Are Survey-Based Self-Employment Income Under-Reporting Estimates Biased? New Evidence from Matched Register and Survey Data

Ana Cinta G. Cabral, Norman Gemmell and Nazila Alinaghi

WORKING PAPER 07/2019
November 2019

Working Papers in Public Finance

VICTORIA UNIVERSITY OF WELLINGTON
TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

Chair in Public Finance
Victoria Business School

The Working Papers in Public Finance series is published by the Victoria Business School to disseminate initial research on public finance topics, from economists, accountants, finance, law and tax specialists, to a wider audience. Any opinions and views expressed in these papers are those of the author(s). They should not be attributed to Victoria University of Wellington or the sponsors of the Chair in Public Finance.

Further enquiries to:
The Administrator
Chair in Public Finance
Victoria University of Wellington
PO Box 600
Wellington 6041
New Zealand

Phone: +64-4-463-9656
Email: cpf-info@vuw.ac.nz

Papers in the series can be downloaded from the following website:
https://www.wgtn.ac.nz/sacl/centres-and-chairs/cpf/publications/working-papers

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

Ana Cinta G. Cabral[*], Norman Gemmell[†] and Nazila Alinaghi[†]

November 2019

This paper is a revised version of *Working Papers in Public Finance*, No. 07/2018

Abstract

Estimates from studies of income underreporting (the proportion of undeclared to true income) by the self-employed using the 'Engel curve' approach of Pissarides and Weber (PW, 1989) have been based predominantly on *survey data* on incomes and expenditures. This paper uses a unique dataset, from New Zealand, that matches survey data on household incomes with administrative tax register data for the same households. This allows us to measure evasion under different incentives for misreporting – official tax returns and an independent statistical survey – and to quantify the impact of measurement error in survey-reported incomes on under-reporting estimates. We find that using tax return 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. This result reflects both measurement error in, and attenuation biases arising from, survey-reported incomes. The former appear to account for much of the difference. If self-employed survey reporting in other countries demonstrates similar differences from equivalent tax records – as seems likely *a priori* – then many previous estimates of self-employment income under-reporting based on the PW approach may be biased downwards.

JEL classification: H26

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

Address: Victoria Business School (SACL), Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand

* Formerly Research Fellow, Inland Revenue, New Zealand.

† Victoria Business School, Victoria University of Wellington.

[Ψ] We are grateful to participants at the International Institute of Public Finance Annual Conference, Glasgow, August 2019, and especially our discussant, Amin Mawani (York University, Toronto) and Mazhar Waseem (University of Manchester, UK) for helpful comments on an earlier draft.

1. Introduction

Measuring unrecorded or 'hidden' income is notoriously complicated due to the willingness to conceal it, with an important, but by no means only, motivation for concealment being avoidance of tax liabilities. Direct measurement of this hidden income, such as via taxpayer audits, is relatively difficult and resource intensive.[1] Indirect measurement of non-compliance via taxpayer surveys run into familiar problems of non-response, selection bias and untruthful answers. These measurement difficulties make it necessary to devise creative ways of tracking the extent of tax non-compliance such as by observing consumers' expenditure patterns to find 'traces' of true income as initially proposed by Pissarides and Weber (PW, 1989) in a UK study, and subsequently applied to a variety of countries by numerus studies.

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 now familiar PW framework, estimates of hidden income have been produced for a number of countries typically based on similar surveys of household incomes and expenditures. Recent challenges to applications of this framework have focussed on such aspects as: how far alternative instruments for permanent income (an important component of the PW model) succeed in removing transitory components (Nygard *et al.*, 2018; Engström and Hagen, 2017; Hurst *et al.*, 2014); the role the public versus private sector employees as the basis for identification of self-employment non-compliance (Paulus, 2015a); and how 'self-employed' households are identified (Kukk and Staehr, 2017).

An important aspect of most PW applications, highlighted by Slemrod and Weber (2012) and Slemrod (2018) but largely untested, concerns the crucial assumption that survey interviewees report the same income values to the survey and to the tax authority. In this paper, with access to unique 'matched' survey and register (tax return) data available for New Zealand, we are able to test this assumption. This also allows us to examine carefully the impact of measurement error on income gap estimates from each data source.

There are several reasons to expect that incomes reported to surveys and tax authorities might differ. Firstly, incomes reported to surveys may 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 reporting.

Secondly, the incentives for reporting to a survey and to the tax administration are different. Misreporting income to the tax administration translates into direct financial savings, while risking penalties and reputational damage if identified, for example by a tax audit. There are no such financial effects from misreporting to a survey however, at least when that survey is conducted by an independent statistical

[1] An example where direct measurement has been attempted is the US Taxpayer Compliance Measurement Programme (TCMP); See Feldman and Slemrod (2007) for discussion.

[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 and third-party reporting makes income visible.

agency. Of course, survey participants may nevertheless fear that their survey-reported income levels are discoverable by the tax authority; in which case they have a greater incentive to report income more honestly to the survey.

This has direct implications for the estimation of income underreporting based on survey information. If survey-reported incomes align better with true income, this will result in a downward bias in estimates of income underreporting using purely survey data. On the other hand, if survey participants regard their survey-reported income as potentially discoverable by the tax authority, they may similarly under-report this income, thus satisfying the Slemrod (2018) 'crucial assumption' as pointed out above. However, *a priori,* we might expect survey-reported incomes to differ from those reported to tax authorities. At a minimum, it is important to test such an implicit assumption.

We address this critical 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.[3] This allows us to investigate the consequences of assuming consistent reporting between income sources for estimates of income underreporting. To do so, we examine different specifications of survey and register income, and analyse the validity of income reported to the survey by both employees and the self-employed. We find that, while the survey reports for third-party employed map closely to their tax register information, this is not the case for the self-employed. This has important consequences for the estimated magnitudes of the income-gap using survey or register sourced data.

Our approach to comparing income reports in the two sources follows the established 'validation studies' found in the labour market literature – where employer records have been used to 'validate' employee self-reports; see for example, Bound *et al*. (1994, 2001), Paulus (2015b), Kreiner *et al.* (2015). However, there is an important point of difference in our case. Traditional validation studies have been concerned to identify 'true' income and focus on *employees only* – given uncertainties over the truthfulness of self-employed income reporting, as stressed by Kreiner *et al.* (2015). Our objective, however, is not to validate 'true' income but rather to validate the extent of *under-reported income* by the self-employed *relative to employees*, for which tax authority income records can be regarded as the 'gold standard'. Hence, we are able to 'validate' survey-based estimates of self-employment underreporting against this benchmark.

We find that self-employment income is reported with more error than employee income. Comparing survey versus register data shows that income underreporting estimates obtained purely from survey data result in a substantial downward bias. We further establish that this is largely due to higher average income reported to the survey than to the register by the self-employed, and to a lesser extent due to measurement error-induced attenuation biases in key regression 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. Hence, a greater degree of honesty by the self-employed when reporting income to official income-expenditure surveys compared with income reports in tax returns appears to be systematic, generating systematically lower underreporting estimates using the former.

[3] The Household Economic Survey in New Zealand is the equivalent of the Living Costs and Food Survey in the UK and the Family Expenditure Survey in the US.

A further advantage of our approach is that it combines the richness of survey information that cannot be obtained from a data register perspective (such as various household economic and demographic characteristics) with the accuracy and reliability of income from administrative sources. This allows us to investigate the heterogeneity of reporting responses by the self-employed across different individual and household characteristics, such as gender, age and regional location. Knowing the heterogeneity of hidden income across demographic groups, for example, can help identify the compliance propensities of particular taxpayer types, potentially important for tax administration design.

In summary, our paper differs from the previous literature in that it tests one of the fundamental premises of the PW approach: the alignment of survey data and tax records (Slemrod, 2018). Previous studies using administrative and survey data have either simply replaced survey-reported income with administrative incomes (Engström and Hagen, 2017), or have not explored the nature of the measurement errors associated with each dataset and which lead to different underreporting estimates (Paulus, 2015a). We carefully examine the implications of measurement error for estimates of underreporting derived from survey and register data. This is important both for interpretation of evidence from prior studies and for future estimation of non-compliance where tax records are not available. The present study therefore contributes to the literature on income measurement error by identifying the characteristics of *self-employment* income error and its relevance for tax non-compliance measurement.

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 compared to register data. Finally, 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.[4] Relevant antecedents of the present paper are studies seeking to identify traces of non-compliance from individual taxpayer data.[5]

Until recently, most researchers outside tax authorities have not had access to the kinds of resources or data available to, for example, the (US) IRS, (UK) HMRC, etc. to measure non-compliance. [6] Slemrod and Weber (2012) therefore 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 shadow economy in the UK from

[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).

[5] Slemrod and Weber (2012), Gemmell and Hasseldine (2012, 2014) and Feige (2016) provide broader reviews and summaries of alternative methods.

[6] With access to 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.

household expenditure data, the structured framework proposed by Pissarides and Weber (1989) to estimate 'traces' of underreporting by the self-employed in the UK has subsequently been applied to a number of other countries. Predominantly, however, these estimates are based on survey-reported incomes and expenditures.

The Pissarides and Weber (PW) approach 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, in part due to the greater prevalence of third-party sourced information for this group. Using the 1982 UK Family Expenditure Survey, PW (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). Results from several subsequent studies using this framework are summarised in Table 1; Kukk *et al.* (2019) also provide a range of estimates for a large number of European Union (EU) countries, using a common PW-based approach.

Table 1 about here

As can be seen in the table, previous results using food expenditures have found a wide range of income underreporting from a low of 11% in Canada to 56% in Estonia, while Kukk *et al*. (2019) find a range of estimates from under 10% to over 40% for their 14 EU countries. Other studies have taken the essence of the PW 'food expenditure' methodology and applied it more widely. 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. They avoided survey-related problems by using tax return data on incomes and charitable contributions, and found US income underreporting by the self-employment on average of 35%.[7]

More recently Artavanis *et al.* (2016) used 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 as high as 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 and matched register data for several years. This allows us to examine the robustness of the more common survey-based underreporting estimates, which establishes the importance of measurement error with such data. The only previous study of which we are aware that has attempted

[7] 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.

to use both survey and register data to measure self-employed underreporting is Paulus (2015a) for Estonia. In a result that echoes our findings, he finds that survey-based income reports generate substantially lower underreporting estimates than register-based estimates, especially when the comparator group is *public sector* employees.[8]

3. Methodology

3.1 The Pissarides-Weber Approach

Since the PW approach is well-established it is summarised briefly here. The 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:

$$Y_E^T = Y_E^D \tag{1}$$

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:

$$Y_{SE}^T = kY_{SE}^D \tag{2}$$

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, κ, defined as the proportion of income that is underreported:

$$\kappa = 1 - \frac{1}{k} \tag{3}$$

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.[9] The baseline for estimation of the self-employment income-gap is the relationship of reported income to expenditure by employees. To the extent that employees also benefit from tax minimisation opportunities, the impact of tax avoidance in the estimate will only capture the differential use by the self-employed of these strategies. These issues are discussed further in section 5.2.

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 difference between the incomes reported by the employed and by the self-employed give us an estimation

[8] However, the Paulus (2015a) study is restricted to housing expenditures such as utilities as the PW expenditure category which, in our New Zealand case, are shown to be strongly affected by the ability to report these under business expenses instead of as private household spending. See Paulus (2015a, pp. 15-16) for arguments regarding the Estonia case. Waseem (2019) provides an interesting alternative approach to identifying self-employment tax evasion responses in Pakistan, comparing responses associated with 'to-zero' tax rate cuts with responses to similar sized but 'not-to-zero' tax cuts.

[9] The income-gap measured by κ captures the deviation of reported income by the self-employed from their 'true' income. These deviations from true income could be via legal tax avoidance mechanisms allowed by the tax system, or illegal tax evasion. The share of legal tax avoidance captured in the estimate will therefore depend on the extent to which tax minimisation opportunities are available specifically to the self-employed or when legal avoidance schemes are more readily exploited by them. Underreported income therefore does not necessarily imply 'underreported tax'.

of income underreporting. Hence the observed Engel curve of the self-employed is expected to lie above that of the employed. The difference between Y_{SE} and Y_E provides an estimate of income underreporting.

Empirically the Engel curve can be translated into an estimating equation for household *i* as:[10]

$$\ln Expenditure_i = \beta_o + \beta_1 \ln Income_i + \gamma SE_i + Demographics_i * \Theta + Wealth_i * \Lambda + \varepsilon \quad (4)$$

where β_1 represents the elasticity of income with respect to expenditure (the slope of the Engel curve), γ 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 Θ and Λ 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), proxies for household wealth to capture wealth-related effects on expenditure. In addition, we construct two variables from the register 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 and capital income proxies.

An estimate of the scaling factor can be obtained using the estimated parameters γ and β_1 as:

$$k = \exp\left(\frac{\gamma}{\beta_1}\right) \quad (5)$$

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 annual recorded income (either from the survey or the register). As a result, even if annual income is reported without error, permanent income is measured with error in both sources. We therefore instrument for recorded income using educational attainment variables and the occupation of the household head: whether 'white-collar' or 'blue-collar'.[11] Equation (4) is therefore estimated using two-stages least squares with OLS equivalents reported in Appendix C.[12]

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 income underreporting 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.

[10] See Hurst *et al.* (2014) for a recent empirical application using a similar specification.
[11] Individuals are classified into white-collar if they occupy the positions of managers or supervisors and blue-collar otherwise.
[12] Kukk and Staehr (2017) and Nygard *et al.* (2018) pursue alternative approaches to estimation of permanent income, based on a survey measure of 'regular income' (Kukk and Staehr) or an average of seven years income data (Nygard *et al.*). Nygard *et al.* (2018, p.1905) show that underreporting estimates are larger when this permanent income measure is used compared to estimates based on annual data. We find that traditional instruments for permanent income (education and occupation variables) perform well but we also include our controls for income stability and volatility in all regressions.

For example, when investigating whether gender affects income reporting behaviour intercepts are included for male and female employees *and* male and female self-employed. This allows identification of 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.[13]

Specifically, we re-write equation (4) as:

$$\ln Expenditure_i = \beta_o + \beta_1 \ln Income_i + \sum_{n=1}^{N} \gamma_n^{SE} SE_i I_n + \gamma_n^{E} (1 - SE_i) \mathrm{I}_n + Demographics_i * \Theta + Wealth_i * \Lambda + \varepsilon_i \quad (6)$$

where *I* is an indicator for the characteristic of interest (such as age or gender), that has categories *n = 1, …, 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:

$$\kappa_n = 1 - \frac{1}{k_n} \qquad n \in N \quad (7)$$

where $k_n = \exp\{(\gamma_n^{SE} - \gamma_n^{E})/\beta_1\}$. 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, 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;
(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).

[13] 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.

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.[14]

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.[15] 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.[16] 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.[17] 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 omission.[18]

(ii) Employees do not underreport their 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. Paulus (2015b) suggests this may be the case in Estonia, for example. If this is the case, here then the estimate of the self-employment income-gap that we obtain 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.[19]

[14] 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.

[15] 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.

[16] 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%.

[17] The coverage ratio in the Living Costs and Food Survey is variable and ranges from 55%-80%.

[18] 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.

[19] Paulus (2015a) finds a larger income-gap when comparing self-employment incomes to public sector employees rather than private sector employees.

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

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 for multi-year income growth, annual income 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 HES 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 50% 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 encompass utility payments, which can be claimed as a business expense. If reclassification of personal spending on utilities as a business expense 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 of the results 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. 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 expenditure 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. Every three years the survey includes a detailed questionnaire on household expenditure. 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.[20] This provides the key variable: expenditure of the household; and a measure of household income as reported to the survey, by aggregating reported incomes of each household member.

4.2 Tax Authority Data

The Inland Revenue Tax Data tables, held within the IDI, collect information on income reported to the tax authority.[21] These contain the universe of annual income tax returns filed by individuals Employer Monthly Schedule (EMS), and the IR3 tax return) suitably anonymised.

Employees' income from wages and salaries are withheld (PAYE) and third-party reported by their employers using the EMS. The IR data tables are built on the basis of the EMS and supplemented with information from the IR3 return. This return is required for individuals who earn income other than salary and wages, dividends and interest. 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.

This dataset 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-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 financial savings; 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 entity. It also helps to avoid a reliance on survey interviewee's self-classification into employment and self-employment income sources in our analysis.

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, allowing us to observe longitudinal records on reported incomes. Each

[20] The types of income and expenditure surveyed in the HES are discussed in more detail in later sub-sections.
[21] For further discussion of the IDI see Appendix A.

individual's (anonymised) income from the register also needs to be aligned with 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. 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. 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 absence of third-party reporting for some income.

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.[22] This second definition, which we refer to as a *25% rule*, seeks to avoid misclassification of households as employed when a substantial proportion of household income comes from self-employment sources. It focuses on the *weight* of self-employment income within household finances.[23]

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. For Estonia, Kukk and Staehr (2017) find that the way in which self-employed households are identified in their analysis, together with the household's share of self-employment income, are important for the magnitude of income underreporting estimates. We therefore complement our 'two definitions' approach above with tests of the effects of allowing for a self-employment income share in underreporting regressions; see Section 5.1.

4.4 Measures of Income

The survey collects separate information on each income source received by each individual within the household. Income sources 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

[22] 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.

[23] A concern with this definition, introduced by PW (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.

contain information on *taxable* income only. Some additional differences in coverage of taxable income in the two datasets are discussed below.[24]

To account for the difference in coverage between sources we constructed an income mapping (in Appendix B). This mapping was essential to construct two equivalent measures of income across survey and 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; see Appendix B. In addition to employment and business income, this includes rental income, taxable benefits, paid parental leave (PPL) and accident compensation commission (ACC) receipts.

Income is recorded gross-of-tax in both the survey and 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.

Capital income is partially covered in register data (see Appendix B): information on interest, dividends and capital gains are not available. 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.[25] The impact of other non-comparable income in the estimation of underreporting is examined in section 5.2.

Self-employment income is often characterised by being more volatile than employment income. This is also observed in our data; see Appendix Table C1. 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 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 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.[26]

[24] 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.
[25] The main types of income that are non-comparable are investment income, non-taxable benefits and family tax credits; see Appendix B.
[26] With only three years of data to construct our income growth variable, we do not use this income data as a permanent income proxy, unlike previous studies that have had access to a longer time-series dimension, or data on 'regular' income (Kukk and Staehr, 2017).

4.5 Measuring Expenditure

We create two measures of household expenditure: food and non-durable goods.[27] 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.[28]

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). As noted above, we 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 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 C2 & C3 compare OLS and IV estimates, and report several tests for endogeneity and instrument validity. If, as argued above, income is endogenous this is expected to bias downward estimates of β_1 in (5). If the self-employment dummy variable, *SE*, is exogenous, then γ in (5) would not be biased downwards directly as a result of endogeneity of γ but could be biased in either direction indirectly as a result of the bias in β_1, the bias depending on the covariance between *SE* and income. If – as we find – any bias in γ is relatively small compared to that for β_1, then from (5) we expect an *upward* bias in the underreporting variables, *k* and κ.

Appendix Tables C2 and C3 confirm that OLS estimates of β_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

[27] 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.

[28] We comment further on the assumption of accurate reporting of food expenditure in Section 5.

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. Various diagnostic tests including Hansen J-test statistics, Anderson-Rubin Wald tests and Kleibergen-Paap under-identification F-tests all strongly support the hypothesis that OLS estimates suffer from endogeneity, and the validity of the instruments.[29] We discuss the implications of measurement error affecting underreporting in IV estimates in section 6.

5.1 Underreporting

Table 2 summarises results for the estimated income-gap of the self-employed using alternative measures of expenditure, income and definitions of a self-employed household, with detailed regression results in Appendix C. Panel A reports estimates of the income-gap classifying the self-employed using the 'opportunity' definition – where a household is treated as self-employed if it receives *any* self-employment 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. [30]

Secondly, unlike most previous studies, 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 assigning a household into employment status where self-employment income is substantially underreported. [31] Kukk and Staehr (2014, 2017) explore self-employment definition issues based on Estonian data.

Columns (1) and (2) of Table 2 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 2 about here

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 results in a similar estimated level of underreporting, on average.

[29] In addition, first stage regression F-statistics and Kleibergen-Paap Wald F-statistics in Appendix Tables C2 and C3 are all strongly supportive of the instruments used: all F-statistics exceed 30.

[30] 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.

[31] PW (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.

We also test whether the measure of income used affects the estimation of income underreporting: lines (2) and (4) of Table 2 Panel A 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%, the income-gap is consistently estimated across both different expenditure and income measures.

Panel B in Table 2 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. 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.

Figure 1 illustrates the data and Engel curve relationships for employees and the self-employed for the case of labour income and food expenditure, based on register data. This shows a clear tendency for self-employed observations to be 'shifted' somewhat higher in the chart (higher expenditure for given income) compared to employees with the statistically significant upward shift in the predicted Engel curve for the self-employed, due to the positive *SE Dummy*, also shown.

Figure 1 about here

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, we note that in Table 2, column (2), using income from the survey consistently leads to a lower estimate of income underreporting, typically by around 6-10 percentage points.

The survey-based estimates in Table 2 are also typically both noisier (slightly larger standard errors) and not always significantly different from zero. Further, treating the estimates in two columns of Table 2 as obtained from two random samples drawn from the taxpayer population, a simple difference in means test confirms that the survey-register differences in underreporting estimates are statistically significant. 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.

As noted earlier Kukk and Staehr (2017) found that how the self-employed were identified using the PW approach for Estonia, was important for income-gap outcomes, with the role of the self-employment income share especially highlighted. Our results using both a zero and 25% income share show little difference (though with some evidence of higher estimated income-gaps with the higher SE income share). To test this aspect further we re-ran the regressions reported in Table 2 (and in Appendix Tables C2 & C3) but where the (0, 1) *SE Dummy* was replaced by a (0, *s*) *Dummy* variable, where *s* is the share of self-employment income in total income (defined using register data).

Results, not shown, were very similar to those in Table 2, Panel A, with all the *s-Dummy* variables statistically significant. These confirm a more general tendency for self-employed with higher *s* values to be associated with larger income-gaps, and provide some confirmation of a systematic component to the heterogeneity in the size of income-gaps across the self-employed, in part related to their SE income share, as found by Kukk and Staehr (2017).

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. These are explored in detail in Appendix E, which examines how income-gap estimates are affected by four aspects in particular.

(i) Do the self-employed demonstrate a preference for eating out?
(ii) Do differences in non-comparable income of the self-employed matter?
(iii) Does allowing for differential access to wealth, capital income or trust ownership make a difference?
(iv) Do the self-employed make use of business expenses for personal consumption?

We summarise results for those four cases in turn in this sub-section.

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.

In the absence of food price data, to examine this issue, we use regressions of the form in (4) but where the dependent variable is the share of food eaten out in total food expenditure. This confirms at most a weak tendency for the self-employed to spend slightly more on food outside the home, accounting for, at most, a small fraction of the 20% income underreporting estimates above; see Appendix Table E1.

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 measure 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.

Regressions in Appendix Table E2, which examine underreporting for sub-samples of households with differing amounts of NCI, confirm that income-gap estimates are little affected by the exclusion/inclusion of households with differing amounts of non-comparable income.

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.

In Appendix E, with limited available wealth and capital gain data, we control for those effects by including soft variables such as housing characteristics, in addition to our income growth and volatility variables. We also consider differences for self-employed households who have access to trusts – since trustee income may also finance household expenditure. However, results in Appendix E confirm that narrowing the sample to households less affected by potential wealth effects or trust ownership, income-gap estimates remain robustly around 20%.

Business versus personal expenses

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.

Appendix E considers whether there is any evidence of this expenses over-reporting phenomenon and confirms that net income underreporting by the self-employed is likely to be at least partially via diversion of those expense types (such as housing utilities, transport, communications), 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 3 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.

Table 3 about here

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. [32]

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.[33] 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

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 3 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.[34] 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

[32] Gender differences have also been extensively documented and studied in criminology with several theories proposed to explain them; see Mears *et al.* (1998).

[33] This age split yielded approximately equal sized age groups; similar results were obtained when we allowed for alternative age group thresholds.

[34] 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, unban self-employment occupations – such as taxis, construction, professional services – may provide similar or greater opportunities for underreporting.

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.[35] 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 β_1 and/or γ 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.

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.

[35] 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.

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):

$$E_i^S = \beta Y_i^R + \varepsilon_i \tag{8}$$

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 ε_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:

$$Y_i^S = Y_i^R + u_i \tag{9}$$

Estimating (9) using only survey data implies:[36]

$$E_i^S = \beta\left(Y_i^S - u_i\right) + \varepsilon_i = \beta Y_i^S + (\varepsilon_i - \beta u_i). \tag{10}$$

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:

$$Y_i^S = Y_i^R + u_i = Y_i^R + \bar{u} + v_i \tag{11}$$

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:

$$E_i^S = \beta Y_i^S - \beta\bar{u} + (\varepsilon_i - \beta v_i) \tag{12}$$

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:

$$\text{plim}\ \hat{\beta} = \lambda\beta \tag{13}$$

where $\lambda = \frac{\sigma_{YR}^2}{\sigma_{YR}^2 + \sigma_u^2}$ is the variance ratio or 'attenuation factor'; see Pischke (2007, p.2). Hence the bias can be given by:

$$-(1-\lambda)\beta = \frac{\sigma_u^2}{\sigma_{YR}^2 + \sigma_u^2}\beta \tag{14}$$

However, if Y_i^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:

$$\text{plim}\ \hat{\beta} = (1 - b_{uYs})\beta \tag{15}$$

where b_{uYs} is the estimated coefficient of a regression of u_i on Y_i^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 a find a positive covariance, σ_{uYs}, and an error variance, σ_u^2, consistently less than half the variance of Y^S (for various income definitions and treatment of outliers).

[36] Of course, E_i^S may also be measured with error but, as noted above, this reduces the efficiency of estimates of β but does not induce bias.

In addition to the prospect of attenuation bias on our income parameter, $\hat{\beta}$ (or β_1 in the estimating equation (4)), there is also the possibility of biases affecting our *SE dummy* parameter, γ. Attenuation bias could arise from mismeasurement of self-employment status using our two proxies (though we have no way of assessing this)[37], 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 γ and β_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 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 4, Panel A, column (4). The standard deviation of the error is however quite large, yielding a variance ratio or attenuation factor, λ, in column (5) of 24% and 32% for labour income and comparable income respectively.

Table 4 about here

[37] 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 – 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.

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^S}, obtained from regressing u_i on survey incomes, Y_i^S. This regression gives an attenuation factor, measured by b_{uY^S} in column (6), lower than the variance ratio in column (5).[38]

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 4, 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^S} continues to be significantly positive (b_{uY^R} 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' 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.[39]

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.[40] 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.

[38] Consistently, column (7) shows that the equivalent regression of the measurement error on register ('true') income yields a negative correlation (b_{vY^R}) between the two. The variance ratio in column (5) is estimated as the variance of the error over the total variance.

[39] 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).

[40] 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.

Note however that Table 4 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 self-employment *income-gaps, relative to employees* (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 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 5. 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 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).

Table 5 about here

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.[41] 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 part 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.

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^2s, 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.[42] 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

[41] Note that mean reversion is higher in self-employment than employment income; this reduces the bias when income is the explanatory variable.

[42] 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.

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 5.[43]

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, $\hat{\gamma}$, 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 6 about here

Table 6 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 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, κ, 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 6. 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 $\beta\bar{u}$ in equation (12) – have a very large effect on estimates of the income-gap, κ. 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 (κ) in particular seems to be substantially mismeasured when survey data are used.

[43] 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.

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: they 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 publicly available (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 for New Zealand to provide empirical magnitudes of those gaps.

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. Using a unique dataset where survey participants are matched to their (suitably anonymised) administrative tax records, we corrected for the potential effects of measurement error known to afflict survey responses in the estimation of income underreporting. Using tax 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 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%.

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, estimated survey-based income-gaps can be up to 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.* (2019) 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. The limited evidence for other countries, such as Estonia (Paulus, 2015a) suggest this 'survey versus register data' aspect of income underreporting estimates may indeed be important more widely.

References

Aguiar, M. and Hurst, E. (2005) Consumption versus expenditure. *Journal of political Economy*, 113, 919–948.

Ahmed, R.A. and Rider, M. (2008) *Pakistan's Tax Gap: Estimates By Tax Calculation and Methodology*. Paper No. 0811, International Center for Public Policy, Andrew Young School of Policy Studies, Georgia State University.

Allingham, M.G. and Sandmo, A. (1972) Income tax evasion: a theoretical analysis. *Journal of Public Economics*, 1, 323-338.

Akee, R. (2011) Errors in self-reported earnings: the role of previous earnings volatility and individuals characteristics. *Journal of Development Economics*, 96, 409–421.

Alm, J. and Torgler, B. (2006) Culture differences and tax morale in the United States and in Europe. *Journal of Economic Psychology*, 27, 224–246.

Andreoni, J., Erard, B. and Feinstein, J. (1998) Tax compliance. *Journal of economic literature*, 36, 818–860.

Artavanis, N., Morse, A. and Tsoutsoura, M. (2016) Measuring income tax evasion using bank credit: evidence from Greece. *The Quarterly Journal of Economics*, 131, 739–798.

Baldini, M., Bosi, P. and Lalla, M. (2009) Tax evasion and misreporting in income tax returns and household surveys. *Politica Economica*, 3, 333-348.

Baum, C., Schaffer, M. and Stillman, S. (2007) Enhanced routines for instrumental variables/GMM estimation and testing. *Stata Journal*, 7, 465–506.

Braithwaite, V. (2003) Dancing with tax authorities: motivational postures and non-compliant actions. *Taxing democracy*, 15–39.

Breusch, T. (2005) *Estimating the underground economy using MIMIC models*. Working Paper, National University of Australia, Canberra, Australia.

Brewer, M. and O'Dea, C. (2012) Measuring living standards with income and consumption: evidence from the UK. *ISER Working Paper Series*, No. 2012–05.

Browning, M. and Leth-Petersen, S. (2003) Imputing consumption from income and wealth information. *Economic Journal*, 113, F282–F301.

Bound, J., Brown, C., Duncan, G.J. and Rodgers, W.L. (1994) Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics*, 12, 345–368.

Bound, J., Brown, C. and Mathiowetz, N. (2001) 'Measurement error in survey data'. In J. Heckman and E. Leamer (eds) *Handbook of Econometrics, vol. 5*. pp. 3705–3830. Amsterdam: North Holland.

Cabral, A.C.G., Kotsogiannis, C. and Myles, G. (2019) Self-employment underreporting in Great Britain: How Much and Who? *CESifo Economic Studies*, 65, 84-107.

Cabral, A.C.G. and Gemmell, N. (2018) Estimating self-employment income gaps from survey and register data. Evidence from New Zealand. *Working Papers in Public Finance*, 07/2018. Victoria Business School, Victoria University of Wellington, Wellington, New Zealand.

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

Engström, P. and Holmlund, B. (2009) Tax evasion and self–employment in a high–tax country: evidence from Sweden. *Applied Economics*, 41, 2419–2430.

Engstrom, P. and Hagen, J. (2017) Income under-reporting among the self-employed. A permanent income approach. *European Economic Review*, 92, 92-109.

Erard, B. and Feinstein, J. (2010) Econometric models for multi-stage audit processes: an application to the IRS national research program. In J. Alm, J. Mártinez-Vazquez, and B. Torgler (eds) *Developing Alternative Frameworks for Explaining Tax Compliance*. New York: Routledge.

Feige, E.L. (2016) Reflections on the meaning and measurement of unobserved economies: what do we really know about the 'shadow economy'? *Journal of Tax Administration*, 2, 5–41.

Feinstein, J.S. (1991) An econometric analysis of income tax evasion and its detection. *The RAND Journal of Economics*, 22, 14–35.

Feld, L.P. and Larsen, C. (2005) *Black Activities in Germany in 2001 and 2004: A Comparison Based on Survey Data*. Study no. 12, Copenhagen: Rockwool Foundation Research Unit.

Feld, L.P. and Larsen, C. (2009) *Undeclared work in Germany 2001-2007–Impact of deterrence, tax policy and social norms: An analysis based on survey data.*

Feldman, N.E. and Slemrod, J. (2007) Estimating tax noncompliance with evidence from unaudited tax returns. *Economic Journal*, 117 (518), 327–352.

Fox, K., Hu, J., Adelsheim, P. (2017). *A comparion of TY99 TCMP and TY01 NRP Results.* Accessible at: https://www.irs.gov/pub/irs-soi/08resconcomparty88.pdf .

Gemmell, N. and Hasseldine, J. (2012). The tax gap: a methodological review. In *Advances in Taxation*, 20, 203–231. Emerald Group Publishing.

Gemmell, N. and Hasseldine, J. (2014). Taxpayers' Behavioural responses and measures of tax compliance 'gaps': a critique and a new measure. *Fiscal Studies*, 35, 275**–**296.

Giles, David .E. (1999) Modelling the hidden economy and the tax-gap in New Zealand. *Empirical Economics*, 24, 71–90.

Gray, P.G. (1955) The memory factor in social surveys. *Journal of the American Statistical Association*, 50, 344–363.

H.M. Revenue & Customs (2016) *Measuring tax gaps 2016 edition. Tax gap estimates for 2014-15*.

Hurst, E., Li, G. and Pugsley, B. (2014) Are household surveys like tax forms? Evidence from income underreporting of the self–employed. *Review of Economics and Statistics*, 96, 19–33.

IRD (2017), *Understanding Schedular Payments*. Available at: http://www.ird.govt.nz/contractors/about/understanding-schedular-payments-contractors.html Inland Revenue, Wellington: New Zealand.

ISWGNA, 2006. *Estimates of unrecorded economy and national accounts.* Retrieved from https://www.oecd.org/ctp/administration/36333066.pdf.

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

Klepper, S. and Nagin, D. (1989) The anatomy of tax evasion. *Journal of Law, Economics, & Organization*, 5, 1–24.

Kleven, H., Knudsen, J. Kreiner, M.B., Thustrup, C., Pedersen, S. and Saez, E. (2011) Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79, 651–692.

Kreiner, C.T., Lassen, D.D. and Leth-Petersen, S. (2014) Measuring the accuracy of survey responses using administrative register data: evidence from Denmark. In *Improving the measurement of consumer expenditures* (pp. 289–307). University of Chicago Press.

Kukk, M. and Staehr, K. (2014) Income underreporting by households with business income: evidence from Estonia. *Post-Communist Economies*, 26, 257–276.

Kukk, M. and Staehr, K. (2017) Identification of households prone to income underreporting: employments status or reported business income? *Public Finance Review*, 45, 599–627.

Kukk, M., Paulus, A. and Staehr, K. (2019) Cheating in Europe: underreporting of self-employment income in comparative perspective. *International Tax and Public Finance*, published online at https://doi.org/10.1007/s10797-019-09562-9.

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

Neter, J. and Waksberg, J. (1964) A study of response errors in expenditures data from household interviews. *Journal of the American Statistical Association*, 59, 18–55.

Nygard, O.E., Slemrod, J. and Thoresen, T.O. (2018) Distributional implications of joint tax evasion. *Economic Journal*, 129, 1894–1923.

Meyer, B.D. and Sullivan, J.X. (2009) *Five decades of consumption and income poverty*. National Bureau of Economic Research, Paper No. w14827.

OECD (2002) *Measuring the non-observed economy.* Available at https://www.oecd.org/std/na/1963116.pdf

OECD (2017) *Self-employment Rate (Indicator).* https://DOI:10.1787/fb58715e-en.

Paulus, A. (2015a) *Income underreporting based on income-expenditure gaps: survey vs tax records*. University of Essex, Institute for Social and Economic Research. Working Paper No. 2015-15.

Paulus, A. (2015b) *Tax evasion and measurement error: An econometric analysis of survey data linked with tax records*. University of Essex, Institute for Social and Economic Research. Working Paper No. 2015-10.

Pedersen, S. (2003) *The shadow economy in Germany, Great Britain and Scandinavia: a measurement based on questionnaire surveys* (No. 10). Rockwool Foundation Research Unit.

Pischke, S. (2007) *Lecture Notes on Measurement Error*. Unpublished paper, London School of Economics. Available at econ.lse.ac.uk/staff/spischke/ec524/Merr_new.pdf.

Pissarides, C.A. and Weber, G. (1989) An expenditure–based estimate of Britain's black economy. *Journal of Public Economics*, 39 (1), 17–32.

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

Scott, C. and Amenuvegbe, B. (1991) Recall loss and recall duration: an experimental study in Ghana. *Inter-Stat*, 4, 31–55.

Slemrod, J. and Weber, C. (2012) Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance*, 19 (1), 25–53.

Slemrod, J. (2018) Tax compliance and enforcement. *NBER Working Paper No. 24799*. Cambridge, MA: National Bureau of Economic Research.

Sudman, S. and Bradburn, N.M. (1973) Effects of time and memory factors on response in surveys. *Journal of the American Statistical Association*, 68, 805–815.

Torgler, B., Schaffner, M. and Macintyre, A. (2007) *Tax compliance, tax morale, and governance quality.* International Studies Program Working Paper, Andrew Young School of Policy Studies, Georgia State University.

Waseem, M. (2018) Taxes, informality and income shifting: evidence from a recent Pakistani tax reform. *Journal of Public Economics*, 157, 41–77.

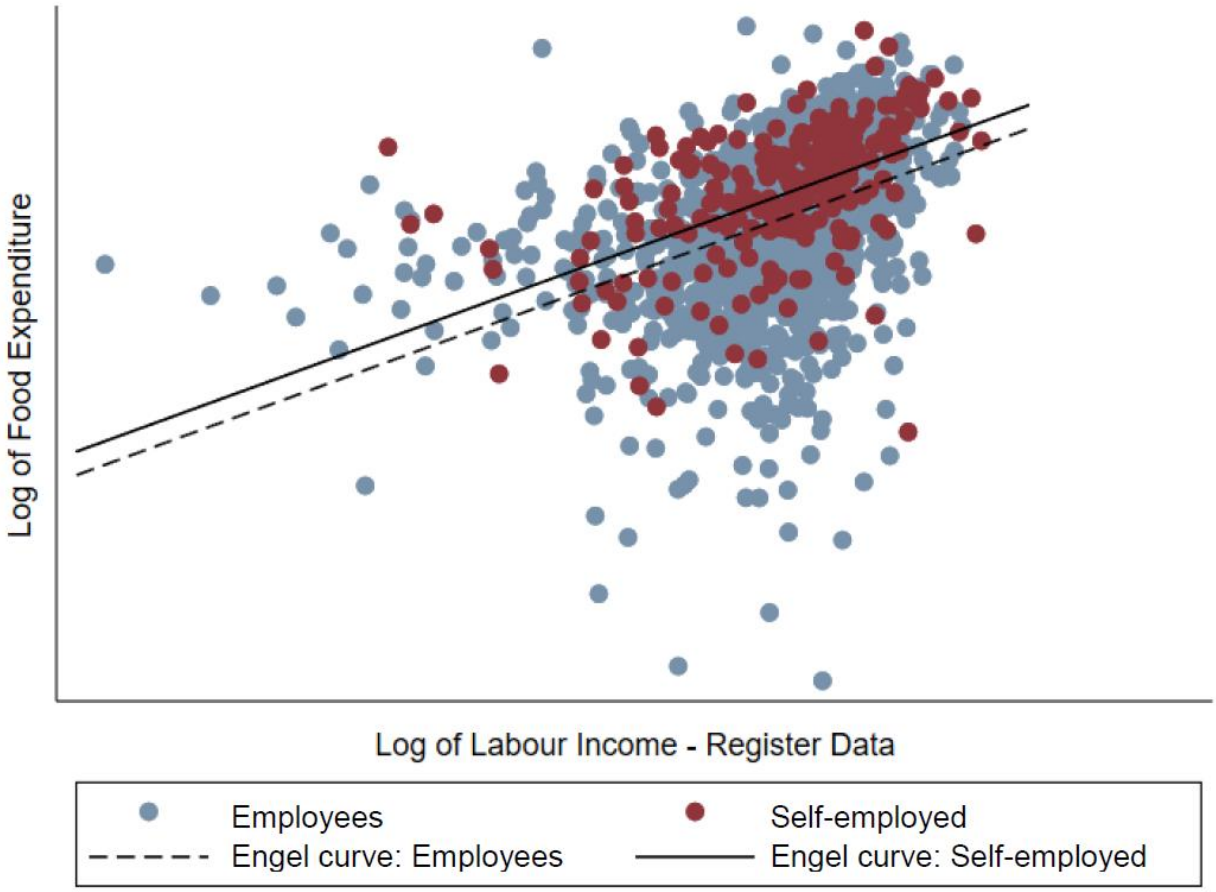
Waseem, M. (2019) Does cutting then tax rate to zero induce behavior different from other tax cuts? Evidence from Pakistan. *Review of Economics and Statistics*, forthcoming. Available at: https://doi.org/10.1162/rest_a_00832.

Acknowledgments and Disclaimers

The authors would like to thank Matt Benge, Sean Comber and John Creedy for comments on previous versions of this paper; and Stephen Challands, Caleb Moses and colleagues for the prompt checking and release of the data. The results presented are the work of the authors, not Statistics NZ; and are not official statistics. They have been created for research purposes from Statistics New Zealand's Integrated Data Infrastructure (IDI). 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 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 confidentialised 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.

Figure 1



Note: Due to taxpayer confidentiality requirements, some outlier observations, and axes values, have been removed from Figure 1.

Table 1 Income Underreporting Estimates[Ψ]

Study	Country	Expenditure type	Underreporting estimate (%)	Income data source[†]
Pissarides & Weber (1989)	UK	Food	36	S
Lyssiotou *et al.* (2004)	UK	Food	22	S
Cabral *et al.* (2019)	UK	Food	19	S
Engström & Holmlund (2009)	Sweden	Food	25-30	S
Johansson (2005)	Finland	Food	17-42	S
Schuetze (2002)	Canada	Food	11-19	S
Hurst *et al.* (2014)	US	Food	25-30	S
Kukk and Staehr (2017)	Estonia	Food	28-56	S
Feldman & Slemrod (2007)	US	Charitable contributions.	35	R
Artevanis *et al.* (2016)	Greece	Bank credit	42-45	Bank data
Paulus (2015a)	Estonia	Housing (e.g. utilities, repairs)	20-25	S
			33, 56*	R

† S = survey, R= register. * Self-employment income-gap of 56% when compared to *public* sector employees; 33% when compared to *private* sector employees (who have an income-gap of 23%). Ψ See Kukk *et al.* (2019) for estimates for 14 EU countries.

Table 2 Income-Gap Estimates and Robustness to Definitions

		(1): Register	(2): Survey
Panel A: Self-Employment: Opportunity			
Expenditure	**Income**	**Income-gap**	
Food	Labour	0.200***	0.114*
		(0.057)	(0.063)
Food	Comparable	0.193***	0.120*
		(0.048)	(0.062)
Non-Durables	Labour	0.204***	0.119**
		(0.047)	(0.051)
Non-Durables	Comparable	0.196***	0.124**
		(0.040)	(0.050)
Panel B: Self-Employment: 25% Rule			
Expenditure	**Income**	**Income-gap**	
Food	Labour	0.216***	0.107
		(0.066)	(0.075)
Food	Comparable	0.206***	0.111
		(0.055)	(0.073)
Non-Durables	Labour	0.254***	0.153***
		(0.053)	(0.059)
Non-Durables	Comparable	0.239***	0.158***
		(0.045)	(0.057)

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 C. Asterisks indicate significance at *** p<0.01; ** p<0.05; * p<0.1.

Table 3 Income-Gap Heterogeneity

	Income type: Labour				
Expenditure category:	Food	Non-Durables		Food	Non-Durables
	Panel A: Gender			Panel B: Region	
Male	0.312***	0.309***	Auckland	0.361***	0.321***
	(0.078)	(0.063)		(0.087)	(0.075)
Female	0.096	0.108	Canterbury	0.333***	0.315***
	(0.083)	(0.070)		(0.099)	(0.091)
			Rest of North	0.047	0.124
				(0.133)	(0.093)
			Rest of South	-0.037	0.001
				(0.149)	(0.119)
			Wellington	0.230*	0.214*
				(0.125)	(0.114)
Test of the equality of income-gaps:					
Chi^2	4.48	5.62		10.06	8.40
p-value	0.03	0.02		0.04	0.08

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. Robust standard errors are in parentheses.

Table 4 Summary Statistics of Reporting Errors

		Mean (Standard Deviation)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Earnings Variables	N	Survey	Register	Error	Variance Ratio (λ)	b_{uYS}	b_{vYR}
				Panel A			
Labour Income	2577	11.036	11.024	0.013	0.242	0.139***	-0.187***
		(0.806)	(0.830)	(0.469)		(0.011)	(0.01)
Comparable: Total Income	2577	11.084	11.092	-0.008	0.318	0.280***	-0.127***
		(0.744)	(0.676)	(0.462)		(0.011)	(0.013)
				Panel B: Omit outliers			
Labour Income	2532	11.041	11.031	0.010	0.152	0.063***	-0.119 ***
		(0.729)	(0.753)	(0.318)		(0.009)	(0.008)
Comparable: Total Income	2526	11.096	11.095	0.001	0.220	0.175***	-0.093***
		(0.647)	(0.616)	(0.327)		(0.009)	(0.008)

Table 5 Summary Statistics of Reporting Errors by Household Type

		Mean (Standard Deviation)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	N	Survey	Register	Error	Variance Ratio (λ)	b_{uYS}	b_{vYR}
Labour Income				*Panel A*			
Self-Employment Income > 0	663	11.215	11.136	0.079	0.255	0.081***	-0.278***
		(0.769)	(0.867)	(0.507)		(0.025)	(0.02)
No Self-Employment Income	1914	10.974	10.984	-0.010	0.236	0.152***	-0.158***
		(0.810)	(0.813)	(0.453)		(0.012)	(0.012)
Comparable: Total Income							
Self-Employment Income > 0	663	11.252	11.168	0.085	0.315	0.190***	-0.295***
		(0.745)	(0.799)	(0.542)		(0.027)	(0.024)
No Self-Employment Income	1914	11.026	11.065	-0.040	0.317	0.305***	-0.041***
		(0.734)	(0.626)	(0.426)		(0.011)	(0.016)
Labour Income				*Panel B: Omit Outliers*			
Self-Employment Income > 0	645	11.192	11.125	0.067	0.253	0.137***	-0.214***
		(0.707)	(0.741)	(0.431)		(0.023)	(0.021)
No Self-Employment Income	1890	10.990	10.999	-0.009	0.111	0.034***	-0.093***
Comparable: Total Income		(0.730)	(0.754)	(0.266)		(0.008)	(0.008)
Self-Employment Income > 0	645	11.232	11.151	0.081	0.276	0.159***	-0.238***
		(0.670)	(0.703)	(0.435)		(0.025)	(0.023)
No Self-Employment Income	1881	11.050	11.075	-0.026	0.184	0.173***	-0.019*
		(0.633)	(0.582)	(0.276)		(0.009)	(0.011)

Table 6 Survey-Register (S-R) Parameter Differences

Dependent var.:	---------- Food Expenditure ----------				---------- Non-Durables Expenditure --------			
Data source:	Register	Survey	Register	Survey	Register	Survey	Register	Survey
Income type:	Labour	Labour	Comp.	Comp.	Labour	Labour	Comp.	Comp.
A: Coefficients								
Income: $\hat{\beta}$	0.460	0.443	0.545	0.443	0.485	0.466	0.573	0.466
S/R ratio	*0.963*		*0.813*		*0.961*		*0.813*	
SE Dummy: $\hat{\gamma}$	0.103	0.0537	0.117	0.0565	0.111	0.059	0.125	0.0619
S/R ratio	*0.521*		*0.483*		*0.532*		*0.495*	
B: Estimates of underreporting								
Multiplier	1.25	1.129	1.239	1.136	1.256	1.135	1.244	1.142
S/R ratio	*0.903*		*0.917*		*0.904*		*0.918*	
Income-gap	0.200	0.114	0.193	0.120	0.204	0.119	0.196	0.124
S/R ratio	*0.570*		*0.622*		*0.583*		*0.633*	

Note: The self-employment definition is 'Opportunity': taxpayers with some self-employment income.

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

Ana Cinta G. Cabral, Norman Gemmell and Nazila Alinaghi

APPENDICES

Contents

Appendix A: Data Matching

A.1 Integrated Data Infrastructure

The Statistics New Zealand (SNZ) Integrated Data Infrastructure (IDI) is a data repository based mainly on administrative data that also contains a number of surveys undertaken by SNZ; see www.stats.govt.nz/integrated-data/integrated-data-infrastructure/. The main structure of the IDI can be described as a central spine to which other datasets are linked. The linked datasets cover a wide range of topics: education, taxation, benefits, health and safety among others. It provides researchers with a unique environment where the benefits of survey and register data can be exploited. These datasets are held securely on SNZ servers and only approved researchers and research projects can access the anonymised datasets under strict confidentiality rules.

The population contained in the IDI can be broadly described as all individuals who have ever been resident of New Zealand. By legislation, no identifier can be used across agencies to identify individuals. However, in some cases such as for administrative data a common identifier – generally the tax registration number (IRD number in NZ) – is readily available. Matching of these datasets are direct such that the quality of the match is very high. The IDI however links other datasets to the IDI spine such as those from selected surveys for which such identifiers are not available.

The IDI matching methodology for surveys is mainly based on probabilistic record matching with careful consideration and monitoring of quality to ensure the reliability of matching is high; that is, ensuring that records from different datasets belong to the same person or entity. Blocking variables are used to ensure that the match is performed subject to the records compared being identical on the blocking variable. Variables are compared in terms of a given set of characteristics and for each characteristic, a measure is computed to assign how closely related the two values are. The sum of those measures, known as a weight, for all given characteristics will form the overall weight for a record pair.

The methodology assigns the pair of records with a weight that shows how reliable the match is and how common the value is. The record pairs with overall weights above a certain threshold are designated as a match. A more thorough explanation of the IDI matching process is given in Statistics New Zealand (2014). This linking methodology is internationally used and is adopted to obtain a high precision rate. Inarguably, linking quality using solely administrative data is superior to the linkage of survey data but several mechanisms are in place to maximise precision. In the absence of a common identifier, linking variables such as the date of birth, first and last names and gender are used. The linking is generally of good quality which may be attributable to the fact that in a small economy like New Zealand, names and dates of birth are more frequently unique.

From the IDI tables, we use the Household Economic Survey and the IR tax year tables. The latter contain the universe of annual income tax returns filed by individuals and businesses. In New Zealand, wages and salary employees' income is withheld and third-party reported by their employers using the Employer Monthly Schedule (EMS). The EMS is a mandatory reporting requirement for all employers with paid employees. The EMS is complemented with the income tax returns filed by individuals (IR3), which is required for those individuals who earn income other than salary and wages, dividends and interests and/or taxable Māori authority contributions which are withheld at source. Particularly interesting within this group are sole traders and other individuals earning self-employment income from a subsidiary job. Those self-employed who are incorporated have the obligation to file an IR4S tax return, and those in a partnership are obliged to file an IR20 tax return.

One of the interesting features of this table is that in order to reclassify payments from EMS that are linked to self-employment, Statistics NZ link payer and payee identifiers which identifies the source of EMS payments. This enables a reclassification of payments from EMS that are attributed to self-employed (in any legal form), who pay themselves a salary, as self-employment income rather than as

income from employment. That is, records that might initially be classified as a regular Wages and Salary (WAS) category, can be classified as self-employment income where identifiers show that the payment is made by the same person receiving the payment.

A.2 Matching and Linking Criteria

This section provides more detail on the matching and linking criteria used to link the Household Economic Survey (HES) and the register data on incomes. Using data on individuals and households that have been matched by Statistics NZ in the IDI, it is possible to combine administrative and survey responses on income. These use a unique Statistics NZ identifier assigned to individuals across data sources. Having identified matched individuals within the IDI, we then have to ensure that the data we select from the survey and administrative sources are linked such that, for example, reported incomes for each individual or household selected from the two sources relate to the same tax years. This linking process involves a number of steps.

In HES, individuals are asked about their income in the reference period. The reference period is set by the interviewer at the beginning of the interview to be the 12 months prior to the day the survey was conducted. We assume all surveys were conducted at the beginning of the month and, since administrative data on wages and salaries are available on a monthly basis, we match for each individual the income reported to the tax administration for the months that exactly match their reference period. That is, if an individual was interviewed in November 2011, we set the reference period November 2010 to October 2011 and we aggregate his/her employment income from the tax record to obtain total income from the register. This should be a 1-to-1 map to the survey for those with a successful match in the absence of measurement errors.

In the case of self-employment income, the survey enquires about the last period for which accounts were prepared or a tax record is available. Due to the restriction of information in the matched dataset we are not certain to which period individuals are referring when reporting their self-employment income in the survey. We impose two criteria to test for the robustness.

The month of March marks the end of the tax year and the deadline for presenting the returns to the tax administration is July. The first linkage criterion hinges on the rationale that those reporting to the survey prior to July will likely report their incomes for the tax year that closed in April of the previous tax period. That is, those reporting in January 2011 will report about the year 2009/10 as their accounts are not yet prepared for the year 2010/11 and that is the last record available. Those reporting after July will have already presented their accounts for the tax year that closes in the year the survey was conducted and will report on that tax year just completed. That is, those interviewed in November 2011 will report on the tax year 2010/11.

This first criterion overlooks the possibility that some overlap is expected especially for the months between March and July. For example, those prompt in complying with their filing obligation may report on the current year rather than the past tax year. Assigning them the past tax year will result in timing errors that relate to the linking process rather than interviewees not reporting accurately.

In order to test whether this criterion provides a satisfactory rationale for how respondents complete the survey, we restrict the sample to those self-employed that report income to the survey that lies within 1% of the reported income to IR. Mapping their survey month to whether they reported their past tax year or their current tax year, we do find the pattern outlined above. Those interviewed prior to July report on the previous tax year to the survey and those after July report on the current tax year. However, the criterion is not perfect and apart from the overlap in March and July some interviewees after July still report on the previous year. To circumvent this issue and allow more flexibility to the criterion trying to ascertain what the respondents are referring to, we establish criterion 1 but allow the match to change from the current tax year or the prior year if reported income to the survey lies within 10% of

the reported income in any of the years, and lies further from the corresponding income following criterion 1.

As an example, consider Person 1 who was interviewed in November 2011, and who reports $50,000 to the survey as self-employment income. Looking at administrative data, for the tax year 2010/11 we have a report of $40,000 and for the tax year 2009/10 a report of $50,000. Following criterion 1, the individual would be assigned such that his income from the register belongs to the tax year 2010/11 as he was surveyed after July. However, his register income matches his survey report for the year 2009/10 very closely. Criterion 2 would then assign self-employment income to the year 2010/11.

Whereas this criterion allows for the flexibility of closely aligning to the income the self-employed report to the survey, the drawback is that some of this mismatches may not be a linking error but in fact measurement error. This is, individuals may be reporting their true income rather than their income reported to the survey, or it can be failure to recall their income tax year from the current year. It does however allow us to minimise the errors in the linkage. Following criterion 2, only 2% of the cases are reclassified.

A.3 Linkage Accuracy

Table A1 Linkage accuracy of individuals in HES, disaggregated by income source.

	Survey and IR	Survey, Not IR	IR, Not Survey	Not HES, not IR	Total
Individuals	49,707	14,052	519	17,268	81,549
Income Source					
Labour Income	34,461	1437	2,856	10,950	49,707
Pensions Income	9,378	93	240	39,999	49,707
Taxable Benefits	1,059	2,712	5,679	40,257	49,707
Student Allowance	837	228	303	48,339	49,707
Paid Parental Leave	345	48	231	49,086	49,707
Rental Income	342	1,998	486	46,884	49,707
Earnings Compensation	591	324	453	48,342	49,707
Comparable Income-All (D2)	43,509	1,047	4,050	1,104	49,707

Appendix B: Concordance of Income Sources - between Inland Revenue and HES data

	Source of Income	**Inland Revenue** (tax return no.)	**HES (**category no.)
Employment income	Wages and Salaries	Wages and salaries from EMS, withholding payments, commissions and bonuses.	Wages and salaries from current and past employment including regular pay, other honoraria, commissions and bonuses. (1.1)
	Self-Employment Income	Sole trader (IR3), company director/shareholder income (IR4S), partnership (IR20), other honoraria, commissions, bonuses.	Self-employment income from current/previous jobs. Net profit, share of profit and loss. (1.2)
	Other		Income from casual jobs and hobbies. (1.3)
Investment income		Rental Income (IR3)	Rental income (2.3)
		-	Interest, dividends, royalties (2.1, 2.2, 2.4)
		-	Overseas interests and dividends and other foreign income. (5.1)
		-	Other investment income (income from trusts, funds).
Pensions	Government	NZ Superannuation and veterans pensions	NZ Superannuation and veterans pension (3.1.)
	Private	-	Private Superannuation income, income from annuities (non-governmental). (4.1, 4.2)
Other government transfers		Benefits	Unemployment benefit, sickness benefit, domestic purposes benefit, invalid's benefit (3.2 – some)
		Student Allowance	Student allowance (3.2)
		Paid Parental Leave	Paid parental leave (3.2)
		-	In-work tax credit, Minimum family tax credit
		-	Other non-taxable benefits.
Other sources of regular and recurring income	ACC	Accident Compensation Corporation (ACC) receipt	Accident Compensation Corporation receipt (1.4)

Appendix C: Summary Statistics & Regression Results

Table C1 Summary statistics of income and expenditure variables (in logs).

Self-Employment Definition	Opportunity		25% rule	
	Employed	Self-Employed	Employed	Self-Employed
Panel A: Expenditure		Mean (standard error)		
Food	8.842	9.145	8.887	9.129
	(0.757)	(0.683)	(0.753)	(0.699)
Non-Durables	9.219	11.150	9.261	9.524
	(0.646)	(0.805)	(0.651)	(0.634)
Panel B: Register		Mean (standard error)		
Labour Income	11.005	11.128	11.041	11.029
	(0.783)	(0.837)	(0.768)	(0.952)
Comparable Income	11.081	11.150	11.108	11.055
	(0.609)	(0.805)	(0.612)	(0.915)
Panel C: Survey		Mean (standard error)		
Labour Income	10.989	11.212	11.031	11.156
	(0.799)	(0.749)	(0.785)	(0.814)
Comparable Income	11.040	11.253	11.079	11.199
	(0.719)	(0.724)	(0.712)	(0.790)
Observations (N)	1914	663	399	2178

Note: Summary statistics of expenditure and income are disaggregated by the two alternative definitions of a self-employed household. The 'opportunity' definition classifies a household into self-employment if it benefits from an income source that comes from self-employment. This definition aims at capturing the 'opportunity' to misreport. A household is classified as employed in the absence of any self-employment income source. Under the 25% rule, a household is classified into self-employment if it derives more than 25% of household income from a self-employment source. The classification is based on register income.

Regression Results

Table C2 IV and OLS estimation of Engel curves using food as the expenditure item

Dependent variable: Food expenditure	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Method:	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Panel A: Coefficients of interest								
Income	Labour and Business (IR)		Labour and Business (Survey)		Comparable Income (IR)		Comparable Income (Survey)	
	0.159***	0.460***	0.170***	0.443***	0.247***	0.545***	0.204***	0.443***
	(0.0228)	(0.0904)	(0.0237)	(0.0868)	(0.028)	(0.101)	(0.0239)	(0.0816)
SE Dummy	0.0739***	0.103***	0.0568**	0.0537*	0.0849***	0.117***	0.0577**	0.0565*
	(0.0286)	(0.0323)	(0.0282)	(0.0291)	(0.0286)	(0.0318)	(0.0282)	(0.0289)
Panel B: Estimations of underreporting								
Multiplier	1.59***	1.25***	1.395***	1.129***	1.411***	1.239***	1.327***	1.136***
	(0.298)	(0.089)	(0.238)	(0.08)	(0.167)	(0.073)	(0.187)	(0.08)
95% CI	1.006	1.076	0.928	0.972	1.084	1.095	0.960	0.980
	2.174	1.424	1.863	1.287	1.737	1.382	1.694	1.292
Income-gap	0.371***	0.200***	0.283**	0.114*	0.291***	0.193***	0.246**	0.12*
	(0.118)	(0.057)	(0.122)	(0.063)	(0.084)	(0.048)	(0.106)	(0.062)
95% CI	0.140	0.089	0.043	-0.009	0.127	0.099	0.038	-0.001
	0.602	0.311	0.523	0.238	0.455	0.286	0.455	0.241
F statistic First Stage		36.2		33.46		48.94		56.07
Hansen J		1.266		1.069		0.409		0.412
Hansen J p-value		0.531		0.586		0.815		0.814
Kleibergen Paap LM		100.2		95.17		131.6		149.2
Anderson-Rubin Wald test		30.7		30.7		30.7		30.73
Anderson-Rubin Wald p-value		0.00		0.00		0.00		0.00
N	2,577		2,577		2,577		2,577	
Employed	1,914		1,914		1,914		1,914	
Self-Employed (Definition = 'Opportunity')	663		663		663		663	

Note: Each column contains the main results of the regression of food expenditure on the different income variables. Summary results are presented in Table 1, Panel A. Columns (1) and (2) regress food on Labour Income built from the register while (3) and (4) use the equivalent measure from the survey. The following covariates are included in the regressions but suppressed here for presentational purposes: demographic characteristics (number of children, marital status, age and gender), proxies for wealth (type of tenure of the dwelling, number of rooms, the number of stories, type of dwelling, the local housing benefit, i.e. Accommodation Supplement area where the house is located, and the region), the annual variability of household income to proxy for income risk and its average growth rate over three years to proxy the stability of household finances); and time dummies. Columns (5), (6) and (7), (8) show the same results as in (1) - (4) but using the comparable income measure. Asterisks indicate significance at the following levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. CI = confidence interval. The table reports several instrument tests of the validity of the instrument set. To test the weak instruments and underidentification, it reports the F-statistic of the first stage (Kleibergen-Paap Wald F) and the Kleibergen-Paap LM test. Critical values for the first are not tabulated. They can be compared to the Stock-Yogo critical values; alternatively Baum, Schaffer and Stillman (2007) suggest a rule of thumb that the F-statistic should be higher than 10 for weak identification issues not to be a problem. Rejection of the null in the Kleibergen-Paap LM tests for underidentification indicates that the smallest canonical correlation between the endogenous variables and the instruments is nonzero. The Hansen J test provides the robust version of the Sargan statistic of overidentification.

Table C3: IV and OLS estimation of Engel curves using non-durables as the expenditure item

Dependent variable: Non-Durables Expenditure	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Method:	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Panel A: Coefficients of interest								
Income	Labour and Business (IR)		Labour and Business (Survey)		Comparable Income (IR)		Comparable Income (Survey)	
	0.159***	0.485***	0.164***	0.466***	0.246***	0.573***	0.200***	0.466***
	(0.0207)	(0.0781)	(0.0214)	(0.0756)	(0.0278)	(0.0856)	(0.0218)	(0.0693)
SE Dummy	0.0793***	0.111***	0.0623***	0.0590**	0.0903***	0.125***	0.0632***	0.0619**
	(0.0246)	(0.029)	(0.0241)	(0.0253)	(0.0246)	(0.0285)	(0.024)	(0.025)
Panel B: Estimations of underreporting								
Multiplier	1.647***	1.256***	1.463***	1.135***	1.444***	1.244***	1.372***	1.142***
	(0.269)	(0.074)	(0.228)	(0.066)	(0.15)	(0.061)	(0.172)	(0.065)
95% CI	1.121	1.111	1.017	1.005	1.150	1.124	1.036	1.015
	2.174	1.401	1.909	1.264	1.737	1.364	1.709	1.269
Income-gap	0.393***	0.204***	0.317***	0.119**	0.307***	0.196***	0.271***	0.124**
	(0.099)	(0.047)	(0.106)	(0.051)	(0.072)	(0.04)	(0.091)	(0.05)
95% CI	0.199	0.112	0.108	0.018	0.167	0.119	0.092	0.027
	0.587	0.296	0.525	0.219	0.448	0.274	0.450	0.222
F statistic First Stage		36.2		33.46		48.94		56.07
Hansen J		1.81		1.533		0.704		0.703
Hansen J p-value		0.405		0.465		0.703		0.704
Kleibergen-Paap LM		100.2		95.17		131.6		149.2
Anderson-Rubin Wald test		48.31		48.31		48.31		48.48
Anderson-Rubin Wald p-value		0.00		0.00		0.00		0.00
N	2,577		2,577		2,577		2,577	
Employed	1,914		1,914		1,914		1,914	
Self-Employed (Definition = 'Opportunity')	663		663		663		663	

Note: Each column contains the main results of the regression of non-durable expenditure on the different income variables. Summary results are presented on Table 1, Panel A. Columns (1) and (2) regress food on Labour Income built from the register while (3) and (4) use the equivalent measure from the survey. Columns (5), (6) and (7), (8) show the same results as in (1) - (4) but using the comparable income measure. The following covariates are included in the regressions but suppressed here for presentational purposes: demographic characteristics (number of children, marital status, age and gender), proxies for wealth (type of tenure of the dwelling, number of rooms, the number of stories, type of dwelling, the local housing benefit, i.e. Accommodation Supplement area where the house is located, and the region), the annual variability of household income to proxy for income risk and its average growth rate over three years to proxy the stability of household finances); and time dummies. Asterisks indicate significance at the following levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. CI = confidence interval. The tables reports several instrument tests of the validity of the instrument set. To test the weak instruments and underidentification, it reports the F-statistic of the first stage (Kleibergen-Paap Wald F) and the Kleibergen-Paap LM test. Critical values for the first are not tabulated. They can be compared to the Stock-Yogo critical values; alternatively, Baum, Schaffer and Stillman (2007) suggest a rule-of-thumb that the F-statistic should be higher than 10 for weak identification issues not to be a problem. Rejection of the null in the Kleibergen-Paap LM tests for underidentification indicates that the smallest canonical correlation between the endogenous variables and the instruments is nonzero. The Hansen J test provides the robust version of the Sargan statistic of overidentification.

APPENDIX D: Testing Income-Gap Covariates

Table D1 Survey and register regressions on external covariates

Income Variable: Labour Income	(1) : Register	(2): Survey	(3): Survey-Register
Age	0.079***	0.055***	-0.025***
	(0.011)	(0.011)	(0.008)
Age^2	-0.001***	-0.001***	0.000***
	(0.000)	(0.000)	(0.000)
Female	-0.112***	-0.134***	-0.022
	(0.026)	(0.026)	(0.019)
Couple	0.833***	0.764***	-0.070***
	(0.031)	(0.03)	(0.022)
Number: Children	-0.202***	-0.156***	0.045***
	(0.024)	(0.024)	(0.017)
Number: House Stories	0.021	0.038	0.016
	(0.032)	(0.032)	(0.023)
House	-0.045	-0.022	0.023
	(0.043)	(0.043)	(0.03)
Other Dwelling	0.172	0.148	-0.024
	(0.267)	(0.267)	(0.191)
Number: Rooms	0.062***	0.051***	-0.011
	(0.01)	(0.01)	(0.007)
Accommodation: Rented	-0.269***	-0.300***	-0.031
	(0.034)	(0.034)	(0.024)
Accommodation: Owned by Trust	0.145***	0.133***	-0.013
	(0.041)	(0.041)	(0.029)
AS: Area 2	-0.205***	-0.227***	-0.022
	(0.056)	(0.056)	(0.04)
AS: Area 3	-0.400***	-0.400***	0.000
	(0.073)	(0.073)	(0.052)
AS: Area 4	-0.527***	-0.512***	0.015
	(0.078)	(0.077)	(0.055)
Region: Canterbury	0.146**	0.214***	0.069
	(0.069)	(0.069)	(0.05)
Region: Rest North Island	0.087	0.099	0.013
	(0.065)	(0.065)	(0.047)
Region: Rest South Island	0.049	0.084	0.035
	(0.066)	(0.066)	(0.047)
Region: Wellington	0.223***	0.274***	0.051
	(0.06)	(0.06)	(0.043)
Year 2007	0.041	0.090**	0.049
	(0.045)	(0.045)	(0.032)
Year 2009	0.049	0.098**	0.049
	(0.046)	(0.046)	(0.033)
Year 2010	0.153***	0.185***	0.032
	(0.044)	(0.044)	(0.031)
Year 2012	0.062	0.131***	0.069**
	(0.046)	(0.046)	(0.033)
Year 2013	0.093**	0.103**	0.01
	(0.046)	(0.046)	(0.033)
Growth (Income)	0.168***	0.084***	-0.084***
	(0.031)	(0.031)	(0.022)
Volatility Income	-0.380***	-0.236***	0.144***
	(0.045)	(0.045)	(0.032)
Self-Employed	**-0.090*** **	**0.009**	**0.099*** **
	(0.031)	(0.031)	(0.022)
Constant	8.961***	9.471***	0.510***
	(0.253)	(0.253)	(0.181)
Observations	2,577	2,577	2,577
R-squared	0.400	0.366	0.039

APPENDIX E: Testing the Robustness of Estimated Income-Gaps

As noted in section 5, results there allowed for a number of alternative income and expenditure definitions and data-related aspects. This Appendix examines the robustness of those results to four other definitional/data issues. In particular, it addresses the following questions.

(i) Do the self-employed have a preference for eating out and does this affect under-reporting estimates?
(ii) Do differences in non-comparable income of the self-employed matter?
(iii) Does allowing for differential access to wealth, capital income or trust ownership make a difference?
(iv) Do the self-employed make use of business expenses for personal consumption?

We examine those four cases in turn below.

(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 2. 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 equation (4) but where the dependent variable is the share of food eaten out in total food expenditure; see Table E1.

Table E1 Testing for Differences in Food Preferences

	(1)	(2)	(3)
Dependent variable: *Share of food eaten out in total food expenditure*			
SE Dummy	0.0203**	0.0252***	0.17
	(0.009)	(0.009)	(0.11)
Income		0.0511***	0.0511***
		(0.006)	(0.006)
SE Dummy∗Income			-0.013
			(0.01)

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.

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.[44]

[44] 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.

(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.[45]

To explore this, we re-estimate Table 2 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 E2, which shows income-gap estimates based on food/non-durable expenditure and labour/comparable income.

Table E2 Allowing for Non-Comparable Income

Self-Employment Definition: Opportunity		(1) NCI<25% Total	(2) NCI<10% Total	Table 1 Full sample
Expenditure	**Income**			
Food	Labour	0.197***	0.175***	0.200***
		(0.05)	(0.051)	(0.057)
Food	Comparable Income	0.201***	0.179***	0.193***
		(0.047)	(0.048)	(0.048)
Non-Durables	Labour Income	0.197***	0.171***	0.204***
		(0.041)	(0.042)	(0.047)
Non-Durables	Comparable	0.201***	0.175***	0.196***
		(0.038)	(0.04)	(0.040)
N		2,376	2,172	2,580

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 2 regressions, though this is only possible for survey-based data. Undertaking this exercise reveals, for example, that, the Table 2 estimate of underreporting using food expenditure/labour income of 0.114, becomes 0.094, and for non-durables

[45] 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.

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.

A potential further source of assets and income not recorded in our dataset arises when individuals own trusts. In New Zealand many individuals hold assets in trusts as a form of asset protection and/or tax minimisation. They are widely held by New Zealand households and not simply by those with high, or self-employment, income. During the years covered by our data the top marginal personal income tax rate and the rate applied to trust income were misaligned (2001-2010), and it was common to observe income flowing through trusts, as opposed to personal income, to benefit from the lower trustee rate.

If this occurred more for the self-employed, it would imply that there is income not being captured in our measure of reported (personal) income that could nevertheless be financing higher expenditure by the self-employed. 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. A similar argument applies to ownership of rental properties which provide a source of wealth and potential tax-free capital gain income.

These sources of capital gains and proxies for wealth might be especially pronounced among the self-employed yielding higher expected income to finance higher expenditure. 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.

Unfortunately, income from trusts is not covered in our register data, but whether the individual receives income from a trust is indicated in the survey. We therefore use a 'flag' for those households receiving trust income, rental income (or a combination of the two) to proxy for this type of asset that could signal higher expected income. As our aim is to understand whether the income-gap we observe is partly due to a wealth effect, using register income (and the survey trust flag) we eliminate households with rental income in column (1) of Table E3, those receiving trust income in column (2) and a combination of the two in column (3). As can be seen, the income-gap results essentially remain unchanged.

While, in principle, receipt of income from trusts is recorded in the survey, the response rate to the question is relatively low. This may be due to recall problems or because income from trusts can remain as trust (as

opposed to personal) income, or can be distributed to beneficiaries. There is however a question in the survey that we can use to instrument ownership of a trust.

Householders are asked about their house ownership and a non-negligible number claim their house is owned by a trust. Responding to the survey that the house is owned by a trust, together with evidence on receipt of trust income, can be treated as an alternative flag for trust *ownership* (as distinct from evidence of *income received* from a trust above). This increases substantially the number of households in our category of potentially accruing capital gains or with indicators of wealth. We use this alternative proxy for trust ownership in Table E4, Panel A, and trust ownership and/or receipt of rental income as a proxy for wealth in Panel B. We run regressions on equation (6) to estimate a within-group comparison of self-employed, with and without potential for capital gains, and their employee counterparts.

A test of the equality of income-gaps reveals that the income-gap between the self-employed with higher wealth (as marked by our flag) and the self-employed with lower wealth is not statistically different implying that even when we control for wealth groups, or estimate income-gaps within each group, the estimates of the income-gap do not significantly vary. To the extent that data allow, it would appear that even narrowing down the sample to households less affected by potential wealth effects, income-gap estimates remain robustly around 20%. Nevertheless, income-gap estimates '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

Overall, it seems reasonable to conclude that our central income-gap estimates are largely uncontaminated by 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 trust or closely-held company) such that the 'underreported' personal income we observe here need not imply a commensurate underpayment of tax.

A final 'wealth' issue we address 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. Alternatively, when income if temporarily higher, they may divert income into saving rather than expenditure. As a further check we run our previous regressions but with total expenditure as the dependent variable. If we obtain a similar estimate of underreporting by the self-employed it would indicate that our previous results are not largely driven by a general tendency of the self-employed to save more or less than their employee counterparts. Table E5 shows the results of those regressions. It is clear that the underreporting estimates are very close to the 20% reported in Table 2 using the same register data, providing further confirmation that our earlier results are robust to caveats around savings differences.

Table E3 The Effect of Wealth

Sample Selection Criterion		(1) No Rental	(2) No Trust	(3) No Rental or Trust
Expenditure	**Income**			
Food	Labour	0.204***	0.226***	0.232***
		(0.062)	(0.067)	(0.073)
Food	Comparable	0.184***	0.218***	0.211***
		(0.053)	(0.055)	(0.061)
Non-Durables	Labour	0.202***	0.201***	0.206***
		(0.05)	(0.053)	(0.057)
Non-Durables	Comparable	0.183***	0.198***	0.189***
		(0.043)	(0.043)	(0.047)
N: Self-Employed		1812	1734	1656
N: Employed		597	492	447

Table E4 Proxying for Wealth and Capital Gains (CG)

Income type:	Labour Income		Comparable Income	
Expenditure category:	Food	Non-Durables	Food	Non-Durables
	Panel A: Proxy for Trust Ownership			
SE vs. E with CG	0.079	0.193**	0.050	0.151*
	(0.129)	(0.098)	(0.114)	(0.089)
SE vs. E with No CG	0.224***	0.205***	0.222***	0.205***
	(0.063)	(0.052)	(0.053)	(0.043)
Test of Income-Gap equality:				
Chi^2	1.190	0.012	2.232	0.339
p-value	0.275	0.914	0.135	0.560
	Panel B: Proxy for Trust Ownership & Rental Income			
SE vs. E with CG	0.120	0.201**	0.050	0.151*
	(0.109)	(0.087)	(0.114)	(0.089)
SE vs. E with No CG	0.222***	0.202***	0.222***	0.205***
	(0.065)	(0.053)	(0.053)	(0.043)
Test of Income-Gap equality:				
Chi^2	0.753	0.000	0.466	0.040
p-value	0.386	0.990	0.495	0.841

Note: Estimates of equation (6) include an intercept for SE and E with and without the proxy for capital gains: proxy for trusts in Panel A and proxy for trusts and rental in Panel B. To obtain the income-gap we use as a benchmark group for the self-employed with and without capital gains and the corresponding group of employed. This allows a within-category (capital gains flag and no capital gains flag) comparison. Observations numbers for each group are shown in Appendix C, Table C5. An alternative specification that compares SE with CG, SE with no CG against a baseline of employed is also estimated; results are shown in Table C6.

Table E5 Testing Robustness to Expenditure-Savings Differences

	(1)	(2)
Income type:	Labour (IR)	Comparable (IR)
Total Expenditure	0.190***	0.183***
	(0.034)	(0.029)
Total Expenditure excl. housing/mortgage[(1)]	0.209***	0.199***
	(0.036)	(0.03)
Total Expenditure excl. housing/mortgage & durables[(2)]	0.212***	0.201***
	(0.037)	(0.031)

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 infrequency of purchase/telescoping errors.

(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 **Error! Reference source not found.**) should be shifted down when expenditures vulnerable to misclassification are used.[46] 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

[46] 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.

Table E6 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 2, 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 E6 Testing Robustness to Business Expenses

Income type:	(1) Labour (IR)	(2) Comparable (IR)
Potential Expenditure Misclassification:		
(1) Utilities, Transport, Communication	0.085	0.096
	(0.086)	(0.067)
(2) Utilities, Transport, Communication, Housing	0.112**	0.118**
	(0.055)	(0.046)

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 infrequency of purchase/telescoping errors.

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 E6 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.

About the Authors

Ana Cinta G. Cabral until recently was a researcher and analyst at Inland Revenue, New Zealand.
Email: anacintagonzalez@gmail.com

Norman Gemmell is Professor of Public Finance at Victoria Business School, Victoria University of Wellington, New Zealand.
Email: norman.gemmell@vuw.ac.nz

Nazila Alinaghi is a Research Fellow in Public Finance at Victoria Business School, Victoria University of Wellington, New Zealand.
Email: nazila.alinaghi@vuw.ac.nz

VICTORIA UNIVERSITY OF WELLINGTON
TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

Chair in Public Finance
Victoria Business School

Working Papers in Public Finance