ERASMUS UNIVERSITY ROTTERDAM

Do Entrepreneurs Under-report Their Income?

An Expenditure-based Estimate of Self-employed Income Concealment in Australia

Martyna Kapociute

357567

2013

Abstract

Entrepreneurs have arguably greater scope compared to wage earners to under-report their incomes. Current paper uses data drawn from the Household, Income and Labour Dynamics in Australia Survey (HILDA[1]) for the years 2001 and 2003-2005 to investigate the degree of self-employment income under-reporting. A method applied is an expenditure based approach developed by Pissarides and Weber (1989). A key idea is that under-reporting of incomes among entrepreneurs would be detectable in the data by using food expenditure equations. The results reveal that households headed by self-employed individuals under-report their incomes by approximately 14-89 per cent. Under-reporting appears to be much more prevailing between entrepreneurs involved in unincorporated business than those involved in incorporated business.

[1] Information about HILDA survey can be found at the official website: http://www.melbourneinstitute.com/hilda/.

Table of Contents

1. Introduction

The last decade entrepreneurship is on rise. Consequently academic researchers as well as society in general show an increased interest in self-employment. Interestingly, while examining the determinants of self-employment, some studies (Blau, 1987; Parker, 1996; Bruce, 2000; Schuetze, 2000) argue that individuals choose entrepreneurship as their career path in order to avoid tax. Hence, in this context it is reasonable to think about the magnitude of self-employed income under-reporting. To measure latter, income and expenditure differences between entrepreneurs and wage workers are explored.

A number of researchers have tried to measure degree of income under-reporting. According to Pissarides and Weber (1989), on average true self-employment incomes are 1,55 times higher than reported incomes. Moreover, there is some evidence that blue-collar workers conceal a little more than white-collar workers – 60 per cent as compared to 50 per cent or in some cases even less. Likewise other scholars (Apel, 1994; Johansson, 2000; Schuetze, 2002; Lyssiotou et al, 2004; Besim and Jenkins, 2005; Engstrom and Holmlund, 2007, etc.) provide similar conclusion – entrepreneurs tend to under-report their incomes relative to wage workers, only there is a dispersion in the magnitude.

However, direct and reliable evidence about the extent of entrepreneurial income under-reporting is difficult to come by. A challenge researchers face, while trying to measure under-reporting accurately, is to find an appropriate data because self-employed individuals are motivated to evade their true income to be recorded in publicly available data. Furthermore, scholars have to adjust an efficient method, since individuals who conceal their incomes cannot simply be determined as those who spend more than they earn, because some of them might dissave and not under-report, while others might under-report and not dissave (Parker, 2009).

The research question that this paper sets to investigate is "Do entrepreneurs under-report their true incomes compared to wage workers in Australia, using Household, Income and Labour Dynamics Survey data from 2001 and 2003-2005?". Answer to this question may give a deeper insight into self-employed earnings concealment problem. Moreover, while entrepreneurial income under-reporting is one of the most common and serious limitation of public-use data and correction factor must be applied when measuring earnings differentials between entrepreneurs and non-entrepreneurs, the obtained results might eliminate latter data limitation.

A method used to answer latter research question was inspired by Pissarides and Weber (1989). Therefore, this paper contributes by replication of Pissarides and Weber (1989) methodology and findings, using a different country sample. In particular, self-employment income

under-reporting is estimated by controlling for food expenditures. Additionally, an analysis of income under-reporting by the legal form of self-employment is carried out.

Taking data from the Household, Income and Labour Dynamics in Australia (HILDA) survey (2001, 2003-2005), two-stages least squares (2SLS) analysis is conducted to investigate entrepreneurial income under-reporting.

The results of current study are in line with the ones presented in Pissarides and Weber (1989) paper – some evidence is found that self-employed under-report their incomes compared to paid employees. Also some evidence is disclosed to support that entrepreneurs with incorporated business conceal less of their true income than entrepreneurs involved in unincorporated business.

The paper is structured as follows: first of all, current knowledge related to the topic is presented, secondly, the data set and method are described enriched by descriptive statistics, thirdly, an empirical analysis is conducted and obtained results are presented. Lastly, the limitations of the analysis are discussed and the conclusions derived.

2. Literature Review

2.1 Tax Evasion and Income Under-reporting

It is widely believed that entrepreneurs have arguably greater opportunities compared to wage workers to under-report their true taxable incomes. First of all, it should be recognised that, working for themselves, self-employed individuals have a greater scope for tax evasion because they are responsible for reporting their incomes to the tax authorities. While employees are less prone to evade tax because of the third-party involved in reporting income (Schuetze, 2002; Engstrom and Holmlund, 2007; Parker, 2009). Consequently, individuals engaged in self-employment may have no incentive to reveal their incomes to the data collectors accurately despite assurances that survey information is totaly confidencial and is never disclosed to the tax authorities (Pissarides and Weber, 1989; Parker, 2009). Hence, measuring income under-reporting is difficult, since the whole motivation behind it is to prevent the real income to be recorded in publicly available data.

However, it is important to distinguish between under-reporting of incomes to the tax authorities and under-reporting to the survey interviewers. By taking into account the difference between former and latter, current literature can be divided into two groups – studies that estimate under-reporting directly based on audits and tax returns and studies which estimate under-reporting indirectly by using household survey data (Feldman and Slemrod, 2007; Lyssiotou et al., 2004).

The most direct and reliable evidence of the degree of income under-reporting to the tax authorities can be derived from data gathered by the Internal Revenue Service's (IRS's) Taxpayer Compliance Measurement Program (TCPM)[2]. Numerous scholars have used this data to capture the extent of tax evasion. According to Feldman and Slemrod (2007), 99,5 per cent of employees' wages and salaries are reported accurately to the IRS, while non-farm proprietors report on average 50-70 per cent of their true incomes. The highest rate of under-reporting belongs to partnerships and "S corporations"[3] – only a little bit more than 40 per cent of true income is reported to IRS. Moreover, there is a noticeable variation between under-reporting rates among industries – self-employed taxi cab drivers are the most likely to under-report their true incomes to the tax authorities (Kesselman, 1989).

Regarding the relation between marginal tax rates and entrepreneurs' tendency to conceal income, Joulfaian and Rider (1998), using TCPM data, find that 10 per cent increase in the marginal income tax increases under-reporting of self-employed by 5 per cent. Besides, latter researchers find a significant negative relation between tax evasion and tax audit rate with an elasticity of approximately 0,7.

It is worth noting that the cost of tax evasion is significant. According to the 1988 TCPM data, sole proprietors are the cause of 30 per cent of concealed tax revenue, however, they comprise only 5,5 per cent of the tax returns filed (Feldman and Slemrod, 2007). More recent data from 2001 TCPM expose that IRS experience 68 billion dollars costs per year because of the non-farm proprietors income concealling. This seeks more than one third of total estimated income under-reporting to the tax authorities for the individual income tax (Slemrod, 2007).

Nonetheless, while using audit-based data sets like the TCPM to estimate the range of self-employed income under-reporting, researchers face few basic problems – first, income sources, such as cash transanctions, are difficult to detect even by the most intensive audit. Secondly, TCPM-type data are not available in most countries (Feldman and Slemrod, 2007). In this respect, other methods of measuring income under-reporting is worthwhile to use.

The overwhelming majority of researchers prefer to use household income and expenditure data rather than audit-based data for income under-reporting estimation. An expenditure-based approach which does not depend on special tax audit programmes was developed by Pissarides and

[2] The TCMP is a program periodically conducted by the Examination Branch of the IRS to estimate compliance with tax laws and revenue lost from non-compliance. Random in-depth audits with intensive probes for under-reported income were performed approximately every three years from 1963 until 1988.

[3] "S corporations" are not obliged to pay any federal income taxes. Instead, the corporation's income, losses, deductions and credits are divided between and passed through to its shareholders. Consequently, the shareholders report the income or loss on their personal income tax returns (Feldman and Slemrod, 2007).

Weber (1989)[4]. In their study researchers investigate the relationship between income and food expenditure for two types of workers – entrepreneurs (self-employed individuals) and employees (individuals in paid-employment). By assuming that, despite employment type, both groups have the same preferences regarding food and report it‘s expenditures correctly, while income is reported accurately only by employees, latter scholars estimate food expenditure equations taking into consideration reported incomes. Differences in the estimated relationship might be assumed as under-reporting of income by self-employed.

According to Pissarides and Weber (1989), expenditure on food is reported most accurate and is the only expenditure item that could be trusted to give precise estimates of income under-reporting. Food expenditure is considered to be seamingly small to cause any suspicion and be concealed for tax purposes or to be considered as business expenses (Pissarides and Weber, 1989; Lyssiotou et al., 2004, Tedds, 2005). Besides, expenditure on other items is less regular (Pissarides and Weber, 1989). In a similar vein, Tedds (2005) outline additional arguments for using food rather than other commodity. The most crucial of them is difficulty to postpone food consumption because food is a necessity. As well "tastes for food are more likely to be uniform across employment groups and over time" (Tedds, 2005, p. 10). Besim and Jenkins (2005) support this theory by stating that the relationship between income and food consumption is the most stable compared to other items of expenditure.

A number of researchers adopted food expenditure-based approach. As a result, literature reveals that, generally, the basic hypothesis, that entrepreneurs are likely to under-report their incomes in comprison with employees, is supported. However, there is a considerable variation in under-reporting estimates within and between countries.

Pissarides and Weber‘s (1989) analysis of 1982 UK Family Expenditure Survey (FES) data shows that on average entrepreneurs tend to under-report their incomes by 55 per cent. It means that, in order to find true incomes, reported incomes have to be multiplied by a factor of 1,55. Consequently, since self-employed account for aproximately 10 per cent of UK GDP, latter estimate implies that the size of the black economy in UK was around 5,5 per cent of GDP in 1982. Moreover, Pissarides and Weber (1989) reveal some evidence regarding the difference of under-reporting rate between white-collar and blue-collar workers. White-collar workers conceal about 28-54 per cent of their true incomes while blue-collar workers tend to conceal a little more – 51-64 per cent. Subsequent work by Baker (1993), who also analysed UK FES data and applied the same methodology, provides lower under-reporting estimates – self-employed individuals conceal from

[4] A more detail explanation of Pissarides and Weber (1989) approach is presented in section X.

20 to 50 per cent of their true incomes. Additionally, latter researcher reports that under-reporting rates differ between occupational and industry groups. In contrast to Pissarides and Weber's (1989) results, Baker (1993) states blue-collar workers to under-report less compared to white-collar self-employed workers. More recent analysis of individual household data drawn from UK FES 1993 by Lyssiotou et al. (2004) suggests even higher under-reporting rates by British entrepreneurs. Lyssiotou et al. (2004) extend Pissarides and Weber (1989) approach by proposing a complete demand system method instead of using a single consumer demand equation for under-reporting estimation. This method is argued to have two advantages: first, it prevents a potential bias in the estimates of under-reporting because it is allowed for self-employed individuals to have diverse preferences and, thus, different consumer demands than wage workers; second, eliminates the need to categorize households by their main source of income. Lyssiotou et al. (2004) estimtate that the size of entrepreneurship-related black economy in the UK amounted to 10,6 per cent of GDP in 1993. This result is nearly twice higher compared to Pissarides and Weber's (1989) findings. Moreover, Lyssiotou et al. (2004) find that households with head in blue-collar self-employment occupation under-report more than households with head in white-collar self-employment occupation – respectively, 118 and 64 per cent. This implies that blue-collar workers report 46 per cent of their true incomes while white-collar workers report 61 per cent.[5] Thus, if more changeable incomes influence entrepreneurs to save more in comparison with employees, these under-reporting rates might be downward-biased (Quadrini, 1999).

Under-reporting rates in Canada appear to be lower. Mirus and Smith (1996), using data from Canadian Family Expenditure Surveys (FAMEX), estimate 12,5 per cent under-reporting rate for entrepreneurs for the year 1990. While more recent analysis of FAMEX data from 1969 to 1992 suggests that households which obtain at least 30 per cent of their income from entrepreneurship tend to under-report their true incomes by 11-23 per cent (Schuetze, 2002). In addition, Schuetze (2002) finds evidence that the degree of under-reporting rate differs significantly among different occupations of self-employed individuals. The highest rate of income concealment is dislosed in the occupations which involve frequent cash transanctions, hence, opportunities to evade tax is high. The highest estimates of under-reporting in Canada are found in construction and services occupations – 38-54 per cent. Further, according to the latter author, households with considerable amount of entrepreneurial income which are headed by younger males tend to under-report income to a greater extent than those households headed by older males. Besides, Schuetze (2002) noticed that couples where both spouses are involved in self-employment conceal less of their income than

[5] Blue collar workers report 46 per cent [=1/(1+1,18)] of their true incomes. White collar workers report 61 per cent [=1/(1+0,64)] of their true incomes (Lyssiotou et al., 2004).

couples where only one of the spouses is self-employed. Also under-reporting rates tend to decrease with age and vary from year to year. Therefore, it is important that researchers use multiple years of data while estimating the extent of self-employed income under-reporting (Shuetze, 2002). Tedds (2005), using the same FAMEX data, adds that in Canada income under-reporting varies respectively to household income levels. Self-employed households at the lower end of the self-employment income distribution present larger gap between reported and true income. At the same time, some of the households in the upper end of income distribution tend to over-report their income (Tedds, 2005).

Applying food expenditure-based approach, Apel (1994) and Engstrom and Holmlund (2007) estimate the under-reporting rates in Sweden. Based on prior Apel's (1994) analysis of 1988 data from Swedish Household Budget Survey (Hushållens utgifter, HUT[6]), in order to arrive at the true self-employment income, reported income should be multiplied by 1,35. Thus, entrepreneurs conceal approximately 35 per cent of their true income. More up-to-date research of HUT data from years 1999-2001, 2003 and 2004 reveals somewhat lower under-reporting estimates. Households with at least one member involved in self-employment tend to under-report their incomes at the average rate of 30 per cent (Engstrom and Holmlund, 2007). Also latter scholars examine whether the rate of under-reporting in Sweden varies depending on the legal form of self-employment. The evidence suggest that self-employed individuals with unincorporated business report significantly lower incomes than those with incorporated business. Engstrom and Holmlund (2007) explain this finding by „indicating higher costs of tax evasion for owners of incorporated businesses since their transactions are more easily exposed to public scrutiny" (Engstrom and Holmlund, 2007, p. 17).

The estimates of income under-reporting by entrepreneurs in Finland appear to be slightly higher compared to Sweden. In his paper, Johansson (2005) uses data from a survey of household expenditures drawn up by Statistics Finland during the years 1994-1996. Depending on how self-employed household is defined, the magnitude of under-reporting varies between 16 and 40 per cent. More precisely, if only the head of the household is involved in entrepreneurship the rate of under-reporting is 16,5 per cent on average. If both spouses are self-employed they conceal approximately 42 per cent of their incomes. Thus, the hypothesis, that the larger household's share of income comes from self-employment, the larger under-reporting rate is, is supported (Johansson, 2005).

Besim and Jenkins (2005) applied the approach pioneered by Pissarides and Weber (1989) to data from 1998-1999 Household Consumption Expenditures Survey (HCES) for North Cyprus.

[6] HUT is the Swedish equivalent to the UK FES.

Latter researchers separated consumption expenditures in three categories – expenditures made by self-employed, private employees and civil servants[7]. Comparison of food expenditures made by these three groups enables to estimate how much self-employed and the private employees understate their incomes as compared to the civil servants. The results indicate that self-employed tend to under-report their income approximately by 11-14 per cent. Moreover, Besim and Jenkins (2005) find that the average degree of under-reporting of private employees is very close to the estimates for the self-employed, with between 9,7 and 14,2 per cent of income being concealed.

More recent paper of Kim et al. (2011) presents a refinement of the Pissarides and Weber approach (1989) for indicating the degree of under-reporting by self-employed. For this purpose scholars use data from two panel surveys – the Korea Labor Income Panel Survey (KLIPS) from 2000-2005 and the Russian Longitudinal Monitoring Survey (RLMS) from 1994-2000. According to Kim et al. (2011), panel data method eliminates the effect of transitory income fluctuations to income under-reporting and provides an exact estimate of the under-reporting rate rather than an interval estimate. The evidence suggests that survey income under-reporting by entrepreneurs in Korea is 20,1 per cent, while in Russia under-reporting rate appear to be slightly lower – 18,7 per cent. Consequently, the true income are 1,25 and 1,23 times the reported income for the households involved in entrepreneurship, respectively.

It is worth mentioning that the degree of income under-reporting can be estimated not only from expenditure and income data, but also from actual tax return income data. Feldman and Slemrod (2007) apply full demand system approach with using US Internal Revenue Service 1999 Statistics of Income Public Use Data, therefore, instead of food expenditures researchers focus on charitable donations included on tax returns for the purpose of claiming tax deductions. As well as food expenditure, charitable contribution is considered unlikely to be misreported by either of occupational groups. Assuming that the source of income is not related to charitable inclinations and that the proportion of true income to taxable income does not depend on income source, Feldman and Slemrod (2007) claim that the compliance rate of self-employment income in the USA varies between 55 and 71 per cent.

The overview of previous literature reveals great interest demonstrated by researchers in estimating income under-reporting by entrepreneurs. The importance of assessing the degree of under-reporting can be explained by two main reasons. First of all, self-employed income under-reporting to the household surveys is one the most serious limitation of public-use data. Hence, raw data provided to interviewers possibly understate true earnings received by entrepreneurs.

[7] A civil servant is a person in the public sector employed for a government department or an agency (Wikipedia, 2013).

Consequently, while comparing income among entrepreneurs and non-entrepreneurs, it is necessary to estimate income under-reporting degree for self-employed respondents and use a correction factor to their reported incomes in order to provide more accurate values of their true incomes (Parker, 2009; Dilnot and Morris, 1981). Secondly, it is one of the methods to measure black economy. Although a number of studies rely on macroeconomic relationships while estimating the size of the black economy, this approach is criticised for not being based on theory and for applying not entirely correct econometric techniques (Thomas, 1999). Consequently, microeconomic aprroaches are in a high interest. Under-reporting rates can be applied in the under-ground economy estimation if the relative importance of self-employment in the entire economy is known (Lyssiotou et al., 2004).

Overall, studies, using expenditure-based approach developed by Pissarides and Weber (1989), conclude that there is a significant under-reporting by self-employed individuals in many countries. However, there is still a debate about the extent of under-reporting and the resulting estimates of under-ground economy. Hence, a further insight to this problem is needed.

The research question of this report derived from the above analysis is formed as follows:

Do entrepreneurs under-report their true incomes compared to wage workers in Australia, using Household, Income and Labour Dynamics Survey data from 2001 and 2003-2005?

To investigate mentioned problem, income and expenditure differences between entrepreneurs and wage workers are explored based on the framework (food expenditure-based approach) previously proposed by Pissarides and Weber (1989). As well as under-reporting by the legal form of self-employment is investigated on purpose to have a deeper insight into this activity.

2.2 A Review of the Pissarides and Weber's (1989) Model

Pissarides and Weber (1989) estimate the rates of income under-reporting by entrepreneurs and employees taking British income and expenditure data from 1982 Family Expenditure Survey (FES). More precise, researchers estimate food expenditure equations depending on reported incomes and household characteristics.

In order to capture the degree of income under-reporting, Pissarides and Weber make three major assumptions:

(1) Expenditure is reported accurate by all households;

(2) Paid-employed households report income accurately though self-employed households might not;

(3) The expenditure function on some items is the same for all households.

According to the researchers, the item of expenditure which is believed to be reported most accurate by both types of households is food. Food expenditure is regarded to be as highly unlikely to be concealed for tax reasons or to be covered as business expenses.[8]

Pissarides and Weber (1989) approach can be represented graphically (Figure 1).

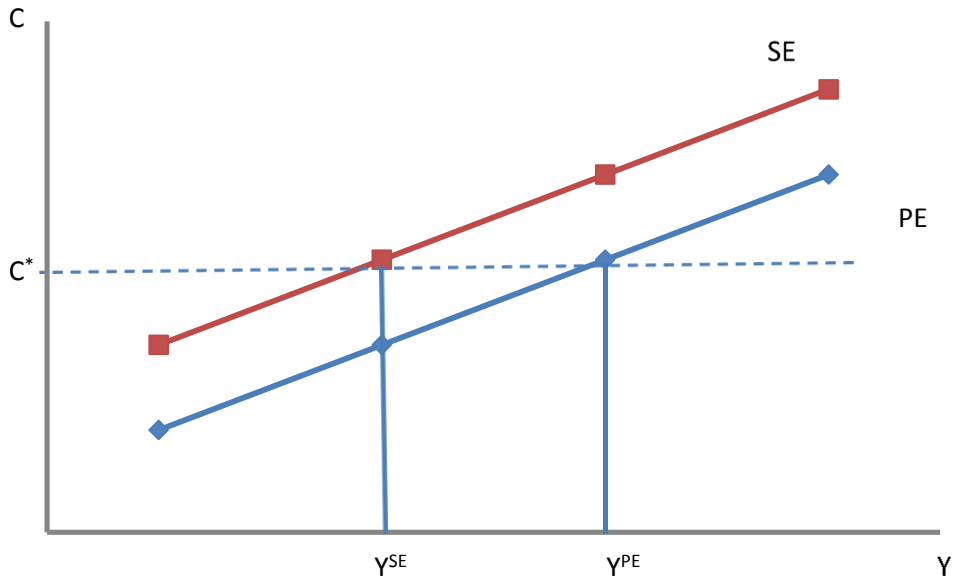

Figure 1. Engel Curves (consumption-income profiles)

Engel curves describe how household expenditure on food varies with household income (Lewbel, 2006). More specifically, Figure 1 illustrates two log-linear consumption-income profiles: one for self-employed households (SE) and one for paid-employed (PE). It is assumed that consumption elasticity with respect to income is equal for both types of households. Hence, Engel curves have the same slope. Furthemore, in order to capture possible under-reporting by self-employed households, intercepts are allowed to vary. In more detail, C denotes logarithmic transformation of food consumption, while Y is a logarithmic transformation of income. Both types of households report the same consumption level of food indicated as C^* while reported income level differs – self-employed household reports income level Y^{SE}, paid-employed Y^{PE}. By assumption, food consumption is reported accurate by self-employed and paid-empolyed households. Though only income reported by paid-employed household corresponds to real income. If preferences are the same, it can be assumed that self-employed household‘s true income is Y^{PE}

[8] Pissarides and Weber (1989) tried to use other items of expenditure (clothing, services and “pre-commitments”) in order to check its’ consistency of the estimates with food, however, the results were mixed and no clear evidence of possible under-reporting was found.

rather than reported Y^{SE}. The extent of under-reporting might be captured by the difference between Y^{PE} and Y^{SE}.

In more detail, Pissarides and Weber's (1989) method consists of two parts. First stage is an estimation of expenditure functions in terms of reported income and household characteristics. In the second step, latter expenditure functions are inverted in order to forecast income from reported expenditure[9]. Hence, food expenditure function can be expressed as fallows:

$$\ln C_i = \alpha Z_i + \beta \ln Y_i^p + \varepsilon_i \quad (1)$$

where $\ln C_i$ is a natural logarithm of food consumption for the household i, Z_i is a set of household characteristics, Y_i^p is the measure of income that effects consumption decisions – permanent after tax income.

Therefore, as Pissarides and Weber (1989) assumed, reported income is likely to differ from permanent income for two reasons: first, true self-employed income might differ from reported income because of under-reporting but not in the wage workers case. The relationship between true (Y_i^t) and reported (Y_i^r) incomes can be written:

$$Y_i^t = k_i Y_i^r \quad (2)$$

where k_i is a random variable which shows the extent of under-reporting by the household i. For the wage workers $k_i = 1$ while for the self-employed $k_i > 1$. A bigger k_i indicates a greater rate of under-reporting.

Secondly, true income (Y_i^t) might differ from permanent income (Y_i^p) because of transitory shocks. As follows, former and latter types of income are related by a random variable p_i:

$$Y_i^t = p_i Y_i^p \quad (3)$$

which measures the effect of aggregate events influencing current income. It is assumed that the mean of p_i is the same for self-employed as well as paid-employed households. Though its variance is likely to differ between each group.

Taking into consideration (2) and (3), food expenditure function (1) can be re-written as follows:

$$\ln C_i = \alpha Z_i + \beta \ln Y_i^r - \beta \ln p_i + \beta \ln k_i + \varepsilon_i \quad (4)$$

It is worth noting that there is no information about p_i or k_i. In order to make estimation possible it is held that p_i and k_i have a log-normal distribution and can be expressed by deviations from their means:

$$\ln p_i = \mu_p + u_i \quad (5)$$

[9] For more details see Pissarides and Weber (1989).

$$\ln k_i = \mu_k + v_i \quad (6)$$

where u_i and v_i have zero means and constant variances σ_u^2 and σ_v^2. Consequently, (5) and (6) can be substituted into (4):

$$\ln C_i = \alpha Z_i + \beta \ln Y_i^r - \beta(\mu_p - \mu_k) - \beta(u_i - v_i) + \varepsilon_i \quad (7)$$

Previously it was noted that for the wage earners $k = 1$, therefore, $\mu_k = 0$. Pissarides and Weber (1989) state that the variance of p_i is bigger for entrepreneurs than for employees. This implies that the mean of $\ln p_i$ (μ_p) is smaller for the self-employed households. Thus, there appear differences in the estimates of equation (7) between two occupational groups because $\mu_p - \mu_k$.

In order to capture the differences in food expenditure between self-employed and wage workers‘ households self-employment indicator is implied:

$$\ln C_i = \alpha Z_i + \beta \ln Y_i^r + \gamma SE_i + \varepsilon_i \quad (8)$$

where SE_i is a dummy variable taking value 1 if household is self-employed and value 0 if household is paid-employed. It is of major importance that reported income Y_i^r here is treated as endogenous and instrumented which allows independently to estimate the variance of residual of reported income for each occupational group. The residual variance is vital for the under-reporting rate calculation. Moreover, Pissarides and Weber (1989) conclude that the coefficient of self-employment indicator is the following:

$$\gamma = \beta[\mu_k + \frac{1}{2}(\sigma_{uSE}^2 - \sigma_{uPE}^2)] \quad (9)$$

where μ_k is the mean of logarithmic expression of k_i and σ_u^2 is the variance of p_i for self-employed and paid-employed respectively.

Furthermore, the most important estimate to capture under-reporting rate is the average value of k_i, the number by which average reported sef-employment income can be multiplied in order to find average true income. Assuming log-normality of k_i, it‘s mean value can be written:

$$\ln \bar{k} = \mu_k + \frac{1}{2}\sigma_{vSE}^2 \quad (10)$$

where $\bar{k}$ is the mean of k_i and σ_v^2 is the variance of k_i. By substituting (9) into equation (10) $\ln \bar{k}$ can be expressed:

$$\ln \bar{k} = \frac{\gamma}{\beta} + \frac{1}{2}(\sigma_{vSE}^2 - \sigma_{uSE}^2 + \sigma_{uPE}^2) \quad (11)$$

However, it is not possible to obtain the estimates of the variances of p_i and k_i. Instead, Pissarides and Weber (1989) estimate a reduced form of income equation (12) to calculate the variance of errors in income for the self-employed ($\sigma_{\gamma SE}^2$) and wage workers ($\sigma_{\gamma PE}^2$) separately.

$$\ln Y_i^r = \delta_1 Z_i + \delta_2 X_i + \zeta_i \quad (12)$$

where X_i is a set of instrumental variables. Further, authors provide assumptions which allow, by using income variance, to estimate upper and lower bounds of the extent of under-reporting. The upper bound of $\bar{k}_u$ can be estimated as follows:

$$\ln\bar{k}_u = \frac{\gamma}{\beta} + \frac{1}{2}(\sigma^2_{\gamma SE} - \sigma^2_{\gamma PE}) \quad (13)$$

An estimate of the lower bound ($\bar{k}_l$) can be obtained by estimating the following equation:

$$\ln\bar{k}_l = \frac{\gamma}{\beta} - \frac{1}{2}(\sigma^2_{\gamma SE} - \sigma^2_{\gamma PE}) \quad (14)$$

3. Empirical Analysis

3.1 Data and Method

To empirically establish current analysis, data are drawn from the Household, Income and Labour Dynamics in Australia Survey (HILDA). The HILDA Survey is a broad social and economic longitudinal survey, with particular attention paid to family and household formation, income and work. Started in 2001 with 10 twelve-month waves (from 2001 until 2010), survey collects data annually and most questions are repeated every year. This data is chosen because it allows to track households‘ expenditure on food (including meals outside) as well as to examine whether the degree of under-reporting varies by the legal form of self-employment (incorporated and unincorporated businesses).

Following previous literature, present study is based on the household level and focuses only on the sample restricted to households containing two adults who are legally married or cohabiting (not legally married but living together in a relationship; same gender couples are included; with or without children). Households are headed[10] by individuals aged 18-65 (respondents are presumed to be retired if aged 65 and over) that are not working in the farm sector. In particular, farmers are excluded because they are likely to have different food expenditure patterns, since they can produce some parts of their food themselves. Moreover, unpaid family workers are dropped from the sample because they do not gain any earnings. These restrictions left 2.280 households (2001).

3.1.1 Self-employed vs. Paid Employees

Most of the previous studies classify households according to the information on income shares from self-employment or paid-employment[11]. However, this approach has several issues: first, it is difficult to choose the boderline and, secondly, it may be sensitive to the legal form of

[10] A person is considered to be “the head of the house“ if in the data set he/she is recorded as the number one member of the household (person number = 1).

[11] Pissarides and Weber (1989) define households as self-employed if income from self-employment accounts for at least 25 per cent of total income.

self-employment. Individual with incorporated business might be formally employed in the company and receive the main part of income in the form of wage (Engstrom and Holmlund, 2007).

Taking into consideration latter issues, self-reported information on employment status of the individual is used in this analysis. Respondents were asked about their current employment status. These answers are used to distinguish between the households which are headed by entrepreneurs (self-employed) or non-entrepreneurs (paid-employees). Self-employed workers are defined as those who identified themselves as employee of own business (incorporated business owners) and those who considered themselves as employers or own account workers (unincorporated business owners). Both, incorporated and unincorporated business owners, are included in self-employment sector. Paid employees are considered to be individuals who described themselves as employees.

Moreover, in this study it is made an assumption that respondents are considered to be self-employed or paid employees only if they referred that in a 12 month period this job activity was their occupation at least the last 3 months.

3.2 Dependent Variable

3.2.1 Food Consumption

Respondents were asked questions related to their expenditure. These answers were used in order to construct the dependent variable – annual food consumption.

This dependent variable is calculated taking into account how much the household spends on groceries in a normal week (it includes food, supermarket and convenience store shopping such as meat and fish, bread and milk, fruit and vegetables, tinned and packaged food, pet food, cleaning products, bathroom products, drinks, etc.; alcohol and tobacco are excluded) and approximately how much the household usually spends per week on meals outside the home (it includes restaurants, bought lunches and snacks; alcohol is excluded). These two types of weekly expenditures are summed and multiplied by 52 weeks in order to get an approximate amount of money spend on food per year. Moreover, the dependent variable is constructed only for the waves 1 and 3 – 5 because information related with household expenditures is provided only in years 2001 and 2003 – 2005.

It might be argued that including meals eaten outside in total food consumption might upward bias the estimates of under-reporting because entrepreneurs simply might eat in the restaurants more often and the prices there tend to be higher. However, Mirus and Smith (1996) as well as Schuetze (2002) found that there is no relationship between the share of income from entrepreneurship and the ratio of restaurant-bought to store-bought food.

The empirical distributions of self-employed and paid-employed households‘ average weekly food expenditures of the year 2001 are presented in Figure 2.

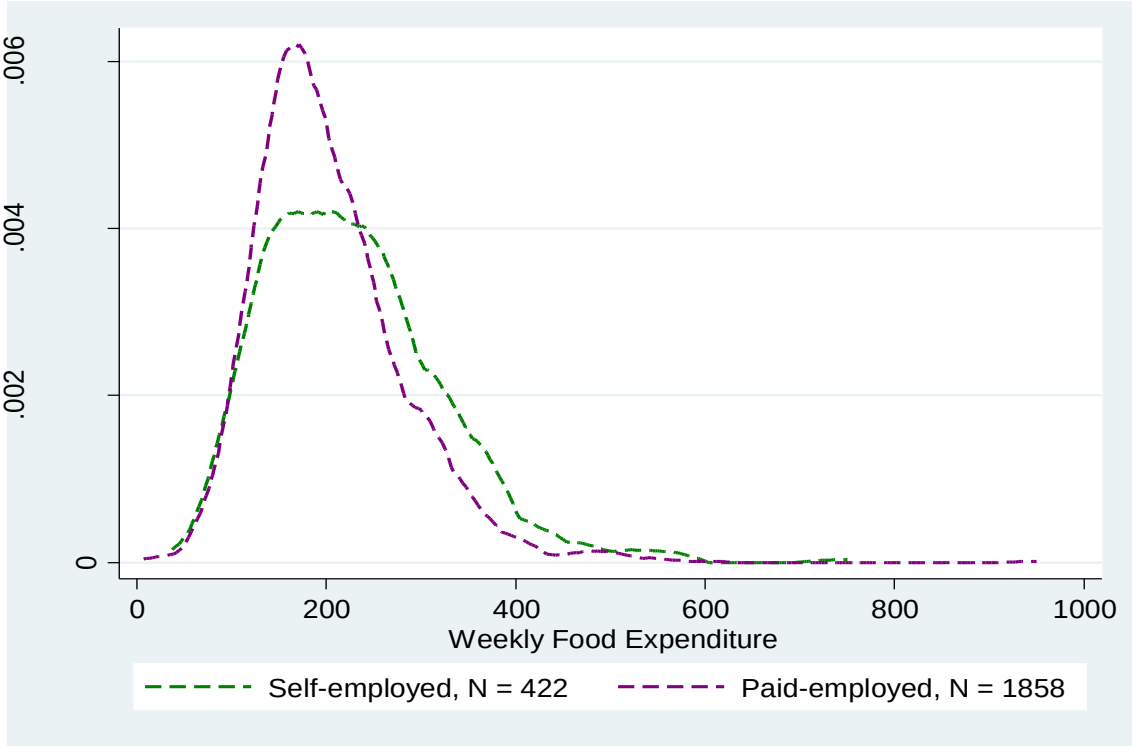

Figure 2. Empirical Distributions, Weekly Food Expenditure, 2001

In the graph it can be seen that the central tendency of the distribution of self-employed households‘ food expenditure is slightly more than that of the paid-employed. The distribution of paid-employed food expenditure has smaller dispersion in comparison with self-employed. However, it is more skewed which means that some of the paid-employed households reported weekly expenditure on food higher than $900.

3.3 Main Independent Variable

3.3.1 Reported Income

Respondents were asked questions related to their earning income. These answers were used in order to construct the main independent variable – annual reported income.

In particular, respondents were asked about their annualy gross wages and salaries before any tax was deducted (employees, incorporated business owners) as well as their total share of profit from their unincorporated business before tax in the last finacial year (unincorporated business owners). The dependent variable reported income is created by taking into account the answers to the latter questions and further recoded so as to obtain the total amount of money

respondents earn per year after taxes. In more detail, total net income is obtained by summing gross reported incomes and than subtracting recorded tax payments (income tax, medicare tax and less offsets). Total net income is assumed to be more than zero.

The empirical distributions of self-employed and paid-employed households‘ reported income are presented in Figure 3 (average weekly self-employed reported income vs. average weekly paid-employed reported income in 2001).

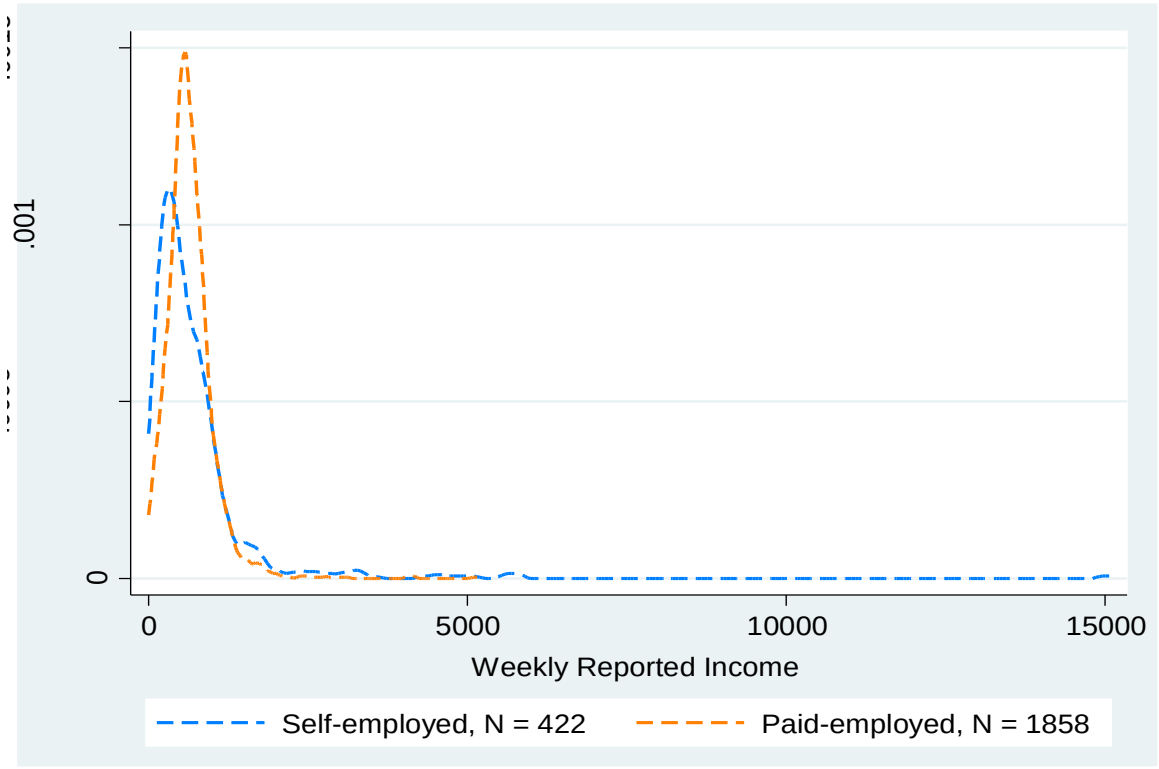

Figure 3. Empirical Distributions, Weekly Reported Income, 2001

The graph displays two important characteristics. First of all, the central point of the distribution of the self-employed reported income is less than the central tendency of the paid-employed reported income distribution. Secondly, paid-employed reported income exposes slightly smaller dispersion compared to self-employed. Furthermore, the distribution of income reported by self-employed households is more skewed – there are respondents who stated to have weekly income higher than $15.000.

More evidence about the differences in income and food expenditure distribution of entrepreneurs and wage workers is depicted in Table 1.

TABLE 1

Summary Statistics: Reported Income, Food Expenditure, by Employment Sector, 2001

	$\ln Y^r$		$\ln C^f$	
	Self-employed	Employees	Self-employed	Employees
STATISTIC				
Mean	10.033	10.229	9.295	9.193
Standard deviation	1.077	0.726	0.422	0.393
10^{th} percentile	8.717	9.473	8.740	8.739
25^{th} percentile	9.559	10.011	9.026	8.962
50^{th} percentile	10.130	10.353	9.345	9.198
75^{th} percentile	10.683	10.639	9.586	9.473
90^{th} percentile	11.051	10.897	9.809	9.655
Kolmogorov-Smirnov test of normality of $\ln Y^r$	0.0584 (0.066)	0.0196 (0.032)		
Observations	*422*	*1858*	*422*	*1858*

Notes: (1) Approximate critical values at the 95 per cent level are in parentheses below Kolmogorov-Smirnov statistics.

Summary statistics of reported income by self-employed and paid-employed households reveal that the mean of self-employment income is lower compared to the mean of income reported by wage workers. However, the mean of food expenditure reported by entrepreneurs is higher than the corresponding entries for the employees in employment. This difference signals about possible under-reporting of self-employment income – even though entrepreneurs state to earn less than wage workers, their expenditures are higher. Only in the upper quantile self-employed income exceeds income of paid-employed households[12]. Moreover, income and food expenditure reported by self-employed households exhibit a greater variation than that of paid-employed. A Kolmogorov-Smirnov test for lognormality of income shows that log-normality is not rejected at the 95 percent level for any occupational group, at the conventional approximation used to derive critical values for the test[13].

[12] This can be explained by so called "superstar theory" (Rosen, 1981) – which says that earnings are influenced by a few entrepreneurial superstars who have a very high income. So this does not characterize the self-employment returns for the majority of business owners.

[13] Lognormality can be rejected if the Kolmogorov-Smirnov statistic exceeds the critical value.

Additionally, the correlation coefficients between food consumption and reported income (overall and reported by self-employed and employees separately) are positive and statistically significant at any significance level (see Table 9 in Appendix). The rank correlation is stronger between food consumption and income reported by paid-employed households.

Table 2 reports trend and change of food expenditure and reported net income over time by employment sector.

TABLE 2

Summary Statistics: Trend of Mean of Food Expenditure and Reported Income, by Employment Sector

	Self-employed				Employees			
	2001	2003	2004	2005	2001	2003	2004	2005
Food expenditure $\ln C^f$	9.29	9.33	9.37	9.41	9.19	9.24	9.27	9.30
Reported income $\ln Y^r$	10.03	10.04	10.06	10.20	10.23	10.30	10.33	10.37
N	*422*	*354*	*374*	*380*	*1858*	*1740*	*1725*	*1818*

The results confirm that mean reported income is higher for the employees than for the self-employed households, although food expenditure imply an opposite pattern. Assuming that respondents report their food expenditure correctly, Engel‘s Law appears to be violated among the two occupational groups, thus, suggesting that there might be a possibility of self-employed income under-reporting.

3.4 Instrumental Variables

3.4.1 Income from Capital

Respondents were asked questions related to their income from capital. These answers were used in order to construct an instrumental variable – annual income from capital.

In more detail, respondents were asked about their total amount of income received from renting properties they own during the last financial year after all expenses were deducted. Also they were asked about financial year ineterest – how much they earn from the sources such as interest from banks or any other financial institution, interest from debentures, bonds, trusts or personal loans to other persons than persons from the same household. The answers to the previous questions were taken into account and futher totalized so to obtain the total annual income from capital. Capital from income is assumed to be more than zero.

3.4.2 Number of Bedrooms

Respondents were asked about their housing situation. More specifically they were asked about the number of bedrooms which are in the house where respondents live, taking into account even bedrooms which are not currently used as such (for example, studies). The answers to the latter question were taken into account and further recoded in order to construct an instrumental variable - total number of bedrooms.

3.5 Control Variables

In order to capture household characteristics control variables[14] are used in under-reporting estimation. The following control variables are believed to affect the budget constraint as well as preferences of the household.

Respondents were asked to provide basic demographic information including *age* (*age squared*) and *gender* where value 1 denotes if the household is headed by male and 0 by female. Furthermore, respondents were asked how many children in total they have ever had including adopted ones and how many of these children are resident and live together with their parents at least 50 per cent of the time. According to the answers to previous questions, the following control variables are constructed – *total number of children* and its squared, *number of resident children* and its squared. Further, variable total number of in-scope *persons in the household* is created. Moreover, respondents were asked questions about their education. First, 10 dummy variables are created each of them presenting the *highest education level achieved*. Additionally, variable *years of education* is constructed taking into account respondents age and how many years passed since they left full-time education. To continue, respondents were asked about their *willingness to take a financial risk*. Entrepreneurs are expected to be more risk takers than wage workers (Parker, 2004). More specifically, respondents were asked whether they are willing to take a financial risk with their spare cash (cash used for savings or investments). Hence, a dummy variable is constructed taking the value 1 if individual wants to take substantial, above average or average risks expecting substantial returns and value 0 if he is not willing to take any financial risk. Following the same logic, dummy variable, related with respondents‘ *saving habits*, is created. It takes value 1 if respondent is likely to save and 0 otherwise. Besides, interviewers recorded their *dwelling type* which let to construct 11 dummy variables where in each of them value 1 denotes a specific type of dwelling and 0 all the other types. Farther, respondents provided answers about their current *job*

[14] For more detail information look Table 10 in Appendix.

industry. Responses to the latter question enabled to creat 16 industry dummies. Finally, 8 state dummies which provide information about the region of respondent‘s residence are constructed.

Moreover, differences between described main characteristics of self-employed and paid-employed households are depicted in Table 3 (sample of 2001). Findings show that paid employees are slightly more educated compared to self-employed workers. However, entrepreneurs tend to be older and have more children (overall and resident). As a consequence, there are more persons in the self-employed household, on average. Besides, self-employed individuals are slightly more likely to save and more willing to take financial risk than paid workers.

TABLE 3

Variable Descriptions and Summary Statistics, by Employment Sector, 2001

	MEAN	
DESCRIPTION	Self-Employed	Paid Employees
Years of education	16.63	16.71
Number of children	2.3	1.8
Number of resident children	1.4	1.3
Age	44.29	40.70
Saving habits	0.76	0.75
Willingness to take financial risk	0.65	0.49
Persons in household	3.5	3.4
Observations	*422*	*1858*

To evaluate the differences in means, presented in Table 3, t-test method is used. Hence, the results are confirmed - the means of age, willingness to take financial risk, the number of children is statistically different from each other at any reasonable significance level. While as expected slight differences between means of the number of persons in household, number of resident children, education level and saving habits of self-employed and paid employees turned out to be insignificant.

3.6 Methodology

In order to make an introduction to the empirical analysis, descriptive statistics are being used. As primary analysis, an arguably straightforward approach to examine income under-reporting by employment status is to check possible differences in reported income among wage

workers and entrepreneurs with similar human capital and other characteristics (Engstrom and Holmlund, 2007). To examine whether households headed by self-employed individuals earn more compared to the households headed by individuals in paid-employment, ordinary least squares (OLS) regressions are being conducted (separate regressions for the years 2001 and 2003-2005). Concretely the following earning function is estimated:

$$\ln Y_i = \delta Z_i + \sigma SE_i + \eta_i \quad (15)$$

where Y_i is gross income (before taxes and transfers), Z_i is a vector of household characteristics, SE_i is a dummy for self-employment status, i is an index of household in a sample of size n (i = 1, ... n) and η_i is a stohastic disturbance term. A negative sign of the estimate of SE_i dummy would be in consistent with self-employed income under-reporting. However, a negative estimate could also reflect compensating income differences which may appear as a result of positive being in self-employment value because of it‘s non-pecuniary benefits such as “being your own boss” (Hamilton, 2000).

Further, following Pissarides and Weber (1989) approach, two-stages least squares (2SLS) estimation is used. Latter scholars as well as some other researchers (Tedds, 2005; Engstrom and Holmlund, 2007, etc.) discussed how income should be interpreted and treated in the estimation of under-reporting. There is a general belief that households base their expenditures on permanent rather than current income. Using current income in the expenditure-based method implementation might bias the estimates of under-reporting. However, data sets usually include only information about current income. Pissarides and Weber (1989) indicate that “…for given permanent income, the measured income of the self-employed may be more variable than the measured income of employees in employment. If this is correct, our measure of income under-reporting by the self-employed will have to be adjusted accordingly.” (Pissarides and Weber 1989, 20). Moreover, it is plausible that income is endogenous in the food consumption equation. Hausman test for endogeneity confirm that the log of household net reported income, given the chosen instruments – income from capital and number of bedrooms – is endogenous (see Table 11 in Appendix). The null hypothesis of income exogeneity in the household food consumption function is rejected at 5 per cent significance level. Thus, to resolve former and latter issues instrumental estimation must be pursued, which also “…enables an independent estimate of the residual variance of reported income for each group which we exploit in the calculation of income under-reporting” (Pissarides and Weber, 1989, p. 22).

In more detail, an instrumented equation 16 (it is an equivalent of eq. 8) is being estimated for each year (2001, 2003-2005) in order to get necessary coefficient estimates for income under-reporting calculation – marginal propensity to consume and coefficient on the self-employment dummy.

$$Ln\ food\ consumption_i = \beta_0 + \beta_1 * ln\ net\ reported\ income + \beta_2 * self\text{-}employed + \beta_3 * gender + \beta_4 * age + \beta_5 * age^2 + \beta_6 * years\ of\ schooling + \beta_7 * financial\ risk + \beta_8 * saving\ habits + \beta_9 * total\ number\ of\ children + \beta_{10} * total\ number\ of\ children^2 + \beta_{11} * number\ of\ resident\ children + \beta_{12} * number\ of\ resident\ children^2 + \beta_{13} * number\ of\ persons\ in\ the\ household + \beta_{14}\text{-}\beta_{21} * region + \beta_{22}\text{-}\beta_{32} * type\ of\ dwelling + \beta_{33}\text{-}\beta_{42} * education\ level + \beta_{43}\text{-}\beta_{58} * industry + \varepsilon_i \quad (16)$$

Instruments used are income from capital and number of bedrooms[15]. Further, equation 16 is being conducted for each occupational group - entrepreneurs and employees - separately to find variance of errors ($\sigma^2_{\varepsilon SE}$ and $\sigma^2_{\varepsilon PE}$) because it is also of interest. In the next step, residual variance of reported income ($\sigma^2_{\gamma SE}$ and $\sigma^2_{\gamma PE}$) is being calculated from reduced form income regression equation 17 (it is an equivalent of eq. 12) separately for self-employed and paid-employed households. The variance of the residual is calculated by dividing the sum of squared residuals by the degrees of freedom.

$$Ln\ net\ reported\ income_i = \beta_0 + \beta_1 * gender + \beta_2 * age + \beta_3 * age^2 + \beta_4 * years\ of\ schooling + \beta_5 * financial\ risk + \beta_6 * saving\ habits + \beta_7 * total\ number\ of\ children + \beta_8 * total\ number\ of\ children^2 + \beta_9 * number\ of\ resident\ children + \beta_{10} * number\ of\ resident\ children^2 + \beta_{11} * number\ of\ persons\ in\ the\ household + \beta_{12}\text{-}\beta_{19} * region + \beta_{20}\text{-}\beta_{30} * type\ of\ dwelling + \beta_{31}\text{-}\beta_{40} * education\ level + \beta_{41}\text{-}\beta_{56} * industry + \beta_{57} * ln\ income\ from\ capital + \beta_{58} * number\ of\ bedrooms + \zeta_i \quad (17)$$

On purpose to indicate the degree of self-employment income under-reporting, lower and upper bounds are being calculated (eq. 13 and eq. 14). Latter estimates are being translated into the mean values of *k*, the number by which mean reported incomes have to be multiplied to arrive at true incomes, by taking antilogs of lower and upper bounds estimates.

For the deeper analysis, it is examined whether the legal form of self-employment matters for the degree of under-reporting by including dummy variables for incorporated or unincorporated businesses. Under-reporting by the legal form of self-employment is being estimated by following the same steps described above.

3.7 Results

Table 4 presents the main estimates of OLS regressions with logarithmic transformation of reported annual gross income as dependent variable (all estimates of the regressions are provided in Tables 12-15 in Appendix).

[15] Pissarides and Weber (1989) use *number of rooms* in the house as one of additional instruments. *Income from capital* as an instrument to achieve identification is used by Engstrom and Holmlund (2007).

TABLE 4

Parameter Estimates from Annual Gross Income Regressions

	OLS REGRESSIONS		
	Dependent variable: Gross Income ($\ln Y_i$)		
	Both genders (1)	Men (2)	Women (3)
2001			
Self-employed (SE)	- 0.451***	- 0.465***	- 0.335*
SE incorporated	0.560***	0.550***	0.584***
SE unincorporated	- 0.068	- 0.131	0.122
2003			
SE	- 0.464***	- 0.429***	- 0.522***
SE incorporated	0.444***	0.422***	0.438***
SE unincorporated	- 0.028	- 0.004	- 0.057
2004			
SE	- 0.447***	- 0.440***	- 0.344**
SE incorporated	0.407***	0.389***	0.390**
SE unincorporated	- 0.054	- 0.126	0.043
2005			
SE	- 0.211***	- 0.292***	- 0.083**
SE incorporated	0.329***	0.372***	0.185
SE unincorporated	- 0.226***	- 0.186***	- 0.347**

Notes: (1) ***: *p*-value less than 0.01; **: *p*-value less than 0.05; *: *p*-value less than 0.10.
(2) Regression include the following explanatory variables: *gender, age, age sq, years of schooling, financial risk, saving habits, number of children, number of children sq, number of resident children, number of resident children sq., number of persons in the household, industry* dummies (16), *education* dummies (10), *type of dwelling* dummies (11) and *state* dummies (8).

The estimates imply that, generally, entrepreneurs earn substantially lower incomes compared to employees. For given characteristics in 2001, the level of reported gross income is approximatelly 36 per cent lower (for men – 37 per cent, for women – 29 per cent)[16]. The estimates of 2003-2005 data confirm latter results. Albeit, in 2005 the difference between self-employed and employees earnings decreases till 19 per cent (for men – 25 per cent, for women – 8 per cent). For self-employed with unincorporated business the negative effect is smaller (in absolute value), thus, it is only significant in 2005. However, self-employed involved in incorporated business appear to have higher incomes. Consequently, the pattern described above provides little evidence for possible self-employed income under-reporting.

Further, instrumented food expenditure regressions of the form of equation 16 are estimated for each year separately. However, before that a heteroscedasticity test run on the latter equation

[16] The estimates in column (1) imply $\exp(-0.451) - 1 \approx -0.36$; in column (2): $\exp(-0.465) - 1 \approx -0.37$; in column (3): $\exp(-0.335) - 1 \approx -0.29$ (Engstrom and Holmlund, 2007).

verify that the errors in the regressions are, as Pissarides and Weber (1989) method requires, heteroscedastic[17] (see Table 16 in Appendix). Hence, the estimation is made under the assumption that its variance takes only two values – one for entrepreneurs and one for wage workers.

Two variables used in IV estimations are income from capital and number of bedrooms. The two instruments are both significant in the first stage regressions, as revealed by their *t*-values (see Table 17 in Appendix). The Sargan test[18] for instrument validity confirm latter results – the null hypothesis that the instrumental variables are uncorrelated to some set of residuals cannot be rejected at 5 per cent significance level, thus instruments are healthy and acceptable (see Table 5).

TABLE 5

Sargan Test Statistics for Validity of Instrumental Variables

Year	Instruments	Sargan statistic	*N*
2001	Income from capital Number of bedrooms	0.960 (0.327)	*563*
2003		2.033 (0.154)	*578*
2004		0.267 (0.605)	*558*
2005		0.076 (0.783)	*653*

Notes: (1) *p*-values are in parentheses.
(2) *N* is a number of observations.

The results of IV regressions are reported in Table 17 in the Appendix. Explanatory variables included in the regressions exhibit patterns that appear reasonable. Age has positive and significant effect, whereas age squared appears to be significant but negative. The number of children as well as total number of persons in the household enter with significant positive impact on food consumption. Moreover, positive saving habits of the household seems to have a negative significant effect. Depending on the year some of the regional variables appear to make significant negative impact on food consumption. While type of housing and industry affect positively. Education level, willingness to take financial risk, number of resident children do not indicate any significant differences in food consumption behavior.

Table 6 presents the key parameter estimates for the calculation of upper and lower bounds of income under-reporting.

[17] Null hypothesis of constant variance is rejected at 10 per cent significance level.
[18] The Sargan test is a statistical test of the over-identifying restrictions. The hypothesis being tested by latter test is that the residuals are uncorrelated with the set of exogenous variables if the instruments are truly exogenous. If the null hypothesis is statistically confirmed (not rejected), the instruments are valid.

TABLE 6

Key Parameter Estimates for Upper and Lower Bounds Estimation

Year	Households headed by self-employed	β	γ	$\sigma^2_{\varepsilon SE}$	$\sigma^2_{\varepsilon PE}$	$\sigma^2_{\gamma SE}$	$\sigma^2_{\gamma PE}$
2001	*N = 422*	0.636 (0.214)	0.178 (0.075)	0.122	0.118	0.737	0.314
2003	*N = 354*	0.284 (0.206)	0.174 (0.061)	0.184	0.127	0.832	0.393
2004	*N = 374*	1.087 (0.638)	0.385 (0.198)	0.124	0.112	0.609	0.376
2005	*N = 380*	0.803 (0.367)	0.106 (0.061)	0.145	0.119	0.471	0.369

Notes: (1) *N* is a number of observations.
(2) Asymptotic standard errors are in parentheses below coefficients estimates.
(3) β (ln reported income) and γ (self-employment dummy) are the coefficients from instrumented eq.16.
(4) $\sigma^2_{\varepsilon SE}$ and $\sigma^2_{\varepsilon PE}$ are residual variances for self-employed and paid-employed households from eq. 16.
(5) $\sigma^2_{\gamma SE}$ and $\sigma^2_{\gamma PE}$ are residual variances for self-employed and paid-employed households from reduced form income regression, eq. 17.

The marginal propensity to consume food β is estimated to vary between 0.284 and 1.087 during 2001 and 2003-2005 years period. The coefficients γ on the self-employment dummy are positive and do not differ in a wide range except in year 2004 the estimate appears to be more than twice higher compared to other years. Both estimates, β and γ, are highly significant in each year. Thus, households headed by entrepreneurs seem to consume more than households headed by wage workers after controlling for net income and houselhold characteristics.

As expected, the residual consumption variance for the self-employed households ($\sigma^2_{\varepsilon SE}$) is always higher than the residual consumption variance for the paid-employed households ($\sigma^2_{\varepsilon PE}$). In a same vein, residual variance of reported income for entrepreneurs ($\sigma^2_{\gamma SE}$) exceeds the residual income variance of wage workers ($\sigma^2_{\gamma PE}$).

Furthermore, estimates of the lower and upper bounds of *k* (see Table 7) are calculated for each year employing the coefficients and residual variances presented in Table 6.

TABLE 7

Lower and Upper Bounds for Mean Under-reporting

Year	Lower bound ($\ln\bar{k}_l$)	Upper bound ($\ln\bar{k}_u$)	Lower bound (value of $\bar{k}_l$)	Upper bound (value of $\bar{k}_u$)	Mean of $\bar{k}$
2001	0.068	0.492	1.07	1.64	1.36
2003	0.393	0.833	1.48	2.30	1.89
2004	0.237	0.471	1.27	1.60	1.44
2005	0.081	0.183	1.08	1.20	1.14

Notes: (1) $\ln\bar{k}_l$ and $\ln\bar{k}_u$ are logarithmic transformations of the mean value of k_i calculated from eq. 13 and eq. 14.
(2) $\bar{k}_l$ and $\bar{k}_u$ are antilogs of $\ln\bar{k}_l$ and $\ln\bar{k}_u$.
(3) Mean of $\bar{k}$ is calculated as follows: $(\bar{k}_l+\bar{k}_u)/2$.

The estimates of lower and upper bounds for 2001 suggest that true income of the households headed by self-employed individuals is higher between 7 and 64 per cent, on average, than reported. Data from 2003 reveal higher bounds – in order to derive true income, reported income has to be multiplied by 1.48 or 2.30. Hence, in the upper bound true self-employed income is more than twice higher compared to reported one. In 2004 entrepreneurs tend to under-report by 27-60 per cent, on average. While in 2005 the range of under-reporting decreases till 8-20 per cent. Implying that reported income must be multiplied by 1.08 in the lower bound and by 1.20 in the upper bound, thus, to derive true income.

Between self-employed households in the sample, approximately 50 per cent runs incorporated business. Entrepreneurs associated with incorporated business are presumably employed by their company and paid by their company as well. Indeed, data reveal that income from wage accounts for a higher share of total income among self-employed households running incorporated business than those involved in unincorporated business.

The results of legal form‘s of self-employment effect on the degree of income under-reporting are given in Table 8.

TABLE 8

Under-reporting by the Legal Form of Self-employment, 2001

	IV REGRESSIONS	
	Dependent variable: Food Consumption ($\ln C_i$)	
Variables	Incorporated business	Unincorporated business
Reported income ($\ln Y_i^r$)	0.837 (2.33)	0.769 (2.72)
SE inc.	0.106 (0.82)	
SE uninc.		0.277 (2.38)
$\sigma^2_{\gamma SEinc(SEuninc)}$ and $\sigma^2_{\gamma SEotherwise}$	0.445; 0.417	0.736; 0.300
k_l and k_u	1.12 – 1.15	1.15 – 1.78
Mean of k	1.14	1.47
t-value for instruments in the first stage	1.79; 1.66	2.33; 1.74
Sargan test	0.367 (0.545)	0.655 (0.419)
R^2	*0.317*	*0.337*
N	*563*	*563*

Notes: (1) Absolute *t*-values based on robust standard errors are in parentheses next to coefficients estimates, except next to Sargan coefficients are *p*-values.
(2) Regression include the following explanatory variables: *gender, age, age sq, years of schooling, financial risk, saving habits, number of children, number of children sq, number of resident children, number of resident children sq., number of persons in the household, industry* dummies (16), *education* dummies (10), *type of dwelling* dummies (11) and *state* dummies (8).
(3) $\sigma^2_{\gamma SEinc(SEuninc)}$ and $\sigma^2_{\gamma SEotherwise}$ are residual variances for self-employed with incorporated, unincorporated business and otherwise employed households from reduced form income regression, eq. 17.
(4) The instruments used are *income from capital* and *number of bedrooms*.

There is a clear pattern in the results suggesting that entrepreneurs involved in incorporated business are less likely to conceal their true income compared to entrepreneurs with unincorporated business. Self-employed running incorporated business under-report from 12 to 15 per cent of their incomes, though these estimates are only marginally significant. However, households headed by self-employed individuals with unincorporated business tend to under-report at the rate of 15 to 78 per cent.

4. Limitations

Current analysis is limited in several respects. First of all, secondary data is used for investigation. The data was not collected for the purpose of this research. The second limitation arises because of a high non-response by self-employed individuals. While interpreting results, it

needs to be taken into consideration that the 20-30 per cent (depends on the year) of respondents who refuses to provide information supposedly engage in under-reporting activity more extensively than those who agree to take part. It is expected that individuals with something to hide, especially those involved in major tax fraud, are more reluctant to respond to the surveys of their income and expenditure. Therefore, the data of current research may under-represent the under-reporting activity.

Moreover, the results of current study represent income under-reporting only by married or cohabiting households with self-employed income. Hence, it should not be interpreted as reflecting a measure of total under-reporting activity. Households with entrepreneurial income but with different characteristics, for instance, single person households, etc., may engage in income under-reporting at variant rate than married or cohabiting households. Consequently, under-reporting estimates must be interpreted and compared in caution.

Finally, the method presented in this paper treats income as independent of the degree of transitory income fluctuations. However, a positive increase of transitory income may enhance income under-reporting by self-employed, which seems probable if entrepreneurs tend to report based on monetary thresholds rather than proportions of true income. Described limitations must be kept in mind while making conclusions regarding the results of current research.

5. Conclusions

This paper explores food expenditure and incomes between self-employed and paid employees of the sample of married or cohabiting households from Australia. In particular, it investigates the extent of entrepreneurial income under-reporting by employing expenditure-based approach developed by Pissarides and Weber (1989). Some evidence is found which support the hypothesis that entrepreneurs tend to under-report their incomes in Australia. More precise, households headed by self-employed individuals conceal approximately 14 to 89 per cent of their true incomes. Furthermore, the results imply that under-reporting is most pronounced among self-employed involved in unincorporated business than those running incorporated business.

Current investigation of self-employed income under-reporting yields results broadly consistent with findings from previous studies. Moreover, it is in accordance with those provided by Pissarides and Weber (1989). This research contributes to the literature by taking a quantitative approach to analyze income under-reporting by self-employed in a different country sample than Pissarides and Weber (1989). However, it is needed a further investigation because this study does not disaggregate under-reporting across demographic characteristics or industry which would shed some light on the determinants of this activity.

The investigation of self-employment income under-reporting is valuable and may illume policy makers about the undergound activities. Furthermore, the evidence that income concealing is likely to depend on the legal form of self-employment is of potential policy relevancy. The details of the regulatory framework of incorporated and unincorporated business might influence the incentive to evade tax. Hence, it should be acknowledged in the policy discussions.

References

Apel, M. (1994) "An expenditure-based estimate of tax evasion in Sweden", in: Apel, M., *"Essays on Taxation and Economic Behavior"*, Economic Studies **18**, Department of Economics, Uppsala University.

Baker, P. (1993) "Taxpayer compliance of the self-employed: estimates from household spending data", *Institute of Fiscal Studies* working paper no. **W93/14**, London.

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

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.

Dilnot, A.W. and C.N. Morris (1981) "What do we know about the black economy", *Fiscal Studies* vol. **2**, 58-73.

Engstrom, P. and Holmlund, B. (2007) "Tax evasion and self-employment in a high-tax country: evidence from Sweden", *Applied Economics, Taylor and Francis Journals* **41 (19)**, 2419-2430.

Feldman, N.E. and Slemrod, J. (2007) "Estimating tax non-compliance with evidence from unaudited tax returns", *The Economic Journal* **117**, 327-352.

Hamilton, B. (2000) "Does entrepreneurship pay? An empirical analysis of the returns to self-employment", *Journal of Political Economy* **108**, 604-631.

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

Joulfaian, D. and Rider, M. (1998) "Differential taxation and tax evasion by small business", *National Tax Journal* **51**, 675-687.

Kesselman, J. R. (1989) "Income tax evasion: an intersectoral analysis", *Journal of Public Economics* **38**, 137-182.

Kim, B., Gibson, J. and Chung, C. (2008) "Using panel data to exactly estimate under-reporting by the self-employed", *Department of Economics* working paper series **15/08**, University of Waikato, Hamilton, New Zealand.

Lewbel, A. (2006) "Engel curves", *Entry for the new palgrave dictionary of economics* vol. **2**, 1-7.

Lyssiotou, P., Pashardes, P. And Stengos, T. (2004) "Estimates of the black economy based on consumer demand approaches", *The Economic Journal* **114**, 622-640.

Mirus, R. and Smith, R.S. (1996) "Self-Employment, tax evasion and the underground economy: micro-based estimates for Canada", *International Tax Program Harvard Law School* working paper 1002, Cambridge.

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

Parker, S. (2009) "The economics of entrepreneurship", in: Parker, S., *"Entrepreneurs' incomes and returns to human capital",* Cambridge: Cambridge University Press, p. 363-385.

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

Quadrini, V. (1999) "The importance of entrepreneurship for wealth concentration and mobility", *Review of Income and Wealth, International Association for Research in Income and Wealth* vol. **45(1)**, 1-19.

Rosen, S. (1981) "The economics of superstars", *A.E.R.* **71**, 845-858.

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

Schuetze, H. (2002), "Profiles of tax non-compliance among the self-employed in Canada: 1969-1992", *Canadian Public Policy* **28**, 219-238.

Slemrod, J. (2007) "Cheating ourselves: the economics of tax evasion", *Journal of Economic Perspective* **21(1)**, 25-48.

Tedds, L. (2004) "Noparametric expenditure-based estimation of income-underreporting and the underground economy", *Department of Economics* working paper **2004:17**, McMaster University.

Thomas, J. (1999) "Quantifying the black economy: measurement without theory yet again", *Economic Journal* vol. **109**, 381–387.

APPENDIX

TABLE 9

Correlation Coefficients Between Food Consumption and Reported Income (Pearson rho), 2001

Variable	Food consumption
Food consumption	1.0000
Reported income	0.2606 0.0000
Reported income (self-employed)	0.2353 0.0000
Reported income (paid-employed)	0.3096 0.0000

TABLE 10

Description of the Variables

Dependent variable		
	Annual food consumption	Continuous variable It is an amount of money spend on groceries and meals eaten outside per year.
Main independent variables		
	Annual reported income (gross)	Continuous variable It is an amount of money generated from wages and salaries as well as from running an enterprise before any taxes were deducted.
	Annual reported income (net)	Continuous variable It is an amount of money generated from wages and salaries as well as from running an enterprise. Net income is given as difference between revenues and taxes.
Control variables		
	Age (squared)	Continuous variable Reveals the age of respondents.
	Gender	Dummy variable 1 – if respondent is male 0 – if respondent is female
	Self-employment	Dummy variable 1 – if respondent is self-employed 0 – if respondent is paid-employee
	Self-employed with incorporated business	Dummy variable 1 – if respondent runs incorporated business 0 – otherwise
	Self-employed with	Dummy variable

	unincorporated business	1 – if respondent runs unincorporated business 0 – otherwise
	Total number of children (squared)	Continuous variable Reveals the number of children.
	Number of resident children (squared)	Continuous variable Reveals the number of resident children. *In year 2003 cannot be constructed due to data limitations.*
	Total number of persons in the household	Continuous variable Reveals the number of persons in the household. *In year 2003 cannot be constructed due to data limitations.*
	Years of schooling	Continuous variable Age at which respondents finished full-time education
	Highest education level achieved	10 dummy variables 1 – post graduate (masters, doctorate); graduate diploma, certificate; bachelor or honours; advanced diploma or diploma; certificate III or IV; certificate I or II; certificate not defined; year 12; year 11 and below; undetermined 0 – otherwise
	Willingness to take financial risk	Dummy variable 1 – if respondent takes substantial, above-average or average risks expecting respectively risks 0 – if respondent is not willing to take financial risks *In year 2005 cannot be constructed due to data limitations.*
	Saving habits	Dummy variable 1 – if respondent saves whatever is left over at the end of the month – no regular plan or spends regular income but saves other income or saves regularly by putting money aside each month 0 – don‘t save *In year 2005 cannot be constructed due to data limitations.*
	Type of dwelling	11 dummy variables 1 – is respondent lives in private house; 1 storey semidetached house; 2 or more storey semidetached house; 1 storey block flat; 2 storey block flat; 3 storey block flat; 4-9 storey block flat; 10 or more storey block flat; attached to a house flat; caravan, tent, cabin, boat house; flat attached to a shop 0 – otherwise
	Job industry	16 dummy variables 1 – if respondent works in mining; manufacturing; electricity; construction; wholesale trade; retail trade; accomodation, cafes and restaurants; transport and storage; communication services; finance and insurance; property and business services; government administration and defence; education; health and community services; cultural and recreational services; personal and other services
	State	8 dummy variables 1 – is respondent lives in New South Wales; Victoria; Queensland; South Australia; Western Australia; Tasmania; Northern Territory; or Australian Capital Territory

		0 – otherwise
Instrumental variables		
	Annual income from capital	Continuous variable It is an amount of money received from renting properties and interest from various sources after all expenses were deducted.
	Number of bedrooms	Continuous variable It is a total number of bedrooms belonging to household members (rooms belonging to any other household in the same dwelling are not counted) including bedrooms which are not currently used.

TABLE 11

Hausman Test Statistics for Endogeneity

Year	Variable	Hausman statistic	*N*
2001	Log net reported income	10.84 (0.0011)	*563*
2003		7.03 (0.0064)	*578*
2004		10.08 (0.0016)	*558*
2005		7.92 (0.0051)	*653*

Notes: (1) *p*-values are in parentheses.
(2) *N* is a number of observations.

TABLE 12

Parameter Estimates from Annual Gross Income Regressions, 2001

	OLS REGRESSIONS		
Variables	Dependent variable: Gross Income ($\ln Y_i$)		
	Both genders (1)	Men (2)	Women (3)
Age	0.089***	0.088***	0.096***
Age sq	- 0.001***	- 0.001***	- 0.001***
Gender	0.539***		
Education	- 0.006	- 0.001	- 0.003
Financial risk	0.183***	0.207***	0.101
Saving habits	0.047	0.086*	- 0.027
Number of children	0.034	0.136**	- 0.284***
Number of children sq	- 0.016	- 0.036**	0.050***
Number of resident children	- 0.127**	- 0.106*	- 0.079
Number of resident children sq	0.032	0.031**	0.009
Number of persons in household	- 0.040**	- 0.003	- 0.113
SE	**- 0.451*** **	**- 0.465*** **	**- 0.335***
SE incorporated	**0.560*** **	**0.550*** **	**0.584*** **
SE unincorporated	**- 0.068**	**- 0.131**	**0.122**
R^2	*0.288*	*0.278*	*0.230*
N	*2147*	*1403*	*744*

Notes: (1) ***: *p*-value less than 0.01; **: *p*-value less than 0.05; *: *p*-value less than 0.10.
(2) Regressions include the following explanatory variables: *industry* dummies (16), *education* dummies (10), *type of dwelling* dummies (11) and *state* dummies (8).

TABLE 13

Parameter Estimates from Annual Gross Income Regressions, 2003

	OLS REGRESSIONS		
Variables	Dependent variable: Gross Income ($\ln Y_i$)		
	Both genders (1)	Men (2)	Women (3)
Age	0.067***	0.057***	0.093***
Age sq	- 0.001***	- 0.001***	- 0.001***
Gender	0.553***		
Financial risk	0.211***	0.236***	0.168***
Saving habits	0.107**	0.112**	0.101
Number of children	- 0.012	0.159***	- 0.362***
Number of children sq	- 0.006	- 0.029***	0.041***
Number of resident children	- 0.118**	- 0.041	- 0.247***
Number of resident children sq	0.029**	0.01	0.039*
Number of persons in household	0.003	- 0.023	0.119**
SE	**- 0.464*** **	**- 0.429*** **	**- 0.522*** **
SE incorporated	**0.444*** **	**0.422*** **	**0.438*** **
SE unincorporated	**- 0.028**	**- 0.004**	**- 0.057**
R^2	*0.338*	*0.302*	*0.276*
N	*1951*	*1219*	*732*

Notes: (1) ***: *p*-value less than 0.01; **: *p*-value less than 0.05; *: *p*-value less than 0.10.
(2) In column (3) education is omitted.
(3) Regressions include the following explanatory variables: *industry* dummies (16), *education* dummies (10), *type of dwelling* dummies (11) and *state* dummies (8).

TABLE 14

Parameter Estimates from Annual Gross Income Regressions, 2004

	OLS REGRESSIONS		
Variables	Dependent variable: Gross Income ($\ln Y_i$)		
	Both genders (1)	Men (2)	Women (3)
Age	0.027*	0.029*	0.077***
Age sq	- 0.001***	- 0.001***	- 0.001**
Gender	0.566***		
Education	0.056***	0.063***	-
Financial risk	0.186***	0.216***	0.121*
Saving habits	0.147***	0.139***	0.174*
Number of children	- 0.028	0.116**	- 0.329***
Number of children sq	- 0.002	- 0.021**	0.041**
Number of resident children	- 0.053	- 0.001	- 0.123
Number of resident children sq	0.010	0.004	0.003
Number of persons in household	- 0.023	- 0.049	0.091
SE	**- 0.447*** **	**- 0.440*** **	**- 0.344** **
SE incorporated	**0.407*** **	**0.389*** **	**0.390** **
SE unincorporated	**- 0.054**	**- 0.126**	**0.043**
R^2	*0.319*	*0.296*	*0.223*
N	*1967*	*1209*	*758*

Notes: (1) ***: *p*-value less than 0.01; **: *p*-value less than 0.05; *: *p*-value less than 0.10.
(2) In column (3) education is omitted.
(3) Regressions include the following explanatory variables: *industry* dummies (16), *education* dummies (10), *type of dwelling* dummies (11) and *state* dummies (8).

TABLE 15

Parameter Estimates from Annual Gross Income Regressions, 2005

	OLS REGRESSIONS		
Variables	Dependent variable: Gross Income ($\ln Y_i$)		
	Both genders (1)	Men (2)	Women (3)
Age	0.095***	0.103***	0.097***
Age sq	- 0.001***	- 0.001***	- 0.001***
Gender	0.540***		
Education	- 0.002	0.002	- 0.008
Number of children	- 0.027	0.064	- 0.245***
Number of children sq	- 0.002	- 0.017	0.041***
Number of resident children	- 0.063	- 0.024	- 0.220
Number of resident children sq	0.003	0.002	0.010
Number of persons in household	0.008	0.018	0.067
SE	**- 0.211*** **	**- 0.292*** **	**- 0.083** **
SE incorporated	**0.329*** **	**0.372*** **	**0.185**
SE unincorporated	**- 0.226*** **	**- 0.186*** **	**- 0.347** **
R^2	*0.273*	*0.242*	*0.181*
N	*2198*	*1343*	*855*

Notes: (1) ***: *p*-value less than 0.01; **: *p*-value less than 0.05; *: *p*-value less than 0.10.
(2) Regressions include the following explanatory variables: *industry* dummies (16), *education* dummies (10), *type of dwelling* dummies (11) and *state* dummies (8).

TABLE 16

Breusch-Pagan/Cook-Weisberg Test Statistics for Heteroscedasticity

Year	Breusch-Pagan/Cook-Weisberg statistic, χ^2
2001	2.73 (0.098)
2003	3.25 (0.063)
2004	2.80 (0.094)
2005	9.63 (0.002)

Notes: p-values are in parentheses.

TABLE 17

Parameter Estimates from IV Regressions

Variables	IV REGRESSIONS			
	Dependent variable: Food Consumption ($\ln C_i$)			
	2001	2003	2004	2005
Reported income ($\ln Y_i^r$)	0.636 (0.214)	0.284 (0.206)	1.087 (0.639)	0.803 (0.367)
Gender	- 0.411 (0.129)	- 0.154 (0.119)	- 0.489 (0.287)	- 0.395 (0.174)
Age	0.025 (0.026)	0.041 (0.019)	-	- 0.056 (0.047)
Age sq	- 0.0002 (0.0003)	- 0.0005 (0.0002)	0.001 (0.001)	0.0006 (0.0004)
Education	0.008 (0.012)	0.023 (0.059)	- 0.079 (0.076)	0.011 (0.035)
Financial risk	- 0.049 (0.059)	0.023 (0.059)	- 0.129 (0.145)	-
Self-employment	0.178 (0.075)	0.174 (0.061)	0.385 (0.198)	0.106 (0.061)
Saving habits	- 0.124 (0.076)	- 0.181 (0.062)	- 0.105 (0.116)	-
Number of children	- 0.075 (0.063)	0.111 (0.035)	- 0.001 (0.077)	- 0.016 (0.051)
Number of children sq	0.015 (0.012)	- 0.013 (0.007)	0.002 (0.015)	- 0.003 (0.010)
Number of resident children	0.088 (0.095)	-	0.085 (0.135)	0.021 (0.089)
Number of resident children sq	- 0.031 (0.019)	-	- 0.002 (0.025)	0.028 (0.019)
Number of persons in the household	0.143 (0.053)	-	0.079 (0.098)	0.114 (0.066)
Region dummies				
New South Wales	- 0.781 (0.482)	- 0.239 (0.185)	0.318 (0.384)	0.316 (0.278)
Victoria	- 0.668 (0.443)	- 0.273 (0.189)	0.202 (0.385)	0.341 (0.289)
Queensland	- 0.719 (0.422)	- 0.315 (0.182)	0.269 (0.392)	0.261 (0.280)
South Australia	- 0.828 (0.444)	- 0.383 (0.185)	0.359 (0.412)	0.214 (0.279)
Western Australia	- 0.639 (0.437)	- 0.355 (0.194)	0.246 (0.395)	0.287 (0.278)
Tasmania	- 0.780 (0.446)	- 0.269 (0.209)	0.326 (0.454)	0.406 (0.345)
Northern Territory	-	-	-	-
Australian Capital Territory	- 0.843 (0.486)	- 0.115 (0.210)	0.699 (0.438)	0.108 (0.286)
Dwelling type dummies				
Separate house	- 0.213 (0.512)	0.576 (0.431)	- 1.368 (1.186)	0.309 (0.618)
Semi-detached house/ row or terrace house/townhouse etc., with one storey	- 0.115 (0.525)	0.635 (0.441)	- 1.352 (1.236)	0.198 (0.602)
Semi-detached house/ row or terrace house/townhouse etc., with two or more storeys	- 0.072 (0.525)	0.658 (0.435)	- 1.353 (1.350)	0.381 (0.611)
Flat/ unit/ apartment:				
In a one-storey block	- 0.095 (0.548)	0.534 (0.457)	- 1.155 (1.054)	0.351 (0.671)
In a two-storey block	- 0.115 (0.538)	0.445 (0.454)	- 1.286 (1.199)	0.307 (0.634)
In a three-storey block	- 0.147 (0.556)	0.867 (0.471)	- 1.729 (1.447)	0.231 (0.606)
In a four to nine storey block	-	0.351 (0.475)	- 1.689 (1.461)	0.112 (0.685)
In a 10 or more storey block	- 0.168 (0.636)	-	- 1.333 (1.490)	0.849 (0.675)
Attached to a house	-	-	- 1.323 (1.332)	- 0.201 (0.813)
Caravan/ tent/ cabin/ houseboat	-	-	-	-
House or flat attached to shop, office, etc.	0.493 (0.586)	0.729 (0.608)	- 1.898 (1.580)	0.631 (0.732)

Education dummies				
Postgraduate – master or doctorate	0.195 (0.329)	0.029 (0.229)	0.008 (0.761)	0.408 (0.410)
Graduate diploma or certificate	0.133 (0.315)	0.099 (0.223)	0.153 (0.779)	0.451 (0.419)
Bachelor or honours	0.294 (0.306)	0.034 (0.217)	0.019 (0.763)	0.430 (0.414)
Advanced diploma or diploma	0.335 (0.308)	- 0.029 (0.223)	0.013 (0.761)	0.592 (0.427)
Certificate III or IV	0.389 (0.304)	0.064 (0.222)	0.197 (0.793)	0.512 (0.426)
Certificate I or II	-	-	0.068 (0.801)	0.593 (0.459)
Certificate not defined	1.069 (0.543)	- 0.165 (0.367)	-	-
Year 12	0.349 (0.305)	- 0.006 (0.219)	0.068 (0.775)	0.466 (0.427)
Year 11 and below	0.332 (0.299)	0.039 (0.224)	0.195 (0.798)	0.557 (0.447)
Industry dummies				
Mining	0.284 (0.420)	0.223 (0.239)	- 0.677 (0.624)	0.692 (0.588)
Manufacturing	0.410 (0.425)	0.324 (0.151)	- 0.353 (0.339)	0.919 (0.586)
Electricity, gas and water Supply	0.326 (0.431)	0.480 (0.191)	- 0.378 (0.345)	0.653 (0.582)
Construction	0.501 (0.427)	0.380 (0.145)	- 0.409 (0.385)	0.905 (0.581)
Wholesale trade	0.385 (0.435)	0.255 (0.181)	- 0.285 (0.329)	0.784 (0.585)
Retail trade	0.599 (0.447)	0.326 (0.132)	- 0.293 (0.311)	0.929 (0.609)
Accommodation, cafes and restaurant	0.448 (0.437)	-	-	1.094 (0.635)
Transport and storage	0.623(0.436)	0.275 (0.159)	- 0.328 (0.315)	0.875 (0.582)
Communication services	0.186 (0.413)	0.099 (0.192)	- 0.395 (0.362)	0.781 (0.582)
Finance and insurance	0.457 (0.411)	0.385 (0.199)	- 0.392 (0.423)	0.519 (0.566)
Property and business services	0.432 (0.425)	0.337 (0.134)	- 0.281 (0.296)	0.899 (0.580)
Government administration and defence	0.321 (0.428)	0.331 (0.152)	- 0.392 (0.359)	0.882 (0.576)
Education	0.487 (0.461)	0.279 (0.156)	- 0.301 (0.315)	0.857 (0.594)
Health and community services	0.379 (0.427)	0.252 (0.152)	- 0.196 (0.253)	0.833 (0.586)
Cultural and recreational services	0.564 (0.472)	0.354 (0.192)	- 0.125 (0.302)	1.159 (0.646)
Personal and other services	0.564 (0.455)	0.407 (0.136)	- 0.294 (0.289)	0.794 (0.601)
t-values for instruments in the first stage	2.78 1.98	1.95 1.98	1.94 1.64	1.74 1.68
R^2	*0.388*	*0.219*	*0.197*	*0.212*
N	*563*	*578*	*558*	*653*

Notes: (1) Standard errors are in parentheses next to coefficients estimates.

(2) The instruments used are *income from capital* and *number of bedrooms*.

(3) Sign "-" marks variables that cannot be constructed due to data limitations or variables that are omitted.