ARE SURVEY-BASED SELF-EMPLOYMENT INCOME UNDER-REPORTING ESTIMATES BIASED?  
NEW EVIDENCE FROM REGISTER AND SURVEY DATA

Ana Cinta G. Cabral* and Norman Gemmell†

September 2018

Abstract
This paper estimates the income-gap (the proportion of undeclared to true income) of the self-employed using traces of expenditure to infer true income holdings following the approach of Pissarides and Weber (1989) and Feldman and Slemrod (2007). This uses the relationship between expenditure and income for the employed (with lower opportunities to evade) to infer the true income of the self-employed. We use a unique dataset from New Zealand that matches individual data including incomes and expenditures in the Household Economic Survey with register incomes declared to the tax administration. This has several advantages in our context including allowing us to measure evasion under different incentives for misreporting, and to assess the impact of measurement error in survey-reported incomes on under-reporting estimates. We find that using register income data leads to robust estimates of income underreporting by the self-employed of around 20% on average. By contrast, estimates are only around half as large when based on survey data. Measurement error in survey-reported incomes would appear to account for much of the difference.

JEL classification: H26,
Keywords: tax evasion, income underreporting, self-employed, measurement error.

Corresponding author: norman.gemmell@vuw.ac.nz
Address: Victoria Business School (SACL), Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand

* Formerly Research Fellow, Inland Revenue, New Zealand.
† Chair in Public Finance, Victoria Business School, Victoria University of Wellington.
1. **Introduction**

Measuring unrecorded or ‘hidden’ income (sometimes referred to as the hidden or shadow economy) is notoriously complicated due to the willingness to conceal it. There are a variety of reasons why income may be received but not recorded, of which avoidance of tax liabilities is an important, but by no means the only, motivation. For tax compliance purposes, some taxpayers’ ability to hide a part of their income means that direct measurement of non-compliance is relatively rare and costly.¹

Surveys of non-compliance run into the caveats of non-response, selection bias and untruthful answers. Intensive audits of tax returns on the other hand, while generally informative, are especially resource intensive. These difficulties of direct measurement of tax non-compliance make it necessary to devise creative ways of tracking its extent such as the observation of expenditure patterns to find traces of true income as initially proposed by Pissarides and Weber (1989) applied to the UK, and more recently by Feldman and Slemrod (2007) for the US.

Pissarides and Weber (1989) used data from the UK Family Expenditure Survey to estimate the amount of non-compliance of the self-employed. The high visibility of employment income, due to the prevalence of third-party information reporting such as PAYE, significantly reduces the scope for employees to underreport their incomes.² Self-employment income on the other hand, being generally self-reported, offers greater opportunities for evasion. Mapping the expenditure and income patterns of employees, and inverting this relationship for the self-employed helps identify the extent of underreporting of the latter.

Using this general framework, estimates of hidden income have been produced for a number of countries largely due to the availability of surveys of family income and expenditures. Importantly however, as highlighted by Slemrod and Weber (2012) and Slemrod (2018), estimates of income underreporting based on survey data rely on the crucial assumption that interviewees report the same income values to the survey and to the tax authority. However, there are several reasons to believe that this might not be the case, which motivates the investigation in this paper.

Firstly, incomes reported to surveys might be subject to omissions of some income sources, or recall errors especially for the group of interest, the self-employed, as their income is more irregular and the accuracy of the reports declines with the length between the occurrence and the reporting. Secondly, the incentives for reporting to the survey and to the tax administration are different. Misreporting income to the tax administration translates into direct tax savings, but risks penalties and reputational damage if identified, for example by a tax audit. There are no such effects from misreporting to the survey however. This has direct implications for the estimation of income underreporting based on survey information: if survey-reported income aligns better with true income, this will result in a downward bias in estimates of income underreporting using purely survey data.

---

¹ An example where direct measurement has been attempted is the US Taxpayer Compliance Measurement Programme (TCMP); See Feldman and Slemrod (2007) for discussion. Non-tax motivations for hiding some incomes include where the income earning activity is illegal (e.g. drugs) or where there are moral/cultural barriers to acknowledging or recording some income sources (e.g. prostitution, even if legal).

² Below we use the terms employee/self-employed rather than employment/self-employment. Employees are those who are hired by a third-party. The self-employed are those employed on their own account. They might also receive a shareholder salary but they still retain discretion over the amount payed (to themselves). The main distinction for our purpose is that self-employment income is not third-party reported and there is discretion over the income declared as opposed to employees where withholding makes income visible.
In this paper we address the criticality of this survey reporting assumption by using a unique dataset that links survey participants in the Household Economic Survey in New Zealand to their tax records for a series of years. This allows us to investigate the consequences of assuming consistent reporting between income sources for estimates of income underreporting in New Zealand. To do so, we use different specifications of survey and ‘register’ income (reported directly to Inland Revenue), and analyse the validity of income reported to the survey by both employees and the self-employed. While the survey reports for third-party employed map closely to their tax register information, this is not the case for the self-employed, with important consequences for the estimated magnitudes of the income-gap using survey or register sourced data.

We find self-employment income is reported with more error than employee income and comparing survey versus register data shows that income underreporting estimates obtained purely from survey data result in a substantial downward bias, in part due to a higher average income reported to the survey than to the register by the self-employed, and in part due to attenuation biases in key parameter estimates. Substantially higher underreporting is uncovered when using the more reliable income measure from the register: an income-gap around 20% versus only 11% when relying solely on survey income data.

A further advantage of our approach is that it combines the richness of survey information that cannot be obtained from a data register perspective with the accuracy and reliability of income from administrative sources. This allows us to investigate the heterogeneity of evasion responses by the self-employed according to different individual and household characteristics, such as gender, age, regional location. Knowing the heterogeneity of hidden income across demographic groups, for example, can help identify the compliance propensities of individual taxpayers – potentially important for tax administration design. Our results suggest that underreporting does not seem to vary with age but males underreport significantly more than females, and urban taxpayers underreport more than rural equivalents. However, differences in legal form of the self-employed (sole trader, director/shareholder, partnership, etc.) – with associated differences in opportunities to underreport, as outlined by Kleven et al. (2011) – are not found to be robustly associated with significantly different underreporting estimates.

The remainder of the paper is organised as follows. Section 2 briefly reviews the literature on the measurement of income underreporting. Section 3 outlines our methodology and its identifying assumptions. Section 4 presents the data and the main variables that feature in the estimation. Section 5 outlines the main results, presents the robustness checks, while section 6 discusses the role of measurement error in the estimation of income-gaps using survey as opposed to register data. Section 7 concludes.

2. Previous Literature

The literature measuring the shadow economy is vast but can be broadly divided into two categories: direct methods that use the individual taxpayer as the compliance decision unit, and indirect, typically more macro-level methods that use proxy relationships to infer the extent of non-compliance. Here we focus on the former micro approach; Slemrod and Weber (2012), Gemmell and Hasseldine (2012, 2014) and Feige (2016) provide broader reviews and summaries of alternative methods. In this section we summarise previous studies that aim to identify traces of non-compliance from taxpayer data.

---

3 The Household Economic Survey is the equivalent in New Zealand to the Living Costs and Food Survey in the UK and the Family Expenditure Survey in the US.
Based on analysing confidential administrative taxpayer data, a number of tax administrations have sought to identify the extent of underreporting of earnings by the self-employed and/or the employed. For the US, for example, the IRS reports underreporting of non-farm incomes by as much as 63% in 2008-10, while in the UK, HMRC (2016, p.50) report a 2014-15 tax-gap of around 14% from self-employed sources. There is no equivalent evidence on the extent of evasion responses for this group in New Zealand using either survey-based or register data.4

Slemrod and Weber (2012) argue that methods that rely on directly measuring traces of non-compliance, including income-expenditure discrepancies, represent a fruitful approach for measurement. Building on Dilnot and Morris’s (1981) first attempt to estimate the size of the black economy in the UK from household expenditure data, Pissarides and Weber (1989) proposed a structured framework for the estimation of underreporting of the self-employed in the UK. Subsequently this ‘traces of non-compliance’ framework has been applied to a number of other countries.

The Pissarides and Weber approach (described in more detail below) measures income underreporting by the self-employed through a comparison of an Engel curve relationship between food expenditure and income for this group to that of employees – who are assumed to be honest reporters. Using the 1982 UK Family Expenditure Survey, Pissarides and Weber (1989) estimated that true self-employment income was on average 1.55 times the income reported by the self-employed, which translated to an income-gap of 36%. They also found underreporting to be higher in blue-collar households than in white-collar households (1.65 versus 1.5).

Using the same approach for 1992, Lyssiotou et al. (2004) estimated an income-gap of 22% on average, also higher for blue-collar households (28%) than white-collar (15%). However, using a complete demand system, they found significantly higher estimates of underreporting, with self-employed blue-collar households underreporting their incomes on average by 54% and white-collar households by 34%. Recent evidence for the UK, from Cabral et al (2016), using the same framework applied to survey data, finds underreporting of around 19% by the self-employed.

Similar studies have been conducted for other countries but care needs to be exercised when directly comparing estimates as the techniques and datasets vary. In Canada, Schuetze (2002), using the equivalent to the Family Expenditure Survey for 1969-1992, found an average underreporting factor of 1.2 (1.12 lower bound and 1.23 upper bound), corresponding to an income-gap of 11% to 19%. More recently, Hurst et al. (2014) for the US estimated self-employment income underreporting in surveys by 25-30%; similar to estimates by Engström and Holmlund (2009) for Sweden. For Finland, Johansson (2005) found that self-employed households with one self-employed individual underreported their incomes by around 17%, but by 42% for households with two self-employed.

The essence of the methodology – using incomes and expenditures in an Engel curve specification – has been applied more widely. For example, Feldman and Slemrod (2007) assume that the relationship between true income and charitable contributions is independent of labour market choices, and estimate the level of underreporting associated with the different sources of income.

---

4 Macro-based estimates put the level of tax evasion in New Zealand between 7% and 11% of GDP (Giles, 1999) and 12% for the period 1999-2006 (Schneider, 2010). However, those macro approaches are known for giving inflated estimates and their use has been widely criticised in the literature (ISWGNA, 2006; Breusch, 2005).
They found income from self-employment in the US to be underreported by an average of 35%. More recently Artavanis et al. (2016), in the same spirit as Pissarides and Weber (1989), use a unique dataset for a large bank in Greece that contains the universe of applications for consumer credit products and mortgages. They invert the relationship between debt and income to estimate income underreporting by the self-employed, finding an underreporting average of 42%-45%.

An advantage of the Feldman and Slemrod (2007) and Artavanis et al. (2016) approaches is that they avoid the potential problems, such as measurement error, associated with the use of survey data. However, as Slemrod (2018, p.24) notes, such studies suffer from the disadvantage that ‘the key assumption – that the conditional charity-income ratio does not vary by employment status – is stronger than the comparable assumption about food’. With access to both survey and register income data, we are able to overcome this more restrictive assumption to estimate Engel curve relationships based on number of food and related expenditure definitions.

Our paper therefore most closely follows that of Pissarides and Weber (1989) and Cabral et al (2016), based on food expenditure-income relationships for employees and the self-employed, which we extend to examine a variety of expenditure categories. As noted earlier, our use of matched register data for several years also allows us to examine the robustness of the more common survey-based underreporting estimates.

3. Methodology

3.1 The Pissarides-Weber Approach

Our approach uses the expenditure capacity for a given level of income of employees and the self-employed to estimate hidden income, based on inferring true income for the self-employed from their expenditure capacity. The model recognises two types of households – self-employed and employed – that, despite being similar in terms of expenditure, differ in their opportunity to underreport incomes.

Letting $Y^D$ be reported (declared) income, and $Y^T$ true income, then for employee households:

$$Y^T_E = Y^D_E$$

(1)

Self-employment income, on the other hand, not being subject to third-party reporting, provides the self-employed with the opportunity of misreporting their earnings, which we specify as:

$$Y^T_{SE} = kY^D_{SE}$$

(2)

where $k$ is a scaling factor by which reported self-employment income needs to be multiplied to obtain true income. The scaling factor $k$ is the coefficient we are interested in estimating. This scaling factor, can be translated into an income-gap, $\kappa$, defined as the proportion of income that is underreported:

$$\kappa = 1 - \frac{1}{k}$$

(3)

Clearly, both variables – the scaling factor and the income-gap – convey the same information expressed in different ways; we focus on income-gap results in Section 5 but report both in Appendix C.

Observing the level of expenditure, $c$, and income, $Y$, of employees gives an indication of the level of income necessary to sustain a particular level of expenditure. That is, for a given level of expenditure, the

---

Note that this survey-based estimate is much lower than the 63% IRS underreporting estimate quoted above for the US. It is unclear how far this is due to the different methods and data used but, as discussed further below, it could be related to a tendency for survey-based estimates to be biased downwards relative to register-based estimates.
The difference between the incomes reported by the employed and by the self-employed give us an estimation of income underreporting.

Graphically, Figure 1 represents this relationship – the Engel curve – for the employed and self-employed. The observed Engel curve of the self-employed is expected to lie above that of the employed. Given a certain level of expenditure, $c$, in Figure 1, the self-employed report an income level, $y^{SE}$, while the employed report income of $y^e$ for the same level of expenditure. The difference between $y^{SE}$ and $y^e$ provides an estimate of income underreporting.

Empirically this can be translated into an estimating equation for household $i$ as:

$$\ln(\ln k) = \beta_0 + \beta_1 \ln \text{Income}_i + \gamma SE_i + \text{Demographics}_i \ast \Theta + \text{Wealth}_i \ast \Lambda + \varepsilon$$

(4)

where $\beta_1$ represents the elasticity of income with respect to expenditure (the slope of the Engel curve), $\gamma$ is the coefficient on a shift dummy variable ($SE = 1$ if self-employed; 0 otherwise) that represents the shift from the employed Engel curve, and $\Theta$ and $\Lambda$ are the coefficients of a vector of household demographic and wealth characteristics respectively.

To capture other important determinants of expenditure levels, (4) includes a vector of household and individual demographic characteristics (number of children, marital status, age and gender) as well as variables that proxy for household wealth. Since wealth measures are not directly observable in our datasets (partly due to the non-taxation of wealth in New Zealand), we seek to capture wealth-related effects on expenditure using two sets of wealth-related or permanent income-related variables, whilst recognising that these are imperfect proxies.

Firstly, we include ‘soft’ survey-reported variables related to housing: type of tenure of the dwelling, number of rooms, the number of stories, type of dwelling, the local housing benefit (Accommodation Supplement) area where the house is located, and the region. Secondly, two variables from the register are used that indicate (i) the annual variability of household income (as a measure of temporary income risk), and (ii) its average growth rate over three years to proxy the stability of household finances. We also conduct extensive sensitivity tests for alternative wealth proxies and capital income measures.

An estimate of the scaling factor can be obtained using the estimated parameters $\gamma$ and $\beta_1$ as:

$$k = \exp\left(\frac{\gamma}{\beta_1}\right)$$

(5)

with the corresponding income-gap computed using equation (3).

A further consideration regarding the estimation of equation (4) is that the measure of income that is expected to influence expenditure decisions is permanent income. However, permanent income is unobservable, and the measure of income we observe is only recorded income (either from the survey or the register). As a result, permanent income is measured with error by both sources. We therefore instrument for recorded income using educational attainment variables and the occupation of the

---

6 See Hurst et al (2014) for a recent empirical application using a similar specification.
7 The accommodation supplement is a transfer from the government to assist people with limited income and cash assets to sustain their accommodation costs. The amount to be received is affected by the area the household is in within the region with higher compensations being received for the North and Central Auckland region. Maximum entitlement rates are only received subject to household circumstances and to an income and asset tests.
3.2 An Extension

The estimating equation can also be modified to allow investigation of the heterogeneity of the income-gap with respect to characteristics of interest. This identifies both the traits that correlate with tax evasion and the types of households most likely to be non-compliant. For each of the characteristics we allow the intercept to vary across employees and self-employed with the same characteristics, hence identifying the ‘pure’ effects associated with self-employment status rather than from heterogeneity in the characteristic itself.

For example, when investigating whether gender affects income reporting behaviour we include intercepts for male and female employees and male and female self-employed. This allows us to identify separate coefficients of underreporting for self-employed males and females. We then test whether gender differences in the estimated income-gaps are significantly different. If so, the characteristic of interest can be signalled as relevant in terms of identifying non-compliance characteristics.

Specifically, we re-write equation (5) as:

\[
\ln \text{Expenditure}_i = \beta_0 + \beta_1 \ln \text{Income}_i + \sum_{n=1}^{N} \gamma_n^{SE} \text{SE}_n I_n + \gamma_n^{E} (1 - \text{SE}_i) I_n + \text{Demographics}_i \cdot \Theta + \text{Wealth}_i \cdot \Lambda + \varepsilon_i
\]

where \( I \) is an indicator for the characteristic of interest (such as age or gender), that has categories \( n = 1, \ldots, N \). The differing intercepts for the self-employed and employed associated with characteristic \( I \) \((N*2)\) can be used to estimate the income-gap as:

\[
\kappa_n = 1 - \frac{1}{\exp \left( \frac{(\gamma_n^{SE} - \gamma_n^{E})}{\beta_i} \right)} \quad n \in N.
\]

A Wald test of the equality of the income-gaps for the different categories of the variable can ascertain whether characteristic \( I \) signals that underreporting significantly varies across the characteristic and thus identifying non-compliers.

3.3 Identifying Assumptions

Before considering results from applying the above approach it is important to be aware of the identifying assumptions on which it relies and associated caveats. There are four key assumptions in particular:

(i) Expenditure is correctly reported by all households on average (but does not preclude the possibility of random measurement error);

(ii) Employees do not underreport their income;

(iii) The (constant) elasticity of expenditure to true income is the same for the self-employed and employees once we control for any confounding factors;

---

8 Individuals are classified into white-collar if they occupy the positions of managers or supervisors and blue-collar otherwise.

9 Ideally, we would like to combine as many characteristics as possible to ascertain which combination of characteristics is differentially associated with underreporting. However, the low number of observations for self-employed households means that dividing them into smaller cells for each defining characteristic results in very low observations per cell making results unrepresentative and/or large regression standard errors.
(iv) The elasticity of expenditure is the same for reported and hidden income.

We discuss each in turn.

(i) **Expenditure is correctly reported by all households.**

Expenditure is the key measure relied on to assess the income capacity of the household and is available only from (HES) survey reports by households. To maximise accuracy, the HES collects expenditure data based on two types of techniques: diary recording and recall questions, with each technique applied to different expenditure items. Recall data has been recognised to suffer from inaccuracy due to some quantities being difficult to remember (Gray, 1955), telescoping errors (Neter and Waksberg, 1964), and progressive amnesia – declining memory with the length of the recall period – (Sudman and Bradburn, 1973; Scott and Amenuvegbe, 1991).

Despite this, recall questions are particularly helpful when the expenditure items of interest are infrequently purchased such as durables; e.g. furniture, household appliances. For more frequent purchases, diaries seek to address accuracy issues by recording respondents’ expenditure on the day it is incurred. In the HES, items such as food expenditure, alcohol, clothing and footwear are recorded in a diary that is kept by members of the household for a period of two weeks.\(^10\)

In the New Zealand case, we have reasons to believe that food and the non-durables basket are reasonably accurately reported. First, because diaries are held only for a limited period of two weeks, this makes ‘diary fatigue’ less of a concern. And as items recorded in the diary are regularly bought, infrequency of purchase is unlikely to substantively affect reports.\(^11\) The basket of non-durables includes a variety of expenditure items that are again recorded in the diary. We exclude from this measure expenditure on alcohol due to established evidence that items such as tobacco and alcohol expenditure are poorly reported in surveys, being sensitive to associated social stigmas.\(^12\) Food expenditure is one of the best covered items of expenditure, with its coverage ranging around 80% in the UK and the US (Brewer and O’Dea, 2012; Meyer and Sullivan, 2010).

Durable goods expenditures (usually recorded via recall questions) are typically less well measured in surveys, and the recognised inaccuracy of this item of expenditure makes it not ideal for inclusion in a dependent variable.\(^13\) Of course, for reliable parameter estimates, measurement errors in the regression dependent variable are a less severe problem than measurement error in an independent variable (coefficient estimates remain consistent but with less precision; see, for example, Pischke, 2007). However, to retain a cleaner and more accurate measure of expenditure, we omit durables expenditures from our dependent variable, but later test for robustness to this definition.\(^14\)

(ii) **Employees do not underreport their income**

---

\(^10\) Diaries, on the other hand, suffer from ‘diary fatigue’ which might affect reports if they pose a high burden on respondents, while short period diaries do not deal well with infrequently purchased items. A mixture of both types of technique are used in the HES to record expenditure data – the approach pursued by Statistical Agencies in a number of countries.

\(^11\) Browning and Leth-Petersen (2003), comparing recall and diary recording of expenditure on food at home for the US, suggest that individuals do a ‘remarkably good job’ when recording food at home as opposed to total expenditure.

\(^12\) The coverage ratio of tobacco and alcohol in the Living Costs and Food Survey in the UK (the equivalent of HES in New Zealand) with respect to the National Accounts is 40%.

\(^13\) The coverage ratio in the Living Costs and Food Survey is variable and ranges from 55%-80%.

\(^14\) Additionally, durables may be used differently between employees and the self-employed as a source of saving or consumption smoothing. For example, the greater volatility of income for the self-employed may encourage greater durable purchases in years of unusually high income.
Wage employees generally have lower opportunities to evade their income, which is typically third-party reported and subject to withholding taxes, thus minimising their scope to underreport. However, there may be cases where wage workers can collude with their employers to negotiate a lower legal wage and receive ‘under the table’ payments. If this is the case, then the estimate of the self-employment income-gap that we observe would correspond to a lower bound estimate. When using employees as a benchmark against which to assess the self-employment income-gap, we are assuming a baseline with low, but not necessarily zero, opportunities for evasion.

(iii) The sensitivity of expenditure to true income is the same for the self-employed and employed.

This assumption is similar to that posed in Feldman and Slemrod (2007) who assume the same relationship between true income and charitable contributions in the US among self-employed and employed. Similarly, Artavanis et al. (2016) assume the credit sensitivity to true income is the same for self-employed and employed. In our case we assume that the pattern of expenditure to true permanent income is unrelated to the selection into self-employment. If there is a different relationship between true income and consumption for both types, then this might be a reflection of hidden risks. In order to control for those risks, we include proxies of income growth and volatility, and asset-related variables.

Another potential source of concern is that some self-employed might be able to treat some personal expenses as business expenses or are able legitimately to claim some business expenses against tax that employees incurring similar expenses cannot similarly offset. In this case the self-employed would appear to have a higher disposable income to spend. The survey explicitly asks about personal expenditure and business expenses in separate parts of the survey. If individuals were treating personal expenses as business expenses, then personal expenditure in the survey on items that can typically be reclassified as a business expense should be lower, *ceteris paribus*, for the self-employed than for the employed. In fact we observe the converse in the data. For example, if the self-employed attribute all of their household fuel expenses to their business, then fuel expenditure of the household would be close to zero, and therefore equivalent employees should have higher household fuel expenditure.

For this reason, we restrict our focus on expenditure categories that are not typically those claimed as business expenses. Food is a small fraction of total business expenses — it can only be claimed if it is an entertainment expense, and usually only 5% of the cost is deductible e.g. if incurred during business trips or promotions. In our broader non-durables measure of expenditure, we include housing costs that would take into consideration utility payments that can be claimed as a business expense. If reclassification of personal spending on utilities as a business was a major feature of the self-employed households then it would not be appropriate to use this type of expenditure in an Engel curve approach to underreporting estimation. In section 5 we test robustness to different expenditure classifications.

(iv) The elasticity of expenditure to hidden and reported income is the same.

This assumption reflects the fact that hidden income has the same capacity to fund spending as reported income. There are reasons to believe that in general this will be the case given that the source of the income is irrelevant to its purchasing power. If, on the other hand, hidden income of the self-employed is more likely to be saved — perhaps to minimise detection via extravagant consumption patterns — then using observed self-employed expenditures would tend to bias downwards estimates of hidden income.

---

15 As in a number of countries, employees in New Zealand are unable to claim business-related expenses even where this may be a legitimate aspect of their employment. In general, this reflects an assumption by the tax authorities that legitimate business expenses in this case would be reimbursed by the employer, though this may not always be the case.
By using such expenditure items as food or non-durables, we would expect to minimise such issues since there is less reason for the self-employed to seek to avoid spending hidden income on this non-conspicuous type of total spending composed of many, relatively low value, individual items.

4. Data

This section describes the administrative and survey data to be used in section 5. The Household Economic Survey (HES) collects information on expenditure and reported income by households in New Zealand and forms the core dataset in the analysis. Taxpayers’ administrative data provide a second source of income reported to the tax administration by each individual. Both data sources are available in Statistics New Zealand’s Integrated Data Infrastructure (IDI), where individuals and households are matched across a number of survey and administrative sources.

4.1 The Household Economic Survey

The Household Economic Survey (HES) collects information on expenditure and income across households in New Zealand. HES is a face-to-face interview where responses are recorded using computer assisted methods. The extended version of the survey that includes a detailed questionnaire on household expenditure is run every three years and an interim shortened version runs in the two intervening years. Each individual in the household is surveyed about income earned in the reference period and expenditure is measured at the household level for a wide array of expenditure categories. This dataset provides the key variable: expenditure of the household; and a measure of household income as reported to the survey, obtained by aggregating income reported by each household member.

4.2 Tax Authority Data

The Inland Revenue Tax Data tables held by Statistics New Zealand within the IDI collect information on income reported to the tax authority. These tables contain the universe of annual income tax returns filed by individuals (Employer Monthly Schedule (EMS) and the IR3 tax return) with anonymised records accessed by authorised researchers using the secure IDI data-lab.

In New Zealand, employees’ income from wages and salaries are withheld (PAYE) and third-party reported by their employers using the EMS. This is a mandatory reporting requirement for all employers with paid employees. The IR data tables are built on the basis of EMS and supplemented with information from the IR3 return. The IR3 tax return is required for individuals who earn income other than salary and wages, dividends and interest and/or taxable Māori authority contributions. Particularly interesting for our analysis are IR3 filing individuals earning self-employment income that can readily be classified into sole traders, director/shareholder of a company, or partners in a partnership.

Register income data also allows us to distinguish a regular EMS payments made by an employer to an employee, from the salary payments that self-employed individuals pay themselves, allowing more accurate classification of individuals into self-employment. Otherwise, directors or shareholders receiving a salary would be misleadingly classified as employees while it is apparent that self-

---

16 Face-to-face has been signalled as an efficient way of minimising non-response and highlighted by the Rockwool Foundation as a desirable survey method to ensure response to sensitive questions (Rockwool Foundation, 2006). Expenditure questions may not in general be considered sensitive but this will apply to certain items of expenditure, e.g. tobacco or alcohol as we discuss below.

17 The types of income and expenditure surveyed in the HES are discussed in more detail in later sub-sections.

18 For further discussion of the IDI see Appendix A.
employment provides them with some discretion regarding the amount of remuneration reported as ‘salary’.

This dataset provides granular information about the individuals in the survey. Unlike previous studies this enables us to: (1) identify the legal form (partners, sole traders, director/shareholders etc.) of the self-employed from a reliable source; (2) observe income as reported to the tax administration, for which there are stronger underreporting incentives in the form of tax savings (but also potentially greater sanctions if discovered); and (3) circumvent potential misclassifications of employment income sources where self-employed individuals pay themselves a salary.

This third aspect is a key issue as the opportunity for underreporting is clearly different when the third-party (‘employer’) that reports income is the self-employed person him/herself compared to when it is a separate individual. It also avoids our analysis relying on survey interviewee’s self-classification into employment and self-employment income sources.

4.3 A Combined Dataset

Respondents to the household economic survey are matched to the register data using a unique identifier assigned by Statistics NZ. Linking the individual in the survey to their tax records allows us to observe longitudinal records on reported incomes to IR. Each individual’s (anonymised) income from the register also needs to be aligned to the income received during the reference period of the survey, since survey interviews are conducted at various points throughout the tax year. Our methods of aligning annual survey and register income data, and tests of alignment accuracy, are described in Appendix A.

To build the dataset we start with the survey. The three-year cycle for the full HES (with an extensive questionnaire on household expenditure) restricts our sample to the years 2006/07, 2009/10 and 2012/13. Interviewees in the survey are matched to their tax records using the unique IDI identifier provided to obtain their income as reported to the tax authorities. We restrict the sample to households where the household reference person (HRP) is in employment and the household receives employment and/or business income. We further restrict the HRP to be below 60 years of age since other studies have found that expenditure patterns vary in retirement; see, for example, Aguiar and Hurst (2005).

Households need to be classified into employee or self-employed households. The availability of register income enables us to distinguish clearly self-employment income sources from employment where the former includes net profit and any PAYE payments or withheld payment received by the sole trader, director/shareholder or partner. We include two alternative definitions of a self-employed household. The first classifies a household as self-employed if it has any income from a self-employed source, and is otherwise treated as an employed household. This definition reflects a household’s opportunity for misreporting self-employment income due to the general absence of third-party reporting.

The second definition classifies a household as self-employed if it derives more than 25% of household income from self-employment; it is classified as an employed household otherwise.19 This second definition, which we refer to as a 25% rule, seeks to avoid misclassification of households as employed

---

19 This definition is applied to register income as it is thought to be a more accurate measure of reported income that can be obtained from a recall question.
when a substantial proportion of household income comes from self-employment sources. It focuses on the weight of self-employment income within household finances.\(^{20}\)

Tax records for interviewees allow us to identify where the self-employed is a sole proprietor, a director/shareholder or part of a partnership, enabling tests of whether different self-employed categories display different underreporting behaviour.

### 4.4 Measures of Income

The survey collects separate information on each income source received by each individual within the household. They can be classified into five groups: labour (employment and business) income; pensions; investment income; welfare benefits; other sources of recurrent income. The main difference between the income sources collected in the survey and those available from the register is that the latter (recorded in IDI tax tables) contain information on taxable income only. Some additional differences in coverage of taxable income in the IDI are discussed below.\(^{21}\)

To account for the difference in coverage between sources we have constructed a mapping (in Appendix B) of income types in the two sources from which we have constructed two measures of income that are comparable across the survey and the register. The first includes employment and business income only. It captures all payments from EMS for employees, net profits and withheld payments received by the self-employed. We refer to it below as Labour Income for short (though recognise that, for the self-employed at least, it may include income best described as a return on capital). The second, which we label as Total Comparable Income, consists of all income types that are reported comparably in both the survey and register. In addition to employment and business income, this includes rental income, taxable benefits, paid parental leave and accident compensation commission (ACC) receipts.

Income is recorded gross-of-tax in the survey and also in the register. However, as the full taxable base is not available to us, we are unable to compute the tax liability to obtain disposable income – without the prospect of introducing more error into the variables. We therefore work with variables in gross terms. Failure to include all sources of household income only matters in our framework if the residual income not included in our measure is held differentially by the self-employed and employed, conditional on their comparable income. In order to test whether there is evidence of this, we compute from the survey a measure of ‘non-comparable income’ – obtained by subtracting the survey income components that are comparable to the register from the total income measure in the survey. We find that, conditional on deciles of comparable income, there is no evidence that the self-employed and the employed differentially hold remaining income not accounted for in our comparable income measure.\(^{22}\)

Self-employment income is often characterised by being more volatile than employment income. This is also observed in our data; see Appendix C. Such volatility of income can be interpreted as a form of income risk, causing current income to deviate from permanent income where the latter is likely to be

\(^{20}\) A concern with this definition, introduced by Pissarides and Weber (1989), is that despite having been used as a strategy to measure the importance of the self-employment source, the rule is applied on the proportion of reported self-employment income to total reported income which differs from ‘true’ income. This definition might therefore entail misclassification of households that underreport their self-employed sources heavily and who, using this rule, are classified as employed, thus mis-measuring the benchmark ‘employees’ group. For this reason, we also explore the alternative ‘opportunity’ based definition.

\(^{21}\) Some sources of taxable income such as dividends and interests are not observable within the IDI. Interests and dividends are subject to withholding taxes and therefore there is no obligation to file a tax return as long as they are withheld at the right tax rate.

\(^{22}\) The main types of income that are non-comparable are investment income, non-taxable benefits and family tax credits; see Appendix B.
more relevant for observed expenditure patterns, and hence should be controlled for. Fortunately, the longitudinal dimension of the register data allows us to control for individuals’ income volatility experienced in the years prior to the survey. We also construct a measure of prior average income growth as a proxy for changes (updates) in expectations of permanent income, and that may subsequently affect expenditure. We construct income risk from the standard deviation of the log of taxable income over the three years prior to the survey and income growth by the average growth of taxable income over the prior years to the survey. Artavanis et al. (2016) follow a similar approach to proxy for local economic conditions and income risk.

4.5 Measuring Expenditure

We create two measures of household expenditure to use as dependent variables: food and non-durable goods.23 Food expenditure is thought to be accurately reported to the survey for several reasons. Firstly, being a necessity, food is an item of expenditure that is less affected by transitory shocks and is not subject to infrequency of purchases. Secondly, it is not an item of expenditure that is associated with a particular lifestyle that non-compliant taxpayers may like to conceal such as expensive holidays or car purchases. Thirdly, expenditure on food is recorded using a two-week diary filled in by survey respondents which ensures more accurate reporting.24 Food has also been shown to be one of the items of expenditure that is better captured in expenditure surveys. Brewer and O’Dea (2012) in the UK, through a comparison of the National Accounts with the Living Costs and Food Survey (LCFS) find that food has a coverage ratio of around 80% for the period 1974-2009. These results compare to the patterns found in the US by Meyer and Sullivan (2009) who find a coverage ratio of 85%. Although there is no equivalent evidence for New Zealand, the similarity of the survey methodology with those in other countries leads us to expect the results to be comparable.

The second expenditure category is a composite measure computed using expenditure on non-durable goods. This basket of non-durables goods is composed of expenditure on food, clothing and utilities, covering a wider spectrum of goods and hence allowing for more heterogeneity. We exclude expenditure on durables as these are potentially affected by problems of infrequency of purchase, telescoping and recall errors as evidenced in the US and the UK by Meyer and Sullivan (2009) and Brewer and O’Dea (2012) respectively.

5. Income Underreporting Estimates

The results presented in this section are obtained from estimation of equation (4). Following Pissarides and Weber (1989) and others, we recognise that annual reported income is likely to contain both transitory and permanent components, where the latter is expected to affect consumption expenditures. We therefore use instrumental variables regressions (IV) with educational attainment and the individual’s occupation used as instruments for reported income. We further control for past income variability and average growth, household and head of the household demographics (age, gender, single/couple, number of

23 Summary statistics of expenditure and income are presented in Appendix C. All income and expenditure variables are deflated to the year 2006 using the quarterly CPI. An alternative specification used the food CPI to deflate food expenditure; results are comparable.

24 We comment further on the assumption of accurate reporting of food expenditure in Section 5.
children), and our previously discussed wealth indicators (type of dwelling, tenure, etc) and region, and controls for each of the three survey years.

Before turning to our underreporting results in detail, Appendix Tables C3 & C4 compare OLS and IV estimates, and reports tests for endogeneity. If, as argued above, income is endogenous this is expected to bias downward estimates of $\beta_1$ in (5). If the self-employment dummy variable, $SE$, is exogenous, then $\gamma$ in (5) would not be biased downwards directly as a result of endogeneity of $\gamma$ but could be biased in either direction indirectly as a result of the bias in $\beta_1$, the bias depending on the covariance between $SE$ and income. If – as we find – any bias in $\gamma$ is relatively small compared to that for $\beta_1$, then from (5) we expect an upward bias in the underreporting variables, $k$ and $\kappa$.

Appendix Tables C3 and C4 confirm that OLS estimates of $\beta_1$ are biased downwards compared to IV estimates (around a third to a half of equivalent IV values) such that underreporting estimates are biased upwards in the OLS cases. The estimated $SE$ dummy parameters appear to be biased downwards in OLS regressions only in the case of the survey data. As a result, using OLS, the upward bias in the underreporting estimate is greater using survey data. In all cases however OLS regressions severely bias underreporting estimates upwards, and Hansen J-test statistics strongly support the hypothesis that OLS estimates suffer from endogeneity. We discuss the implications of measurement error affecting underreporting in IV estimates in section 6.

5.1 Underreporting

Table 1 summarises results for the estimated income-gap of the self-employed using alternative measures of expenditure, income and definitions of a self-employed household; detailed regression results are in Appendix C. Panel A estimates the income-gap classifying the self-employed using the ‘opportunity’ definition – where a household is classified as self-employed if they receive any self-employment income stream even if it is not their main source of income. This ‘opportunity’ definition has two main advantages. Firstly, the direct observation of income sources from IR’s tax register avoids having to classify individuals based on their survey response reports of self-employment and employment income sources.  

Secondly, unlike previous literature, this definition avoids using the magnitude of the self-employment income stream. Previous papers, such as Pissarides and Weber (1989) and Hurst et al. (2014) have used the share of reported self-employment income in total income to classify households into self-employment. However, classifying households based on reported, rather than true, income risks incorrectly assign a household into employment status where self-employment income is substantially underreported.

Columns (1) and (2) of Table 1 respectively show income-gap results when using income measured from the register, and from the survey. Results are reported for each combination of expenditure (food and non-durable) with income (‘labour’ and ‘comparable’) as discussed in section 4.

Table 1 about here

---

25 The US Internal Revenue Service (IRS) in their tax gap reports document that underreporting is concentrated in categories of income with limited information reporting, and underreporting generally decreases across income categories with greater information reporting.

26 Pissarides and Weber (1989) instrument the self-employment dummy variable in order to correct for the misclassification. However, there are no clear instrumental variables to correct for this bias such that weak instruments can introduce a larger bias than not instrumenting.
Using food expenditure as the dependent variable and labour income as reported in the register, we estimate the self-employed underreport on average 20.0% of their income to the tax authority. The estimate is similar (20.4%) when we use the broader set of non-durable expenditure items, demonstrating that using the different expenditure items does not seem to affect the measurement of the income-gap. This is despite the possibility that non-durables expenditure contains some items that could potentially be claimed as business expenses such as housing costs (utilities, rent). If reclassifying such personal expenses as part of the business is important within the data then the estimated income-gap using the non-durables basket should be biased downwards relative to the food-based estimate. That is, the self-employed individual’s apparent personal consumption of non-durables – and hence income estimate – is lower, *ceteris paribus*. However, including non-durables in regressions appears to result in a similar estimated level of underreporting, on average.

We also test whether the measure of income used affects the estimation of income underreporting: lines (2) and (4) report results using the wider income definition which includes rental income, taxable benefits, and other regular payments (paid parental leave, student allowances) in addition to labour income. With values of 19.3% and 19.6%, this demonstrates that the income-gap is consistently estimated across both different expenditure and income measures.

Panel B in Table 1 tests the robustness of the specification to the definition of a self-employed household, by instead applying the 25% rule to household incomes calculated from the register (a household is self-employed if more than 25% of household income comes from self-employment sources). Despite possible misclassification of households due to the rule being based on reported, rather than true, income it can be regarded as a way of refining those households for whom self-employment represents a substantial share of their household finances and hence greater incentive and/or opportunity to underreport. If this is the case we would expect a higher level of underreporting to be observed once households with a small share of self-employment income are deselected. The results in Panel B, column (1) show that the income-gap estimates are robust to this change of definition: with food expenditures, the estimated income-gap is 21.6% (using labour income) and 20.6% (using comparable income) which are close to the results obtained in Panel A. The level of underreporting uncovered is only estimated to be larger when non-durables are chosen as the measure of expenditure: equivalent percentages are 25.4% and 23.9%. Since all of those register-based income-gap estimates in Panel B are larger (to varying degrees) than their Panel A equivalents, there is some evidence here that those with larger opportunities to evade – namely more than 25% of their total income from self-employment – display somewhat greater underreporting than the broader category of self-employed.

Turning to survey-based estimates, while the income components in both the register and the survey are conceptually equivalent, the variables differ in a number of respects, including due to measurement error in the survey variable as individuals are asked to recall the income from the previous 12 months or, for the self-employed, the last time accounts were prepared. We defer discussion of the validity of incomes reported in the survey and its impact on income-gap measurement to section 6. For now, it is sufficient to note that in Table 1 column (2) using income from the survey consistently leads to a lower estimate of income underreporting, typically by around 6-10 percentage points. In addition, the survey based estimates are typically both noisier (slightly larger standard errors) and not always significantly different from zero. Given the prevalence of survey-based estimates in the existing literature, this raises the
important question of whether measurement error especially within survey data may inhibit identification of the full extent of income underreporting.

5.2 Robustness to Identifying Assumptions

Earlier we acknowledged that the method used to identify the magnitude of underreporting depends on a number of identifying assumptions. Below we consider, to the extent that the available data allow, how far those assumptions might influence our income-gap estimates.

(i) Do the self-employed have a preference for eating out?

Though food expenditure might reasonably be thought to be purchased similarly by the employed and self-employed for a given income, one caveat to this is that the self-employed may have a higher propensity to eat food out of the home rather than within (our food expenditure data include both). For example, if the self-employed work longer hours there may be a preference for substituting commercially-prepared, for home-prepared, food. If this is the case, the equivalent food consumption would be more expensive for the self-employed.

To explore this issue, we could use a ‘food eaten in’ variable to replace our food dependent variable in Table 1. However, since food eaten in and out are likely to be close substitutes, this would bias estimation of underreporting as food-in expenditures by the self-employed, for any given level of income, would not be similar to employees due to preference differences rather than income reporting differences. It is therefore better to use total food expenditures whilst recognising that differences in food prices associated with eating in or out may raise expenditure on the latter, other things equal.

In the absence of food price data, to examine this issue – albeit indirectly – we repeat regressions of the form in (4) but where the dependent variable is the share of food eaten out in total food expenditure; see
Table 2. Regression (2) suggests that, as expected, the share of food eaten out (by both groups) is positively related to their income levels, but both regressions (1) and (2) confirm that there is only a weak tendency for the self-employed to spend relatively more on food outside the home – by around 2 percentage points. This is clearly too small to account for the 20% income underreporting estimates above though it could contribute a small fraction.\(^{27}\)

#### (ii) Non-comparable income differences

As noted earlier, we have sought to measure income comparably across the employed and self-employed by restricting our income definitions to ‘labour’ income (= business income for the self-employed) and the broader ‘comparable’ income including rental income, taxable benefits etc. However, it is possible that to the extent that non-comparable income (NCI) – mainly interest and dividends – is held differentially by the self-employed, this could give rise to different food or non-durable expenditures not captured by the earlier results.

For example, if the self-employed have higher investment income, this could give them a higher expenditure capacity that we are otherwise treating as underreporting, whereas it results from mismeasurement of their full income stream. Importantly for our results however, any other income that is excluded from our estimation only matters for biases in income underreporting estimates if it is differentially held by employees and self-employed.\(^{28}\)

To explore this, we re-estimate Table 1 regressions for restricted samples of households where NCI is less than 10%, and less than 25%, of household income, as reported in the survey (the only source of such information). Results are reported in Table 3, which shows income-gap estimates based on food/non-durable expenditure and labour/comparable income.

Table 3 about here

In fact income-gap estimates are little affected by the exclusion/inclusion of households with differing amounts of non-comparable income. For example, using food expenditure/labour income, we obtain slightly lower gap estimates of 0.197 and 0.175 compared to our earlier estimate of 0.20. In general, the greater restriction on non-comparable income (to 10% rather than 25%) leads to income-gap estimates around 2 to 2.5 percentage points lower than with the full sample. That is, the ability of the self-employed to consume more for the same level of comparable income due to higher levels of non-comparable income, at most accounts for a small fraction of the previous underreporting estimates.

Notwithstanding these results, it is important that the income measure used as a right-hand-side variable in these income-gap regressions is specified to be as comparable as possible. This can be seen by substituting total income (comparable plus non-comparable) in Table 1 regressions, though this is only possible for survey-based data. Undertaking this exercise reveals, for example, that, the Table 1 estimate of underreporting using food expenditure/labour income of 0.114, becomes 0.094, and for non-durables

---

\(^{27}\) Regression (3) allows for an interaction dummy variable with income that tests whether the self-employed consume a higher share of food out that is related to their income? This is clearly rejected by the data.

\(^{28}\) Unlike New Zealand, many OECD countries levy social security (SS) taxes on employees’ wages and salaries (paid by employee and employer). This provides an additional incentive in some countries towards self-employment both due to avoidance of employer SS contributions and the exemption of capital income (e.g. dividends) from SS taxation. This is likely to make the ‘non-comparable’ income category in New Zealand somewhat smaller for the self-employed than otherwise.
becomes 0.098 instead of 0.119. Use of total income therefore has a further tendency to reduce the already smaller underreporting estimate obtained from survey data, as opposed to register data.

(iii) Wealth, capital gains and trusts
A possible reason for observing lower reported income for given expenditure compared to employees could be due to wealth effects, where business assets owned by the self-employed enable higher consumption, for example by providing collateral for higher borrowing or allowing higher consumption in the face of a more volatility income stream. Similarly, if the self-employed have greater assets, anticipated accrued or realised capital gains may help fund higher expenditure, and imply higher expected income than that captured in our model. In this case, our estimates of self-employed underreporting would be biased upwards.

As noted above, given limited available wealth and capital gain data, we have sought to control for those effects by including soft variables such as housing characteristics, in addition to income growth and volatility variables. In addition, our use of education level and occupational variables as instruments is designed to eliminate the impact of volatility in annual income – for both the employed and self-employed – and our IV results confirm that this source of endogeneity would seem to be adequately dealt with by this process.

An important further source of assets and income not recorded in our dataset arises when individuals own trusts. In New Zealand many individuals, not simply the self-employed, hold assets in trusts as a form of asset protection and/or tax minimisation; see Tax Working Group (2010). However, if the self-employed make greater use of trusts to channel some personal income, this non-reported (personal) income could be financing higher expenditure by the self-employed.29 A similar argument applies to ownership of rental properties which provide a source of wealth and potential tax-free capital gain income. If these are also especially pronounced among the self-employed they could yield higher expected income to finance higher expenditure compared to employees.30

Unfortunately, income from trusts is not covered in our register data, but whether the individual receives income from a trust, or owns a trust, is indicated in the survey. In Appendix E we therefore use a number of indicators of trust ownership and trust or rental income receipt to test whether this leads to alternative income-gap estimates. Results there confirm that, to the extent that data allow, narrowing the sample to households less affected by potential wealth effects, income-gap estimates remain robustly around 20%.31

It would seem reasonable to conclude therefore that our central income-gap estimates are largely independent of whether or not some self-employed households have access to additional income not recorded in our datasets, or participate in trusts. It is important to note, however, that if some households do have access to such additional income, it may be taxed appropriately under a different tax (e.g. as a

29 Trusts are also frequently used as a means of protecting assets in New Zealand such that ownership of a trust might provide some indication of households’ asset holdings.

30 This issue might be especially pronounced for the incorporated self-employed, as the practice of creating a trust is especially widespread among this group. Dividends are also not separately identified in our data although we do test for its implications in terms of restricting the impact of the non-comparable income which includes the latter. As one of our robustness checks, we repeat the estimation of income underreporting but eliminate the incorporated from our sample. This results in a similar level of underreporting to that obtained for the full sample.

31 Nevertheless, income-gap results for those ‘with capital gains’ yield relatively lower point estimates using food expenditures, and standard errors are generally larger. This would seem to reflect more ‘noise’ in the limited data available here and hence suggest caution in assessing the magnitude of income-gap estimates.
trust or closely-held company) such that the ‘underreported’ person income we observe here need not imply a commensurate underpayment of tax.

A final ‘wealth’ issue we address in Appendix E concerns the possible greater use of savings by the self-employed to finance their expenditures. If the self-employed have more volatile income they may use savings to fund current expenditure when income is temporarily low, or divert income into saving when income is temporarily higher. Again, results in the Appendix for various expenditure/savings proxies confirm underreporting estimates very close to the 20% reported in Table 1 using the same register data.

(iv) Business versus personal expenses

It is well-known that the self-employed may reduce their tax liabilities either through reporting lower gross income or by inflating expenses; the latter especially providing opportunities to the self-employed. If some personal expenses can be claimed as business expenses then total personal expenditure should appear lower for the self-employed for given income, or alternatively ‘true’ income may be higher for the self-employed than would be inferred from their observed expenditure.

In our dataset self-employment income is recorded as net profit; that is, after deduction of any relevant expenses. Hence to the extent that the self-employed underreport gross profit or over-report expense deductions, both have a symmetric effect in our dataset on their observed income (net profit). As a result, our evidence on self-employment ‘income underreporting’ includes both underreported gross income and over-reported deductions.

Nevertheless, it is interesting to consider whether there is any evidence of expenses misreporting. We do so by examining personal expenditure categories that can be expected especially to provide opportunities for misclassification as business expenses. If this occurs by the self-employed, these personal expenses should appear lower for the self-employed for given income compared to items of expenditure for which deductibility is not granted. As a result, if we run regressions using only these business expense categories, instead of food, as the dependent variable, the self-employed should look more similar to employees, yielding lower estimates of underreporting. That is, the Engel curve for the self-employed (which lies above the employee equivalent in Figure 1) should be shifted down when expenditures vulnerable to misclassification are used. Indeed, with sufficient diversion of those expenses from ‘personal’ to ‘business’ the Engel curve for the self-employed could even lie below that of employees; that is, we could in principle obtain negative estimates of underreporting.

In Table 4 we consider those potentially misclassified expenses. In row (1) we examine expenditure on three categories: household utilities (electricity, gas, water etc.), transport and communications (postage, telephone etc.); in row 2 we add housing expenditures. Again using register data it can be seen that the income underreporting estimates (around 8-11%) are indeed much lower than reported in Table 1, around 20%. This evidence therefore suggests that using the Engel curve methodology to identify underreporting is less reliable if the expenditure variable used for this exercise is vulnerable to reclassification as business expenses by the self-employed.

Table 4 about here

---

Footnote: Consistent with the underreporting methodology used throughout this paper, this exercise assumes that reports of such personal expenses to the survey, by households diverting some personal expenses to their businesses, capture only those personal expenditures and not those assigned to their business accounts. For example, if a self-employed sole trader buys a computer entirely from business funds, it is assumed that this is not reported to the HES as personal computer expenses by this individual, even if the computer is partly used for personal purposes.
Specifically, in our case, implicitly (and erroneously) treating these three or four vulnerable expense categories ‘as if’ they were the same across employees and self-employed, continues to generate a substantive income-gap estimate but this is around half as large in Table 4 as that obtained using more comparable expenditure categories as dependent variables. The extent of such re-classification cannot be identified directly from these results however since, as noted earlier, large diversion of personal-to-business expenses could even lead to negative income-gap estimates. It does however suggest that net income underreporting by the self-employed is likely to be at least partially via diversion of those expense types, and is worthy of further investigation.

5.3 Characteristics of Underreporting

In this sub-section we consider the heterogeneity of the estimated income-gap associated with different characteristics of interest that may help inform how those gaps vary across household types. Since this analysis involves various self-employed sub-samples, we focus on the larger sample size obtained using the ‘opportunity’ classification of self-employment income. The analysis uses equation (6), following the same method as in sub-section 5.1, using IV methods to help correct for measurement error in the income variable. Income is measured using information from the register. We present a summary of results below.

Gender and age

One of the main advantages of using a combination of survey and register data is that we can investigate the demographics of non-compliance using variables that are not typically available to the tax administration (since they are not required for tax purposes). Panel A in Table 5 documents our finding that males underreport more than females and this is observed consistently across income and expenditure variables. Note that the specification in equation (6) isolates the effect of gender from the confounding effect of the opportunity to underreport. That is, self-employed males and females are compared to their male and female employee counterparts.

Similar gender effects have been documented in the tax evasion literature based on experiments (Spicer and Becker, 1980) and surveys (Torgler and Schneider, 2007). Recently, Kleven et al. (2011) using a randomised audit experiment also found that being female is negatively associated with non-compliance. This evidence is not however uncontested. Schuetze (2002), for example, found no difference in male/female compliance behaviour, and Baldini et al. (2009), who use discrepancies between survey and register income as a proxy for tax evasion fail to find a gender difference.\(^{33}\)

Non-compliance is also typically found in the literature to be inversely related to age. In our analysis, we created three age brackets splitting the age distribution into: <35 years of age, 35-50 and >50, based on the age of the head of household. We find that the estimated size of the income-gap does not vary with age.\(^{34}\) This is consistent with Kleven et al. (2011) who also fail to find an age effect. Feinstein (1991), using TCMP data for the US, documents the inverse effect of age on compliance only for one of the two years covered by his study.

Regional variation and self-employment legal form

---

\(^{33}\) Gender differences have also been extensively documented and studied in criminology with several theories proposed to explain them; see Mears et al. (1998).

\(^{34}\) This age split yielded approximately equal sized age groups; similar results were obtained when we allowed for alternative age group thresholds.
The spatial variation of the income-gap can be informative about the concentration of the income-gap in certain regions, though we can only examine regional effects at a fairly high level of aggregation to avoid small sub-sample sizes. Nevertheless, this method can identify broad ‘hot-spots’ where non-compliance is concentrated. Based on the six main New Zealand regions, Panel B in Table 5 reports that underreporting appears to be concentrated in Auckland, Canterbury and Wellington. These three urban regions have income-gaps that are not statistically different from each other, but that are significantly higher than in the Rest of the North and South Islands. Clearly there could be many factors underlying these regional results, but they do indicate that underreporting is concentrated in more densely populated and economically active urban regions.

From our register dataset, we can identify whether a self-employed person is registered as a sole trader, a partnership or a director/shareholder of a company. These different legal forms of self-employment could potentially affect non-compliance via the extent of, for example, cooperation among business partners, or the extent of external reporting/oversight such as via requirements for public registering of financial accounts. In addition, the granularity of our register data allows us to observe a sub-set of the self-employed for whom withholding taxes are applied: in New Zealand this applies to contractors within a specified list of occupations.

Since our analysis is performed at the household level, self-employment income is first aggregated into categories for the household. For this estimation, the benchmark is composed of households with no self-employment income and four different self-employment dummy variables relating to each category are introduced. However, we were unable robustly to estimate the impact of different legal forms on underreporting. Using food expenditure, the income-gap was found to vary significantly with the legal form of the self-employed, with underreporting concentrated among sole traders and those that received schedular payments. However, using the non-durables expenditure basket, underreporting was not found to vary significantly in association with the legal form; hence unable to reject the null hypothesis that all legal forms of self-employment underreport to a comparable extent. The lack of consistency of these results suggests caution in interpreting legal form effects and could be attributable to a low number of observations within each category.

6. Measurement Error and Underreporting Estimates

As noted earlier, classical measurement error within income data could potentially yield biased income-gap estimates due to downward ‘attenuation’ bias in estimates of the parameters \( \beta_1 \) and/or \( \gamma \) from regressions such as (4). From equation (5), the direction and size of bias in the income-gap will depend on the relative size of any biases in the two parameters. If, in addition, the size of the error is correlated with income leading to a case of non-classical measurement error, any attenuation bias may be reinforced or counteracted by this correlation. Hence, our income-gap estimates could be biased towards, or away from zero if these measurement error effects are important.

---

\(^{35}\) Further disaggregation into regions is not possible due to small cells. This result (higher underreporting in urban areas) is in contrast to some presumptions that rural areas are more prone to underreporting – for example, because self-employed farming activity dominate rural areas and the personal/business boundary can be hard to monitor. However, urban self-employment occupations – such as taxis, construction, professional services – may provide similar or greater opportunities for underreporting.

\(^{36}\) See Cabral and Gemmell (2018) for classification details. Where a household receives self-employment income from more than one source, e.g. sole trader and director shareholder; we calculate the primary source of self-employment income and classify the household accordingly.
Our results reported below reveal that measurement error has a mean close to zero for employees while being severely biased for the self-employed. The latter report higher incomes on average to the survey than they do to the register which contributes substantially to estimates of income underreporting using survey incomes, yielding a lower estimate than those obtained using the income reported to the register.

We interpret this finding as reflecting the role of incentives for reporting to each income source. For the employed, reporting consistently to the survey and the register is easier due to the high frequency of payments and third-party withholding of their income yielding limited possibility for non-compliance. However, for the self-employed, the time span between the presentation of their tax return and the preparation of business accounts, together with greater opportunities to misreport their incomes, allows larger discrepancies when comparing the register and the survey.

Evidence below also suggests that attenuation biases associated with measurement error appear to be an important source of the lower level of self-employment underreporting when those estimates rely on survey data alone. This suggests the possibility that income underreporting estimates for other countries – which are largely based on survey data – could be substantively downward-biased if measurement error properties are similar to those in New Zealand data. At a minimum, other countries’ survey-based estimates should not be compared directly with register-based estimates for New Zealand, without first assessing the size of possible attenuation biases on estimated parameters for both income and self-employment variables in income-gap regressions.

6.1 Measurement Error and Attenuation Bias

To assess the extent of attenuation biases in our context, consider the ‘true’ Engel curve relationship in (8):

\[ E_i^S = \beta Y_i^R + \varepsilon_i \]  

(8)

where \( E_i \) is reported expenditure by individual \( i \), \( Y_i \) is \( i \)'s income; ‘\( S \)' and ‘\( R \)' superscripts refer to Survey and Register sources respectively, and \( \varepsilon_i \) is a random error term. Both incomes and expenditures are measured in natural logarithms. However, where there is measurement error in observed survey incomes, then:

\[ Y_i^S = Y_i^R + u_i \]  

(9)

Estimating (9) using only survey data implies:

\[ E_i^S = \beta (Y_i^S - u_i) + \varepsilon_i = \beta Y_i^S + (\varepsilon_i - \beta u_i). \]  

(10)

Where there is measurement error in reported survey incomes, and allowing for the possibility that the mean error is non-zero, \( Y_i^S \) can be expressed as:

\[ Y_i^S = Y_i^R + u_i = Y_i^R + \bar{u} + v_i \]  

(11)

where \( v_i = (u_i - \bar{u}) \), \( E(v_i) = 0 \), and \( \bar{u} \neq 0 \). Thus estimating (11) based only on survey data gives:

\[ E_i^S = \beta Y_i^S - \beta \bar{u} + (\varepsilon_i - \beta v_i) \]  

(12)

This reveals two sources of bias: (i) attenuation bias associated with the error term, \( (\varepsilon_i - \beta v_i) \), and (ii) systematic ‘mean’ bias of expenditures by \( \beta \bar{u} \), which are downward biased if \( \bar{u} > 0 \), and upward biased if \( \bar{u} < 0 \). For the standard classical measurement error case where \( Y_i^R \) and \( u_i \) are uncorrelated, the attenuation bias can be summarised by:

\[ \text{plim} \hat{\beta} = \lambda \beta \]  

(13)

\(^{37}\) Of course, \( E_i^S \) may also be measured with error but, as noted above, this reduces the efficiency of estimates of \( \beta \) but does not induce bias.
where \( \lambda = \frac{\sigma^2 \hat{\beta}}{\sigma^2 \beta + \sigma^2 u} \) is the variance ratio or ‘attenuation factor’; see Pischke (2007, p.2). Hence the bias can be given by:

\[
-(1 - \lambda) \beta = \frac{\sigma^2 \hat{\beta}}{\sigma^2 \beta + \sigma^2 u} \beta
\]

(14)

However, if \( Y^R \) and \( u_i \) are correlated – as might be expected if survey income reports for higher (register) income taxpayers are subject to more, or less, reporting error – then it can be shown that (14) becomes:

\[
\text{plim} \hat{\beta} = (1 - b_{uv}s) \beta
\]

(15)

where \( b_{uv}s \) is the estimated coefficient of a regression of \( u_i \) on \( Y^S \). Pischke (2007) shows that larger covariances imply an increase in the attenuation factor, or a decreased bias in estimates of \( \hat{\beta} \), if more than half of the variance in \( Y^S \) is measurement error. Otherwise larger covariances lead to an increased bias in estimates of \( \hat{\beta} \). In our survey data, discussed below, we find a positive covariance, \( \sigma_{uv}s \), and an error variance, \( \sigma^2 u \), consistently less than half the variance of \( Y^S \) (for various income definitions and treatment of outliers).

In addition to the prospect of attenuation bias on our income parameter, \( \hat{\beta} \) (or \( \beta ; \) in the estimating equation (4)), there is also the possibility of biases affecting our \( SE \ dummy \) parameter, \( \gamma \). Attenuation bias could arise from mismeasurement of self-employment status using our two proxies (though we have no way of assessing this)\(^38\), and/or via covariance of self-employment status with income which we pursue in the discussion to follow. While we recognise that some measurement error may be present in the dummy representing self-employment status, the main source of the bias appears to channel through the misreporting of income. In addition, to the extent that there are different mean income errors, \( \bar{u} \), for the employed and self-employed, this mean ‘shift’ effect will be captured within the \( SE \ dummy \) parameter via the term in \( \beta \bar{u} \) in (12). As a result, estimates of the income-gap using (5) could be subject to several sources of bias in \( \gamma \) and \( \beta_1 \), when using survey reported data.

As noted earlier, our interest is not in validating how accurately survey income captures ‘true’ income of New Zealand taxpayers. Rather it is to assess how far survey reports may overestimate incomes reported to the tax authority, since it is the latter that is crucial, together with household expenditure data, in determining the extent to which income is underreported for tax purposes. This approach to underreporting estimation relies on separately identifying Engel curves for employees and the self-employed. Hence, validating incomes separately for both taxpayer groups will be important. However, first we consider the extent of measurement error in survey incomes for the taxpayer sample as a whole.

In line with our earlier analysis, we conduct this validation exercise for both our labour income and ‘comparable income’ measures. We first assess the measurement error for both income variables for the overall sample (including all types of employment). The sample is then split between self-employed and

---

\(^{38}\) The use of self-reported employment status to classify self-employed may be subject to larger measurement error due to for example the misperception or misunderstanding of the difference between self-employment and third-party employee status. Our second definition of self-employment defined as the availability of a self-employment source is less likely to be subject to measurement error, but not free from it. Failure to report a self-employment income source – be it for evasion or failure to take reasonable care – would lead to classification errors in this variable. The definition of self-employment used throughout this exercise has relied on the latter. Given that we cannot fully assess the extent of misclassification in this variable, the two alternative specifications of the self-employed variable were used to estimate underreporting for the purpose of robustness.
employed to analyse measurement error for each group. Finally we study the effect of the measurement error of income on the covariates introduced in the estimation.

This latter step is crucial to understanding the sources of bias in the estimation of underreporting using Engel curves obtained from incomes reported to the survey and the register. Since the estimation of the extent of income underreporting is obtained using both the elasticity of consumption to income (the slope of the Engel curve) and the self-employment dummy variable (intercept), and given the presence of measurement error in survey incomes, the covariance of the measurement error of income with the self-employment dummy variable could impact the ultimate estimated value of the intercept of the Engel curve, and thus affect estimates of the extent of underreporting.

For the sample as a whole, we observe that survey reports are accurate ‘on average’ with the mean of the error, \( u_i \), of 0.01 for labour income, an -0.01 for comparable income, as shown in Table 6, Panel A, column (4). The standard deviation of the error is however quite large, yielding a variance ratio or attenuation factor, \( \lambda \), in column (5) of 24% and 32% for labour income and comparable income respectively.

Table 6 about here

In the absence of correlation between the error and the true value, this is the classical measurement error case, and the variance ratio should equal the coefficient, \( b_{uy}^{\varepsilon} \), obtained from regressing \( u_i \) on survey incomes, \( Y_i \). This regression gives an attenuation factor, measured by \( b_{uy}^{\varepsilon} \) in column (6), lower than the variance ratio in column (5).\(^{39}\)

The bias is therefore estimated to be 0.139 for labour income and 0.280 for comparable income. In both cases these are lower than the variance ratios of 0.242 and 0.318 respectively, due to the negative correlation of the error with the register income value, as documented in column (7). These coefficients show the expected magnitude of the attenuation bias on income parameters from a regression where survey income is used as an independent variable instead of the register measure.

Alternatively, the reliability of the data can be presented using the reliability ratio, obtained as one minus the results in column (6). That is, the reliability ratio equals 0.86 (1-0.139) for labour income and slightly lower, 0.72 (1-0.28) for comparable income estimates, compared to the register. Note however that the bias due to covariates is not considered here.

To test for the robustness of the result for our measurement error estimates, we consider whether this result is derived from the presence of outliers, which can potentially influence mean measurement errors. We follow Bound et al. (1994) and define outliers to be those for which the true reported value, or measurement error, lie four or more standard deviations from the variable mean. The results are displayed in Table 6, Panel B, and it can be seen that the Panel A evidence of non-classical measurement error is not driven by outliers. That is, though parameter magnitudes differ when outliers are omitted, \( b_{uy}^{\varepsilon} \) continues to be significantly positive (\( b_{uy}^{\varepsilon} \) significantly negative).

This non-classical character of the measurement error for income has previously been documented in the income validation literature. Bound et al. (1994), comparing a US company’s records of their employees’

\(^{39}\) Consistently, column (7) shows that the equivalent regression of the measurement error on register (‘true’) income yields a negative correlation (\( b_{uy}^{\varepsilon} \)) between the two. The variance ratio in column (5) is estimated as the variance of the error over the total variance.
earnings to the employees’ self-reports, find evidence of non-classical measurement error in the latter, who estimate biases of between 0.13 and 0.24 – implying reliability ratios of 0.76 to 0.87.\(^{40}\)

Kreiner et al. (2015) report reliability ratios for a measure of gross income from a telephone survey of taxpayers using tax records as a validation dataset.\(^{41}\) They document large biases between the register and the survey with a considerable attenuation bias: the reliability ratio is as low as 0.57. The construction of their income variable for Denmark is closest to our comparable income measure for New Zealand. While we also observe a lower reliability ratio for comparable income, 0.72, this is nevertheless higher than found in Kreiner et al. (2015). However, the low reliability in their dataset seems to be driven by a few outliers in the size of the measurement error. When Kreiner et al repeat the exercise restricting the measurement error window to be ±2 standard deviations (s.d.), this yields a reliability ratio of 0.82. In our case, the reliability ratio when adopting a similar ±2 s.d. control for outliers, increases to 0.83 – very close to the Danish result. Note however that Table 6 refers to both employees and self-employed in our case; while in the Danish case their estimates refer solely to employees.

### 6.2 Heterogeneity of the Measurement Error

The common practice in validation studies has been to eliminate the self-employed, mainly because they are interested in identifying ‘true’ income, and validation requires independent, reliable, preferably third-party reported, information. This is most likely for employees using tax return or employers’ records. The self-employed are generally excluded because their income is self-reported to the register and tax evasion motives and opportunities are greater for this group than among employee counterparts.

We argued above that our objective is different; namely to assess how far estimated income-gaps (based on expenditure-income relationships) from survey data are validated by income-gaps obtained from reports to the tax authority, rather than ‘true income’ (reported plus unreported). Hence we might reasonably treat income-gaps obtained from register (income) data as the ‘gold standard’ for our underreporting method, including both employees and the self-employed, against which survey income data may be assessed.

However, given the acknowledged greater incentive and opportunity for the self-employed, relative to employees, to deliberately underreport their income to the register via their tax returns, it is particularly interesting to examine how survey reports for this group compare to register equivalents. Register-survey differences can be the result of the usual more widely recognised causes of measurement error such as recall mistakes; but evasion also stands out as a likely underlying explanation for the discrepancy. We therefore repeat our validation exercise for the full sample but separating the sample into self-employed and employed households.

Results are reported in Table 7. Of particular note, self-employment income on average is larger in the survey than in the register, resulting in a positive error. The mean of the error is far from zero, at 0.079 for households receiving self-employment income, while the mean of the measurement error is close to zero, −0.01, for households not receiving self-employment income. Measurement error for both

---

\(^{40}\) In an earlier exercise, Bound and Krueger (1991) compared late-1970s earnings data in the US Current Population Survey (CPS) and Social Security payroll tax records, finding reliability ratios of 0.82 (men) and 0.92 (women). Interestingly, these reliability ratios fell substantially when the income data was assessed in first differences (over two years) – to 0.65 (men) and 0.81 (women). Bingley and Martinello (2017), however, find measurement errors for income to be classical in a Danish study (but not for a ‘length of schooling’ variable).

\(^{41}\) They asked respondents to report their gross annual income in 2009, including earnings such as pension contributions and payments, unemployment insurance payments, cash benefits and other forms of transfer income.
household types are non-classical as they are correlated with the register value of the variable in column (7). As a result, the reliability of the data is lower than implied by the variance ratio in (5). That is; $b_{uys}$ in column (6) is less than $(1 - \lambda)$ obtained from column (5).

Also, in Panel A self-employment income appears to have higher reliability than employment income. However, this is affected by some outliers in the data as shown in Panel B. We define outliers in terms of the discrepancy between register and survey values; hence an outlier could be due to tax avoidance as well as resulting from any transcription or other reporting errors.

Controlling for outliers in Panel B continues to identify a large bias in terms of differences between survey and register reports. The estimated bias from a regression on income from the self-employed is estimated at 0.137 for households with self-employment income and 0.034 for households with only employment income. The bias is therefore very low in households where third-party reported income is available. Hence in studies using income from employment as an explanatory variable, the extent of the bias due to the use of survey data would not be expected to be substantial.42 These results are in line with the literature on third-party reported incomes (Bound et al., 1994; Kreiner et al., 2011). Intuitively it is plausible that income sources that are more irregular and not subject to third-party reporting, such as self-employment income, are observed to have lower reliability and a mean further from zero.

In estimating income underreporting, the difference in mean reported income in the survey and the register is a core contributor, since the calculation of the underreporting parameter is based on differences in average income of employed and self-employed households for given levels of expenditure, as captured by differences in the term $\beta \bar{u}$ in (12) for the employed and self-employed. The higher estimate of income in the survey as opposed to the register therefore necessarily affects the estimation of underreporting by making self-employed households appear more similar to employed households given other characteristics in common. The resulting estimated difference in mean expenditure would appear smaller, causing the estimate of income underreporting to be understated. By instead using the register source, we are able to match accurately employed and self-employed expenditures with declared incomes, allowing us to estimate underreporting that reflects the ‘true’ income component that the tax administration fails to observe. Graphically, this means that the Engel curve for the self-employed in the survey appears closer to the Engel curve of the employed than in the register, implying a lower underreporting coefficient being estimated from survey incomes.

6.3 External Covariates

It is possible that some of the unconditional difference observed in mean reported incomes between survey and register is related to differences in the relationship between external covariates in our Engel curve based estimates of underreporting. One way to assess the role of these contributory factors to the size of our measurement error is to observe their impact on those Engel curves. In Appendix D we therefore regress each income measure – from the survey and the register – against a set of covariates relevant to the underreporting analysis. These include: age, gender, housing and household characteristics, geographic location, year dummies etc., and, crucially, the self-employment identifier. Results are shown in Appendix Table D1.

42 Note that mean reversion is higher in self-employment than employment income; this reduces the bias when income is the explanatory variable.
From columns (1) and (2) of the Appendix table, the covariance for most regressors can be seen to be similar across both measures of income (survey or register) though individual coefficients vary: $R^2$s, for example, are close at 0.400 (register) and 0.366 (survey). The most interesting difference for our purpose is between the parameter estimates for the self-employment dummy variable. This is significantly negative for register data ($-0.090$; s.e. = 0.031), but small and statistically insignificant for the survey ($0.009$; s.e. = 0.031). Thus, even when we control for a large set of characteristics, the self-employed continue to report significantly lower incomes to the register compared to the employed, even though similar income levels are reported to the survey for both groups.\footnote{Note that in this regression, the benchmark variable for measuring underreporting, namely expenditure, is omitted. The fact that self-employed and employed with similar characteristics report similar values of income to the survey while lower to the register simply illustrates the upward bias in the survey. To measure underreporting, however, the relevant measure is relative to expenditure capacity, not simply differences between income reports.}

Indeed the conditional impact of self-employment, relative to employees, on income in column 3 of Appendix Table D1, at $(-0.099)$, is very similar to the unconditional income (survey-register) errors between employed and self-employed, around 0.097 (0.087 minus $-0.010$) to 0.125 (0.085 minus $-0.040$), from column 4 of Table 7.\footnote{Table 7 column 4, shows that self-employment (log) labour income is 0.079 higher in the survey compared to the register, while the equivalent value for employees is -0.10. For comparable income the values are 0.085 and -0.04 respectively.}

Returning to our Engel curve underreporting specification in which expenditure is a function of income, the $SE$ dummy and covariates, the impact on underreporting estimates from measurement error within other covariates will depend on the covariance between the error term and these covariates. The positive, significant covariance between the error and the $SE$ dummy in column (3) of Appendix Table D1 confirms that the bias to estimates of the parameter on the $SE$ dummy in a regression using income from the survey, $\hat{\beta}$, remains downwards. That is, estimates of the difference in mean expenditure between the self-employed and the employed, conditional on income (and other covariates) would be smaller than when using register data, thus generating a lower underreporting estimate.

Table 8 about here

Table 8 repeats the earlier income parameters from Engel curve regressions using survey and register incomes, of $\hat{\beta}_S$ and $\hat{\beta}_R$, and shows values of the ratio, $\hat{\beta}_S/\hat{\beta}_R$, below relevant parameters. These ratios are around 0.90 for labour income and 0.80 for comparable income. Comparing this with the downward attenuation bias estimates for $\hat{\beta}_S$ above, suggests that the Table 10 differences between the $\hat{\beta}_S$ and $\hat{\beta}_R$ estimates are dominated by measurement error in the survey income data. Equivalent ratios for the $SE$ dummy parameters, $\hat{\gamma}_S/\hat{\gamma}_R$, are around 0.78 (labour) and 0.67 (comparable) where $\hat{\gamma}_S$ ($\hat{\gamma}_R$) is the estimated parameter on the $SE$ dummy in regressions including survey (register) incomes. The bias in the $SE$ dummy parameter is therefore substantially larger.

How these two biases interact to affect biases in income-gap estimates is not straightforward since the income-gap, $\kappa$, is measured as $\kappa = 1 - (1/k)$, where $k = 1 - \exp(\hat{\gamma}/\hat{\beta})$. To see how these underreporting measures are affected by these biases, consider the parameter estimates in Table 8. For the four combinations of expenditure (food, non-durables) and income (labour, comparable) in Panel A, it can be seen that the survey and register values for $\hat{\beta}$ are quite close – S/R ratios are around 0.81 – 0.96. However, for $\hat{\gamma}$, S/R ratios are much lower at 0.48 – 0.53; that is, larger attenuation biases in the $SE$ dummy parameter.

Panel B shows that, while this generates values of the scaling factor, $k$, that are relatively close (S/R ratios around 0.91), equivalent income-gap S/R ratios are in the 0.57 – 0.63 range. That is, the net outcome for
the income-gap, shown in Panel B, is that survey-based estimates are between a half and two-thirds of equivalents calculated using register data.

Clearly then, the mean error in the survey-based incomes, which downwardly bias \( \hat{\gamma} \) estimates – associated with the term \( \beta \hat{u} \) in equation (12) – have a very large effect on estimates of the income-gap, \( \kappa \). Additionally however, the attenuation biases in estimates of \( \hat{\gamma} \) and \( \hat{\beta} \) also appear to contribute a non-trivial element. Hence, for tax authorities assessing the amount of ‘missing income’ or evasion, the extent to which survey-sourced measures deviate from their register-based equivalents depends heavily on the particular measure adopted. At least in this New Zealand case, the income-gap measure (\( \kappa \)) in particular seems to be substantially mismeasured when survey data are used.

In order to deal with such measurement error, IV estimation has been suggested as a possible method (Pischke, 2007; Bound et al., 2001). In particular, the bias in the estimates could be corrected if we could find an instrument that is correlated with measured income, uncorrelated with the transitory component of income (that is in the error term – in order to solve for the endogeneity problem that affects any estimates of consumption functions), and also uncorrelated with the measurement error. In the context of this research, the higher mean value of income in the survey compared to the register could also be interpreted as an indication of possible unreported income, thus adding another layer to the requirements of the instruments: the instrument used should also be uncorrelated with the underreported income component. It is however hard to find an instrument that could meet all of these characteristics and that is therefore able to correct for the measurement error in the survey data in order to obtain consistent estimates across sources. The use of register information for comparison, when available, therefore seems a reasonable approach to benchmark the estimation of income underreporting using public (survey) sources.

7. Conclusions

This paper has sought to contribute to the literature on the methods of measuring self-employment income-gaps, to assess the role of register and survey data in income-gap estimation, and to use a unique dataset to provide empirical magnitudes of those gaps for New Zealand.

In particular, we used specified expenditure levels within the Engel curve framework of Pissarides and Weber (1989) to infer the true income of the self-employed using employees as a benchmark. We used a unique dataset where survey participants are matched to their (suitably anonymised) administrative tax records to correct for the potential effects of measurement error known to afflict survey responses in the estimation of income underreporting. Using register data, we found that the self-employed in New Zealand underreported on average around 20% of their income. This estimate was found to be robust to alternative specifications of the relevant expenditure and income variables and to alternative definitions of self-employed households. Nevertheless, 95% confidence intervals around this central estimate give an underreporting range of approximately 10-30%.

In New Zealand, it is sometimes claimed that the total income available to some households is not adequately captured in personal income data (whether survey or tax register), due to the relative ease with which households, perhaps especially the self-employed, can earn income through trusts or closely-held companies. Whatever the merits of this claim which is hard to verify from existing data, our evidence suggests it has little impact on the estimated size of personal income underreporting by the self-employed relative to employees. Of course, our data to test for effects on household expenditures is limited by the
partial data available. Nevertheless, underreporting estimates appear robustly similar when we differentiate between those with, and without, access to such income sources.

The opportunity to combine survey and register information is rare but in this case has substantial benefits. Firstly, we were able to correct for the impact of measurement error within survey income reports on estimates of the income-gap. We found that using survey-based income leads to average income-gap estimates as much as 6-10 percentage points lower than when register income is available; that is, the survey-based income-gap can be estimated at up to almost half the value obtained from register data. This appears to be due to both high mean reporting errors by the self-employed and to attenuation biases in survey-based regression parameter estimates due to non-classical measurement error.

Secondly, combining the two data sources enabled us to access a wider set of demographic variables than is typically available for tax purposes. This brings particular insights for compliance policy by identifying characteristics of non-compliant individuals or households. We found that the income-gap varies significantly by gender and region. Ceteris paribus, males significantly underreport more than females and income-gaps are higher in urban regions where population and economic activity are more concentrated. We did not find any significant underreporting effects of age or in association with the specific legal form of self-employment, though the latter result may be due to data limitations.

Given various data and other differences across countries, comparisons of our results with other studies using similar methods to estimate the extent of non-compliance by the self-employed should be treated with caution. However, to the extent that published estimates are comparable, New Zealand would appear to be at the lower end of the range of values found across countries. Our register-based results are similar to underreporting estimates for the UK and Canada as documented by Cabral et al. (2016) and Schuetze (2002) respectively, but less than those obtained for the US (Feldman and Slemrod, 2007) and Greece (Artavanis, et al., 2016). More importantly, since most other countries’ income-gap estimates are based on survey data, their estimates could represent significant underestimates (if similar survey measurement errors are repeated there) and may be more suitably compared with the much lower values obtained here when using New Zealand survey data.
References


**Acknowledgments and Disclaimers:**

The authors would like to acknowledge the contributions of Matt Benge and Sean Comber to previous versions of this paper; participants of the Public Economics Seminar at Victoria University of Wellington, in particular to John Creedy; and Stephen Challands, Caleb Moses and colleagues for the prompt checking and release of the data. The results presented in this study are the work of the authors, not Statistics NZ; they are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI), managed by Statistics New Zealand. The opinions, findings, recommendations, and conclusions expressed in this paper are those of the authors, not Statistics NZ, or Inland Revenue.

Access to the anonymised data used in this study was provided by Statistics NZ under the security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation, and the results in this paper have been confidentialized to protect these groups from identification and to keep their data safe. Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI.

The matching of different data sources on the IDI spine is done by Statistics NZ. These datasets are anonymized thereafter and made available to researchers. Further information on the IDI is provided in Appendix A.
Table 1  Income-Gap Estimates and Robustness to Definitions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>Labour</td>
<td>0.200***</td>
<td>0.114*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.057)</td>
<td>(0.063)</td>
<td></td>
</tr>
<tr>
<td>Food</td>
<td>Comparable</td>
<td>0.193***</td>
<td>0.120*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.048)</td>
<td>(0.062)</td>
<td></td>
</tr>
<tr>
<td>Non-Durables</td>
<td>Labour</td>
<td>0.204***</td>
<td>0.119**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.047)</td>
<td>(0.051)</td>
<td></td>
</tr>
<tr>
<td>Non-Durables</td>
<td>Comparable</td>
<td>0.196***</td>
<td>0.124**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.040)</td>
<td>(0.050)</td>
<td></td>
</tr>
</tbody>
</table>

Panel B: Self-Employment: 25% Rule

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>Labour</td>
<td>0.216***</td>
<td>0.107</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.066)</td>
<td>(0.075)</td>
<td></td>
</tr>
<tr>
<td>Food</td>
<td>Comparable</td>
<td>0.206***</td>
<td>0.111</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.055)</td>
<td>(0.073)</td>
<td></td>
</tr>
<tr>
<td>Non-Durables</td>
<td>Labour</td>
<td>0.254***</td>
<td>0.153***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.053)</td>
<td>(0.059)</td>
<td></td>
</tr>
<tr>
<td>Non-Durables</td>
<td>Comparable</td>
<td>0.239***</td>
<td>0.158***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.045)</td>
<td>(0.057)</td>
<td></td>
</tr>
</tbody>
</table>

Note: This table contains the coefficients of interest, the multiplier and the income-gap. Statistics on the quality of the instruments are provided in Appendix B. Asterisks indicate significance at *** p<0.01; ** p<0.05; * p<0.1.
Table 2  Testing for Differences in Food Preferences

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable: Share of food eaten out in total food expenditure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE Dummy</td>
<td>0.0203***</td>
<td>0.0252***</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td>0.0511***</td>
<td>0.0511***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>SE Dummy*Income</td>
<td></td>
<td></td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01)</td>
</tr>
</tbody>
</table>

Note: The regression uses labour income (survey) and is conditional on the usual set of covariates. Self-Employment Definition: Opportunity. Robust standard errors are in parentheses. Asterisks indicate significance at the following levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 3  Allowing for Non-Comparable Income

<table>
<thead>
<tr>
<th>Self-Employment Definition: Opportunity</th>
<th>(1) NCI&lt;25% Total</th>
<th>(2) NCI&lt;10% Total</th>
<th>Table 1 Full sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expenditure</td>
<td>Income</td>
<td>Expenditure</td>
<td>Income</td>
</tr>
<tr>
<td>Food</td>
<td>Labour</td>
<td>0.197***</td>
<td>0.175***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Food</td>
<td>Comparable Income</td>
<td>0.201***</td>
<td>0.179***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.047)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>Non-Durables</td>
<td>Labour Income</td>
<td>0.197***</td>
<td>0.171***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.041)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>Non-Durables</td>
<td>Comparable</td>
<td>0.201***</td>
<td>0.175***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.038)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>2,376</td>
<td>2,172</td>
</tr>
</tbody>
</table>

Table 4  Testing Robustness to Business Expenses

<table>
<thead>
<tr>
<th>Income type:</th>
<th>(1) Labour (IR)</th>
<th>(2) Comparable (IR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential Expenditure Misclassification:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Utilities, Transport, Communication</td>
<td>0.085</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>(0.086)</td>
<td>(0.067)</td>
</tr>
<tr>
<td>(2) Utilities, Transport, Communication, Housing</td>
<td>0.112**</td>
<td>0.118**</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.046)</td>
</tr>
</tbody>
</table>

Notes: Total expenditure is the broader set of non-durables and durable expenditure of the household. It includes: food, alcohol, clothing, housing and mortgage, communication, transportation, health, recreation and a miscellaneous category. Two alternative specifications of the total expenditure variables are included.

(1) Total expenditure excluding housing and mortgage expenditure to test for the robustness to the impact on the estimation of the potential mismeasurement of rental and mortgage expenses (as a meaningful comparison would require imputing rental costs for owner occupied housing).

(2) Total expenditure also excluding expenditure on durables to test for the impact in frequency of purchase/telescoping errors.
Table 5  Income-Gap Heterogeneity

<table>
<thead>
<tr>
<th>Expenditure category</th>
<th>Food</th>
<th>Non-Durables</th>
<th>Food</th>
<th>Non-Durables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.312***</td>
<td>0.309***</td>
<td>Auckland</td>
<td>0.361***</td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
<td>(0.063)</td>
<td></td>
<td>(0.087)</td>
</tr>
<tr>
<td>Female</td>
<td>0.096</td>
<td>0.108</td>
<td>Canterbury</td>
<td>0.333***</td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.070)</td>
<td></td>
<td>(0.099)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rest of North</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.133)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rest of South</td>
<td>-0.037</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.149)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Wellington</td>
<td>0.230*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.125)</td>
</tr>
</tbody>
</table>

Test of the equality of income-gaps:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi²</td>
<td>4.48</td>
<td>5.62</td>
<td>10.06</td>
<td>8.40</td>
</tr>
<tr>
<td>p-value</td>
<td>0.03</td>
<td>0.02</td>
<td>0.04</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Note: Estimation of these Engel curves to investigate the income-gap heterogeneity follows equation (6). The income-gap is computed as in equation (7). Income-gaps reported in Panel B for the regions of Auckland, Canterbury and Wellington are not statistically different from each other in any of the specifications but they are significantly different from the level of the income-gap in the other two regions: Rest of North and Rest of South. Full results are reported in the Appendix including tests of the equality of the income-gaps. Robust standard errors are in parentheses.

For Tables 6 & 7 – see next page.

Table 8  Survey-Register (S-R) Parameter Differences

<table>
<thead>
<tr>
<th>Dependent var.:</th>
<th>Food Expenditure</th>
<th>Non-Durables Expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data source</td>
<td>Register Survey</td>
<td>Register Survey</td>
</tr>
<tr>
<td>A: Coefficients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income: $\hat{\beta}$</td>
<td>0.460</td>
<td>0.443</td>
</tr>
<tr>
<td>S/R ratio</td>
<td>0.963</td>
<td>0.813</td>
</tr>
<tr>
<td>SE Dummy: $\hat{\gamma}$</td>
<td>0.103</td>
<td>0.0537</td>
</tr>
<tr>
<td>S/R ratio</td>
<td>0.521</td>
<td>0.483</td>
</tr>
<tr>
<td>B: Estimates of underreporting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiplier</td>
<td>1.25</td>
<td>1.129</td>
</tr>
<tr>
<td>S/R ratio</td>
<td>0.903</td>
<td>0.917</td>
</tr>
<tr>
<td>Income-gap</td>
<td>0.200</td>
<td>0.114</td>
</tr>
<tr>
<td>S/R ratio</td>
<td>0.570</td>
<td>0.622</td>
</tr>
</tbody>
</table>

Note: the self-employment definition in Table 5 is ‘Opportunity’: taxpayers with some self-employment income.
Table 6: Summary Statistics of Reporting Errors

<table>
<thead>
<tr>
<th>Earnings Variables</th>
<th>(1)</th>
<th>(2) Mean (Standard Deviation)</th>
<th>(3) Survey</th>
<th>(4) Register</th>
<th>(5) Error</th>
<th>(6) Variance Ratio ($\lambda$)</th>
<th>(7) $b_{\nu R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labour Income</td>
<td>2577</td>
<td>11.036 (0.806)</td>
<td>11.024</td>
<td>0.013</td>
<td>0.242</td>
<td>0.139***</td>
<td>-0.187***</td>
</tr>
<tr>
<td>Comparable: Total Income</td>
<td>2577</td>
<td>11.084 (0.744)</td>
<td>11.092</td>
<td>-0.008</td>
<td>0.318</td>
<td>0.280***</td>
<td>-0.127***</td>
</tr>
<tr>
<td>Panel B: Omit outliers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labour Income</td>
<td>2532</td>
<td>11.041 (0.729)</td>
<td>11.031</td>
<td>0.010</td>
<td>0.152</td>
<td>0.063***</td>
<td>-0.119***</td>
</tr>
<tr>
<td>Comparable: Total Income</td>
<td>2526</td>
<td>11.096 (0.647)</td>
<td>11.095</td>
<td>0.001</td>
<td>0.220</td>
<td>0.175***</td>
<td>-0.093***</td>
</tr>
</tbody>
</table>

Table 7: Summary Statistics of Reporting Errors by Household Type

<table>
<thead>
<tr>
<th>Earnings Variables</th>
<th>(1)</th>
<th>(2) Mean (Standard Deviation)</th>
<th>(3) Survey</th>
<th>(4) Register</th>
<th>(5) Error</th>
<th>(6) Variance Ratio ($\lambda$)</th>
<th>(7) $b_{\nu R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labour Income</td>
<td>663</td>
<td>11.215 (0.769)</td>
<td>11.136</td>
<td>0.079</td>
<td>0.255</td>
<td>0.081***</td>
<td>-0.278***</td>
</tr>
<tr>
<td>No Self-Employment Income &gt; 0</td>
<td>1914</td>
<td>10.974 (0.810)</td>
<td>10.984</td>
<td>-0.010</td>
<td>0.236</td>
<td>0.152***</td>
<td>-0.158***</td>
</tr>
<tr>
<td>Comparable: Total Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labour Income</td>
<td>663</td>
<td>11.252 (0.745)</td>
<td>11.168</td>
<td>0.085</td>
<td>0.315</td>
<td>0.190***</td>
<td>-0.295***</td>
</tr>
<tr>
<td>No Self-Employment Income &gt; 0</td>
<td>1914</td>
<td>11.026 (0.734)</td>
<td>11.065</td>
<td>-0.040</td>
<td>0.317</td>
<td>0.305***</td>
<td>-0.041***</td>
</tr>
<tr>
<td>Panel B: Omit Outliers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labour Income</td>
<td>645</td>
<td>11.192 (0.707)</td>
<td>11.125</td>
<td>0.067</td>
<td>0.253</td>
<td>0.137***</td>
<td>-0.214***</td>
</tr>
<tr>
<td>No Self-Employment Income &gt; 0</td>
<td>1890</td>
<td>10.990 (0.730)</td>
<td>10.999</td>
<td>-0.009</td>
<td>0.111</td>
<td>0.034***</td>
<td>-0.093***</td>
</tr>
<tr>
<td>Comparable: Total Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labour Income</td>
<td>645</td>
<td>11.232 (0.670)</td>
<td>11.151</td>
<td>0.081</td>
<td>0.276</td>
<td>0.159***</td>
<td>-0.238***</td>
</tr>
<tr>
<td>No Self-Employment Income &gt; 0</td>
<td>1881</td>
<td>11.050 (0.633)</td>
<td>11.075</td>
<td>-0.026</td>
<td>0.184</td>
<td>0.173***</td>
<td>-0.019*</td>
</tr>
</tbody>
</table>