Who Does Not Share their Tax Information? And its Implication for Income Inequality

Jean-Michel Billette* Pierre Brochu† Louis-Philippe Morin‡

August 2012§

Abstract

This paper examines the economic implication of the decision to give 2006 Census respondent the option of letting Statistics Canada access their income tax files rather than answering income related questions directly. We find that the consent decision does matter when it comes to family income inequality, particularly for the bottom tail of the distribution. The consent decision does not, however, materially affect the estimation of standard wage equations.

Keywords: Census, Inequality Measurement, Information Sharing

JEL classification: C31, I32, J31.

*Statistics Canada

†Department of Economics, University of Ottawa

‡Corresponding author. Department of Economics, University of Ottawa, Ottawa, Ontario, Canada, K1N 6N5. Telephone: (613) 562-5800 ext. 4867. lmorin@uottawa.ca

§This work has benefited from comments from Miles Corak, Gilles Grenier, Andrew Heisz and conference and seminar participants at the 2012 CEA Meeting and the University of Ottawa. The data used in this study were accessed on site at the Carleton, Ottawa, Outaouais Local Research Data Centre (COOL RDC). This study reflects the views of the authors, not of Statistics Canada.

1 Introduction

The recent decision to replace the compulsory long-form Canadian Census with a voluntary survey (starting in 2011) has received much press, and its implications for sample representativity are well understood among economists (e.g. Dillon (2010); Green and Milligan (2010); Thompson (2010); Veall (2010)).

What has gone relatively unnoticed, however, is the new method of income information collection introduced in the previous Census, i.e. the 2006 Census. For the first time, respondents were given the option of 'sharing' their income-tax information.[1] That is, respondents could now let Statistics Canada access their income tax files instead of self-reporting their income. The sharing option proved popular; income information now comes from administrative records for the large majority (about 80 percent) of individuals (henceforth referred to as 'sharers'), with the remaining 20 percent coming from self-reported information (from 'non-sharers').[2] Considering the extensive literature that suggests that measurement error in income matters (e.g. Bound, Brown, and Mathiowetz (2001); Gottschalk and Huynh (2010)), this significant change in the way income information is obtained may introduce a non-trivial self-selection problem and have important implications for economic research.

In this paper, we explore the economic implications for family income inequality of the decision to give 2006 Census respondents the option of letting income information be directly accessed from their tax file. Numerous studies have looked at the implication of using income from tax versus survey data (e.g. Frenette, Green, and Picot (2006); Burkhauser, Feng,

[1] Prior to 2006, income data was self-reported.

[2] By giving consent, the respondent did not have to answer 13 earnings related questions. This reduction in respondent burden is probably the main reason why the greater majority of respondents gave Statistics Canada permission to access their tax data.

Jenkins, and Larrimore (2012)),[3] but to our knowledge the economic literature is silent with respect to the consent issue.[4]

From an estimation perspective, the advantages of using the Canadian Census are numerous. First, the fact that coverage is almost complete in the Census,[5] allows us to more cleanly identify the effect of the consent decision than studies using survey data where self-selection into surveys is an important confounder.[6] Second, the sample size of the long-form Census is large enough (about twenty percent of Canadian households) to look at fine portions of the income distributions. One can, for example, get very precise estimates of average incomes by vingtiles. Finally, unlike the available Canadian tax data (e.g. the Longitudinal Administrative Databank (LAD)), the Census contains a rich set of socioeconomic characteristics (e.g. educational attainment, gender/ethnicity characteristics, and labour force status). Observing the respondent's socioeconomic characteristics will allow us to see whether these characteristics can explain the differences in income distributions between sharers and non-sharers, shedding more light on potential measurement issues related to self-reporting.

Our paper finds important differences in the income distributions of sharer and non-sharer families, particularly at the bottom of the distribution. These results have important implications for measures of inequality. If, for example, one looks at the top/bottom decile

[3]Although there is much evidence that suggests important differences when using income from tax versus survey data, there are some exceptions (Burkhauser, Feng, Jenkins, and Larrimore (2012); Frenette, Green, and Picot (2006)). Burkhauser, Feng, Jenkins, and Larrimore (2012), by and large, do not find drastic differences between using survey self-reported data or administrative tax data in computing U.S. income shares—except at the top end of the distribution. Frenette, Green, and Picot (2006) examine trends in income inequality over the 1980s and 1990s using three different sources of data (i.e. Survey data (Consumer Finance (SCF) and Survey of Labour and Income Dynamics (SLID)), Census data and tax data). They argue that under-coverage in survey data (i.e. SCF/SLID) is more problematic, particularly at the bottom of the distribution, than the source of income (self-reported versus administrative).

[4]There has been technical papers both in Canada (e.g. Michaud, Dolson, Adams, and Renaud (1995); Abraham, Rivard, Giles, and Lathe (2001)), and abroad (e.g. Pascale (2011); Sakshaug and Kreuter (2011)) that have looked at the issue of consent bias. There is some evidence, albeit mixed, that consent varies according to socioeconomic characteristics, like gender, age, and education. These papers do not, however, explore its implication for questions of economic interest.

[5]Net under-coverage was 2.67 percent in the 2006 Census (Statistics Canada 2010).

[6]The high coverage of the Census data is a key reason why many Canadian studies (e.g. Frenette, Green, and Picot (2006); Frenette, Green, and Milligan (2007) and Boudarbat, Lemieux, and Riddell (2010)) argue that the Canadian Census is the better source of data to study income inequality and the evolution of the wage structure in Canada.

ratio for total family income, we would find 13.3 and 64.9 ratios for sharers and non-sharers, respectively. Not surprisingly, the top/bottom decile ratios are lower when focusing on after-tax income, and this is true for both groups. However, the top/bottom decile ratio is still much larger for non-sharers (50.0 versus 10.1). The (very) high top/bottom decile ratios of sharers suggest that self-selection may be an issue. The relatively low sharers' ratios may also be cause for concern.[7] If self-reported income tends to underestimate true income levels at the bottom of the distribution, it may lead researchers to underestimate the true extent of income inequality.[8]

We do find that sharers and non-sharers differ in their socioeconomic characteristics, but that these differences do not fully explain the income distribution differences. Just as suggested by the unconditional income distribution, our RIF-OLS regression results suggest that the differences in revealed income are concentrated at the lower tail.

When we focus on individual wages of full-time workers, as is typically done in the human capital literature, the consent decision has, for the most part, little impact. The coefficients of sharers and non-sharers tend to be very close in magnitude. The one exception is for visible minorities, but even then, the difference does not seem to affect the estimate based on all full-time workers. Our RIF-OLS results confirm our OLS findings.

Given that policies regarding consent are still in flux (Pascale 2011), understanding the implications of a move towards a mixed mode of income information collection is important. By relying more on administrative data, statistical agencies can reduce costs and respondent fatigue; the latter being an often cited explanation for the declining survey response rates.[9] If one is, however, to move towards a greater reliance on administrative data in economic research, as many researcher would like to see happen (e.g. Card, Chetty, Feldstein, and Saez

[7]The sharers' top/bottom decile ratio for total income is lower than for any of the five earlier censuses (Frenette, Green, and Picot 2006).

[8]Fortin, Green, Lemieux, Milligan, and Riddell (2012), which provides a current overview of the Canadian inequality literature, suggest that income inequality has increased in Canada since the early 1990s.

[9]There is a long history in the U.S. of linking surveys (e.g. CPS and SIPP) with administrative datasets. What makes the Canadian case unique, however, is that giving consent actually reduces the time needed to complete the survey, i.e. there are fewer questions that need to be answered when giving consent. This is in direct contrast to American surveys where the linked administrative data complement the survey data.

(2010)), one should also recognize its potential challenges. Our findings have implications as to whether one should give the respondent choice of consent or simply implicitly impose it, as non compulsory (or non random) consent seems to introduce another confounding factor for inequality measurement.

The rest of the paper is divided as follows. The next section describes the 2006 Census data, and how it can shed new light on income measurement issues. The income distributions of sharers and non-sharers is analyzed in Section 3. In section 4, we look at socioeconomic characteristics of sharers and non-sharers. Section 5 investigates whether the differences found in Section 3 can be explained by differences in socioeconomic characteristics. Section 6 focuses on standard wage equations, and Section 7 concludes.

2 Data

This study is based on the Master Files of the 2006 Census long-form questionnaire. Until 2006, the Census long-form questionnaires were targeting approximately 20 percent of Canadian households.[10] For social scientists, there are a number of advantages of using long-form Census files over other Canadian surveys: 1) Census master files were the largest Canadian data sets available to researchers containing detailed information on individuals' socioeconomic status, and; 2) completing the questionnaire was compulsory to ensure representability of the gathered information.

For the purpose of our study, the 2006 long-form Census files present a third crucial advantage: The 2006 Census gave each respondent aged 15 and up the option of letting Statistics Canada access their income tax file. Therefore, the 2006 Census was composed of three types of economic families: 1) families where all individuals consented to share their income tax information, 2) families where all individuals refused to share their income tax information, and 3) families where some individuals refused while others consented. The

[10]The (compulsory) Census long-form questionnaire was abolished and replaced in 2011 by a voluntary survey, the National Household Survey (NHS).

Census also asks respondent income tax paid (which was not the case in 2001). Given that tax paid is more challenging to recall (or estimate) than total income, the consequences of allowing the use administrative data may be exacerbated for after-tax inequality measures. Having a clear picture of both pre- and after-tax income inequality is critical if equity is a goal of public policy.

As with many other income inequality studies, the bulk of our analysis is concentrated on economic families as opposed to individuals (e.g. Frenette, Green, and Milligan (2007); Milligan (2008); Heisz (2007)).[11] To get a clearer view of the differences between sharers and non-sharers, we exclude economic families where some economic family members gave consent, and others did not. This ensures that the family level income measures do not come from a hybrid of self-reported and administrative information. Fortunately, the willingness to share tax information differs across family members in only 10 percent of the households. We also concentrate on private dwellings and therefore exclude all collective households/dwellings.[12] Finally, as with Frenette, Green, and Picot (2006), we drop individuals in economic families that live in the territories (about 1.1 percent of the original 2006 Census sample), and also those at the bottom and top 0.1 percent of each income distribution.

We focus on three measures of income: market income (which includes employment income, investment income, retirement pensions, superannuations and annuities, and other money income), total income (which is market income plus government transfers) and after-tax income (which represents total income minus personal provincial and federal taxes paid). The income measures are adjusted to account for economies of scale in larger families, i.e. family income is divided by the square root of the family size, but the unit of observation is

[11] In some studies (e.g. Burton and Phipps (2011); Lu, Morissette, and Schirle (2011)), the focus has been the *census* family.

[12] A collective household/dwelling refers to a dwelling of a commercial, institutional or communal nature. They include lodging or rooming houses, hotels, motels, tourist homes, nursing homes, hospitals, staff residences, communal quarters (military bases), work camps, jails, missions, group homes, and so on. They may be occupied by usual residents or solely by foreign and/or temporary residents.

the individual.[13] As such, our measures of inequality is with respect to all individuals, but based on economic family income (and household size).[14] An economic family is formally defined as *"two or more persons who live in the same dwelling and are related to each other by blood, marriage, common-law or adoption"* (Statistics Canada 2007). But, given that we want to cover the full population, we treat unattached individuals as economic families of size one.

3 Evidence

Figures 1 and 2 present the family-adjusted and individual total income distributions, respectively. Each figure displays the income distributions for sharers, non-sharers, and the overall population. Although there are important differences between sharers and non-sharers, the distributions overlap significantly. This suggests that the decision not to share income tax information is not localized to a specific part of the income distribution. Both figures, however, suggest important differences at the bottom tail. There appears to be relatively more non-sharers with total income of zero. Importantly, the proportion of non-sharers at the bottom of the income distribution is important enough to ensure that the densities for sharers and for the overall distribution diverge significantly as we get close to 0.[15] To better understand these patterns we explore mean income by vingtiles.

Table 1 presents the mean incomes by vingtile for economic families of sharers, non-sharers, sharers and non-sharers combined, and for all individuals.[16] The striking finding from Table 1 is that the income distributions of non-sharers are very different from the ones of sharers. The mean incomes are systematically smaller for non-sharers. This is true

[13] This approach was also favoured by Frenette, Green, and Picot (2006), Frenette, Green, and Milligan (2007), Frenette, Green, and Milligan (2009), and Heisz (2007).

[14] Alternatively we can think of having one "adult equivalent adjusted" income per family with the weight being adjusted to account for family size.

[15] Although not presented here, the picture for after-tax income is very similar.

[16] We replicated the mean income vingtiles for the 2001 Census where we again removed individuals in economic families that live in the territories, and those at the bottom and top 0.1 percent of the income distributions; recall that there was no consent question in the 2001 Census. We got vingtile results that are essentially the same as those of Table 3.3 in Frenette, Green, and Picot (2006).

for all vingtiles and all income measures (except for the first three vingtiles of the wage distributions which are bounded at 0).[17] Despite representing approximately 14 percent of the total sample, the income distributions of non-sharers seem to have a non-negligible impact on the overall distributions.

The differences in mean incomes between sharers and non-sharers is not homogeneous across vingtiles. The differences at the top four income vingtiles are relatively stable; the mean incomes for sharers being between eight and twelve percent higher for all income measures—a difference that is by no means trivial. For the lower vingtiles, the income gaps between sharers and non-sharers become even larger (in relative terms). A mean total income difference of $7,600 at the bottom decile is very large considering that non-sharers only averaged $1,724.[18] Finding larger differences at the bottom tail of the distribution has important implications for measures of inequality. If, for example, one looks at the top/bottom decile ratio for total income, as in Frenette, Green, and Picot (2006), we would find 13.3 and 64.9 ratios for sharers and non-sharers, respectively. The ratio for the combined sharers and non-sharers distribution is 16.2 (for the overall population this ratio is 15.9) which is significantly above 13.3. Clearly, despite representing a small proportion of the population, non-sharing families affect income inequality measures. Not surprisingly, the top/bottom decile ratios are lower when focusing on after-tax income, and this is true for both sharers and non-sharers. However, the top/bottom decile ratio is still much larger for non-sharers (50.0 versus 10.1).

If one compares the mean vingtiles across 2001 and 2006 censuses (see Table 2), there is evidence that suggests that sharers and non-sharers are different. The top/bottom decile ratio was 16.2 in the 2001 Census—when all income information was self-reported,[19] which

[17]Note that one should not see the vingtile mean incomes of the total population distributions as weighted averages of the sharers, non-sharers, and mixed-response mean incomes as, for example, individuals in the X vingtile of the sharers distribution might not be in the X vingtile of the total population distribution.

[18]The mean incomes for the top and bottom deciles are the averages of the bottom two and top two income vingtiles, respectively.

[19]Frenette, Green, and Picot (2006) found that the top/bottom decile ratio ranged from 14.1 to 17.0 over the 1981 through 2001 censuses.

is remarkably different from the 64.9 we found for non-sharers in the 2006 Census. It is possible that inequality may have increased over the 2000s, but the difference is too big to believe that this is the only explanation.

4 Who Does Not Share their Income-Tax Information?

The fact that the income distributions of sharers and non-sharers differ significantly does not imply that non-sharers misreport their income per se. It does suggest however that the two groups may be systematically different. This section looks at socioeconomic characteristics of each group to investigate whether these characteristics can explain the income distribution differences presented above. Table 3 presents descriptive statistics. The first two columns of Table 3 are for individuals where all (economic) family members agreed on the consent decision (all consented or all refused), while the last column presents the same information, irrespective of whether other family members agreed on the consent decision or not.

In terms of sharers versus non-sharers, we see small differences across gender and age groups. In particular, females are slightly more important among sharers. The more significant differences are for visible minorities, aboriginals, and immigrants—they are more prevalent among non-sharers. There is also a clear pattern for educational attainment with sharers tending to be more educated. Finally, while Ontario and British Columbia have larger representation among non-sharers, the opposite is true for Quebec.

We turn to regressions to explore which characteristics seem to be driving the consent decision. More specifically, we regress the consent binary variable on individual characteristics. Table 4 presents the results from estimating a linear probability model using the full adult population (all individuals aged 15 and up).[20]

[20] The consent decision is highly correlated within economic families. 93 percent of individuals aged 15 and up were in economic families in which all adults answered the same way to the consent question–they all consented or all refused. Therefore, it is possible that one person in the family decided for the other members. For this reason, we also estimated our model on the sample of reference persons. The results are very similar to the ones present here. We also found very similar results when we focused on individuals where all economic family members agreed on the consent decision. These findings are available upon request.

For the most part, the regression results are in line with the descriptive statistics. Visible minorities and Aboriginals are 3.9 and 5.9 percentage points, respectively, less likely to share their income-tax information than whites, while immigrants are 2.5 percentage points less likely to share than non-immigrants.[21] There is an education 'effect' that is mainly driven by less educated individuals; dropouts are 4.6 percentage points less likely to share than high-school graduates. This finding is in line with previous literature linking education to trust and citizenship (e.g. see Dee (2004), Milligan, Moretti, and Oreopoulos (2004) and Oreopoulos and Salvanes (2011)).[22] One noticeable difference with the summary statistics, however, is that there is a clearer age profile emerging from the regression results; the relationship is concave and peaks at the mid- to late-50s depending on the specification.[23]

5 Do Observable Characteristics Explain the Income Differences Between Sharers and Non-Sharers?

To investigate the presence of systematic income difference between sharers and non-sharers, we want to regress the economic family income measures on a dummy variable equal to 1 if the economic family consented to share their income tax information, controlling for the number of adults relative to the size of the economic family and the characteristics of these adults. To motivate the functional form of our (economic) family-level regression, we first start with a more or less standard model describing the link between individual income and individual characteristics. Specifically, we can imagine the revealed income of individual i

[21] Using German data, Sakshaug and Kreuter (2011) found that foreign citizens were less prone to provide consent than their German counterpart.

[22] For the technical literature that looks at consent bias, the education results are mixed. Pascale (2011) found that consent increases with the level of education, with the largest difference being for high school dropouts. However, Abraham, Rivard, Giles, and Lathe (2001) found the opposite. They found that more educated individuals were less likely to grant permission to access their tax data.

[23] Using 5-year age dummies instead of imposing a quadratic age profile leaves the estimates for the other parameters unchanged.

from family f can be linked to his/her personal characteristics in the following way:

$$y_{if} = \alpha + \beta Share_{if} + \boldsymbol{X_{if}\gamma} + \varepsilon_{if} \tag{1}$$

where $Share_{if}$ is a dummy variable equal to 1 if the individual consents to Statistics Canada using his/her income tax file as information source. $\boldsymbol{X_{if}}$ is a vector of personal characteristics (e.g., gender, visible minority status, immigrant status, education, age, and provincial fixed effects). Now, since we observe a fair number of adults with zero or negative income, we could not follow the common practice of regressing the log of income on personal characteristics (in level). We can, however, minimize the impact of choosing this particular functional form on the estimated coefficient of interest by relying on binary righthand-side variables. In particular, we transformed our continuous age variable into a series of dummy variables covering 5-year intervals.

We use equation (1) to express the family-adjusted income as a function of family characteristics. The economic-family total income is simply the sum of total incomes of all adults in the family:

$$\sum_{i=1}^{A_f} y_{if} = \alpha A_f + \beta A_f \times Share_f + \sum_{i=1}^{A_f} \boldsymbol{X_{if}\gamma} + \sum_{i=1}^{A_f} \varepsilon_{if}$$

where A_f is the number of adults in family f. Note that since we will be estimating the regression equations on families where all individuals shared or all refused, we can replace the sum of adult sharers by the number of adults times the consent dummy $A_f \times Share_f$. By dividing both sides of this last equation by the square-root of the economic-family size ($\sqrt{N_f}$), we can express the adjusted family income as:

$$\frac{1}{\sqrt{N_f}} \sum_{i=1}^{A_f} y_{if} = \alpha \frac{A_f}{\sqrt{N_f}} + \beta \frac{A_f}{\sqrt{N_f}} Share_f + \frac{1}{\sqrt{N_f}} \sum_{i=1}^{A_f} \boldsymbol{X_{if}\gamma} + \frac{1}{\sqrt{N_f}} \sum_{i=1}^{A_f} \varepsilon_{if}$$

or

$$\widetilde{y_f} = \widetilde{\alpha_f} + \beta \widetilde{Share_f} + \widetilde{\boldsymbol{X_f}} \boldsymbol{\gamma} + \widetilde{\varepsilon_f}. \tag{2}$$

The main difference between equations (1) and (2) is that the latter does not restrict $\boldsymbol{X_{if}\gamma}$ and ε_{if} to only affect y_{if}. One could imagine that, in families where there is some coordination, one member's characteristics could affect another member's income. If individuals were randomly assigned to families, and there was no interaction between family members, the results from estimating equations (1) or (2) should be very close (when using the appropriate weights).[24] It should be noted that we control for family size in equation (2) using the ratio of the number of workers on the square-root of the total family size (using $\widetilde{\alpha_f}$).

It is important to recognize that when interpreting the *Share* coefficient estimate, we do not claim to be capturing a causal relation from consent to income. It is, for example, possible that one's true income affects the likelihood to share one's tax information—the causality link could go either way. The coefficient estimate simply represents a difference in revealed income that is not explained by individual (or family) characteristics.

Below, we estimate equations (1) and (2) using OLS to capture potential differences in average income between sharers and non-sharers, conditioning on a series of covariates. We are, however, also interested in observed distributional differences (over and above the mean). We therefore use the RIF-OLS method proposed by Firpo, Fortin, and Lemieux (2009) to investigate whether distributional differences are non monotonic across the income population.[25]

Table 5 presents OLS results from estimating equation (1), where individual total income

[24] The family-level regressions (based on equation (2)) are estimated using only one observation per family while individual-level regressions (based on equation (1)) are estimated using all adults. The family-level regression weights are therefore adjusted by multiplying the individual-level weights by the family size.

[25] Two excellent introductions to the RIF regression methodology are Fortin, Lemieux, and Firpo (2011) and Heywood and Parent (2012), which apply the methodology to the female-male and to the white-black wage gaps, respectively.

is the dependent variable. Controls are added sequentially. Table 5 shows that the results are robust across specifications; there remains important differences in total income between sharers and non-sharers even when one adds controls. Specification (1) simply computes the difference in average total income. So, the estimate for the constant term in column (1) (i.e. $28,553) can be used as a reference point to judge the economic significance of the income difference. The difference ranges from $3,889 to $5,277 across specifications—a difference that is both economically and statistically significant. All other estimates have the expected sign. Females have lower incomes, as do visible minorities, Aboriginals, and immigrants. There is also a clear education profile; those with at least a bachelor degree earn more than high school graduates. Although not presented here, the (5-year) age fixed effects suggest a concave relationship with individuals aged 45 to 49 having the largest average total income difference.

Table 6 presents OLS results from estimating equation (2), where the adjusted family income is the dependent variable, and the economic family is the unit of observation. As is done in Table 5, controls are added sequentially. While the coefficient estimates for some of the controls vary in magnitude when aggregating individuals to families (e.g. the female coefficient estimates is about three times smaller when looking at economic families), the results for the consent coefficient estimates remain stable. The difference across sharers and non-sharers varies between $3,573 and $5,443 for the family-size adjusted income, and between $3,889 and $5,277 for individual total income. In the end, aggregation at the family level does not affect our findings.

Figure 3 compares the family-level OLS coefficient estimate for Sharer (based on specification (4) of Table 6) to the RIF-OLS estimated coefficients for the 1st to the 99th quantiles of the family-adjusted total income distribution.[26] Just as suggested by the unconditional income distribution, the RIF-OLS regression estimates suggest that the differences in re-

[26]Since the RIF-OLS estimation requires the inclusion of a constant term in the estimated equation, we compare the OLS and RIF-OLS results based on Specification (4) of Table 6, but including a constant term. Doing so does not affect our results. The OLS Sharer coefficient estimate is $3,770 when including a constant term versus $3,751 when not.

vealed incomes are concentrated at the bottom end of the wages distribution. The estimates seem to increase at the very top of the income distribution, but they become significantly less precise. Focusing on after-tax, or individual-level income provides similar results (and are available upon request).

6 Consent Decision and Standard Wage Equations

In the previous sections, we have shown that the consent decision may distort poverty measures, especially bottom sensitive ones, and that the income differences between consenting and non-consenting families cannot be explained by observable characteristics. We now investigate whether allowing people to share their income-tax information can also affect the estimation of standard wage equations. Given that sharers and non-sharers differ significantly in terms of race, immigrant status and education, allowing individuals to share their income-tax information could have some non-trivial implications for researchers interested in the link between these characteristics and wages.

Since most of the differences in wages between sharers and non-sharers are found at the bottom of the distribution, it is not obvious whether we will still observe a difference once we focus on workers. It is also not clear whether the wage equations will yield different coefficient estimates, and if so, whether these differences are large enough to be potentially problematic for researchers.

A common practice in labour economics (e.g. see Katz and Murphy (1992), Card and Lemieux (2001), and Boudarbat, Lemieux, and Riddell (2010)) is to concentrate on full-time workers and exclude individuals making less than the equivalent of about half the minimum wage. In order to gauge the potential seriousness of the wage self-reporting issues, we follow the previous literature and concentrate on individuals who worked mainly full-time in 2005 and earned more than $75 per week (in 2000 dollars), as is done in Boudarbat, Lemieux, and Riddell (2010)—a study that also looks at the 2006 Census data. Since the earnings of individuals who worked only a few weeks in 2005 might be more volatile, we also look, as a robustness check, at 'full-time-full-year' workers (i.e. individuals who worked mainly full-time and for 48 weeks or more in 2005). Following most of the literature, we also exclude self-employed from the analysis.

Table 7 presents the results from estimating a 'standard' wage equation on full-time workers. The estimation is done separately for sharers, non-sharers and for all full-time workers. These three sets of regression results allow us to 1) test for differences in coefficient estimates between sharers and non-sharers and, 2) gauge whether the differences (if there are any) matter for estimating a standard wage equation on all full-time workers.

Overall, Table 7 suggests that, despite being for the most part statistically different, the coefficient estimates for sharers and non-sharers are (in some cases surprisingly) close in magnitude. The one noticeable difference is for visible minority, but even this large difference does not seem to affect the coefficient estimate based on all full-time workers. One factor that could explain this finding is that the proportion of full-time workers (86.2 percent) that share their income-tax information is even larger than in the population in general (82.6 percent).[27]

As a robustness check, we verify whether the observed differences in estimates vary across quantiles. Figure 4 presents RIF-OLS estimates (from the 1st to 99th quantile) for the female, visible minority, immigrant and no diploma variables—the four variables for which the parameter estimates differed the most across sharers and non-sharers. In all cases, the differences in estimates are relatively stable across quantiles. The only exception is for immigrant where the difference seems larger at the bottom of the wage distribution; the difference is about 5 percentage points at the lowest vingtile and close to zero and not statistically significant at the top of the distribution. In the end, the RIF-OLS results suggest

[27]As a robustness check, we also estimated the same regression, but concentrating on individuals who worked full-time and 48 weeks as in Morin (2011). The results were very close to the ones for full-time workers.

that the Table 7 results, which focus on the mean, are not missing much when it comes to the wages of full-time workers.

7 Conclusion

In this paper, we explore the economic implications of the new method of income information collection introduced in the 2006 Census—both for family income inequality and the estimation of standard wage equation. The fact that respondents could now let Statistics Canada access their income tax files, instead of self-reporting their income, matters for family income inequality. More precisely, we find large differences between sharer and non-sharer families at the bottom tail of the family income distribution. We also find that the two groups differ in their socioeconomic characteristics, but these differences cannot fully account for the income distribution differences. This holds true whether one focuses on the mean, or expand our analysis to other part of the distribution.

Interestingly, the consent decision has little impact for the estimation of standard wage equations. The coefficients estimates for sharers and non-shares are, for the most part, similar in magnitude. When they do differ, as is the case for visible minority, it is not large enough to affect the coefficient estimate based on all full-time workers.

References

Abraham, E., M. Rivard, P. Giles, and H. Lathe (2001): "Results of the Tax Permission Question in the Survey of Labour and Income Dynamics," Discussion Paper 75F0002MIE-01002, Statistics Canada.

Boudarbat, B., T. Lemieux, and W. C. Riddell (2010): "The Evolution of the Returns to Human Capital in Canada, 1980-2005," *Canadian Public Policy*, 36(1), 63–89.

Bound, J., C. Brown, and N. Mathiowetz (2001): "Measurement Error in Survey Data," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 5 of *Handbook of Econometrics*, chap. 59, pp. 3705–3843. Elsevier.

Burkhauser, R. V., S. Feng, S. P. Jenkins, and J. Larrimore (2012): "Recent Trends in Top Income Shares in the United States: Reconciling Estimates from March CPS and IRS Tax Return Data," *Review of Economics and Statistics*, 94(2), 371–388.

Burton, P., and S. Phipps (2011): "Families, Time, and Well-Being in Canada," *Canadian Public Policy*, 37(3), 395–423.

Card, D., R. Chetty, M. S. Feldstein, and E. Saez (2010): "Expanding Access to Administrative Data Research in the United States," American Economic Association, ten years and beyond: Economists answer NSF's call for long-term research agendas.

Card, D., and T. Lemieux (2001): "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis," *Quarterly Journal of Economics*, 116(2), 705–746.

Dee, T. S. (2004): "Are There Civic Returns to Education?," *Journal of Public Economics*, 88(9-10), 1697–1720.

Dillon, L. (2010): "The Value of the Long Form Canadian Census for Long Term National and International Research," *Canadian Public Policy*, 36(3), 389–393.

Firpo, S., N. M. Fortin, and T. Lemieux (2009): "Unconditional Quantile Regressions," *Econometrica*, 77(3), 953–973.

Fortin, N. M., D. A. Green, T. Lemieux, K. Milligan, and W. C. Riddell (2012): "Canadian Inequality: Recent Developments and Policy Options," *Canadian Public Policy*, 38(2), 121–145.

Fortin, N. M., T. Lemieux, and S. Firpo (2011): "Decomposition Methods in Economics," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 4A, chap. 1, pp. 1–102. Elsevier.

Frenette, M., D. A. Green, and K. Milligan (2007): "The Tale of the Tails: Canadian Income Inequality in the 1980s and 1990s," *Canadian Journal of Economics*, 40(3), 734–764.

——— (2009): "Taxes, Transfers, and Canadian Income Inequality," *Canadian Public Policy*, 35(4), 389–411.

Frenette, M., D. A. Green, and G. Picot (2006): "Rising Income Inequality in the 1990s: An Exploration of Three Data Sources," in *Dimensions of Inequality in Canada*, ed. by D. A. Green, and J. R. Kesselman. UBC Press, Vancouver, Canada.

Gottschalk, P., and M. Huynh (2010): "Are Earnings Inequality and Mobility Overstated? The Impact of Nonclassical Measurement Error," *Review of Economics and Statistics*, 92(2), 302–315.

Green, D. A., and K. Milligan (2010): "The Importance of the Long Form Census to Canada," *Canadian Public Policy*, 36(3), 383–388.

Heisz, A. (2007): *Income Inequality and Redistribution in Canada: 1976-2004*. Statistics Canada Catalogue no. 11F0019MIE2007298, Ottawa.

Heywood, J. S., and D. Parent (2012): "Performance Pay and the White-Black Wage Gap," *Journal of Labor Economics*, 30(2), pp. 249–290.

Katz, L. F., and K. M. Murphy (1992): "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 107(1), 35–78.

Lu, Y., R. Morissette, and T. Schirle (2011): "The Growth of Family Earnings Inequality In Canada, 1980-2005," *Review of Income and Wealth*, 57(1), 23–39.

Michaud, S., D. Dolson, D. Adams, and M. Renaud (1995): *Combining Administrative and Survey Data to Reduce Respondent Burden in Longitudinal Surveys*vol. 95-19, pp. 11–20. Statistics Canada.

Milligan, K. (2008): "The Evolution of Elderly Poverty in Canada," *Canadian Public Policy*, 34(s1), 79–94.

Milligan, K., E. Moretti, and P. Oreopoulos (2004): "Does Education Improve Citizenship? Evidence from the United States and the United Kingdom," *Journal of Public Economics*, 88(9-10), 1667–1695.

Morin, L.-P. (2011): "Cohort Size and Youth Earnings: Evidence from a Quasi-Experiment," CLSRN working papers, UBC Department of Economics.

Oreopoulos, P., and K. G. Salvanes (2011): "Priceless: The Nonpecuniary Benefits of Schooling," *Journal of Economic Perspectives*, 25(1), 159–84.

Pascale, J. (2011): "Requesting Consent to Link Survey Data to Administrative Records: Results from a Split-Ballot Experiment in the Survey of Health Insurance and Program Participation (SHIPP)," Survey methodology study series 2011-03, Center for Survey Measurement Research and Methodology, Census Bureau, Washington D.C.

Sakshaug, J., and F. Kreuter (2011): "Assessing the Magnitude of Administrative Non-Consent Biases in the German Pass Study," Manuscript.

Statistics Canada (2007): *2006 Census Dictionary*. Catalogue no. 92-566-XWE, Ottawa.

——— (2010): *2006 Census Technical Report: Coverage*. Catalogue no. 92-567-X, Ottawa.

Thompson, D. (2010): "The Politics of the Census: Lessons from Abroad," *Canadian Public Policy*, 36, 377–382.

Veall, M. R. (2010): "2B or Not 2B? What Should Have Happened with the Canadian Long Form Census? What Should Happen Now?," *Canadian Public Policy*, 36(3), 395–399.

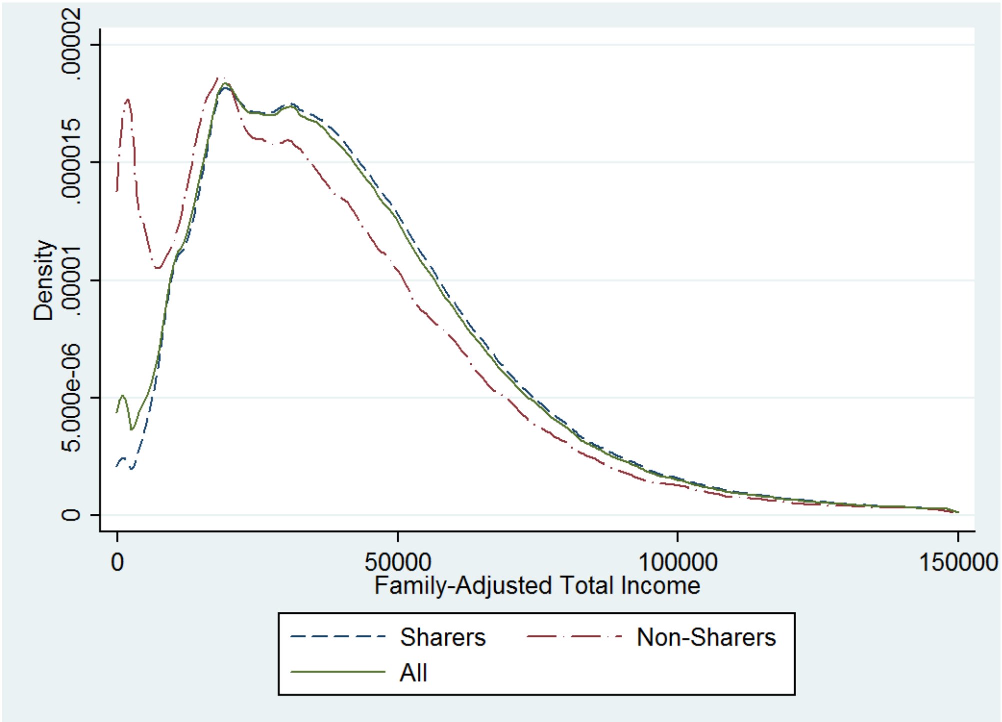

Figure 1: Family-Adjusted Total Income Distributions of Sharers and Non-Sharers

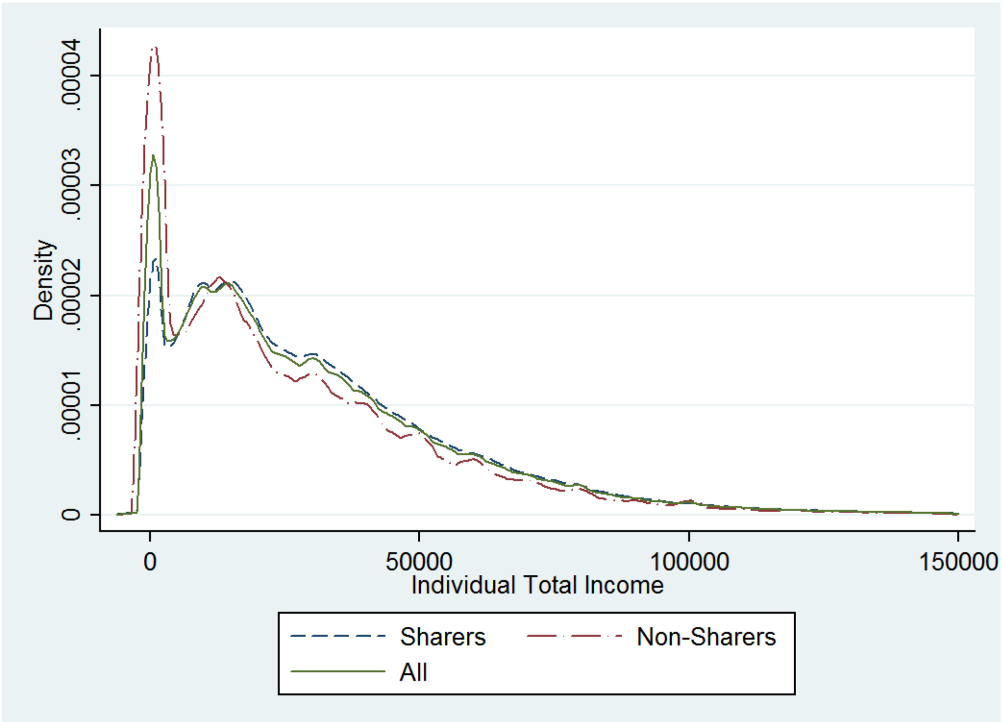

Figure 2: Individual Total Income Distributions of Sharers and Non-Sharers

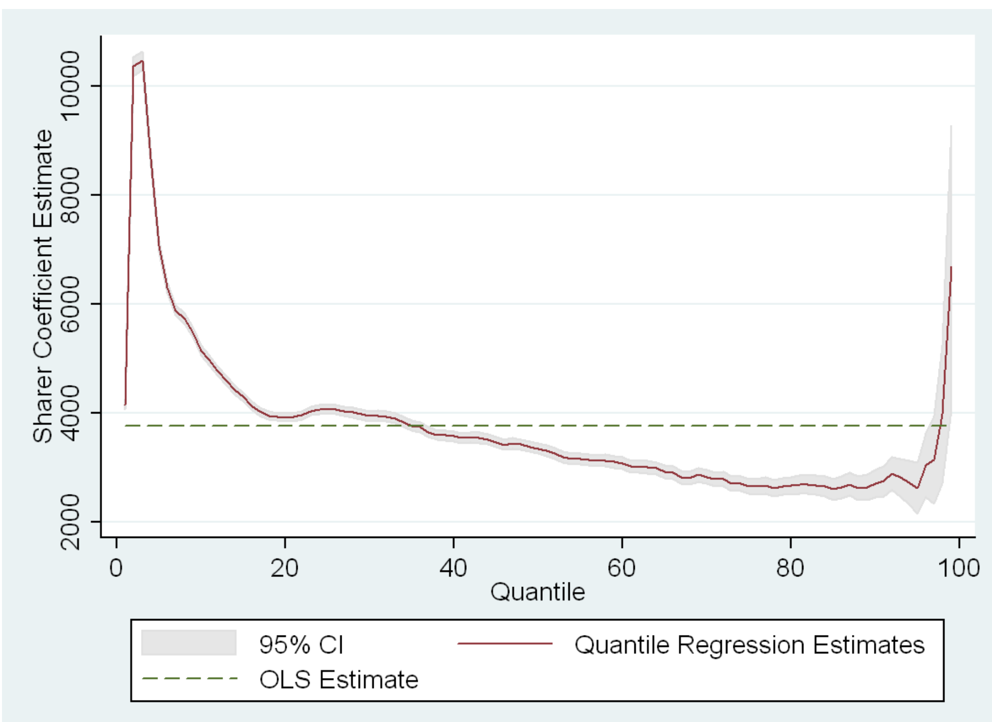

Figure 3: Sharer Coefficient Estimate by Quantile (Family-Adjusted Total Income)

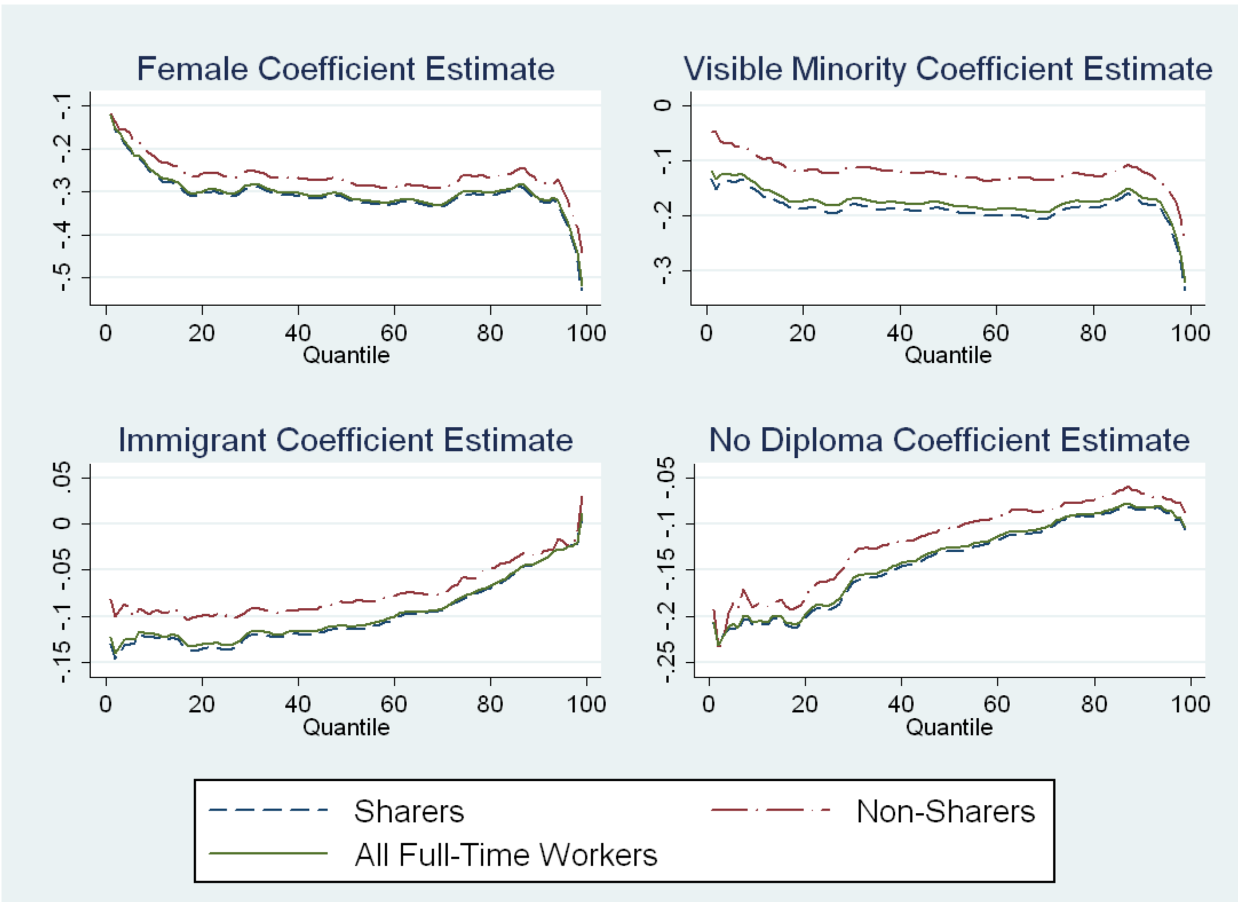

Figure 4: Coefficient Estimates by Quantile (Full-Time Workers)

Table 1: Mean Income by Vingtile, Economic Family (Adjusted for Family Size, $2005)

	Bottom	2nd	3rd	4th	8th	10th	12th	17th	18th	19th	Top
A. Wages											
Sharers	0	0	0	476	17,758	25,572	33,193	57,368	65,391	77,683	127,801
Non-Sharers	0	0	0	0	8,577	18,063	26,616	51,830	59,966	72,070	116,771
Sharers & Non-Sharers	0	0	0	139	16,529	24,523	32,263	56,612	64,633	76,903	126,120
All	0	0	0	197	16,858	24,690	32,321	56,456	64,439	76,662	125,923
B. Market Income											
Sharers	-49	2,482	7,001	11,096	25,425	32,141	39,167	63,667	72,509	86,959	157,257
Non-Sharers	-70	0	0	1,237	17,690	25,231	32,769	57,829	66,414	80,144	139,897
Sharers & Non-Sharers	-73	1,108	5,410	9,677	24,344	31,174	38,279	62,841	71,656	85,994	154,648
All	-71	1,277	5,652	9,878	24,411	31,186	38,243	62,664	71,438	85,725	154,332
C. Total Income											
Sharers	6,316	12,331	16,121	19,054	30,725	36,653	43,011	65,980	74,533	88,746	158,847
Non-Sharers	307	3,140	7,433	11,848	23,693	30,057	36,658	60,203	68,513	82,000	141,767
Sharers & Non-Sharers	4,166	10,915	14,939	18,033	29,735	35,716	42,131	65,170	73,692	87,795	156,278
All	4,275	11,003	14,999	18,085	29,730	35,683	42,071	64,991	73,485	87,531	155,952
D. After Tax Income											
Sharers	6,144	12,145	15,703	18,226	27,292	31,788	36,590	53,781	60,046	70,112	114,282
Non-Sharers	294	3,093	7,350	11,667	21,880	26,766	31,841	49,584	55,710	65,389	103,896
Sharers & Non-Sharers	4,031	10,767	14,627	17,369	26,534	31,086	35,936	53,194	59,435	69,457	112,725
All	4,141	10,854	14,679	17,406	26,536	31,071	35,896	53,062	59,278	69,248	112,433

Notes. The summary statistics are weighted.

Table 2: Evolution of Mean Family-Adjusted Income by Vingtile ($2000)

	Bottom	2nd	3rd	4th	8th	10th	12th	17th	18th	19th	Top
A. Wages											
2001	0	0	0	77	14,969	21,594	28,006	47,784	54,296	64,283	99,981
2006	0	0	0	175	15,030	22,013	28,817	50,335	57,453	68,351	112,272
B. Market Income											
2001	-10	514	3,747	7,828	21,092	26,913	32,821	52,890	60,096	71,656	119,780
2006	-63	1,138	5,039	8,807	21,765	27,805	34,097	55,871	63,693	76,431	137,601
C. Total Income											
2001	3,129	8,895	12,676	15,518	25,471	30,500	35,851	54,945	61,985	73,410	121,403
2006	3,812	9,810	13,373	16,124	26,507	31,815	37,510	57,945	65,518	78,042	139,045

Notes. The summary statistics are weighted.

Table 3: Summary Statistics, Means (individuals 15 years and up)

	Sharers	Non-Sharers	Overall Population
A. Gender			
Female	.516	.510	.515
B. Age			
Age 15 to 29	.230	.234	.241
Age 30 to 44	.268	.262	.265
Age 45 to 64	.336	.337	.335
Age 65+	.165	.167	.159
C. Educational Attainment			
Less than High School	.225	.254	.237
High School	.256	.257	.256
Post Secondary	.332	.318	.326
Bachelor's and up	.187	.171	.182
D. Immigrant Status			
Yes	.226	.291	.237
E. Visible Minority Status			
Caucasian	.830	.761	.814
Aboriginal	.027	.039	.030
Non-Caucasian, Non-Aboriginal	.143	.200	.156
F. Family Size			
1	.175	.226	.170
2	.318	.300	.304
3	.184	.165	.183
4	.196	.174	.201
5+	.128	.134	.142
G. Province/Territories			
Newfoundland	.017	.012	.017
Prince Edward Island	.004	.004	.004
Nova Scotia	.030	.024	.030
New Brunswick	.024	.019	.024
Quebec	.251	.215	.242
Ontario	.377	.409	.384
Manitoba	.035	.036	.035
Saskatchewan	.030	.028	.030
Alberta	.104	.094	.102
British Columbia	.126	.159	.133
Observations	3,989,945	759,705	5,119,850

Notes. The summary statistics are weighted. The number of observations are rounded to a base of 5. The overall population contains individuals from 'all-sharers', 'all-non-sharers', and 'mixed-response' economic families.

Table 4: Consent Decision: Linear Probability Model

	(1)	(2)	(3)
Age	.007***	.005***	.005***
	(.0000)	(.0000)	(.0000)
Age^2	-.000***	-.000***	-.000***
	(.0000)	(.0000)	(.0000)
Female	.005***	.004***	.004***
	(.0003)	(.0003)	(.0003)
Visible Minority	-.041***	-.042***	-.039***
	(.0006)	(.0006)	(.0006)
Aboriginal	-.073***	-.062***	-.059***
	(.0010)	(.0010)	(.0010)
Immigrant	-.029***	-.031***	-.025***
	(.0005)	(.0005)	(.0005)
No Diploma		-.044***	-.046***
		(.0005)	(.0005)
Some Post Secondary		.005***	.003***
		(.0004)	(.0004)
Bachelor and Up		.018***	.018***
		(.0005)	(.0005)
Constant	.677***	.716***	.709***
	(.0010)	(.0011)	(.0010)
Province Fixed Effects	No	No	Yes
R^2	.011	.014	.015
N	5,119,850	5,119,850	5,119,850

Notes. The estimations were done using Census weights. The number of observations are rounded to a base of 5. Robust standard errors are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5: Tax Information Sharing and Total Income: OLS Results, Individual level

	(1)	(2)	(3)	(4)
Sharer	5,277***	4,525***	3,889***	4,052***
	(45.5)	(42.1)	(40.4)	(40.2)
Female		-14,032***	-14,000***	-13,967***
		(30.5)	(29.3)	(29.1)
Visible Minority		-6,736***	-7,880***	-8,290***
		(54.3)	(52.2)	(52.0)
Aboriginal		-10,040***	-6,021***	-6,692***
		(91.9)	(88.5)	(89.1)
Immigrant		-1,666***	-3,200***	-4,587***
		(45.9)	(44.1)	(44.5)
No Diploma			-7,114***	-6,608***
			(43.6)	(43.4)
Some Post Secondary			4,882***	5,216***
			(39.1)	(38.9)
Bachelor and Up			23,181***	23,227***
			(45.9)	(45.6)
Constant	28,553***	48,289***	42,909***	45,518***
	(41.8)	(63.0)	(65.9)	(68.2)
Age Fixed Effects	No	Yes	Yes	Yes
Province Fixed Effects	No	No	No	Yes
R^2	0.00	0.15	0.22	0.23
N	4,740,330	4,740,330	4,740,330	4,740,330

Notes. The estimations were done using Census weights. The number of observations are rounded to a base of 5. Robust standard errors are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 6: Tax Information Sharing and Total Income: OLS Results, Economic Family Level

	(1)	(2)	(3)	(4)
Sharer (adj)	5,443***	4,225***	3,573***	3,751***
	(46.8)	(45.0)	(42.6)	(42.2)
Female (adj)		-3,425***	-4,411***	-4,354***
		(77.8)	(73.7)	(72.9)
Visible Minority (adj)		-7,146***	-7,601***	-7,761***
		(60.4)	(57.2)	(56.8)
Aboriginal (adj)		-13,238***	-7,841***	-8,726***
		(107)	(102)	(103)
Immigrant (adj)		-3,289***	-5,598***	-7,592***
		(58.8)	(55.9)	(56.4)
No Diploma (adj)			-8,876***	-7,918***
			(62.9)	(62.5)
Some Post Secondary (adj)			3,642***	4,391***
			(58.1)	(57.6)
Bachelor and Up (adj)			25,207***	25,446***
			(64.1)	(63.4)
Constant (adj)	27,493***	44,168***	39,952***	42,460***
	(43.0)	(91.3)	(94.8)	(95.7)
Age Fixed Effects (adj)	No	Yes	Yes	Yes
Province Fixed Effects (adj)	No	No	No	Yes
R^2	0.59	0.63	0.66	0.67
N	2,490,890	2,490,890	2,490,890	2,490,890

Notes. The estimations were done using Census weights adjusted for family size. The number of observations are rounded to a base of 5. Robust standard errors are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 7: Consent Decision and Log Weekly Wages (Full-Time Workers)

	Sharers	Non-Sharers	Difference	All Full-Time Workers
Age	.081***	.080***	.001**	.081***
	(.0002)	(.0006)	(.0007)	(.0002)
Age^2	-.001***	-.001***	.000	-.001***
	(.0000)	(.0000)	(.0000)	(.0000)
Female	-.305***	-.263***	-.042***	-.299***
	(.0008)	(.0021)	(.0022)	(.0008)
Visible Minority	-.187***	-.120***	-.067***	-.175***
	(.0016)	(.0034)	(.0038)	(.0014)
Aboriginal	-.133***	-.145***	.011*	-.136***
	(.0026)	(.0057)	(.0063)	(.0024)
Immigrant	-.099***	-.075***	-.025***	-.096***
	(.0014)	(.0031)	(.0034)	(.0013)
No Diploma	-.140***	-.117***	-.023***	-.137***
	(.0015)	(.0036)	(.0038)	(.0013)
Some Post Secondary	.146***	.145***	.001	.146***
	(.0011)	(.0027)	(.0029)	(.0010)
Bachelor and Up	.459***	.437***	.022***	.456***
	(.0012)	(.0030)	(.0032)	(.0011)
Constant	4.80***	4.79***	.010	4.80***
	(.0047)	(.0117)	(.0130)	(.0044)
Province Fixed Effects	Yes	Yes		Yes
R^2	.27	.23		.26
N	2,009,140	329,160		2,338,305

Notes. The estimations were done using Census weights. The number of observations are rounded to a base of 5. Robust standard errors are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.