient for self-employment is positive, as expected. In terms of magnitudes (i.e. the antilog of γ / β), the results imply a degree of underreporting of 24% for the regression where the head of the household is self-employed, and some 41% for the regression where both spouses are self-employed. The difference between these estimates and the rough estimates of about 10.5% and 18.5% from above can be interpreted as the effect of controlling for other exogenous factors affecting household food consumption, such as household size, number of children etc.

In this household food consumption equation, it is not unlikely that income is endogenous. This may be due to the fact that food consumption affects labour supply, and thereby household income. Consequently, we tested whether the log of household income was endogenous, given our chosen instruments, for both definitions of a self-employed household. According to a Hausman-type test, the hypothesis that the log of income was exogenous in the household food consumption function was rejected. As instruments we utilised information on spouse's labour supply, which affects household income, but not necessarily household food consumption (see table 2 for further details).

As we rejected the hypothesis that income is exogenous in the household food consumption function, we went on and re-estimated (6) using two-stages least squares estimation. The results of those estimations are shown in columns 2 and 4 of table 2, for the two definitions of a self-employed household. Although the parameter estimates have changed somewhat, they do not qualitatively alter the results. In terms of magnitudes of underreporting of self-employment income, the 2SLS results imply a degree of underreporting of 16.5% and 42.0% for the two definitions of a self-employed household. The difference between these estimates the estimates from the OLS estimations can be interpreted as simultaneity bias. It should also be emphasized, that the regression-based results do not alter the notion that if both spouses are self-employed, more underreporting occurs.

As discussed in the theoretical section, it is perhaps more appropriate to treat p_i and k_i as random variables. It is then not possible to estimate an average for income underreporting, but it is possible to compute upper and lower bounds for underreporting. In order to do that, we need the marginal propensity to consume, β, the coefficient of the self-employment dummy, γ, the residual variance of reported income for the self-employed and the employees, $\sigma^2_{\zeta SE}$ and $\sigma^2_{\zeta EE}$, and the variance of η_i from the expenditure function for each group, $\sigma^2_{\eta SE}$ and $\sigma^2_{\eta EE}$. These estimates

3. The working assumption in this paper is that the expenditure function for food for the self-employed and the employees are the same. We 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 –0.136 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 -0.432, and therefore we reject the hypothesis of non-linearity in the effect of income on food expenditure.

Table 2: Regression results for household food expenditure

	Head of household self-employed		Both spouses self-employed	
	(OLS)	(2SLS)	(OLS)	(2SLS)
Self-employed	0.041	0.047	0.105	0.114
	(1.71)	(1.93)	(3.26)**	(3.52)**
Log of household income	0.191	0.307	0.194	0.325
	(6.98)**	(5.57)**	(7.08)**	(5.69)**
Age	-0.001	-0.008	-0.003	-0.011
	(0.12)	(0.91)	(0.35)	(1.24)
Age squared	0.000	0.000	0.000	0.000
	(1.02)	(1.69)	(1.25)	(2.01)*
Number of children	0.258	0.251	0.258	0.249
	(17.00)**	(16.11)**	(17.01)**	(16.00)**
Children squared	-0.012	-0.012	-0.012	-0.011
	(4.10)**	(4.01)**	(4.05)**	(3.94)**
Children younger than 7	-0.067	-0.069	-0.066	-0.067
	(4.47)**	(4.58)**	(4.40)**	(4.51)**
Constant	7.345	6.115	7.354	5.965
	(21.85)**	(10.03)**	(21.95)**	(9.50)**
Observations	2053	2053	2053	2053
R-squared	0.33	0.33	0.33	0.33
F-test for excluded instruments, p-value		0.000		0.000
Wu-Hausman-test for exogeneity of log of household income, p-value		0.016		0.009
Sargan-test for over-identification, p-value		0.265		0.806

T-statistics in parentheses. * significant at 5%; ** significant at 1%. Regressions also include a constant, and 5 controls for regions, and 13 controls for time of data collection. Additional instruments in the first stage of the 2SLS regressions: House ownership, number of rooms in dwelling, Spouse's months of work, Spouse's months of self-employment, Spouse's months of unemployment, Spouse's months of part-time work, and the product of the self-employment dummies with Age, Age squared, Children, Children squared, Children younger than 7, Spouse's months of work Own house, and the regional dummies.

Table 3:
Estimates needed for calculation of upper and lower bounds of income under-reporting

Sample	β	γ	$\sigma^2_{\eta SE}$	$\sigma^2_{\eta EE}$	$\sigma^2_{\zeta SE}$	$\sigma^2_{\zeta EE}$
Head of household self-employed	0.307	0.047	0.158	0.139	0.170	0.053
	(5.57)	(1.93)				
Both spouses self-employed	0.325	0.114	0.143	0.116	0.131	0.064
	(5.69)	(3.52)				

Note: Numbers in parentheses are t-values are given in table 3. Using (11) and (12) we can then estimate the lower $\bar{k}_l$ and upper $\bar{k}_u$ bounds of income underreporting. For households where the head is self-employed, the upper and lower bounds can then be calculated to be 23.5% and 10.0%. For households where both spouses are self-employed, the upper and lower bounds can be calculated to be 46.9% and 37.3%.

Concluding remarks

In this paper the extent of income underreporting by self-employed households was estimated by the expenditure based approach developed by P & W (1989). Two different definitions of a self-employed household were used, one where only the head of the household was required to be self-employed, and another where both spouses were required to be self-employed. For the first definition we found self-employment income to be underreported by some 16.5% on average, and for the second definition by some 42% on average. This supports the hypothesis that the larger a household's share of income that comes from self-employment, the larger is the share of income that is concealed. Using these estimates, and assuming that underreporting of income occurs in the self-employment sector of the economy only, it is possible to gauge overall underreporting of income in the Finnish economy. Since income from self-employment is about 8% of all income in Finland, the estimate of the black economy in Finland is about 1.3% of GDP, using the first definition of a self-employed household, and about 3.2% using the second.

It is likely that both estimates are underestimates 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. One can easily think that farmers, students, and retired individuals receive income from work that do not show up in any official records. However, since the purpose of tax evasion is to conceal income from the authorities, it is of course very difficult to make reliable estimates because of the lack of data. This expenditure-based approach can, therefore, be seen as an interesting attempt to measure something that otherwise would be very hard to get an estimate of.

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 (2002) provides some evidence on these matters, but the literature is still very much in its infancy.

References

Apel, M. 1994. *Essays on Taxation and Economic Behaviour*, Ph.D. thesis, Economic Studies 18, Uppsala University.

Baker, P. 1993. Taxpayer compliance of the self-employed: estimates from household spending data, IFS Working Paper, No W93/14.

Blanchflower, D. 2004. "Self-employment: more may not be better", *Swedish Economic Policy Review,* 11:15-73.

Blau, D. 1987. "A time-series analysis of self-employment in the United States", *Journal of Political Economy*, 95:445-468.

Bruce, D. 2000. "Effects of the United States tax system on transitions into self-employment", *Labour Economics*, 95:545-574.

Cullinan, J. 1997. "Self-employment income underreporting and estimates of the black economy: Britain –1987 & 1992", Working Paper. National University of Ireland, Maynooth.

Goldstein H., Hansen W., Ognedal T. and Ström S. 2002. Svart arbeid fra 1980 til 2001.Oslo: Ragnar Frisch Centre for Economic Research, Rapport

3/2002.

Income distribution statistics 1994, Edita Oy, Helsinki, Finland, 1996.

Income distribution statistics 1995, Edita Oy, Helsinki, Finland, 1997.

Income distribution statistics 1996, Edita Oy, Helsinki, Finland, 1998.

Isachsen, A. and Ström, S. 1980. "The hidden economy, the labour market and tax evasion" *Scandinavian Journal of Economics* 82:304-311.

Isachsen, A. 1985. "The size and growth of the hidden economy in Norway" *Review of Income and Wealth* 31:21-38.

Johansson, E. 2000. "Self-employment and the predicted earnings differential – evidence from Finland", *Finnish Economic Papers* 13:45-55.

Lackó, M. 1996. Hidden economy in East European countries in international comparison. Working Paper. International Institute for Applied Systems Analysis, Laxenburg, Austria.

Lackó, M. 1997. Does power consumption data tell the story? Electricity intensity and the hidden economy in post-socialist countries. Working Paper. International Institute for Applied Systems Analysis, Laxenburg, Austria.

Lackó, M. 1998. The hidden economies of Visegrad countries in international comparison: a household electricity approach. Working Paper. International Institute for Applied Systems Analysis, Laxenburg, Austria.

Lackó, M. 1999. Hidden economy an unknown quantity? Comparative analyses of hidden economies in transition countries in 1989-1995. Working Paper 9905. Economics Department, University of Linz, Austria.

Mirus, R. and Smith R. .1996. Self-employment, tax evasion, and the underground economy: micro based estimates for Canada. Working Paper 1002. International tax program, Harvard Law School.

Mogensen, G., Kvist, H., Körmendi, E., and Pedersen S. 1995. The shadow economy in Denmark 1994: measurement and results. Study no. 3, Rockwool Foundation Research Unit, Copenhagen.

Parker, S. 1996."A time series model of self-employment under uncertainty", *Economica*, 65:459-75.

Pissarides, C and Weber, G. 1989. "An expenditure based estimate of Britain's black economy", *Journal of Public Economics*, 39:17-32.

Schneider, F. 1998. Further empirical results on the size of the shadow economy of 17 OECD countries over time. Discussion Paper. University of Linz, Austria.

Schneider, F. and Enste, D. 2000. "Shadow economies: size, causes, and consequences" *Journal of Economic Literature,* 38:77-114.

Schuetze, H. 2002. Profiles of tax noncompliance among the self-employed in Canada: 1969-1992, *Canadian Public Policy*, 28:219-237.

Schuetze, H. 2000. "Taxes, economic conditions and recent trends in male self-employment: a Canada-U.S. comparison", *Labour Economics*, 7:507-544.

Appendix: Derivation of upper and lower bounds of income underreporting

The income equation needed to calculate income variance of errors for the self-employed ($\sigma^2_{\zeta SE}$) and for the employees ($\sigma^2_{\zeta EE}$) can be written as follows:

(A1) $\ln Y_i' = Z_i\delta_1 + X_i\delta_2 + \zeta_i$

where is a set of identifying instruments. Suppose then that the unexplained variations in permanent income in (A1) 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 (A1). Remembering that $\sigma^2_{vEE} = 0$ we have that:

(A2) $\mathrm{var}\zeta_{SE} - \mathrm{var}\zeta_{EE} = \mathrm{var}(u-v)_{SE} - \mathrm{var}u_{EE}$

and if expanding the right hand side:

(A3) $\sigma^2_{\zeta SE} - \sigma^2_{\zeta EE} = \sigma^2_{uSE} + \sigma^2_{vSE} - 2\mathrm{cov}(uv)_{SE} - \sigma^2_{uEE} = \sigma^2_{uSE} + \sigma^2_{vSE} - 2\rho\sigma_{uSE}\sigma_{vSE} - \sigma^2_{uEE}$

where ρ is the partial correlation coefficient between u_{SE} and v_{SE}. If $\rho=0$ then σ^2_{vSE} and σ^2_{uSE} are negatively related, so (A3) gives a lower bound when σ^2_{uSE} takes its lowest value. The lowest value σ^2_{vSE} can take is 0 which implies, using (A3) and (9), that the lower bound is:

(11) $\ln \bar{k}_l = \mu_k + \frac{1}{2}\sigma^2_{vSE} = \frac{\gamma}{\beta} - \frac{1}{2}(\sigma^2_{\zeta SE} - \sigma^2_{\zeta EE})$

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

(12) $\ln \bar{k}_l = \mu_k + \frac{1}{2}\sigma^2_{vSE} = \frac{\gamma}{\beta} + \frac{1}{2}(\sigma^2_{\zeta SE} - \sigma^2_{\zeta EE})$

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 gives 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. However, as P&W show by example, even relatively values of , the upper bound does not rise very much.