BANK OF CANADA
BANQUE DU CANADA

Working Paper/Document de travail
2015-22

Sheltered Income: Estimating Income Under-Reporting in Canada, 1998 and 2004

by Geoffrey R. Dunbar and Chunling Fu

www.bank-banque-canada.ca

Bank of Canada Working Paper 2015-22

June 2015

Sheltered Income: Estimating Income Under-Reporting in Canada, 1998 and 2004

by

Geoffrey R. Dunbar[1] and Chunling Fu[2]

[1]Currency Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
and
Department of Economics
University of Ottawa
Ottawa, Ontario, Canada K1N 6N5
geoffreydunbar@bankofcanada.ca

[2]Vancouver School of Economics
University of British Columbia
Cheryl.Fu@ubc.ca

ISSN 1701-9397

Acknowledgements

We thank Kim Huynh, Ben Fung, Maren Hansen, Brian Peterson, Katya Kartashova, Kevin Milligan, Krishna Pendakur, Simon Woodcock, Pierre Duguay, and seminar participants at the Bank of Canada, Finance Canada and the Canadian Economics Association Annual meetings for helpful comments and suggestions. We thank Statistics Canada for access to the data at the Research Data Centres of the University of British Columbia and the University of Ottawa. All remaining errors are our own.

Abstract

We use data from the Survey of Financial Security and the Survey of Household Spending to estimate the incidence and extent of income under-reporting in Canada in 1998 and 2004. We estimate that the proportion of households under-reporting income is roughly 35 to 50 per cent in both years. Our estimates also suggest that the amount of under-reported income rose by roughly 40 per cent between 1998 and 2004 and remained stable as a proportion of GDP of 14 to 19 per cent. We find evidence that income under-reporting is pervasive and is not confined to households that report self-employment income in the survey data. We also find that poverty measures that rely on reported income appear unreliable because under-reporting necessarily implies a lower reported income. Thus, households that under-report appear to be poorer. We propose a simple ratio method of identifying households that under-report income using the household's budget share on shelter.

JEL classification: H26, I32, K42
Bank classification: Domestic demand and components

Résumé

Les auteurs utilisent des données de l'Enquête sur la sécurité financière et de l'Enquête sur les dépenses des ménages pour mesurer l'incidence et l'ampleur de la sous-déclaration des revenus au Canada en 1998 et en 2004. Ils estiment que la proportion des ménages déclarant des revenus inférieurs à ceux qu'ils perçoivent réellement se situe, en gros, entre 35 et 50 % pour ces deux années. De plus, selon leurs estimations, la valeur totale des revenus sous-déclarés s'est accrue de quelque 40 % entre 1998 et 2004, et elle est demeurée stable en proportion du PIB, soit entre 14 et 19 %. Les auteurs constatent que la sous-déclaration des revenus est très répandue et qu'elle ne se limite pas aux ménages qui, selon les données des enquêtes, déclarent des revenus d'un travail indépendant. Ils soulèvent également des doutes quant à la fiabilité des statistiques sur la pauvreté établies à partir du revenu déclaré, étant donné que les ménages qui sous-déclarent leurs revenus paraissent nécessairement plus pauvres. Les auteurs proposent une méthode simple fondée sur des ratios pour déterminer quels ménages sous-déclarent leurs revenus, en s'appuyant sur la part du budget consacrée au logement.

Classification JEL : H26, I32, K42
Classification de la Banque : Demande intérieure et composantes

Non-Technical Summary

Quantifying the size of the underground economy in Canada is difficult and there are typically wide discrepancies in estimates depending on the methodology. Statistics Canada uses national account discrepancies to estimate the level of unreported income at roughly 3 per cent of GDP. Previous studies using microeconomic consumption data, Schuetze (2002) and Fortin *et al.* (2010), estimate the level at roughly 5 per cent. Studies using macroeconomic data, Mirus and Smith (1981), Mirus *et al.* (1994), Giles and Tedds (2002), and Schneider (2012), generally estimate the size at roughly 15 per cent.

Choosing between estimates has been challenging because of the lack of agreement between methodologies. This study helps to resolve the difference between macrodata and microdata estimates. Its findings suggest that the methodology used by Statistics Canada may be misleading.

Previous studies using microdata, Schuetze (2002) and Fortin *et al.* (2010), relied upon the assumptions that only self-employment income can be under-reported and that all individuals truthfully report all their sources of income. This paper shows that these assumptions are at odds with the data and lead to an underestimate of unreported income.

This paper uses household-level income data and some consumption data from the Survey of Financial Security (SFS) and imputes missing consumption data for each household from the Survey of Household Spending. The income data in the SFS are taken from administrative (Canada Revenue Agency) data for 80 per cent of respondents. We conduct many robustness exercises for the imputation without changing the results. We acknowledge that our results are sensitive to the imputation, and best viewed as statistical evidence.

Roughly 30-40 per cent of households that report salaried income appear to spend more than their after-tax income. For households that do report some self-employment, the percentages are 10-20 points higher. Overall, roughly 35-50 per cent appear to under-report some income.

Income under-reporting skews statistics based on reported income because a household under-reporting income necessarily reports that it is poorer than it actually is. Sixty to seventy per cent of households reporting income of less than $20,000 a year appear to be under-reporting income.

Economic theory predicts that households that under-report income should appear to have higher budget shares of consumption than comparable households that do not under-report. This prediction is confirmed in the estimates. This paper reports thresholds at which a household is likely to be an under-reporter.

The data used provide no guidance as to how households under-report income or what methods of payment are used to avoid reporting income.

1 Introduction

It is well understood that individuals have an incentive to lower their income-tax obligations to a tax collector because doing so provides more discretionary income. One method to lower taxes is to under-report income. Depending on the tax and transfer system, under-reporting taxable income may also trigger eligibility for public transfers. As a consequence, most tax collectors are granted enforcement powers that can include the power to audit, levy penalties and, in some jurisdictions, pursue criminal charges including incarceration. Measuring unreported income is typically challenging for tax authorities because the precise purpose of under-reporting is to conceal income from these authorities.

In this paper we provide estimates of the proportion of all households that under-report income (the extensive margin) and the estimated level of tax evasion (the intensive margin) in Canada. Our estimates suggest that roughly 35 to 50 per cent of all households did not report at least some taxable income to the Canada Revenue Agency (CRA) in both 1998 and 2004. While large, these estimates are similar to estimates of tax evasion for the United States. Andreoni *et al.* (1998) report that estimates from the Taxpayer Compliance Measurement Program (TCMP), a randomized audit conducted by the Internal Revenue Service (IRS) in the United States, suggest that roughly 40 per cent of households under-reported their income to the IRS in 1988. Our estimates suggest that an upper-bound estimate of the amount of under-reported income for households that under-report was roughly $15,000 ($22,000) to $20,0000 ($30,000) in 1998 (2004), and depended in part on the level of reported income. Aggregated to the national level, we find that under-reported income as a percentage of GDP was 14-19 percent in both years, which is in line with estimates reported by Mirus and Smith (1981), Mirus *et al.* (1994), Giles and Tedds (2002) and Schneider (2012) for Canada using macroeconomic data. Indeed, we believe that our study is the first to reconcile estimates of under-reporting using microdata to those obtained using macrodata for Canada.

Similar to Tedds (2010), we find evidence that the probability that a household under-reports income is a non-linear and decreasing function of income. The reason is straightforward. By under-reporting income, a household's reported income looks lower than its true income, and so the relative proportion of under-reporting rises as the level of reported income falls. Our results do suggest that anti-poverty initiatives based on reported incomes may not always benefit the intended target group and may, in fact, transfer income from poorer households that truthfully report income to richer households that do not. Finally, we note that the data used in our study provide no evidence on the mechanisms or payment methods used to earn unreported income. There are likely many methods that can be used to hide income and our data do not permit investigation of these channels.

Previous studies using microdata for Canada, Schuetze (2002) and Fortin *et al.* (2010), typically suggest that the level of unreported income is roughly 5 per cent of GDP. These studies use a methodology pioneered by Pissarides and Weber (1989), which assumes that only some types of income, usually self-employment, can

be under-reported and that all sources of income are reported by households. Our second contribution is to challenge this assumption by showing that under-reporting of taxable income does not appear to be confined to households that report self-employment. In contrast, we find that roughly 30-40 per cent of households that report only salaried income under-report income in our sample. We argue that there are many methods of earning income that may not be reported or identified as self-employment on survey questionnaires. Also, there would appear to be little reason for an individual to report being self-employed if the sole purpose of that self-employment was to earn additional, unreported, income.

Nevertheless, we do find that the *relative* incidence of under-reporting for workers who self-identify as self-employed is roughly 10-20 per cent higher than for salaried workers, which is similar to the relative magnitudes obtained by Schuetze (2002) using a methodology proposed by Pissarides and Weber (1989). As emphasized by Slemrod (2007), the methodology of Pissarides and Weber (1989) identifies the relative incidence of under-reporting by the self-employed, and thus we view the similarity of our relative results with those of Schuetze (2002) as corroborative evidence for our measure of under-reporting. We note that a related approach suggested by Lyssiotou *et al.* (2004) uses a complete demand system specification and thus does not need to rely on a sharp classification of self-employed and not-self-employed to invert budget shares. However, this approach is sensitive to households truthfully reporting their income sources (see our caution above), and yields estimates for under-reporting relative to a benchmark for some income source that is assumed to be truthfully reported. Fortin *et al.* (2010) apply this methodology to income under-reporting in the province of Quebec, Canada, and estimate under-reported income to be roughly 4.6 per cent of provincial GDP in 1997, rising to 5.6 per cent in 2002. We note that in the sample used by Fortin *et al.* roughly 9 per cent of households have income earned from self-employment, and so our estimate of the (implied) aggregate intensive margin of under-reporting by the self-employed (roughly 50 per cent) appears to be similar.

Gervais (1994) adopts a different approach using discrepancies in national accounts, combined with an account-specific multiplier to adjust for the ease of tax avoidance, to yield a measure of missing income. The OECD likewise recommends a similar methodology and this is the methodology underlying Statistics Canada's measure of the underground economy.[1,2] Terefe *et al.* (2011) follow this approach and estimate unreported income as roughly 3 per cent of GDP. One key issue with the account-specific multiplier approach is that it assumes that certain proportions of income or consumption can be misstated and then applies these proportions to national accounts data. However, estimating the proportions that are misstated requires knowledge of the true income data, which is precisely the object that is unknown.

Our third contribution is to propose a ratio statistic for identifying the probability of income under-reporting by a household based on budget shares of consumption items. We propose a ratio of shelter costs (mortgage payments or rent) to income as an indicator of under-reporting. Our basic intuition is that

[1] See Colledge (2002).

[2] See http://www.cra-arc.gc.ca/nwsrm/fctshts/2012/m09/fs120927-eng.html, doi 11/1/13.

households spend according to their true permanent income and not their reported income. In our data, we find that most households that under-report their income have mortgage-payment-to-income ratios (MIR) or rent-to-income ratios (RIR) in excess of households that do not under-report.[3] Our method appears related to the theoretical literature on tax evasion. Richter and Boadway (2005) argue that taxing an observable good in the Allingham and Sandmo (1972) model improves the efficiency of taxation. Rather than introduce a new tax, our empirical methodology provides an easy method to detect possible tax evasion provided that governments collect consumption expenditures – in our example, shelter costs. Thus, similar to Richter and Boadway, we base our approach on the notion that consumption decisions depend on true permanent income, not reported income. We note that income-tax filers in Canada are not required to report consumption data to the CRA as a normal part of their tax filing, so using a MIR or RIR is beyond the CRA's scope.

Our strategy to identify income under-reporting is to compare our estimate of a household's total annual expenditure to its reported annual income. The data for our study come from the 1999 and 2005 waves of the Survey of Financial Security (SFS) and the 1998 and 2004 waves of the Survey of Household Spending (SHS) (the SFS data are lagged one year by construction and thus this correspondence is the appropriate match).[4] The SFS collects data on household income including, for most households, the income actually reported to the CRA and some data on household consumption, which account for roughly a 30 per cent share of a typical household's budget. To calculate total expenditure for a household, we impute the missing consumption data from the SHS into the SFS data, conditioning on the demographic structure, location and expenditure items that are reported in both surveys. We do this in two different approaches, to highlight the sensitivity of our results to the imputation. In the first, we impute non-durable consumption from the SHS and add this to the expenditure data reported in the SFS to calculate our first measure of total expenditure. This approach is conservative in that it omits semi-durables and durables that are not financed (and thus not reported in the SFS). Our second approach is to impute the total consumption for the household from the SHS, again conditioning on the demographic structure, location and expenditure items that are reported in both surveys. This approach may exaggerate under-reporting, since it implicitly adds at least some durable consumption to all households. We view our measures as the likely lower bound and likely upper bound on under-reported income. We also repeat both approaches, adding the household's reported income to the conditioning set, and compare our results. Finally, for both approaches, we account for imputation error by adding random draws from the weighted distribution of imputation errors to construct five implicates per household. These implicates provide some evidence as to the sensitivity of our results to the imputation procedure. We find our estimates of the average incidence and level of underreporting are very similar across

[3]Indeed, a non-negligible fraction of households in our data report positive savings and a MIR or RIR greater than 1. A related question not addressed in this paper is exactly how such households qualify for mortgages from financial institutions or pay their rent.

[4]The SFS is an irregular survey conducted by Statistics Canada in 1999, 2005 and 2012. The most recent 2012 survey data were not available to researchers at the time this study was undertaken.

the implicates. This suggests that the imputation procedure is not driving our findings. Nevertheless, we caution that our approach is of a statistical nature and is not unequivocal evidence of the level of income under-reporting in Canada.

In effect, our approach relaxes the assumption that an identifiable set of households can be defined as truthful income reporters at the cost of assuming that survey coverages, adjusted for sampling weights, are in some sense close. The detailed description of the SHS and our imputation methods can be found in Section 4. We note that our imputation procedures yield aggregate moments in the SFS that are quite similar to those in the SHS. To the extent that there is some under-reporting of consumption or savings, our approach will be conservative both on the intensive margin and the extensive margin for households (for instance, we do not observe illegal drug expenditures or money laundering in the SHS). At the most basic level, our results suggest it is difficult to reconcile data on household incomes and household expenditures in Canada.

Imputing consumption is not without precedent. Skinner (1987) imputes consumption from the Survey of Consumer Expenditure (CEX) into the Panel Study on Income Dynamics (PSID), considers a range of control variables, and proposes two approaches: a reduced form and an extended form. Palumbo (1999) extends Skinner's imputation approach and also proposes a structural model of household expenditure. Blundell *et al.* (2004) invert a food demand equation estimated from the CEX to yield total consumption in the PSID. They compare their approach to that of Skinner (1987) and find that the latter, while underestimating the level of consumption, does match the variance of log consumption reasonably well. Unfortunately, the SFS does not collect food expenditure and we are unable to exploit the micro-founded approach of Blundell *et al.* (2004) because none of the expenditure items observed in both surveys is sufficiently normal to yield confidence in a demand estimation. However, the results of Blundell *et al.* (2004) suggest that our imputed consumption is biased lower and thus our results are likely to be conservative. Finally, Fisher and Johnson (2006) impute consumption from the CEX to the PSID using a broader range of control variables (mainly demographic) than Skinner (1987) and compare their approach to the latter and to Blundell *et al.* (2004). They suggest that imputing using a wide range of demographic information yields the most plausible estimates. Since the SFS data do not allow us to follow Blundell *et al.* (2004), we impute consumption from the SHS to the SFS, similar to Fisher and Johnson (2006). We note that we do not use imputed values to explain a household behaviour, but instead use our imputed values to calculate simple averages of outcomes across all households. Thus, while we carefully consider how imputation error may affect our averages, we are agnostic over issues such as how imputation error would affect estimated coefficients from a regression relating the imputed values to household behaviour.

Our paper proceeds as follows. In Section 2 we describe our proposed methodology to identify income under-reporting households. In Section 3 we describe the two data sources, SFS and SHS, used in our study. In Section 4 we describe our imputation approach for consumption, savings and dissavings. In Section 5

we present results for the extensive and intensive margins of income under-reporting for 1998 and 2005. In Section 6 we illustrate the effect of income under-reporting on poverty statistics and income tax revenues. In Section 7 we derive the theoretical relationship between budget shares and under-reported income and illustrate empirically how budget shares on shelter map to our measures of income under-reporting. Section 8 concludes.

2 Methodology Overview

One method to identify income under-reporting is to compare reported income from both the employee and employer. The CRA does compare the income reported by an employer and an employee as one measure of reliability, but such an approach is only feasible if both an employer and an employee report income. Indeed, in Canada, contractual employment may not necessitate payroll reporting by the employer if such employment is deemed to be the contracting of a self-employed worker. For such workers, there is therefore no employer record to which one can compare reported income even if the employee is a sole-source contractor for a given employer. Moreover, in such relationships it is unclear how an employee would report their self-employment status on survey questionnaires.

There are also many methods of earning income that may not even be reported or identified as self-employment. It would seem unlikely, for example, that individuals who moonlight in illegal service industries, such as the provision of drugs, even report being self-employed as such on survey questionnaires. Certainly in the SFS data, we did not find households self-identifying as sellers or producers of marijuana, for example. Yet Easton (2004) estimated 17,500 illegal marijuana cultivators in British Columbia, producing $2 billion in annual revenues. Diplock *et al.* (2013), using more recent data, estimated the size of marijuana cultivation in British Columbia at 13,000 commercial operations producing $4 billion in annual revenues. More generally, it would seem unlikely that a household earning unreported income from self-employment (legal or otherwise) would report being self-employed while simultaneously choosing not to report the associated income.[5] Finally, some individuals may earn unreported offshore income which, given Canadian income tax rules, is required to be reported to the CRA. This last category may help explain why domestic expenditures are higher than earnings for some households, though we acknowledge that it is not clear that such income necessarily implies loss of income tax receipts if it is reported elsewhere.

We propose a reliability check for household income that does not rely on comparing income reports from multiple sources. Define Y as reported household income, C as household consumption, S as household savings, D as household debt payments and T as household tax payments. Then a simple accounting identity holds for after-tax income and household expenditures:

[5]This statement should not be confused with the converse – many households do in fact report both salaried and self-employment income.

$$Y - T = C + S + D.$$

Our reliability check for household income is to compare our estimates of household expenditures, $C+S+D$, to the after-tax income, $Y - T$, of the household (mostly taken from the CRA records). Throughout this study, we focus on households for which $C + S + D > Y - T$, whom we define as having unreliable income.[6] In particular, we use the survey weights to calculate the share of the total population for which $C + S + D > Y - T$. If Y, T, C, S and D are observable then our reliability check would be a tautology. Because our estimates of household expenditure may include estimation error (statistical noise) from the imputation we use, we examine how sensitive our calculated shares are to random perturbations by drawing four additional imputation samples, and find that the shares are robust across these samples. Although our reliability check is necessarily statistical in nature, we conduct several robustness exercises which we believe allow the reader to examine how sensitive our results are to departures from our baseline case. The income necessary for the accounting identity to hold for every household in the data is also a lower-bound estimate of unreported income, given that we are unlikely to measure every household consumption item.

3 The Data

The primary data sources for our study are the 1999 and 2005 Survey of Financial Security (SFS) and the 1998 and 2004 Survey of Household Spending (SHS), both collected by Statistics Canada.[7] The SHS and SFS have many of the same demographic, geographic and expenditure questions in common, which aids our imputation approach.

The SFS is a self-report survey of the assets and debts of Canadian households at the time of the survey and the income and expenses for the previous calendar year, 1998 and 2004, respectively. The SFS comprises two subsamples. The first subsample is drawn from the Labour Force Survey (LFS) sampling frame and reports households across the ten provinces excluding those households on Indian Reserves or located on federal institutions (such as military bases). The second subsample is drawn from high-income neighbourhoods to account for the disproportionate wealth held by these households. Survey weights are provided to balance the unequal selection probabilities and response rates so that the survey is representative of the Canadian population. Survey weights are provided for both years; however, we note that the weighting methodology did change between the 1999 and 2005 SFS. The 1999 SFS weights are based on province – age – sex groups to match population counts. In 2005, the weights were adjusted to include administrative

[6] In this study, we do not focus on households over-reporting income, because doing so would lead them to pay higher taxes and would not imply any loss of revenue for the tax authority.

[7] The 2012 SFS microdata were not released to researchers in time for our study. However, the microdata apparently do not include some of the questions we use for our study, notably the question regarding the comparison of a household's consumption to income.

data from the CRA and a second survey, the Survey of Labour and Income Dynamics. We discuss below the implications of the reweighting.

The SFS collects asset and liability information from each surveyed household, and income and demographic information from each adult (15+) respondent for the household. As pointed out in the Introduction, the income data we report are the same as the income data reported to the CRA (the federal government department responsible for taxation) for a subset of households and so are free from measurement error to the extent that reported income is our object of interest. Moreover, the data reported are both the household's gross income for the year, including government transfers, and the household's net after-tax income. Thus, the effect of tax shelters or tax credits (such as the investment tax credit) on household income is captured in the latter. Perhaps more importantly for our study, if the individual permits the CRA record linkage, then Statistics Canada considers these data complete and does not impute or alter them. For individuals that do not link their records to the CRA data, some income data may have been imputed by survey technicians, according to the SFS report. Since our estimates of under-reporting are sensitive to the reported after-tax income, we examine the sensitivity of our results to whether households self-report their income or link their income responses to the CRA records. We do not find much difference.

There are three notable advantages to using the SFS data set. The first advantage is that the sample design of the SFS is stratified into two samples: a main sample drawn from the LFS sampling frame, and a second sample drawn from high-income areas. The second sample specifically over-represents high-income households relative to the LFS sample frame and the SHS data. Thus, the SFS sample is less likely to underestimate (or overestimate) the extensive margin of income under-reporting for wealthy households because of small (sub) sample concerns. Therefore, to the extent that wealthy households may have either a higher incentive or an easier opportunity to evade taxes than relatively poorer households, the SFS sample should better identify the incidence of under-reporting for these households. Roughly 9 per cent of the 1999 SFS and 17 per cent of the 2005 SFS were drawn from the high-income subsamples.

A second reason for using the SFS is the breadth of information the SFS collects on household wealth and financial security. The SFS also asks households whether, excluding investments, automobile purchases and home purchases, a household's income is greater than, less than or equal to its expenses. We use the household's response to this question as a robustness check to ensure that we are not incorrectly classifying households that finance household consumption by increasing household debt as under-reporting their household income. We also interpret this question as asking how the household's typical consumption expenses, which must be less than or equal to its expenditure including investments, automobile purchases and/or home purchases, relate to its income. The distribution of responses to this question is reported in Table 1. We note that this measure of expenditure is, by definition, a lower bound on the household's total expenditure, since it excludes investments, automobile purchases and home purchases. We assume that the response

to this question is, therefore, transitive in that it also holds true for total expenditures which would include these items. We use the response to this question to decide whether to impute savings or dissavings (in addition) to consumption from the SHS for a particular household. If the household's typical consumption expenses are greater than or equal to its reported income but the household answers to the contrary, then we interpret this household as under-reporting. Because the SFS contains information on whether a household has purchased a home during the reference year, we also compare our measure of under-reporting both for new and existing homeowners. We also use information about household investments to determine whether a household is, in fact, a saver. We note that if a household reports income greater than spending, then that household should hold some assets as savings.

Table 1: Self-Reported Income vs. Spending

		1999 per cent	2005 per cent
(1) Dissavers:	Income < spending	16	19
(2) Paycheck:	Income = spending	43	36
(3) Savers:	Income > spending	42	45

Using just the expenditures reported in the SFS, we count the fraction of households for which expenditures are greater than income. Perhaps surprisingly, we find that roughly 4 per cent of households in the SFS appear to under-report their income by this measure. For self-employed households, the percentage of under-reporting households by this measure rises to 9.5 per cent. As well, we find that under-reporting is lowest when only one member of the household links their income data to the CRA records. At least two possible conclusions can be inferred from this finding. First, these households may have either misreported their income and expenses or the responses may have been miscoded. Certainly, it would be surprising if the survey was entirely free from error, particularly for self-reported after-tax income. However, a second conclusion is simply that these households are in fact reporting truthfully and either ignore or do not care that their responses are at odds with the data they provide. We are unable to differentiate between these interpretations. Nevertheless, the fraction of households that fall into this category are small and do not seem to affect the overall qualitative conclusions we draw in this paper.

We conduct robustness checks using other survey questions in the SFS regarding asset sales, monetary gifts and pawnbroking. The SFS asks households whether they have needed to sell an asset or deposit an item at a pawnshop in order to pay off a bill, whether the household is behind on a debt repayment or whether the household has received any gift money. We use responses to these questions to construct a financial stress indicator for households. Conditioning our measures of income under-reporting on the financial stress indicator does not affect our results.

The third advantage of using the SFS data set (for at least the CRA-linked income) is that there does

not seem to be an attempt by the survey methodologists to reconcile apparently contradictory income and expenditure entries, as appears to be the case with the SHS. Even setting aside concerns that a household may condition its responses to the SHS questions to avoid any inconsistencies between income and expenditure, the survey codebook for the SHS reports that "The senior interviewer reviews the questionnaire and attempts to balance the household expenditures and income within 15 per cent of each other." Leaving aside the question of exactly how the senior interviewer accomplishes this balancing, which is clearly problematic, we are also concerned that interviewers may not remain neutral if households appear to offer inconsistent responses. Indeed, as a practical observation, we found that a number of mortgage payments in the SHS appeared to be imputed when the original reported payment exceeded roughly 33 per cent of household income. An additional concern with the SHS is that in both 1998 and 2004 the survey weights are chosen in part to match the income distribution reported to the CRA which, as this paper argues, may be unreliable.

The SHS is a self-report annual survey of detailed spending and income of Canadian households across all provinces and territories.[8] The SHS includes data on the savings (dissavings) flows of households and detailed consumption data. One concern is that the SHS is a recall survey for the 1998 and 2004 survey years. Thus, households are asked to recall their consumption expenditures some period after those consumption decisions were made. This introduces concerns that our measures of household consumption suffer from recall error, which also may be correlated with the likelihood of a household under-reporting its income. We have no direct evidence of survey recall error for the 1998 and 2004 SHS data. Tremblay *et al.* (2010) report, using data from a pilot survey for a major survey redesign in 2009, that the estimates of total household expenditure for diary respondents and recall interview respondents were, on average, within a range of -3.0 to +0.2 per cent. Thus, for the purposes of our exercise, survey recall error does not appear to be a significant concern.

To ensure that the samples from each survey are comparable, we remove part-year households, multi-family households and households living in the territories from the SHS data. We also remove households with extremely low income (before-tax income of less than $300).[9] In addition, we remove any households with missing data from the SFS. Our working sample consists of 13,358 and 13,272 cases for the 1998 and 2004 SHS, and 14,191 and 4,340 cases for the 1999 and 2005 SFS, respectively. Our sample selection procedures remove approximately 10 per cent of the records for both the 1999 and 2005 SFS using the survey weights. We note that the survey size for the 2004 SFS is much smaller than for 1998 because Statistics Canada reduced the survey sample for the second wave.

We report the demographic characteristics of households in the SFS and SHS in Table 13 by comparing the weighted means and standard deviations of some of the household characteristics from these two data

[8]The territories are covered only in selected years.

[9]We notice that there are quite a few households with an identical low-income level in the SFS, which appear to be possibly imputed by survey technicians. There are no such observations in the SHS. We remove these income outliers from both the SHS and SFS data since they will likely bias our imputation results.

sources by year, including demographic characteristics, type of dwellings, size of the area of residence, home ownership status and vehicle ownership status. As the table indicates, most of the characteristics of these two data sets are very similar, regardless of the inclusion of the 'high-wealth' subsample in the SFS. In the SHS, the reference person, defined as the person with the most knowledge of the family's financial situation, is slightly older than the reference person in the SFS. There is also a higher percentage of married households in the SHS and this affects the average spousal ages (in our code, zero is recorded as the spousal age when no spouse is present). Comparing spousal ages across the two surveys only for households with spouses present yields higher average ages, roughly three years younger than the average reference person. Spouses in the SHS are also roughly two years older than spouses in the SFS. Home ownership rates are also slightly higher in the SHS. Finally, there appear to be more households drawn from large urban areas in the SFS as compared to the SHS (this is most noticeable in 2005).

4 Imputing Expenditure from the SHS

Our proposed methodology for identifying under-reported income requires us to estimate the household's expenditures – consumption, savings flows and dissavings flows. Both the SFS and the SHS have identical expenditure data for rent; mortgage payments; electricity, water, oil and gas, and vehicle registration expenses; property taxes; condo fees; childcare expenses; and spousal support payments. (Hereafter, we refer to these common expenditure items as ongoing expenses.) Although the share of such expenses does vary widely across households, for a typical family these expenses are roughly 30 per cent of household expenditure. Table 12 compares the mean and standard deviations of these ongoing expenses by year. Most of the individual items and the total ongoing expenses in the two data sets are similar. The two exceptions are the mortgage and rent payment: households in the SFS on average pay more on mortgages and rent. This is consistent with the fact that there are more families (and more homeowners) in the SHS. In our imputation procedures we take into account this difference in sample composition of the SHS and the SFS and estimate expenditures for families and individuals, renters, and owners separately. We use these consumption items to help control for household consumption preferences. For instance, some households may prefer to spend a larger fraction of their income on housing by reducing their consumption of other items such as a vehicle. A classic example is that one couple may prefer to live in an expensive urban condo and take public transit, while another household may prefer to live in a suburban house and drive an SUV.

One advantage of having to impute expenditure into the SFS is that households should have less incentive to misreport their answers in the SFS, since they should have little reason to believe that their responses can be verified through a straightforward comparison of expenditure and income. Nevertheless, our approach is clearly susceptible to errors in our imputed estimates of household expenditure. For example, for most households there is some value of consumption that can be assigned such that the household's expenditures

are less than or equal to its income. Whether these consumption values are reasonable for that household is the key question. To ensure that our imputed consumption distributions match as closely as possible the distributions observed in the SHS, we separate households into four distinct types: single renters, single homeowners, family renters and family homeowners. Then, conditioning on the demographic structure, location and common expenditure items (including some polynomial terms, as we discuss in this section), we estimate a linear regression to predict consumption (or non-durable consumption) for each type using the survey weights for the SHS data. We note that for the SHS regressions, the R^2 values are all above 0.8 and in most cases over 0.9. Conditional on the survey weights and covariates, we use these regression coefficient estimates to predict the consumption for that household type in the SFS. This step effectively matches our distributions for the conditional means for each type. We then construct four additional implicates (values of imputed consumption) for each household by drawing a random error from the SHS linear regression and adding this to the value imputed in the SFS. Thus for each household we have five imputed values of consumption, at least four of which effectively match the cross-sectional conditional distributions for each type in the SFS.[10] The sample correlations, adjusting for the SFS survey weights, between the household consumption implicates are typically around 0.7. Thus, the implicates provide variation at the margin across the sample and therefore any conclusions we draw from our sample are robust to reasonable perturbations of the household's estimated consumption.[11] We also test our measures of under-reporting to errors in imputation by assuming that our estimates are inaccurate for ranges of $2,000 to $10,000 per household, which accounts for non-zero mean errors in the imputation. While these obviously affect our estimates of the level of under-reporting, they do not change our qualitative conclusions regarding the extensive and intensive margins.

Using the SHS data, we estimate the household's consumption expenses as follows:

$$C_{i,j} = \alpha_{i,j} + P'_{i,j}\beta_{i,j} + X'_{i,j}\gamma_{i,j} + e_{i,j}, \tag{1}$$

where the dependent variable, $C_{i,j}$, is our measure of the household's gross consumption; $P_{i,j}$ are the consumption items reported in both the SFS and the SHS as listed in Table 12; $X_{i,j}$ are the socio-demographic and geographic characteristics of households; and $e_{i,j}$ are residuals for household type $i = \{renter, homeowner\}$ and family type $j = \{single, family\}$. More specifically, $X_{i,j}$ includes a significant

[10] In earlier work, we did 500 replications of our consumption imputation, adding random draws from the error terms from our consumption-imputation regression to see how sensitive our results were to imputation errors. We found that our results were not very sensitive to possible imputation errors. Although related, this approach was somewhat different, since our bootstrapping procedure in this case was not designed to match the conditional distributions by type.

[11] Our approach does not uniquely identify under-reporting households but it is designed to identify the aggregate level of under-reporting of similar household types. For instance, one household may appear to under-report income using one implicate but not another implicate. At the same point, a second observably similar household may appear to under-report using the second implicate but not the first. Conditioning on observables, our approach yields an estimated binary distribution of under-reporting of households. If these estimated distributions shift across the implicates, then this is an indication that our results are sensitive to the exact mapping of consumption to households. If they do not, then this is an indication that our imputations by household type are sufficiently close to yield reliable aggregate estimates.

variety of socio-demographics, including: age of the main income earner (MIE); age of the spouse; married MIE; male MIE; weeks worked by MIE and spouse; major source of income; number of adults, youth, child and income earners in the household; home mortgage free; vehicle ownership; province of residence; urban size; type of dwelling; and in the regressions including income, the before-tax income of the household. One difference between the control variables across these two years is that we are only able to include education levels in the SFS 2005 imputation, since they are not reported in the 1998 SHS. For the age variables we also include polynomials of order 3 and we include quadratic terms for the numbers of adults and children. We also interact the ongoing expenses items with an equivalence scale measure (the square root of the total household size) and with dummy variables for the three largest cities in Canada (Montréal, Toronto and Vancouver). When we also include income variables, we include the logarithm of income in addition to the level of income.[12]

Our first approach is to sum all current consumption expenditures for the household as reported in the SHS. We term this consumption measure C^{TOT}. For each consumption regression, we save the regression coefficients and then use these estimated coefficients to predict consumption for households in the SFS using the $P_{i,j}$ and $X_{i,j}$ for each household type and family type observed in the SFS. Our predicted consumption measures in the SFS are, conditional on the observables, equal to the average consumption observed in the SHS by construction for each household type and family type. We also examine our results both using household income covariates to help predict consumption and excluding household income.[13]

Our second imputation approach is to calculate the total non-durable consumption for the household using the data in the SHS. We term this consumption measure C^{NDC} and then add this imputed value to the ongoing expenses recorded in the SFS for each household. Table 18 in the appendix lists all the items that are included in the non-durable consumption measure C^{NDC}. This approach avoids a "frequency" bias in that we exclude large purchases that are infrequent to all households in our sample. We note that if a large purchase is financed by debt, then we should already capture the flow cost of that expense in our consumption measures. What is not included is durable consumption that is fully paid at the time of purchases or semi-durables. We argue that this omission biases our results downward and thus we are likely to underestimate the true amount of income under-reporting by this method. To calculate debt-servicing and interest costs for the 2005 SFS data, we use the aggregate annual interest expense reported for each household. However, annual interest expenses are not available for the 1999 survey. To construct the interest expenses for households in the 1999 survey, we use the reported data on the levels of household liabilities, such as student loan debts, credit card debts, etc., to construct estimates of the total interest burden faced

[12]Our full regression specifications are available upon request. We observe that the consumption regressions have an R^2 above 0.8 in almost all cases, and above 0.9 in many.

[13]One problem with ordinary least squares is that imputed consumption levels are unconstrained and so may, depending on the distribution of the covariates, yield implausible consumption estimates. To account for extreme values of imputed consumption, we winsorize values in our bottom and top 1 percentiles to be the value of the 1st percentile and 99th percentile, respectively.

by the household. One complication is that the liability-level data are for the time of the data collection (May to July 1999), and thus the annual interest cost is sensitive to when the debt is incurred. An additional complication is that households may not face similar interest rates. In an attempt to be conservative in our estimate of the total interest cost for households, we choose to set interest rates that seem to be near the lower range of available data (Table 11). We include amortization payments assuming a 10-year amortization for student loans. We note that the tax treatment of student loan interest payments changed between the 1999 and 2005 surveys, but, since we use after-tax income from CRA records, we do not need to adjust for these changes in the taxation of student loan interest payments. We assume an interest rate of 9 per cent for credit card debt on the assumption that some households shift balances from high interest cards to low interest cards. Other interest rates (in the -3 to $+3$ per cent range) are tested and our results do not appear to be sensitive to the interest rate selected. Our imputed values of non-durable consumption constructed in this way are lower than the imputed total expenditure by roughly $17,000 for the average household in 1998, and roughly $23,000 in 2004 (although this difference obviously varies across the distribution of expenditure).

For households that report spending more than their income (dissavings) or report spending less than their income (savings), we also impute savings and dissavings in a similar fashion. We acknowledge that the responses to this question in the SFS are not exactly accurate, since the question asks households to evaluate their expenditure omitting investments, automobile purchases or home purchases. This omission is not critical for those households spending more than their income, because these households must, by transitivity, be spending more than their income, including expenditures for investments, automobile purchases or home purchases. However, as we note above, this transitivity may not apply for households that report spending equal to or less than their income. Thus we impute savings for households that report spending less than their income and impute savings for households that report spending less than their income and also report investments. This approach should be conservative, since households that report spending equal to their income ignoring their savings should be treated in our approach as having no savings, and thus we should understate their household expenditure. To impute savings and dissavings, we specify a regression similar to the consumption regression, equation (1), with one exception: we specify savings and dissavings in logarithms. This avoids predicting negative (dis)savings for these households. We then predict (dis)savings in the SFS data using the regression coefficients from the SHS, and take the exponent to yield (dis)savings in dollar terms.

There are two measures of savings in the SHS: money flows and changes in registered retirement savings plans (RRSPs). Money flows measure the net changes in households' assets and liabilities during the survey year, including the contributions to and withdrawals from the RRSP.[14] By definition, this variable is intended

[14]Items included in money flows: net changes in bank balances; money on hand; money owed to the household; money owed by the household; purchase and sale of stocks and bonds; personal property and real estate; expenditures on home additions, renovations and new installations; and contributions to and withdrawals from RRSPs.

to measure household savings. As a robustness check, we also estimated results using net changes in RRSPs as an alternative measure of households' savings. The results are very similar to using the money flows measure, and thus we report only the results using the money flows measure.[15]

It is almost certain that our imputation regression, equation (1), has endogenous covariates. In particular, it is unlikely that the covariance of P and e is zero, perhaps because of measurement error, so that our coefficient estimates β are likely to be biased and inconsistent.[16] However, the endogeneity is not a concern to our imputation approach. What we seek is to evaluate conditional means and to use these covariates to predict the conditional means. In other words, we are interested in ordinary least squares (OLS) as a statistical technique to predict $C_{i,j}$ and do not require any of the assumptions of the typical OLS regression for consistency of the regression parameter estimates, which are not of interest.[17] Indeed, the tendency of endogenous regressions to overfit the conditional mean is, in our case, helpful.

We note that the covariance between predicted expenditure and e is zero by construction in OLS. This follows from an application of the Frisch-Waugh-Lovell Theorem (OLS splits $C_{i,j}$ into orthogonal components – one conditional and one unconditional). It follows that the best predicted value of an observation is the fitted value (conditional) plus the expected error term (unconditional). Therefore, the error term does not bias the conditional expected value and hence the conditional expected value plus a random draw from the error distribution is an unbiased estimate of the true value of an observation. Thus, for each household we draw, using the survey weights to account for the sample composition, four residuals from the appropriate consumption regression (and (dis)savings regression, if appropriate) in the SHS and add this error term to the imputed consumption value (and (dis)savings value, if appropriate) for each household. Consequently, for each household we construct five imputed measures of each measure of consumption: our baseline imputation that does not add draws from the SHS errors and four implicates that each include a different random error draw.[18] By construction, the additional four implicates are designed to match the distribution of consumption in this household type/family type cell in the SFS with the distribution of consumption in the same cell in the SHS, conditional on the observables.[19] Because we are concerned about the sensitivity of our results to small changes in the imputed values, we also consider an additional robustness exercise in which we add small amounts of income to households. For each household, we add either $2,000, $5,000 or $10,000 to

[15]The results using RRSP savings are not reported in the paper but are available upon request from the authors.

[16]The detailed imputation results are available from the authors upon request subject to the release conditions of the Statistics Act. However, we can provide our code to allow researchers to replicate our imputations exactly within Statistics Canada Research Data Centres.

[17]See, for instance, Davidson and MacKinnon (1993, Chapter 1 and pages 209-10).

[18]Four error draws appears to be within the typical range of roughly 3-5 in this literature when the goal is to reproduce an estimate of the distribution. See Little and Rubin (2002) for a discussion of imputation methods.

[19]In earlier work, we drew 500 Monte Carlo replications using random draws from the imputation residual distribution to evaluate the sensitivity of our main results to imprecision in our consumption measures and found very little sensitivity to these draws. In this current study, we instead draw four additional implicates per household and examine the sensitivity of all of our results.

assess how sensitive our under-reporting results are to small changes in income.[20]

Tables 14 and 17 in the appendix report our imputed expenditure (consumption and (dis)savings) measures by percentiles and compare those imputed values to the observed expenditure values for the same percentiles in the SHS. For the 1998 survey year, our baseline expenditure measures not using income variables in the imputation tend to overstate average expenditure for the lower percentiles and understate average expenditure for the upper percentiles. Our estimate of average expenditure using our baseline is within $1,000 of the SHS reported value. We note that finding higher expenditure values for the SFS is not unreasonable, since the survey coverage for the SFS differs slightly from that of the SHS: the SFS oversamples households from what are thought to be high-income neighbourhoods. The same pattern is mostly observed for the expenditure measure imputed using income variables, with the exception that the upper 99th percentile of expenditure is higher than that observed in the SHS. For non-durable consumption (also including (dis)savings), our lower percentiles are much higher for the imputed values in the SFS than for the SHS. Turning to our implicates, we find that the values imputed in the SFS compare quite closely to those in the SHS for 1998, except for the lowest percentile. The remaining percentiles are generally well within our $2,000 error band. For the lowest percentile, our implicates imply negative household expenditures due mainly to large draws from the imputation errors. Although negative expenditures would appear unlikely – they would require substantial dissavings – negative expenditures work against our approach because they cannot imply unreported income. Given that they account for only 1 percentile of our sample, we do not choose to make any further adjustments. We also note that the implicates appear to match the upper percentiles very closely, with the possible exception of the top 5 percentiles of non-durable consumption (which appear to be understated). For the 2005 survey year, our imputed values again appear to match the SHS quite well, with the exception of those imputed using income variables. Our implicates appear to match less well for the upper percentiles in 2005. We conjecture that this may be for two reasons. The first is that the smaller sample for the 2005 SFS implies fewer households in these bins, and so small-sample effects may be present. However, more probable in our view is that an even larger proportion of the 2005 SFS is drawn from the high-income subsample, and this increases our estimates in the tails of the distribution, because of the correlations between expenditure and the high-income covariates (as can be seen in the baseline results), and because the survey weights in 2005 are based, in part, on income (unlike in the 1999 SFS). Since our implicates add random draws to household expenditure prior to applying the SFS weights, the larger proportion of high-income households implies that relatively more high-income households will get a positive error draw, thus increasing expenditure in the upper percentiles. We note again that finding higher consumption expenditures, particularly in the higher percentiles, for the SFS is not unreasonable.[21]

[20] Alternatively, one could view the under-reporting results for these income amounts as the extent of under-reporting of at least this magnitude.

[21] Ideally, one might prefer to develop survey weights that are more appropriate to our task by not matching on reported incomes. However, this task requires raw data to which we do not have access, and is beyond the scope of this study.

5 Income Under-Reporting

In this section, we present our estimates of the extensive margin of income under-reporting for Canada. As we have discussed, we compare the households' reported income to the imputed expenditure to identify those households that are under-reporting their income (*i.e.*, expenditure exceeds income). We first report our estimates for 1998 for our full sample of households using our measure of imputed total expenditure, and for each of the incomes added to household income, to assess how sensitive our results are to minor discrepancies. We then disaggregate our estimates by self-employment status, business ownership status and also by whether a household reports savings or dissavings to determine whether our estimates of savings and dissavings are driving our main results. We next consider two alternative measures of household expenditure: (i) our measure of total expenditure imputed using income variables, and (ii) our measure of household expenditure using imputed non-durable consumption added to ongoing expenses reported in the SFS. Finally, we report our estimates for the 2005 survey year.

5.1 Total Expenditure

We begin by examining income under-reporting by two measures of household self-employment using our estimates of total expenditure imputed not using income variables. The first measure is whether the reference person or the spouse reports being self-employed. The second measure is whether the reference person or spouse owns a business. For each measure we then report our estimated proportion of households under-reporting income: *i.e.*, the proportion of households for which our estimate of expenditure and (dis)savings is greater than the households' reported income. We also disaggregate our estimates according to whether the household linked its reported income to the CRA data. For each household, either none, half or all of the reference person and spouse (if applicable) link their income to the CRA reports.

Turning first to households in which both the reference person and spouse (if applicable) are salaried workers, we find relatively little difference in our results across CRA reporting status for imputed total expenditure, not using income variables (see Table 2). If we assume that our estimated total expenditure is exactly accurate, then roughly half of salaried households appear to under-report income. This finding appears true across our additional implicates. In Tables 15 and 16 in the appendix, we report the correlations between our implicates and between our implicates and our baseline imputation for the 1998 survey year (the correlations are similar for 2004). By design, our implicates match the distributions observed in the SHS conditional on the observables but, as we note above, there is not a single mapping that achieves this end. Hence, our implicates assess the sensitivity of our under-reporting estimates to expenditure patterns observed in the SHS that are uncorrelated with household observables. We report the average under-reporting estimate across our implicates in Table 2.[22]

[22]We acknowledge that reporting the range of estimates across our implicates might also be of interest. However, we do

As we note above, it is unlikely that our estimated total expenditure is exactly accurate and so we also consider allowing a margin of error to our estimates. If we add $2,000 to household income, we see that the incidence of income under-reporting for salaried households falls to slightly above 0.4 for both our baseline and implicates. Adding $5,000 to income reduces our estimated income under-reporting to roughly 0.3 for our baseline sample and slightly higher, 0.35, for our implicates. Adding $10,000 to income (20 per cent of the average household income) reduces our estimate of income under-reporting to approximately 0.2 for our baseline and 0.23 for our implicates. (Our implicates tend to fall at a slightly lower rate than our baseline sample because the SHS imputation errors are skewed to the right.) We also observe that in all cases the incidence of income under-reporting is roughly 0.1-0.2 higher for the self-employed than for salaried households. The same general pattern is present for households in which neither the reference person nor the spouse owns a business. Indeed, the overall pattern of the results is nearly qualitatively and quantitatively identical. We also note that our estimates are unlikely to be invalidated by our adjustments for imputed values of consumption in the 1st and 99th percentiles, since these account for at most 2 per cent of our sample. We conclude that it appears to be extremely difficult to reconcile an assumption of truthful income reporting by salaried households with the data from the SFS for Canada.

Turning next to households with either the reference person or the spouse self-employed, we find higher incidences of income under-reporting. If we assume that our estimated total expenditure is exactly accurate, then we find that the incidence of income under-reporting lies between 0.54 and 0.62, depending on the household's CRA reporting status. The proportion falls as we add income to households, mimicking the pattern observed for the salaried households. One difference we observe between the salaried and self-employed households is that the CRA reporting status appears to matter for the self-employed. In virtually all cases, the incidence of income under-reporting appears higher for households that link their income to the CRA report (although this does not appear true for households differentiated by business-ownership status). We observe the same difference for households differentiated by business-ownership status.

We are naturally concerned that our measure of income under-reporting may rely too heavily on the household's declared income and that traditional methods of (dis)savings may not accurately reflect the source of funds for some households' expenditures. For example, households may borrow money from other family members, households may sell durable goods to finance expenditures, households may consume using credit that goes unpaid, and/or households may use pawnshops to obtain collateralized loans. The SFS asks households whether they have needed to sell an asset or deposit an item at a pawnshop in order to pay off a bill, whether the household is behind on a debt repayment or whether the household has received any gift money. We use responses to these questions to construct a financial stress indicator for households. Conditioning our measures of income under-reporting on the financial stress indicator yields estimates of

not observe much variation across our implicates and so choose to report the average. The detailed results are available upon request.

under-reporting that are roughly quantitatively identical to those reported in Table 2.

Our initial estimates suggest that the incidence of income under-reporting is higher for self-employed households, but would seem to cast doubt on the assumption that only non-salaried households under-report income. By extension, our results suggest that either salaried workers and their firms under-report remuneration or that salaried workers earn income via unreported self-employment. Currently, we cannot distinguish between either interpretation, although we note that there are plausible incentives for both. For instance, workers and firms may agree to under-report income to avoid both income taxes and payroll taxes. And salaried workers that earn additional income through self-employment are unlikely to report being self-employed if the activity is illegal.

Table 2: Income Under-Reporting and Self-Employment Status, 1998: Total Expenditure, No Income Imputation

				$2,000 Added		$5,000 Added		$10,000 Added	
	CRA	Baseline	IA	Baseline	IA	Baseline	IA	Main	IA
Salaried	0	0.48	0.48	0.41	0.42	0.31	0.35	0.19	0.24
	0.5	0.52	0.48	0.44	0.43	0.34	0.35	0.20	0.23
	1	0.48	0.48	0.41	0.43	0.29	0.35	0.17	0.23
Self-employed	0	0.54	0.54	0.51	0.50	0.44	0.46	0.37	0.37
	0.5	0.59	0.59	0.56	0.56	0.53	0.52	0.45	0.38
	1	0.62	0.59	0.58	0.56	0.53	0.51	0.44	0.40
Not business owner	0	0.48	0.48	0.41	0.42	0.30	0.34	0.18	0.23
	0.5	0.52	0.48	0.44	0.42	0.34	0.35	0.20	0.23
	1	0.48	0.48	0.40	0.42	0.28	0.34	0.16	0.22
Business owner	0	0.52	0.52	0.48	0.49	0.42	0.44	0.34	0.35
	0.5	0.55	0.55	0.51	0.52	0.46	0.47	0.36	0.34
	1	0.57	0.54	0.53	0.51	0.46	0.46	0.36	0.34

Notes: CRA is the proportion of household earners who link their income to the CRA records; Baseline refers to the baseline imputaton; IA refers to the average of the implicate samples and 'X' Added refers to our robustness exercises of adding 'X' to the household income.

5.2 Savers and Dissavers

Our estimates of total household expenditures include our estimates of household savings and dissavings. To examine whether income under-reporting is prevalent for savers and dissavers, we next disaggregate our measures of income under-reporting for households that we classify as Paycheck, Savers or Dissavers according to the definitions above. Our sample sizes across these classifications differ and so our implicates tend to exhibit more variability within each classification. We present our estimates of under-reporting for each classification in Table 3, assuming no added income.

For salaried households, there appears to be little difference worth noting in under-reporting estimates across the Paycheck, Savers and Dissavers classifications for the baseline estimates. For the implicates,

there is some evidence that the income under-reporting incidence is lower for Dissavers households than for Paycheck and Savers households. (This appears to reflect a left skew in the regression residuals for Dissavers in the SHS imputation regressions.)

For self-employed households, income under-reporting appears to be higher for Paycheck households than for Savers and Dissavers households if at least one household member links income to the CRA report. In general, Dissavers households appear to have a lower incidence of income under-reporting than Savers and Paycheck households.

We also consider adding income to households and find the general patterns roughly identical to those reported in Table 2 for the pooled group of households. As an additional robustness exercise, we redefine our Savers and Paycheck households depending on whether households that are classified as Savers do in fact hold assets. As a final robustness exercise, we also consider the incidence of income under-reporting depending on whether a household purchased a home in the previous year. We find very little difference across these robustness checks.[23] We conclude that estimates of the incidence of income under-reporting do not appear unduly sensitive to the imputed values of savings or dissavings.

Table 3: Income Under-Reporting and Saving, 1998: Total Expenditure, No Income Imputation

		Paycheck		Savers		Dissavers	
	CRA	Baseline	IA	Baseline	IA	Baseline	IA
	0	0.50	0.49	0.47	0.50	0.46	0.41
Salaried	0.5	0.57	0.53	0.47	0.50	0.46	0.41
	1	0.50	0.49	0.48	0.51	0.45	0.40
Self-	0	0.52	0.55	0.61	0.61	0.43	0.38
employed	0.5	0.68	0.67	0.56	0.61	0.42	0.38
	1	0.68	0.64	0.59	0.60	0.58	0.48

Notes: CRA is the proportion of household earners who link their income to the CRA records; Baseline refers to the baseline imputaton and IA refers to the average of the implicate samples.

5.3 Imputed Total Expenditure Using Income Variables, 1998

The estimates of the incidence of income under-reporting presented thus far rely on values of total expenditure imputed from the SHS without using the households' reported income. We prefer to present the results that do not include expenditure values that depend in part on income, for two reasons. The first reason is that household income in the SHS is both self-reported by the household and it appears to be edited by survey technicians. Of the reference person and spouse interviewed in 1999, roughly two-thirds linked their income variables in the SFS to the CRA data. We are concerned that respondent recall errors or survey edits for household income in the SHS may bias the link between income and expenditure imputed into the SFS

[23] These results are also available upon request.

in unknown ways. Our second reason is that we wish to avoid using what may be perceived as a circular argument when we examine the reliability of budget shares for identifying under-reporting. This criticism may be particularly pronounced if some income is under-reported, since the linear regression underlying the imputation would effectively pool the budget shares of both the truthfully reporting households and the under-reporting households. If, on average, households truthfully report their income, then including income variables would tend to lower our estimates of the incidence of income under-reporting.

Nevertheless, adding the income variables to the consumption regressions with the SHS data does improve their goodness-of-fit. Thus, to ensure that our results are not overly sensitive to the exclusion of additional conditioning information, we estimate the incidence of income under-reporting using total expenditure imputed including income variables. Our estimates are presented in Table 4. Confirming our intuition, our estimates are in general slightly lower than those presented in Table 2 but exhibit the same patterns. One difference is that the implicates tend to yield higher estimates than our baseline due to the skew in the SHS regression residuals. Similar to the results using imputed values not including income variables, we conclude that it appears to be extremely difficult to reconcile an assumption of truthful income reporting by salaried households with the data from the SFS for Canada. We also conclude that omitting or including income variables in the imputation regressions does not meaningfully change our results.

Table 4: Income Under-Reporting and Self-Employment Status, 1998: Total Expenditure, Using Income Imputation

				$2,000 Added		$5,000 Added		$10,000 Added	
	CRA	Baseline	IA	Baseline	IA	Baseline	IA	Main	IA
Salaried	0	0.46	0.48	0.36	0.40	0.23	0.31	0.13	0.20
	0.5	0.49	0.48	0.39	0.42	0.26	0.34	0.14	0.23
	1	0.45	0.47	0.34	0.40	0.20	0.30	0.11	0.19
Self-employed	0	0.52	0.53	0.48	0.48	0.37	0.40	0.21	0.29
	0.5	0.63	0.53	0.55	0.49	0.44	0.43	0.26	0.32
	1	0.57	0.56	0.50	0.51	0.41	0.44	0.24	0.32
Not business owner	0	0.46	0.48	0.36	0.40	0.23	0.31	0.12	0.20
	0.5	0.50	0.48	0.40	0.42	0.26	0.34	0.14	0.23
	1	0.45	0.47	0.34	0.40	0.21	0.30	0.11	0.19
Business owner	0	0.47	0.49	0.42	0.44	0.32	0.37	0.19	0.28
	0.5	0.54	0.49	0.46	0.45	0.37	0.39	0.21	0.29
	1	0.47	0.50	0.41	0.45	0.30	0.37	0.17	0.27

Notes: CRA is the proportion of household earners who link their income to the CRA records; Baseline refers to the baseline imputaton; IA refers to the average of the implicate samples and 'X' Added refers to our robustness exercises of adding 'X' to the household income.

5.4 Imputed Non-Durable Expenditure, 1998

As we have already discussed, imputing total expenditure may overinflate our estimates of the incidence of income under-reporting because it effectively adds some durable expenditures to each household. Given that durable expenditures are often lumpy, this smoothing effect may lead us to overestimate in some years and underestimate in others if, in the aggregate, there is some correlation across durable-good expenditures across households. Since Canada is a small open economy that imports durable goods, exchange rate movements may be a candidate correlation device for aggregate durable expenditures. We are also concerned that we might be double counting some expenditures that are financed via debt payments, to the extent that households report the total expenditure rather than the flow expense.

To help address these concerns, we consider an approach that imputes only non-durable expenditures from the SHS into the SFS. We then add these imputed non-durable expenditures to the ongoing expenses that are recorded in the SFS to calculate the household's total expenditure. We term this expenditure estimate C^{NDC}. By construction, C^{NDC} is a lower bound on household expenditure, since it omits all durable and semi-durable expenses that are not explicitly financed (and thus included in our interest expenses).

In Table 5 we report our estimates of the incidence of income under-reporting using our imputed non-durable expenditures, and (dis)savings if appropriate, added to the household's reported ongoing expenses in the SFS. We note that our estimated incidence of income under-reporting is lower than the estimates obtained using imputed total expenditure, as expected. However, we still find evidence of income under-reporting across all our specifications, for both salaried and self-employed households. The same patterns are evident for non-business-owning households and business-owning households. We view the estimated incidence of income under-reporting using C^{NDC} as a plausible lower bound for income under-reporting in Canada in 1998.

Perhaps curiously, our estimates for the incidence of income under-reporting for salaried households, not adding income, are higher than those for the same household type using our imputed total expenditure and adding $10,000 of income. Since, for the average household, non-durable consumption plus ongoing expenditures are roughly $17,000 lower than total expenditures, it must be that the distribution of income under-reporting is unequally distributed over reported household income. We return to this point in Section 6.

We conclude that our estimates of the incidence of income under-reporting in Canada suggest that it is relatively widespread and not confined to self-employed households (or at least households that indicate being self-employed on survey questionnaires). Our robustness exercises have not cast doubt on this conclusion. Nevertheless, our estimates suggest that a fair degree of the extensive margin of income under-reporting is of relatively small magnitude and may reflect small errors in the distribution of imputed consumption (although our implicates do not support this contention). For example, considering our estimates presented

Table 5: Income Under-Reporting and Self-Employment Status, 1998: Non-Durable Expenditure, No Income Imputation

				$2,000 Added		$5,000 Added		$10,000 Added	
	CRA	Baseline	IA	Baseline	IA	Baseline	IA	Main	IA
Salaried	0	0.35	0.37	0.28	0.31	0.19	0.24	0.12	0.16
	0.5	0.37	0.38	0.29	0.33	0.20	0.25	0.11	0.16
	1	0.31	0.35	0.24	0.29	0.17	0.22	0.09	0.14
Self-employed	0	0.42	0.41	0.38	0.38	0.32	0.34	0.20	0.26
	0.5	0.52	0.53	0.50	0.48	0.46	0.40	0.27	0.32
	1	0.50	0.49	0.46	0.45	0.37	0.39	0.28	0.31
Not business owner	0	0.36	0.37	0.28	0.31	0.19	0.24	0.11	0.15
	0.5	0.38	0.39	0.30	0.33	0.21	0.25	0.11	0.16
	1	0.32	0.35	0.24	0.29	0.17	0.22	0.09	0.13
Business owner	0	0.36	0.37	0.33	0.34	0.26	0.30	0.17	0.23
	0.5	0.41	0.44	0.37	0.40	0.32	0.33	0.22	0.27
	1	0.40	0.42	0.36	0.38	0.29	0.32	0.20	0.24

Notes: CRA is the proportion of household earners who link their income to the CRA records; Baseline refers to the baseline imputaton; IA refers to the average of the implicate samples and 'X' Added refers to our robustness exercises of adding 'X' to the household income.

in Table 2, we observe that the estimated incidence of income under-reporting falls roughly 7 percentage points for salaried and self-employed households if we add $2,000 in annual income. For the purposes of this study, and yielding the results that follow, we choose to report statistics based on our measure of income under-reporting adding $2,000 to each household. This value of added income when applied to household consumption ensures that the distributions of imputed consumption averages by percentile are lower at each percentile than the same consumption reported in the SHS for that percentile. From a different perspective, while this leads us to overstate annual gross domestic income (GDI) by roughly $38 billion given the survey weights for our sample, we feel it is also a reasonable compromise considering the likelihood of consumption recall misspecification. As we note above, Tremblay *et al.* (2010) suggest that consumption recall errors for the SHS are not larger than 3 per cent. Since aggregate consumption expenditures in Canada were slightly over $700 billion in nominal terms in 1998, adding $38 billion to income roughly equates to 6 per cent of consumption and 4 per cent of GDI, which appear to be reasonable aggregate error bands.

Using these estimates of income under-reporting, our results suggest that the incidence of income under-reporting for salaried households in 1998 was between 0.24 (using non-durables) and 0.44 (using total expenditure). In comparison, our estimates for self-employed households ranged between 0.41 and 0.56. Our overall estimate of income under-reporting ranged between roughly 0.27 and 0.42 using the sample proportions of salaried and self-employed households.

We note that our results for 1998 are, in general, comparable in magnitude to those of Andreoni *et al.* (1998), who found an incidence of approximately 0.4 for the United States using TCMP data. We also

compare our estimates of tax evasion with a small-scale Canadian survey that directly asked the respondents questions related to income-reporting behaviour. A Financial Post/Compas poll in 1995 of 820 Canadian adults reported that 20 per cent of the respondents admitted hiding income to avoid paying tax. Given the small size of the survey, it appears reasonable to conclude that our estimates are comparable. Finally, our results suggest that a relative incidence of under-reporting by the self-employed is between roughly 12-17 percentage points. This is similar to Schuetze (2002), who effectively estimates the same relative incidence given his assumption that salaried workers do not under-report. He finds that the range of self-employed under-reporting is roughly 15-30 per cent of the self-employed. We note that our estimates of the relative incidence of under-reporting by the self-employed are typically all within or close to this range.

5.5 Income Under-Reporting in 2004

In this section, we report our estimates of income under-reporting by the two measures of household self-employment noted above using our estimates of total expenditure imputed not using income variables. We choose this approach because the estimates of total expenditure presented in Table 17 suggest that including income variables in the imputation will lead to higher levels of expenditure, particularly in the upper percentiles of expenditure.[24] Table 6 presents our estimates of the incidence of income under-reporting for our baseline specification and our implicates. There appear to be two qualitative differences with respect to the results reported for 1998. First, in general, our estimates of income under-reporting are roughly 10 percentage points higher for households that do not link their income to their CRA reports. This appears true across each of our added income specifications and implicate averages. Second, the incidence of income under-reporting does not decrease as much as we add income, and remains 5-10 percentage points higher for the estimates with $10,000 added income. While this may partly reflect inflation effects, it does suggest an increase in the nominal value of under-reporting between 1998 and 2004. With the exception of these two differences, the remaining picture is consistent with that reported for 1998.

Overall, our results for 2004 suggest that the incidence of income under-reporting for salaried households was between 0.33 (using unreported results for non-durables) and 0.48 (using total expenditure). In comparison, our estimates for self-employed households ranged between 0.49 and 0.57. Our overall estimate of income under-reporting ranged between roughly 0.35 and 0.49 using the sample proportions of salaried and self-employed households.

6 Inequality and Tax Evasion

Our estimates of the extensive margin of income under-reporting suggest that it was relatively widespread in Canada in both 1998 and 2004. Given that we find evidence of income under-reporting, one may question the

[24] The results for our specifications are available upon request.

Table 6: Income Under-Reporting and Self-Employment Status, 2004: Total Expenditure, No Income Imputation

				$2,000 Added		$5,000 Added		$10,000 Added	
	CRA	Baseline	IA	Baseline	IA	Baseline	IA	Main	IA
Salaried	0	0.56	0.53	0.48	0.48	0.40	0.42	0.29	0.33
	0.5	0.46	0.46	0.41	0.42	0.34	0.37	0.24	0.30
	1	0.45	0.47	0.40	0.43	0.33	0.37	0.23	0.29
Self-employed	0	0.67	0.63	0.66	0.63	0.58	0.60	0.51	0.53
	0.5	0.67	0.62	0.61	0.58	0.56	0.56	0.56	0.49
	1	0.52	0.52	0.48	0.50	0.45	0.46	0.37	0.39
Not business owner	0	0.56	0.53	0.48	0.48	0.39	0.42	0.29	0.33
	0.5	0.45	0.46	0.40	0.42	0.33	0.37	0.23	0.30
	1	0.45	0.46	0.39	0.42	0.31	0.36	0.21	0.28
Business owner	0	0.63	0.60	0.61	0.58	0.56	0.54	0.44	0.48
	0.5	0.63	0.55	0.56	0.53	0.55	0.52	0.54	0.46
	1	0.52	0.53	0.50	0.50	0.46	0.46	0.39	0.40

Notes: CRA is the proportion of household earners who link their income to the CRA records; Baseline refers to the baseline imputaton; IA refers to the average of the implicate samples and 'X' Added refers to our robustness exercises of adding 'X' to the household income.

extent to which income under-reporting households exhibit predictable behaviour. One predictable pattern is that the incidence of income under-reporting households should rise as reported incomes fall because the action of under-reporting income necessarily lowers reported income. A related question is whether there are under-reporting households across the distribution of reported income or whether they are clustered in a particular range of reported income.

In Table 7 we list the proportion of income under-reporters by reported income levels using our estimates of the income necessary to match imputed total expenditures. We choose to use total income because these estimates should be upper-bound estimates of the incidence of under-reported income. We find that in the lowest range of income ($0-$20,000 per year), approximately 65 per cent of households under-report their income. The percentages in higher income ranges are less staggering but suggest that income under-reporting is pervasive across income categories. For the upper income categories we find that our implicates tend to suggest higher incidences of under-reporting than our baseline estimates. We note that the non-linearity in under-reporting is similar to that found by Tedds (2010). We caution that our results may understate the incidence of under-reporting in the higher income brackets, particularly in our baseline sample, because we may underestimate savings for these households for two reasons. The first is that we impute savings for households that report expenditure less than income excluding investments, home purchases and auto purchases. To the extent that wealthier households may report expenditure equal to income excluding investments, home purchases and auto purchases, but yet may have expenditure on investments, homes, or autos, we may underestimate their expenditure by this latter amount. The second reason is that households

in the high wealth survey subsample for the SFS may have higher savings than households with savings in the SHS sample, which would affect our imputations, and these high-wealth households are likely concentrated in the higher income bracket. Underestimating their expenditures would lower the intensive and extensive estimates of under-reported income for this group.

Table 7: Income Under-Reporting by Reported Income Level

		1998		2004	
Income	CRA	Baseline	IA	Baseline	IA
0-20	0	0.681	0.638	0.645	0.665
	0.5	0.706	0.613	0.695	0.567
	1	0.657	0.620	0.558	0.557
20-40	0	0.494	0.500	0.652	0.564
	0.5	0.585	0.508	0.544	0.558
	1	0.571	0.539	0.568	0.550
40-60	0	0.362	0.391	0.536	0.537
	0.5	0.363	0.409	0.399	0.424
	1	0.402	0.409	0.484	0.481
60-80	0	0.248	0.284	0.353	0.351
	0.5	0.226	0.298	0.376	0.378
	1	0.225	0.315	0.366	0.419
80+	0	0.086	0.155	0.209	0.267
	0.5	0.072	0.164	0.107	0.206
	1	0.100	0.188	0.156	0.242

Notes: CRA is the proportion of household earners who link their income to the CRA records; Baseline refers to the baseline imputation; and IA refers to the average of the implicate samples.

We next reconstruct the distribution of income in both 1998 and 2004 by adding our estimates of the amount of under-reported income to the household's reported income if our estimates imply that the household is under-reporting income. We do this by using both our conservative measure of household expenditure, C^{NDC}, and our measure of total household expenditure, C^{TOT}. For both measures, we use our baseline estimates of under-reporting assuming \$2,000 in additional household income. The general patterns are qualitatively identical, and so we choose to report our estimates using total household expenditure.[25] We compare our estimate of the true income distribution to the reported income distribution to determine how the income distribution is affected by unreported income.[26] We find that the true income profile is skewed to the right and that the true incidence of income poverty at any absolute level is lower than reported (see Figures 1 and 2). We also find that the proportion of households in what could be characterized as the middle income range is larger than reported income measures would suggest.

In additional but unreported profiling, we find that tax non-compliance occurs across all occupations, regions, education and income levels. We do not find these results surprising. In addition to marginal costs

[25]The results using non-durable consumption are available upon request.

[26]We estimate the kernel density of true income and reported income using an Epanechnikov kernel and select the optimal bandwidth as the bandwidth that minimizes the mean integrated squared error if the data were Gaussian and a Gaussian kernel was used. Given the skewed nature of our data, this bandwidth may be too smooth.

Figure 1: Distribution of Estimated True and Reported Income, 1998

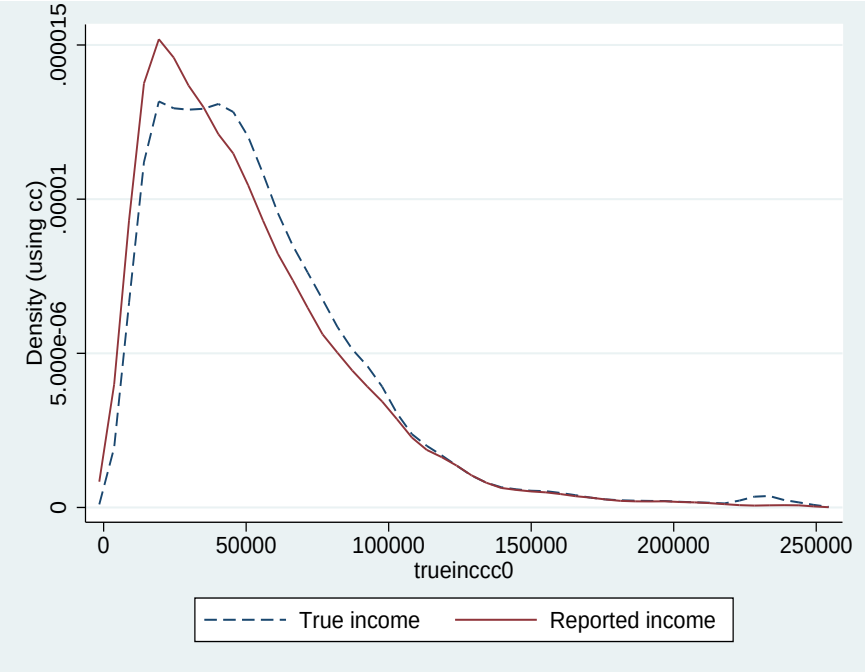
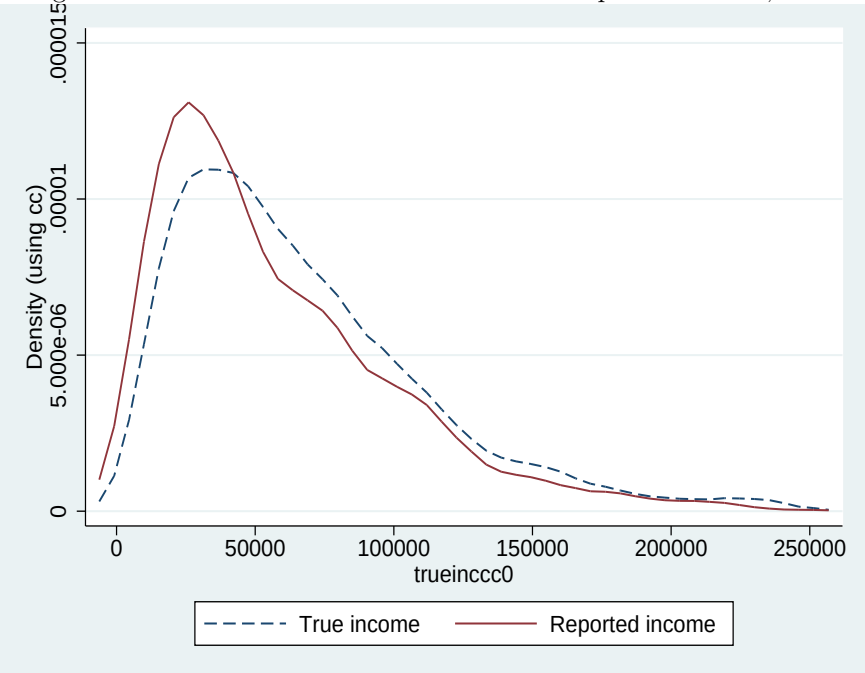


Figure 2: Distribution of Estimated True and Reported Income, 2004



and benefits associated with tax evasion that are specific to observable characteristics such as occupational status (*i.e.*, it is easier to hide gratuities than salary), individuals also likely face marginal costs and benefits

that differ across personal characteristics (*e.g.*, a mother may be less likely to risk incarceration than a single male). Therefore, it does not seem surprising that an occupational indicator, such as self-employment status, is not a reliable instrument for income under-reporting.

6.1 Tax Loss: A Rough Estimate

Our analysis has mainly focused on the extensive margin of income under-reporting. As we argue in the previous section, the extensive margin is important for the design of social policy. However, the intensive margin is also important for policy, since it reflects the amount of unreported income. The intensive margin matters for both tax revenues and the size of government social transfers.

Using our estimates of total household expenditure and the sample weights, we estimate that total under-reported income in 1998 ranged from $131.0 billion (baseline sample) to $177 billion (implicate average). In 2004, the estimated under-reported income using the two measures was $184 billion and $254 billion, respectively. As a percentage of GDP, this is between 14 and 19 per cent in both 1998 and 2004.[27] We interpret these estimates as the upper bound of the level of under-reported income in Canada for these years. As we noted in the Introduction, our estimates of the relative size of unreported income are in line with those estimated using macroeconomic models and data; *e.g.*, Mirus *et al.* (1994), Giles and Tedds (2002) and Schneider (2012). Previous estimates obtained using microdata (*e.g.*, Schuetze 2002) and using national accounts data (*e.g.*, Terefe *et al.* 2011) estimated a much lower relative amount of unreported income, roughly 2-5 per cent of GDP. Our estimates help to reconcile the estimated differences between the methodologies. As we discuss above, the microeconometric approach employed by Schuetze (2002) is founded on an assumption that appears at odds with the data used in our study. Terefe *et al.* (2011) use an upper-bound approach suggested by Gervais (1994) and the OECD (2002) that applies ad-hoc multipliers to national accounts data to estimate an upper bound of the relative size of the underground economy (of which under-reported income is a subset). Our results add to the existing literature which suggests this approach provides very misleading estimates of the relative size of under-reported income. Finally, our estimates suggest that unreported income increased roughly 40 per cent between 1999 and 2005, mirroring almost exactly the growth in nominal GDP.

In Tables 8 and 9 we present estimates of the average value of missing income for each of the levels of reported income we examined above. Our results suggest a non-linear relationship between reported income and under-reported income, mirroring the findings of Tedds (2010), who uses a different methodology. We find that the highest amounts of average under-reported income are typically at the lowest and highest income brackets. There is also no apparent difference in our estimates whether we use our under-reporting thresholds of $2,000 or $5,000 added income.

[27] If the remaining 10 per cent of the population that is not accounted for by the survey weights is similar, then these estimates would be roughly 1-2 percentage points higher.

Table 8: Estimated Under-Reported Income (Expenditure), $2,000 Added Income

1998

	Average Amount		Total Amount	
Income	Baseline	IA	Baseline	IA
0-20	-$16,697.55	-$20,534.13	-$57,143,505,959	-$65,337,652,167
20-40	-$14,745.09	-$19,587.12	-$38,284,894,683	-$48,870,069,787
40-60	-$15,576.13	-$22,257.58	-$19,299,168,319	-$29,048,848,545
60-80	-$17,750.21	-$23,250.26	-$8,339,868,439	-$14,707,292,241
80+	-$24,987.80	-$31,829.03	-$8,628,951,304	-$19,486,956,841
Total			-$131,696,388,704	-$177,450,819,579

2004

	Average Amount		Total Amount	
Income	Baseline	IA	Baseline	IA
0-20	-$26,439.99	-$30,678.91	-$52,312,348,557	-$64,104,243,011
20-40	-$17,538.43	-$23,079.49	-$56,774,272,259	-$69,075,790,641
40-60	-$24,518.33	-$30,962.31	-$31,189,535,665	-$40,372,922,076
60-80	-$21,484.87	-$32,805.79	-$17,377,814,796	-$29,329,905,752
80+	-$29,942.93	-$38,247.31	-$26,243,594,001	-$48,806,852,768
Total			-$183,897,565,278	-$251,689,714,248
Percentage Change 1998-2004			40%	42%

Notes: Baseline refers to the baseline imputation and IA refers to the average of the implicate samples.

Table 9: Estimated Under-Reported Income (Expenditure), $5,000 Added Income

1998

	Average Amount		Total Amount	
Income	Baseline	IA	Baseline	IA
0-20	-$21,691.45	-$24,805.19	-$54,524,257,852	-$63,848,623,590
20-40	-$17,891.28	-$22,625.30	-$36,993,924,854	-$47,671,965,610
40-60	-$19,273.59	-$25,459.23	-$18,033,805,903	-$28,524,379,578
60-80	-$22,197.89	-$26,616.34	-$8,126,424,789	-$14,475,693,234
80+	-$30,617.80	-$36,892.44	-$9,035,572,290	-$19,720,419,302
Total			-$126,713,985,688	-$174,241,081,314

2004

	Average Amount		Total Amount	
Income	Baseline	IA	Baseline	IA
0-20	-$32,423.51	-$36,490.42	-$51,416,129,729	-$63,671,324,344
20-40	-$20,268.05	-$25,874.42	-$55,984,324,413	-$69,613,167,786
40-60	-$27,396.85	-$34,507.56	-$30,383,000,885	-$39,777,971,315
60-80	-$25,697.79	-$35,753.92	-$17,200,064,647	-$29,116,549,975
80+	-$34,836.27	-$41,253.49	-$27,151,561,547	-$48,600,792,277
Total			-$182,135,081,222	-$250,779,805,696
Percentage Change 1998-2004			44%	44%

Notes: Baseline refers to the baseline imputation and IA refers to the average of the implicate samples.

Estimating the lost tax revenue requires information on the value of each income item and eligible tax credit that is not readily available. And, as we have cautioned above, it is possible that some income is earned and taxed offshore and that tax treaties would obviate taxation in Canada. We are therefore cautious about over-interpreting our estimates of forgone tax revenues. Nevertheless, one straightforward approach is to apply a uniform average or marginal rate to calculate the lost revenue from under-reported income. Using this method, we apply a 36 per cent marginal tax rate (the top rate for these years) to the under-reported income for our sample, which implies an upper bound for missing tax revenue for the 1999 survey (1998 tax year) in the range of $47-$63 billion, and for the 2005 survey (2004 tax year) in the range of $66-$90 billion.[28]

We stress that our estimates of tax loss to the government cannot be viewed as a proxy for the social burden of tax evasion. Our estimates of the tax loss do not incorporate social efficiency considerations resulting from transfers from poorer households to richer households (in income terms) through a misapplied tax and transfer system. We note that our estimates of the income distribution suggest that such transfers do occur.

7 Budget Shares and Income

In this section, we propose a simple theoretical test of income under-reporting. Our approach borrows intuition partly from the literature on tax compliance and tax evasion, reviewed in Andreoni *et al.* (1998). While we are not directly interested in the design of an optimal tax policy, we are interested in the effect of tax evasion on the composition of the household's consumption bundle. We propose a stylized model of the income reporting function for an individual and analyze the effect of tax evasion on the composition of a ratio of individual consumption to reported income. In effect, what we propose is a simple ratio test to detect income under-reporters in existing survey data.

Consider an individual with income y; a single consumption good, c, at (exogenous) prices p; savings, s, which earn an (exogenous) after-tax rate of return (1+r); a time discount factor $0 < \beta < 1$; and a utility function $u(c)$ that satisfies the Inada conditions.[29] Let subscript t indicate time and assume that the individual faces a known end-of-life in period T. For simplicity, we assume that the individual faces a voluntary tax rate τ_t on all reported income in period t. The problem facing the individual is

[28] In earlier work, we also used the tax simulator developed by Milligan (2008) and found that the tax simulator estimates were nearly identical to those assuming a 36 per cent tax rate for that sample. We choose not to use the tax simulator for this study because we are concerned that we may not observe relevant variables for tax deductions for households that under-report income. For example, small-business tax credits or self-employment expenses may be applicable to some of the under-reported income we find for households that report only salaried income.

[29] We assume here that income from savings cannot be hidden, but this assumption is simply for exposition. Modelling c as a vector of consumption goods does not affect our argument, and thus we choose a scalar valued consumption index for expositional simplicity.

$$\max_{c_t,\tau_t} \sum_{t=0}^{T} \beta_t u(c_t) \quad \text{s.t.} \\ p_t c_t + s_t = y_t - \tau_t y_t + (1 + r_t) s_{t-1}. \tag{2}$$

The solution to the individual's problem is characterized by $\tau_t = 0$. This result is trivial and simply serves to motivate that individuals typically prefer to pay no tax and do not when it is costless to avoid it.

Instead, suppose that taxation is involuntary but households pay a utility cost to hide income from taxation, $\gamma(y)$, where we assume that γ is differentiable and $\gamma'(y) > 0$. We interpret the utility cost as any or all of the effort cost to the individual to hide the income, the expected punishment if caught, or the moral cost to the household of cheating. We choose not to model the risk to the individual of being audited, as is often analyzed in the tax evasion literature (*e.g.*, Allingham and Sandmo 1972, Yitzhaki 1974, Richter and Boadway 2005), for two reasons. First, our arguments described in this section doubtlessly hold in these models under typical expected utility and punishment assumptions – *i.e.*, as long as individual consumption rises when income is under-reported. Second, we do not seek to explain why individuals under-report or what mechanisms a taxation authority could feasibly design for social efficiency (or any other policy goal).

Define $\hat{y}_t$ as unreported income so that $\bar{y}_t = y_t - \hat{y}_t$ is the individual's reported income. The individual's problem thus becomes:

$$\max_{c_t,\hat{y}_t} \sum_{t=0}^{T} \beta_t [u(c_t) - \gamma(\hat{y}_t)] \quad \text{s.t.} \\ p_t c_t + s_{t+1} = y_t - \tau_t(\bar{y}_t) + (1 + r_t) s_t. \tag{3}$$

We rewrite the problem as

$$V(y_t, s_t) = u(c_t) - \gamma(\hat{y}_t) + \beta_t V(y_{t+1}, s_{t+1}). \tag{4}$$

For simplicity, we focus on the periods $t < T$. Only the first-order conditions with respect to $\hat{y}_t$ are relevant for our discussion:

$$\gamma'(\hat{y}_t) = u'(c_t) \frac{\partial c_t}{\partial \hat{y}_t} \frac{\tau_t}{p_t} + \beta_t \frac{\partial V(y_{t+1}, s_{t+1})}{\partial s_{t+1}} \frac{\partial s_{t+1}}{\partial \hat{y}_t} (1 + r_{t+1}). \tag{5}$$

That is, the household will choose to under-report income $\hat{y}$ until the marginal cost of under-reporting equals the expected marginal benefit of doing so. A second implication of equation (5) is that an individual's consumption, c_t, is increasing in $\hat{y}_t$:

$$\frac{\partial c_t}{\partial \hat{y}_t} = [\gamma'(\hat{y}_t) - \beta_t \frac{\partial V(y_{t+1}, s_{t+1})}{\partial s_{t+1}} \frac{\partial s_{t+1}}{\partial \hat{y}_t} (1 + r_{t+1})] \frac{p_t}{u'(c_t)\tau_t} > 0, \tag{6}$$

by our assumptions on utility, γ and that equation (5) holds. Next, consider a consumption-expense-to-income ratio, CIR, defined as

$$CIR = \frac{c_t}{\bar{y}_t}. \tag{7}$$

The CIR is increasing in unreported income, since

$$\frac{dCIR}{d\hat{y}_t} = \frac{\frac{dc_t}{d\bar{y}_t}}{\bar{y}_t} + \frac{c_t}{(\bar{y}_t)^2} > 0. \tag{8}$$

Thus, individuals that under-report income should have higher consumption-to-reported-income ratios (budget shares) than households that truthfully report at a given income level.[30] We note that the consumption measure, c_t, examined here is an ongoing consumption expense. Ideally, the ongoing consumption expense should have no fixed costs, no 'necessity' welfare implications (such as subsistence food), or be an infrequent expense, such as a refrigerator purchase. Fortunately, individuals typically have one ongoing expense that, we argue, does largely satisfy these conditions – shelter costs. Mortgage payments and rent have no infrequency problem, there is typically an active market in shelter and, at least in relative terms, the fixed costs of shelter can be small, *e.g.*, for renters. In the empirical work that follows, we show that mortgage-payment-to-income ratios (MIR) and rent-to-income ratios (RIR) can be used to identify tax evasion. We note that preference heterogeneity, which scales consumption by a multiplicative factor, implies a distribution of CIRs so, for example, a household with a high CIR may either be under-reporting income or have higher-than-average preferences for consumption. Thus, as a practical matter, we examine this trade-off below. Finally, we acknowledge that MIR may fail to identify under-reporting households if those households own their homes outright. One possibility would be to use imputed rents in place of the mortgage, but we leave this for future research.

7.1 Income Under-Reporting

Once we have a direct measure of income under-reporting households, we can compare the relative MIR and RIR of households by their estimated income-reporting status, as our theory suggests. We term households true-reporters if, by our measure, their income is greater than or equal to our estimate of their expenditures, using our baseline imputation with $2,000 added income. We term households under-reporters if their income is less than our estimate of their expenditures, again using our baseline imputation with $2,000 added income. The terminology chosen is merely illustrative and not a definitive characterization of the household income-reporting behaviour given the uncertainty implicit in our estimates. We caution that the results we present here are best viewed as suggestive evidence, given the imprecision inherent in our imputation strategy. Nevertheless, if we do not see a difference in the relative MIR and RIR as predicted by theory, then this would cast some doubt on the accuracy of our imputation approach.

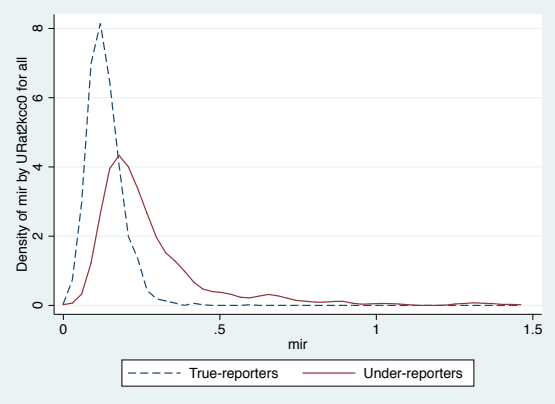
In Figure 3 we first present a kernel density estimate of distribution of MIR, by income-reporting status, for 1999, selecting only those households that have purchased their principal residence within the past five years and only considering imputed consumption.[31]

[30] These findings mirror, in some respects, the intuition that governments can reduce income tax evasion by taxing a consumption good, *e.g.*, Richter and Boadway (2005).

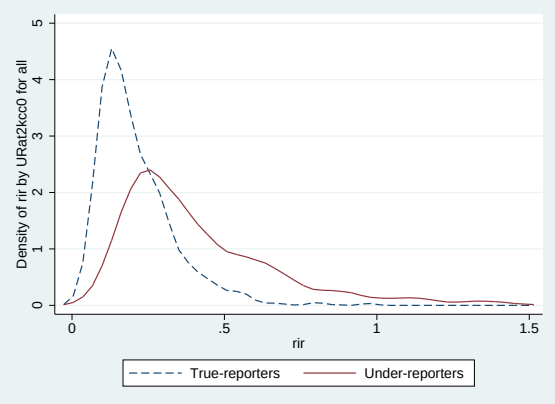
[31] We ignore the effects of imputed savings, since we include only households that reported income equal to spending.

The density estimates strongly suggest that the distribution of MIR is different across groups. This suggests that a MIR threshold can differentiate between true-income reporters and income under-reporters.[32] The results for RIR, also shown in Figure 3, show a similar pattern. Since the debt payment data for 1999 is an estimate, we also compare kernel density estimates of the distribution of MIR and RIR for true-income and income-under-reporting households in the 2005 SFS. Figure 4 demonstrates that the distribution of MIR and RIR also differs across households in 2005, and suggests that our findings for 1998 are, in all likelihood, robust to the debt payment estimate in 1998.

Figure 3: Kernel Density of MIR and RIR, by Income Reporting Status, 1999



Mortgage-payment-to-income ratio (MIR)



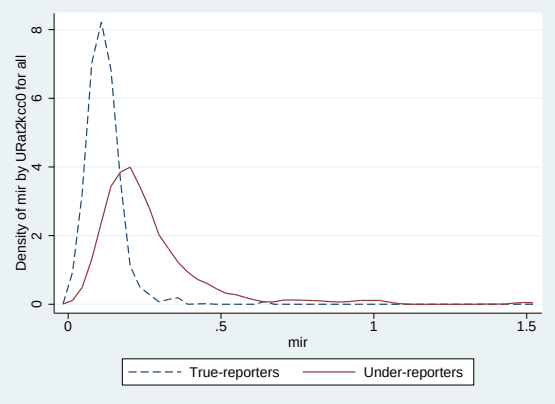
Rent-to-income ratio (RIR)

7.2 MIR and RIR as Indicators

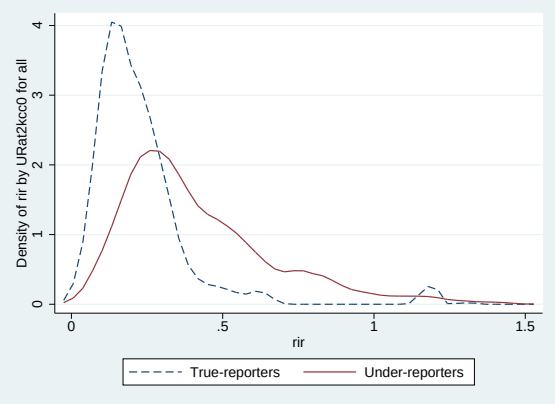
One question is whether there is a MIR or RIR threshold at which one can infer that a household is, in all likelihood, under-reporting income. We calculate the MIR and RIR levels at which 90 per cent and 95 per cent of households are found to be under-reporting income using our baseline results with $2,000 added income (the implicate results for this case are similar). The results are reported in Table 10. These numbers imply that in the 1998 tax year, if a family had mortgage payments that were greater than 0.36 times their reported gross income, then there was a 95 per cent chance that they were an income under-reporter. For the 2004 tax year, the corresponding MIR cut-off threshold was 0.36. For renters, we find that with a RIR

[32] We note that Canadian banks tend to grant a total mortgage of three times a family's gross income (or approximately 36 per cent debt-service ratio). Almost all the non-under-reporters' MIR fall below this percentage, suggesting that banks are applying this policy to these people. However, a significant proportion of under-reporters have a MIR well beyond this threshold and yet still are able to receive a mortgage. We can only speculate as to why they would be approved for a mortgage when their reported income is clearly outside the norm.

Figure 4: Kernel Density of MIR and RIR, by Income Reporting Status, 2005



Mortgage-payment-to-income ratio (MIR)



Rent-to-income ratio (RIR)

above 60 per cent in the 1998 tax year, or 63 per cent in the 2004 tax year, 95 per cent of households under-reported their income. We note that these MIR and RIR levels are basically constant across years.

Table 10: Thresholds of MIR and RIR

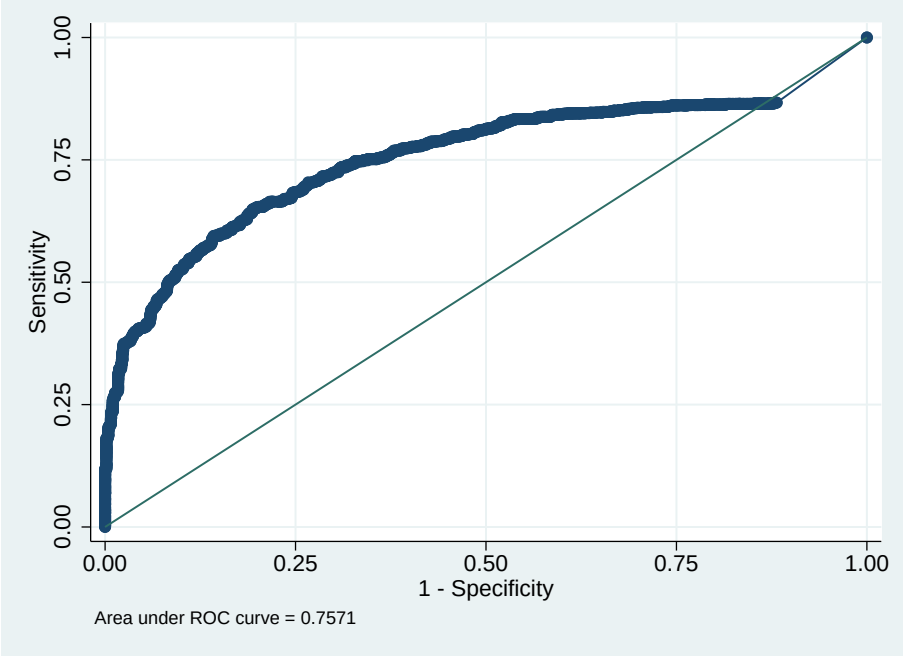
	1999		2005	
	MIR	RIR	MIR	RIR
90 per cent cut-off	0.24	0.53	0.20	0.56
95 per cent cut-off	0.36	0.60	0.36	0.63

If we continue to increase the MIR or RIR threshold, eventually we can approach a 100 per cent hit rate; that is, eventually, the threshold will be so high that there will only be a few households left in the category and they will most likely be under-reporters. Intuitively, a perfect MIR or RIR threshold would catch all the under-reporters to be audited, but not categorize any true-income reporters to be audited. We use a receiver operating characteristic (ROC) curve as an analytic tool to illustrate the benefit of using either MIR or RIR as a test of under-reporting. ROC is widely used in the engineering, medicine and psychology literature to select optimal cut-off values (Egan, 1975) and to illustrate. It is the plot of the true positive rate (TPR=TP/(TP+FN), *i.e.*, sensitivity) against the false positive rate (FPR=FP/(FP+TN), *i.e.*, 1-specificity). In our context, TPR is the proportion of correctly identified under-reporters (*i.e.*, with MIR/RIR above the thresholds) out of all under-reporters, whereas the FPR is the proportion of true-income reporters who are misidentified as under-reporters (*i.e.*, with MIR/RIR above the thresholds). Of particular interest for our discussion is that a diagonal line from the origin to the upper-right corner at (1.0, 1.0) is a random-guess line or a line of no discriminability that represents, essentially, flipping a coin to decide whether someone should be audited. As well, the upper-left corner of the ROC curve (0,1) represents a 100 per cent hit rate and a 100 per cent correct rejection rate, which is a perfect threshold with the maximum

discriminability. If the ROC curves lie above the random-guess lines for our MIR and RIR, then this implies that they have power as an indicator. In addition, the area under the ROC curve (above the random-guess line) is a measure of the probability that an under-reporter will have a MIR or RIR higher than a randomly selected non-under-reporter.

For our data, we plot the ROC curve for MIR and RIR thresholds for our baseline results in Figures 5 through 8.[33] We also note on each graph the estimated area under the ROC curve. As the graphs indicate, both MIR and RIR have predictive power when compared against the random-guess line, since the estimated areas are roughly 0.7-0.8 (the standard errors of the estimates of these areas are very small). We conclude that revealed shelter consumption behaviour may be a reliable indicator of income under-reporting.

Figure 5: ROC for MIR, 1999

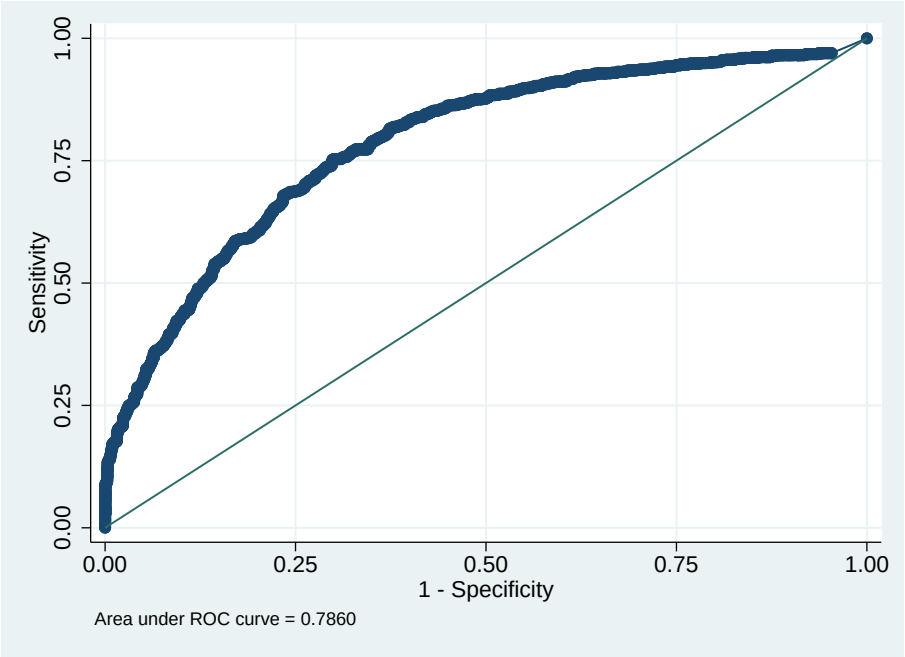


8 Conclusion

Using data from the Survey of Financial Security and the Survey of Household Spending, we demonstrated that self-employment is a poor indicator of possible income under-reporting. In addition, we illustrated that reported income as a poverty measure is misleading. We find that about 65 per cent of the households in the low-income range (less than $20K) under-report their income. Thus, tax and transfer policies based on reported income may have unwelcome, and regressive, social-efficiency costs.

[33] The results across the implicates are similar.

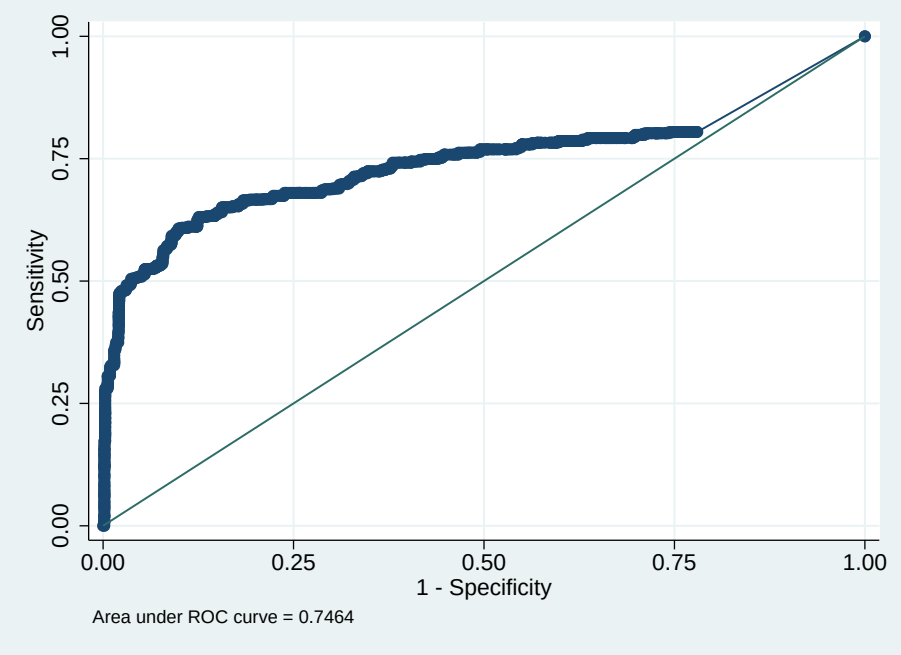
Figure 6: ROC for RIR, 1999



We are understandably cautious about our results. We subject our measures to robustness exercises and find that our results appear, by and large, to be robust. Our estimates of unreported income suggest that, as an upper bound estimate, roughly 14-19 per cent of Canada's GDP is unreported, and that approximately 35-50 per cent of households under-report at least some income. We also find that under-reported income as a fraction of GDP was stable between 1998 and 2004. Nevertheless, our results are best interpreted as statistical estimates of the aggregate intensive and extensive margins of income under-reporting, because they rely on matching data from two surveys. As we report, our estimates of the aggregate intensive and extensive margins do not appear to depend very much on the exact imputation, since our results are similar across the implicates we generate. To gain confidence in the under-reporting measures at the level of an individual household, one approach might be to generate an empirical distribution for each household using a large number of error draws from the imputations. Since we are not interested in statistically identifying the identities of probable income under-reporting households in this study, we do not undertake this exercise. At the most basic level, our results suggest that it is difficult to reconcile data on household incomes and household expenditures in Canada.

We also proposed a simple and intuitive indicator of income under-reporting. Our method is a straightforward application of the permanent income hypothesis: households make consumption decisions based on their true expected lifetime income, not reported income. Thus, the relationship between reported income and spending must be systematically different across income under-reporting households and the true-income

Figure 7: ROC for MIR, 2005



reporting households. In particular, we studied the shelter-costs-to-reported-income ratio and found that most households that under-report their income have mortgage-payment-to-income ratios (MIR) or rent-to-income ratios (RIR) well in excess of households that do not under-report. To the extent that social policies are intended to assist the truly poor, our results suggest that linking transfers to shelter costs may be deserving of further study.

References

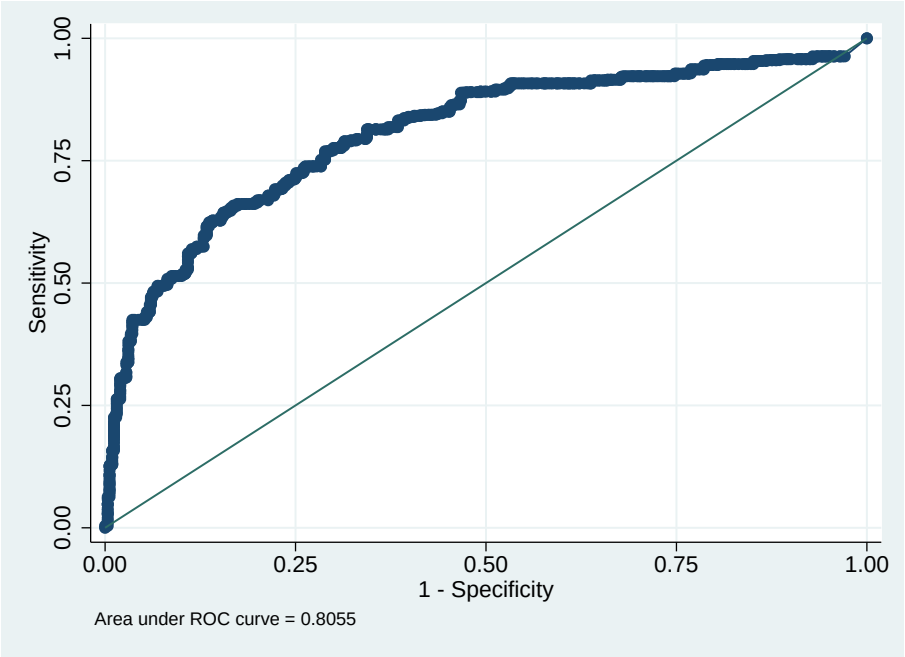
[1] **Allingham, M., and Sandmo, A.** 1972. "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1(3/4), 323-38.

[2] **Andreoni, J., Erard, B. and Feinstein, J.** 1998. "Tax Compliance," *Journal of Economic Literature*, 36(2), 818-860.

[3] **Blundell, R., Pistaferri, L. and Preston, I.** 2004. "Imputing Consumption in the PSID Using Food Demand Estimates from the CEX," *IFS Working Papers*, W04/27, Institute for Fiscal Studies.

[4] **Colledge, M.**, ed. 2002. "Measuring the Non-Observed Economy - A Handbook," Organisation for Economic Co-operation and Development, International Monetary Fund, International Labour Organisation and CIS STAT, OECD, Paris

Figure 8: ROC for RIR, 2005



[5] **Davidson, R. and MacKinnon, J.** 1993. *Estimation and Inference in Econometrics*, Oxford University Press, New York.

[6] **Diplock, J., Plecas, D. and Garis, L.** 2013. "Targeting Marihuana Growing Operations in British Columbia: A Summary Report Highlighting Current Research Findings," mimeo.

[7] **Easton, S. T.** 2004. "Marijuana Growth in British Columbia," *Public Policy Sources*, Vancouver, BC: The Fraser Institute.

[8] **Egan, J.P.** 1975. *Signal Detection Theory and ROC Analysis*, Academic Press, New York, USA.

[9] **Fisher, J. and Johnson, D.** 2006. "Consumption Mobility in the United States: Evidence from Two Panel Data Sets," *Topics in Economic Analysis and Policy*, Berkeley Electronic Press, 6(1), Art. 16.

[10] **Fortin, B., Lacroix, G. and Pinard, D.** 2010. "Evaluation of the Underground Economy in Quebec: A Microeconomic Approach," *International Economic Journal*, 24(10).

[11] **Gervais, G.** 1994. "The Size of the Underground Economy in Canada," Statistics Canada Catalogue no 13-603-MPE1994002. Ottawa, Ontario.

[12] **Giles, D.E.A. and Tedds, L.** 2002. *Taxes and the Canadian Underground Economy*, Canadian Tax Paper No. 6, Toronto: Canadian Tax Foundation.

[13] **Little, R.J.A. and Rubin, D.B.** 2002. *Statistical analysis with missing data, 2nd edition.* New York: Wiley and Sons.

[14] **Lyssiotou, P., Pashardes, P. and Stengos, T.** 2004. "Estimates of the Black Economy Based on Consumer Demand Approaches," *Economic Journal*, 114, 622-639.

[15] **Milligan, K.** 2008. Canadian Tax and Credit Simulator. Database, software and documentation, Version 2008-1.

[16] **Mirus, R. and Smith, R.** 1981. "Canada's Irregular Economy," *Canadian Public Policy*, 7(3), 444-453.

[17] **Mirus, R., Smith, R. and Karoleff, V.** 1994. "Canada's Underground Economy Revisited: Update and Critique," *Canadian Public Policy*, 20(3), 235-252.

[18] **Palumbo, M.** 1999. "Uncertain Medical Expenses and Precautionary Saving Near the End of the Life Cycle," *Review of Economic Studies*, 66(2), 395-421.

[19] **Pensions and Wealth Surveys Section, Income Statistics Division** 2003. "1999 Survey of Financial Security Public Use Microdata File User Guide," Catalogue no. 13F0026MIE No. 002, Statistics Canada.

[20] **Pissarides, C. and Weber, G.** 1989. "An Expenditure-based Estimate of Britain's Black Economy," *Journal of Public Economics*, 39, 17-32.

[21] **Richter, W. and Boadway, R.** 2005. "Trading Off Tax Distortion and Tax Evasion," *Journal of Public Economic Theory*, 7(3), 361-381.

[22] **Schneider, F.** 2012. "The Shadow Economy and Work in the Shadow: What Do We (Not) Know?," IZA Discussion Paper No. 6423.

[23] **Schuetze, H.** 2002. "Profiles of Tax Non-compliance Among the Self-Employed in Canada: 1969 to 1992," *Canadian Public Policy*, University of Toronto Press, vol. 28(2), pages 219-237, June.

[24] **Skinner, J.** 1987. "A Superior Measure of Consumption from the Panel Study of Income Dynamics," *Economic Letters*, 23, 213-216.

[25] **Slemrod, J.** 2007. "Cheating Ourselves: The Economics of Tax Evasion," *Journal of Economic Perspectives*, 21(1), 25-46.

[26] **Tedds, L.** 2010. "Estimating the Income Reporting Function for the Self-Employed," *Empirical Economics*, 38(3), 669-687.

[27] **Terefe, B., Barber-Dueck, C. and Lamontagne, M.-J.** 2011. "Estimating the Underground Economy in Canada, 1992-2008," Statistics Canada Report.

[28] **Tremblay, J., Lynch, J. and Dubreuil, G.** 2010. "Pilot Survey Results from the Canadian Survey of Household Spending Redesign," Joint Statistical Meetings.

[29] **Yitzhaki, S.** 1974. "A Note on Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 3(2), 201-02.

Appendix

Table 11: Base Interest Rate Assumptions for 1998 Liabilities

Type of debt	**Interest rate assumed (%)**
Student loan	7
Credit card debt	9
Home equity loan	6
Line of credit	7
Other debt	8

Table 12: Ongoing Expenses Comparison of the SFS and SHS

	1998				2004			
	SHS		SFS		SHS		SFS	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Mortgage	2743	4893	3438	7398	3359	5542	4429	10134
Rent	1757	2989	2116	3602	1820	3286	2344	4136
Condo Fees	75	437	114	648	108	608	171	916
Property Tax	1024	1111	1362	2122	1266	1351	1917	3156
House Insurance	288	306	389	563	385	407	593	1235
Oil and Gas	482	570	540	870	802	917	909	1226
Electricity	964	712	1010	1032	1154	848	1258	2308
Water	169	295	144	279	219	405	189	371
Vehicle	973	827	1099	1223	1351	1171	1502	1630
Childcare	313	1255	350	1601	326	1317	436	2476
Support Payments	218	1728	249	2153	268	3152	181	1336
Nobs	13558		14191		13272		4445	

Table 13: Demographic Comparison of the SFS and SHS

	1998				2004			
	SHS		SFS		SHS		SFS	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Age of Reference Person	49.45	16.45	47.73	16.86	50.88	16.32	49.11	16.82
Age of Spouse	30.13	25.26	27.06	24.63	31.03	26.16	29.29	26.07
Married	0.64	0.48	0.61	0.49	0.64	0.48	0.61	0.49
Number of Adults	1.84	0.71	1.80	0.74	1.83	0.72	1.80	0.75
Number of Youth	0.18	0.48	0.18	0.48	0.17	0.48	0.16	0.46
Number of Children	0.58	0.96	0.52	0.93	0.49	0.89	0.45	0.89
Number of Earners	1.35	1.08	1.34	1.03	1.37	1.08	1.39	1.03
Weeks Full-Time Reference Person	25.07	24.40	30.77	24.37	25.68	24.53	30.20	24.63
Major Source of Income -Wage	0.59	0.49	0.60	0.49	0.60	0.49	0.60	0.49
Major Source of Income - Self Employment	0.06	0.24	0.06	0.23	0.07	0.26	0.06	0.24
Single Detached House	0.65	0.48	0.60	0.49	0.66	0.47	0.60	0.49
Lowrise Apartment	0.14	0.35	0.15	0.36	0.14	0.35	0.18	0.38
Highrise Apartment	0.06	0.23	0.07	0.26	0.05	0.22	0.06	0.24
Large Urban (>500K)	0.34	0.47	0.39	0.49	0.33	0.47	0.44	0.50
Small Urban (<500K)	0.40	0.49	0.41	0.49	0.48	0.50	0.39	0.49
Own with no Mortgage	0.36	0.48	0.32	0.47	0.36	0.48	0.34	0.47
Own with Mortgage	0.34	0.47	0.32	0.47	0.35	0.48	0.31	0.46
Own Vehicle	0.80	0.40	0.80	0.40	0.80	0.40	0.78	0.41
	13558		14191		13272		4445	

Table 14: Imputed Expenditure, 1998

	SHS		SFS Baseline				SFS Implicates Average			
	Actual		Imputed, no Income		Imputed, Income		Imputed, no Income		Imputed, Income	
	CC	NDC	CC	NDC	CC	NDC	CC	NDC	CC	NDC
Mean	39594	21829	38562	20493	39718	21448	39508	21003	40253	22071
Percentile										
1 per cent	5600	-15745	7235	464	6629	-3455	-1060	-21865	-10631	-24694
5 per cent	10094	-162	10602	5323	10118	4718	9939	3768	9176	3098
10 per cent	12769	4063	13928	6960	13009	6239	13192	6115	12259	5622
25 per cent	20366	8993	22031	11414	21543	10718	21584	10732	21171	10356
50 per cent	33385	17276	35500	18829	35578	18587	35269	18980	35385	18652
75 per cent	50503	28808	50163	28112	51126	28504	51765	29665	52452	29914
90 per cent	71362	43608	66690	37605	69056	41084	70665	42094	72749	44838
95 per cent	86861	56260	78537	45966	83348	51289	86818	51537	90294	55648
99 per cent	147090	104274	130426	70812	166423	94462	144922	83449	182894	127195
Nobs	13558				14191					

Notes: CC refers to imputed Total Expenditure and NDC refers to imputed non-durable consumption and (dis)savings. We report the 1 and 99 percentiles only for illustrative purposes as the raw data do not support these percentiles as valid estimates at this level of detail.

Table 15: Under-Reporting Status Correlations Between Implicates, 1998

			Implicate					Implicate				
			Baseline	1	2	3	4	Baseline	1	2	3	4
		Baseline	1									
Imputed	Implicate	1	0.6013	1								
Without		2	0.6078	0.462	1							
Income		3	0.5862	0.4451	0.4573	1						
Variables		4	0.6036	0.4674	0.4764	0.4506	1					
		Baseline						1				
Imputed	Implicate	1						0.5398	1			
Using		2						0.5388	0.4036	1		
Income		3						0.5275	0.3904	0.3979	1	
Variables		4						0.5304	0.3777	0.396	0.3856	1

Table 16: Household After-Tax Income - Expenditure Correlations Between Implicates, 1998

			Implicate					Implicate				
			Baseline	1	2	3	4	Baseline	1	2	3	4
		Baseline	1									
Imputed	Implicate	1	0.8862	1								
Without		2	0.8988	0.818	1							
Income		3	0.8949	0.8387	0.8218	1						
Variables		4	0.8958	0.8209	0.833	0.8331	1					
		Baseline						1				
Imputed	Implicate	1						0.8836	1			
Using		2						0.8718	0.8037	1		
Income		3						0.8841	0.7983	0.7883	1	
Variables		4						0.8357	0.7624	0.777	0.7585	1

Table 17: Imputed Expenditure, 2004

	SHS		SFS Baseline				SFS Implicates Average			
	Actual		Imputed, no Income		Imputed, Income		Imputed, no Income		Imputed, Income	
	CC	NDC	CC	NDC	CC	NDC	CC	NDC	CC	NDC
Mean	48605	26236	49656	26094	53549	29691	51093	27036	55021	31001
Percentile										
1 per cent	5388	-20663	3310	-8380	4815	-11349	-7109	-29374	-14896	-36583
5 per cent	11216	-372	10221	4857	11172	4471	9003	-71	8748	1415
10 per cent	14990	4652	14564	6802	15383	6792	13253	5713	13989	6016
25 per cent	24479	10637	25950	12768	25172	12540	25149	11992	24530	11924
50 per cent	40180	20563	43570	22963	42044	22517	43185	23527	41504	22832
75 per cent	61233	34399	64918	35607	64539	34997	67909	38277	66897	37116
90 per cent	87730	52850	88559	49756	89621	50199	97494	57205	96494	56635
95 per cent	110033.8	71221	113221	59834	119287	60688	122253	67474	129871	74355
99 per cent	193466	129281	178103	103413	404893	321615	193632	101121	461967	344907
Nobs	13272					4340				

Notes: CC refers to imputed Total Expenditure and NDC refers to imputed non-durable consumption and (dis)savings.

Table 18: Consumption Items from the SHS Included in the Non-Durable Consumption Measure

F001		food at home and at restaurants
H001		household operation
	less:	childcare payment
	less:	domestic and other custodial services
J001		clothing
K008		rented and leased auto
K019		operation of auto
	less:	auto insurance
	less:	auto registration fees
K031		public transportation
L101		health care expenses
L201		personal care expenses
M101		recreation
	less:	purchase of recreational equipment
M201		reading materials
M301		education
	less:	tuition fee
N101		alcohol and tobacco