

CAHIER DE RECHERCHE #1306E
Département de science économique
Faculté des sciences sociales
Université d'Ottawa

WORKING PAPER #1306E
Department of Economics
Faculty of Social Sciences
University of Ottawa

Opting or Not Opting to Share Income Tax Information with the Census: Does it Affect Research Findings?*

Pierre Brochu[†], Louis-Philippe Morin[‡] and Jean-Michel Billette[§]

November 2013

* 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.

[†] Department of Economics, University of Ottawa, 120 University Private (9031), Ottawa, Ontario, Canada, K1N 6N5; Email: pbrochu@uottawa.ca.

[‡] Department of Economics, University of Ottawa, 120 University Private (9056), Ottawa, Ontario, Canada, K1N 6N5; Email:lmorin@uottawa.ca.

[§] Statistics Canada.

Abstract

This paper examines the 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 giving respondents the option to share their income tax files (or not) adds a confounding factor when it comes to measuring 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.

Key words: *Census, record linking, consent, income inequality, wage equation*

JEL Classification: C31, I32, J31.

Résumé

Ce document examine l'implication de la décision de donner aux répondants du recensement de 2006 l'option de donner permission à Statistique Canada de consulter leur déclaration d'impôt plutôt que de répondre directement aux questions portant sur le revenu. Nous trouvons que donner l'option aux répondants de partager (ou pas) leur rapport d'impôt ajoute un facteur de confusion lorsqu'il s'agit de mesurer l'inégalité du revenu des familles, particulièrement dans le bas de la distribution. Toutefois, la décision de consentement ne touche pas de façon importante l'estimation des équations salariales standard.

Mots clés : *recensement, couplage des données, consentement, inégalités des revenus, équation de salaire*

Classification JEL : C31, I32, J31.

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 (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.ⁱ That is, respondents could now let Statistics Canada access their income tax files instead of self-reporting their income. The sharing option proved popular; in 2006, income information came 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 (‘non-sharers’).ⁱⁱ Considering the extensive literature that suggests that measurement error in income matters (Bound, Brown and Mathiowetz 2001), and particularly that the problem differs across survey and administrative data (e.g. Abowd and Stinson (forthcoming); Gottschalk and Huynh (2010); Kapteyn and Ypma (2007)),ⁱⁱⁱ this significant change in the way income information is obtained may introduce a non-trivial self-selection problem, and as such, have important implications for empirical research.

In this paper, we explore the implications for family income inequality (and for wage-equation estimation) of the decision to give 2006 Census respondents the option of letting income information be directly accessed from their tax file. In doing so, we also investigate who decided not to share their income-tax information. Numerous studies have looked at the implications of using income from tax versus survey data (e.g. Burkhauser et al. (2012); Frenette, Green and Picot (2006)), but to our knowledge the literature is silent with respect to the consent issue.^{iv}

The fact that coverage is almost complete in the Census^v 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. The high coverage of the Census data is also a key reason why many previous Canadian studies (e.g. Frenette, Green and Milligan (2007); Frenette, Green and Picot (2006)) have relied on the Canadian Census to study income inequality; it ensures a better representativity of the bottom tail of the income distribution. The Census has additional features that are useful for our analysis. 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. Finally, 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.

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 ratio for total family-adjusted 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 non-sharers suggest that self-selection may be an issue. The relatively low sharers' ratios may also be cause for concern.^{vi} If self-reported income tends to underestimate true income levels at the bottom of the

distribution, then moving from self-reported to administrative income information could lead researchers to falsely conclude that inequality is decreasing.^{vii}

We 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 income distributions, our regression results (that control for differences in characteristics across sharers and non-sharers) also suggest that the differences are concentrated at the lower tail of the distribution. The evidence leads us to believe that there is something (unobservable) about these people that makes them non-sharers, and at the same time leaves them in the lower part of the income distribution. Given that measurement error varies across data sources, the presence of self-selection will be a confounding factor when looking at income inequality.

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 coefficient estimates 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.

Given that policies regarding consent are still in flux (Pascale 2011), understanding the implications for researchers 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.^{viii} If one is, however, to move towards a greater reliance on administrative data in empirical research, as many researchers would like to see happen (e.g. Card et al. (2010)), one should also recognize its potential challenges.

The rest of the paper is divided as follows. The next section describes the 2006 Census data, and how they can shed new light on income measurement issues. The income distributions of sharers and non-sharers are 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 discusses the implications of our findings and then 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. 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. Finally, 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.

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); Heisz (2007); Milligan (2008)). 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 ten percent of the households. We also concentrate on private dwellings and therefore exclude all collective households/dwellings. 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 four measures of income: wages (wages and salaries), market income (which includes wages and salaries, total self-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 the individual.^{ix} As such, our measures of inequality are with respect to all individuals, but based on economic family income (and household size). 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 individual and family-adjusted total income densities, respectively. Each figure displays the income densities for sharers and non-sharers. The densities of sharers and non-sharers overlap significantly. This suggests that the decision not to share income tax information is not localized to a specific part of the income density. However, there are some important differences across sharers and non-sharers. First, Figure 1 suggests that there is

significant heaping in the non-sharer density. In particular, we can see clear ‘blips’ at \$30,000 and \$50,000 and at other multiples of ten thousand. This is not surprising. Statistics Canada (2008) shows that the new response method drastically decreased the amount of heaping in employment income. For example, in the 2001 Census 14.4 percent of positive employment income ended in ‘0,000’ (a multiple of ten thousand); this fraction dropped to 2.6 percent in 2006 (Statistics Canada 2008). This reduction in heaping suggests that the tax-data may reduce the amount of measurement error due to rounding.^x Note that when we turn to family-adjusted income, in Figure 2, heaping seems to disappear almost completely. This can be explained by the fact that there is significant variation in (square-root) family size, and that the individual incomes are ‘averaged out’. The second difference between the densities of sharers and non-sharers (and the most crucial one in our opinion) is that there appears to be relatively more non-sharers at the bottom of the income density. To further explore this issue (and its impact on inequality) we focus on the cumulative distribution function (CDF).

[insert Figures 1 and 2 here]

Figure 3 presents the family-adjusted CDFs for wages, market income, total income and after-tax income. Figure 3 clearly highlights what was suggested by Figures 1 and 2: there are important differences in the income distributions of sharers and non-sharers. Not surprisingly, a Kolmogorov-Smirnov test would reject the hypothesis that the distributions are the same at the one percent significance level. Of particular interest is the fact that the sharers’ CDF stochastically dominate the non-sharer’s CDF (in a first-order sense), if we abstract from the few negative income values. The differences between sharers and non-sharers are more pronounced in relative terms at the bottom percentiles. A mean total income difference of \$7,600 at the bottom decile is, for example, very large considering that non-sharers only averaged \$1,724.^{xi} Finding larger

differences at the bottom tail of the distribution has important implications for measures of inequality. If, for example, we look 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 ratio is still much larger for non-sharers (50.0 versus 10.1).

[insert Figure 3 here]

The fact that the sharers' top/bottom decile ratios are so large, as compared those of sharers and those of previous censuses, suggests that these individuals may be different from the rest of the population. We would not expect to observe such large differences in top/bottom decile ratios if people were randomly assigned to non-sharers or sharers, and the main consequence of not sharing was to introduce a rounding error. Clearly, self-selection into the sharer and non-sharer groups is a potential confounding factor when looking at measures of inequality---an issue that we further explore in subsequent sections.

In Figure 4, we compare the 2000 and 2005 CDFs for family-adjusted wages, market income, and total income (after-tax income is not available in the 2001 Census). The two distributions are similar. The top/bottom decile ratio, for example, was 16.2 in the 2001 Census---when all income information was self-reported, as opposed to 15.9 in the 2006 Census. If one were to ignore the change in the income-information collection, one could conclude that, if anything, inequality (as measured by the top/bottom decile ratio) decreased between 2000 and 2005. This

would seem to contradict some of the evidence from Fortin et al. (2012). Although they find that the Gini coefficient has fallen in the 2000s prior to the recent downturn, they do find some evidence of increased inequality when focusing on individual wages.^{xiii} Now, if people at the bottom of the income distribution tend to underestimate their income when self-reporting - something we will come back to in Sections 5 and 7 - then the appearance of a decrease in inequality in the Census between 2000 and 2005 could be simply due to the change in the source of income-information collected. The top/bottom decile ratio could have been over-estimated in 2001 simply because the 5th and 10th percentiles were under-estimated (due to self-reporting). Turning to a mix of tax and self-reported data could push up the 5th and 10th percentiles (if the tax data are more accurate for small income values), which would in turn decrease the top/bottom decile ratio. If this is the case, the 2006 Census data could wrongly suggest that inequality decreased between 2000 and 2005, and this is not even accounting for the (potentially) confounding effect of the self-selection problem mentioned above.

[insert Figure 4 here]

In the next two sections, we will show that sharers and non-sharers differ in terms of observable characteristics, and investigate whether these characteristics can explain the observed differences in income. We will then turn to the income measurement issue.

4. Who Does Not Share their Income-Tax Information?

Aside from the heaping issue, 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 1 presents descriptive

statistics. The first two columns of Table 1 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.

[insert Table 1 here]

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 are associated with the consent decision. More specifically, we regress the consent binary variable on individual characteristics. Table 2 presents the results from estimating a linear probability model using the full adult population (all individuals aged 15 and up).^{xiii}

[insert Table 2 here]

For the most part, the regression results are in line with the descriptive statistics. Visible minorities and Aboriginals are 3.7 and 6.1 percentage points, respectively, less likely to share their income-tax information than whites, while immigrants are 4.1 percentage points less likely to share than non-immigrants (see column (3)).^{xiv} There is an education ‘effect’ that is mainly driven by less educated individuals; dropouts are less likely to share than high-school graduates. This finding is in line with previous literature linking education to trust and citizenship (e.g. Dee

(2004)).^{xv} Although not presented here, the quartic age coefficient estimates suggest that sharing increases with age until around age 35 and then plateaus.

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 regress the economic family income measures on a dummy variable equal to one 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 traditional model describing the link between individual income and individual characteristics. Specifically, the revealed income of individual i from family f can be linked to his/her personal characteristics in the following way:

$$y_{if} = \alpha + \beta \text{Sharer}_{if} + X_{if}\gamma + \varepsilon_{if}, \quad (1)$$

where Sharer_{if} is a dummy variable equal to one if the individual consents to Statistics Canada using his/her income tax file as information source. X_{if} is a vector of personal characteristics (e.g., gender, visible minority status, immigrant status, education, age, rural-residence status, and provincial fixed effects). In particular, the relationship between age and income is captured by a quartic function.^{xvi} 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).^{xvii}

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:

$$\sum_{i=1}^{A_f} y_{if} = \alpha A_f + \beta A_f \times \text{Sharer}_f + \sum_{i=1}^{A_f} X_{if} \gamma + \sum_{i=1}^{A_f} \varepsilon_{if},$$

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 ($A_f * \text{Sharer}_f$). 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:

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

or

$$\tilde{y}_{if} = \tilde{\alpha}_f + \beta \tilde{\text{Sharer}}_f + \tilde{X}_f \gamma + \tilde{\varepsilon}_f. \quad (2)$$

The main difference between equations (1) and (2) is that the latter does not restrict $X_{if} \gamma$ and ε_{if} to only affect y_{if} . One could imagine that, in families where there is 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).^{xviii} 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 $\tilde{\alpha}_f$).

It is important to recognize that when interpreting β , 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. Given that we are also interested in observed distributional differences over and above the mean, we use the Recentered Influence Function (henceforth referred to as RIF) methodology proposed by Firpo, Fortin and Lemieux (2009). The RIF regression approach identifies impacts at different points of the unconditional income distribution, which is in direct contrast to (conditional) quantile regression which focuses on the conditional distribution. We favor the former approach since the unconditional income distribution is the object of interest in this study.^{xix}

Table 3 presents OLS results from estimating equation (1), where individual total income is the dependent variable. Looking across the columns, we observe that our main results are robust to the inclusion of different sets of control variables; there remain important differences in total income between sharers and non-sharers. 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 \$3,840 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 age quartic terms suggest an age-income profile very close to a concave function (peaking around 45 years of age according to specification (2)).

[insert Table 3 here]

Table 4 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 3, 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 adjusted *Sharer* coefficient estimate varies between \$3,624 and \$5,443, which is in line with individual total income results. In the end, aggregation at the family level does not affect our findings. It should be noted that the family-level regression equation (equation (2)), does not include a constant. As such, the R-squared values presented in Table 4 are not comparable to those of Table 3.

[insert Table 4 here]

Figure 5 compares the family-level OLS coefficient estimate for *Sharer* (based on specification (4) of Table 4) to the RIF estimated coefficients for the 1st to the 99th quantiles of the family-adjusted total income distribution.^{xx} Just as suggested by the unconditional income distribution, the RIF regression results suggest that the differences are concentrated at the bottom of the distribution. More precisely, raising the portion of the population that are sharers would be associated with higher incomes at lower percentiles of the income distribution. Given that the OLS and RIF 95 percent confidence bands overlap only in a few places (i.e., between the 18th and 38th quantiles and above the 95th quantile), one can conclude that the OLS estimate does not capture all of the heterogeneity.

[insert Figure 5 here]

The parameter estimates presented in Tables 3 and 4 were based on the combined samples of sharers and non-sharers. It is quite possible that self-reporting one's total income affects the link between observed characteristics and income, which could in turn affect the estimated income difference between sharers and non-sharers. We now present the results from an

alternative exercise that attempts to circumvent this potential problem. The exercise consists in using the sample of sharers (only) to predict the income of each non-sharer, based on their observable characteristics. By looking at the difference between the predicted and self-reported incomes, we can get an estimate of the 'reporting error' (if we assume that the income-tax information is error free) or of the 'unexplained difference'.^{xxi}

Figure 6 presents the results from such an exercise for the family-adjusted income. Results from individual total income are very similar, and are available upon request. The top panel of Figure 6 presents the actual and predicted family-adjusted income distributions for both sharers and non-sharers, while the bottom panel presents the corresponding estimated residual (observed minus predicted income) distributions. Two points come out of Figure 6: 1) the top panel shows that the distribution of predicted income of non-sharers lies to the left of the predicted one for sharers, suggesting that part of the observed family-adjusted income difference (between sharers and non-sharers) is due to observed characteristics, and 2) the bottom panel suggests that these differences in observed characteristics can only explain a small portion of the observed difference in family-adjusted income. In fact, if we compare the average predicted and observed differences in family-adjusted total income (\$1,958 versus \$7,006), we realize that family characteristics can only explain 28 percent of the observed income difference between sharers and non-sharers. If we do the same exercise with individual total income, individual characteristics only explain 25 percent of the observed income difference. The average estimated residuals (based on the *Sharer* coefficient estimates) for the non-sharers are very much in line with the results from Tables 3 and 4. In particular, their average estimated residual for individual total income is 3,934 while the *Sharer* coefficient estimate in Table 3 is 3,977 (in column (4)). The family-adjusted average residual difference^{xxii} is also very close to the adjusted *Sharer* coefficient

estimate in column (4) of Table 4 (3,778 and 3,759, respectively). The similarity of these estimates could be explained by 1) non-sharers represent a small fraction of the total population and/or 2) the regression coefficient estimates may not differ significantly across groups. Overall, both the regression results presented in Tables 3 and 4, and the above exercise suggest that a large fraction (about three quarters) of the observed income differences between sharers and non-sharers cannot be accounted for by personal characteristics.

[insert Figure 6 here]

6. Consent Decision and Standard Wage Equations

In the previous sections, we show that the consent decision may affect inequality measures, particularly those that are sensitive to changes in the bottom of the income distribution, and that the income differences between sharer and non-sharer 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. Boudarbat, Lemieux and Riddell (2010); Card and Lemieux (2001); Katz and Murphy (1992)) 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 5 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.

[insert Table 5 here]

Overall, Table 5 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.^{xxiii} 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).^{xxiv}

7. Discussion and Conclusion

The fact that respondents could now let Statistics Canada access their income tax files, instead of self-reporting their income, would seem to affect 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 income, or expands the analysis to other part of the distribution.

On the other hand, allowing individuals to share their income-tax information does not have a significant impact on the estimation of standard wage equation parameters. Whether we look at OLS or RIF estimates, the magnitude of the estimate differences between sharers and non-sharers is never large enough to have a noticeable impact on the overall population parameter estimates.

What should we make of these findings? Our wage equation findings would indicate that giving respondents the option to share their tax information is not a confounding factor when it comes to the estimation of standard wage equations. As such, our findings support using the 2006 Census in combination with earlier censuses when looking at the evolution of the returns to human capital in Canada, as is done Boudarbat, Lemieux and Riddell (2010).

As for our inequality findings, the answer is less clear. While the regression results suggest that sharers and non-sharers with similar characteristics report total incomes that are significantly different, it is clear that the coefficient estimate for *Sharer* cannot be seen as the causal impact of forcing individuals to share their income-tax data on reported income. The positive partial correlations between *Sharer* and observed income found in Tables 3 and 4, and most importantly in the RIF estimated coefficients for the lower quantiles (Figure 5), are too large to simply reflect heaping in the non-sharers' distribution. Similarly, the increased reporting of small amounts due to the new response mode (Statistics Canada 2008) cannot account for the

large RIF estimated coefficients for the lower quantiles. The evidence therefore leads us to believe that there is simply something (unobservable) about these people that makes them non-sharers, and at the same time leaves them in the lower part of the income distribution.

The presence of self-selection in the consent decision would be of no consequence if there was no systematic difference across data sources (administrative versus self-reported);^{xxv} one would simply focus on the overall income distribution - the one that includes both sharers and non-sharers. There is, however, much evidence in the literature that administrative and tax data differ, particularly in the tails (e.g. Frenette, Green and Picot (2006); Gottschalk and Huynh (2010); Meyer, Mok and Sullivan (2009)). Frenette, Green and Picot (2006), for example, find that the bottom decile of the Census income distribution is 8.2 percent and 39.2 percent higher than in the tax data in 1995 and 2000, respectively. As such, the presence of self-selection of non-sharers into the bottom of the distribution will be a confounding factor when looking at income inequality.^{xxvi}

Would forcing individuals to share their income-tax data (using, say, an implied consent approach) have resulted in observing significantly different income distributions? There are a few reasons to think that the income distribution could have looked similar to the one we observe (i.e. where respondents are given the choice to share their income tax data). It is quite possible that a significant portion of non-sharers simply do not file their income tax. It could also be the case that individuals who refuse to share do so because they file their income tax late, something that is particularly problematic for the Census. The Census must be returned by post or online by May 16th, (i.e. census day). Yet, the tax filing deadline for most Canadians is April 30th, the exception being the self-employed whose deadline is June 15th. As such, some late tax filers might not offer to share their non-existent tax records. Although we cannot identify late-filers in the Census, there

exists a tax literature that looks at the non-compliance issue. In particular, Boame (2009) looks at the characteristics of tax filing compliers (i.e. individuals timely filing their required return) using Canadian Revenue Agency data. His findings seem somewhat in line with our non-sharers. For example, males, younger individuals, and those living in Ontario and British Columbia (as opposed to Quebec) are all more likely to be non-compliers and non-sharers.^{xxvii}

So even if these non- and late-filing individuals were to be forced to share their income-tax information with Statistics Canada, the Census and tax-income information might not be matched. Forcing these individuals to share their income-tax information would probably result in Statistics Canada imputing their income information. 89.1 percent of individuals who agreed to share their income-tax information had their Census information matched to a Canada Revenue Agency tax record (Statistics Canada (2008), and this number could very well be significantly lower if we forced non-sharers to share their income-tax information.

References

- Abowd, J. M., and M. H. Stinson. Forthcoming. "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data," *Review of Economics and Statistics*.
- 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.
- Barreca, A. I., J. M. Lindo, and G. R. Waddell. 2011. "Heaping-Induced Bias in Regression-Discontinuity Designs," Working Paper No. 17408, NBER.
- Blank, R. M., K. K. Charles, and J. M. Sallee. 2009. "A Cautionary Tale about the Use of Administrative Data: Evidence from Age of Marriage Laws," *American Economic Journal: Applied Economics* 1(2): 128-49.
- Boame, A. K. 2009. "A Panel Analysis of Behavior Change in Individual Income Tax Compliance," in *Recent Research on Tax Administration and Compliance*. IRS Research Bulletin.
- 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.

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.

Frenette, M., D. A. Green, and K. Milligan. 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.

Kapteyn, A., and J. Y. Ypma. 2007. "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data," *Journal of Labor Economics* 25(3): 513-551.

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.

Meyer, B. D., W. K. C. Mok, and J. X. Sullivan. 2009. "The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences," Working Paper No. 15181, NBER.

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.

Morin, L.-P. 2011. "Cohort Size and Youth Earnings: Evidence from a Quasi-Experiment," Working Paper No. 85, CLSRN.

Murphy, K. M., and F. Welch. 1990. "Empirical Age-Earnings Profiles," *Journal of Labor Economics* 8(2): 202-229.

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.

Rehavi, M. M. 2010. "Partial Reporting: An Example from Charitable Giving," Manuscript.

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.

Statistics Canada. 2008. "Census Questions on Income: New Features and Important Changes," <http://www12.statcan.gc.ca/census-recensement/2006/ref/info/income-revenu-eng.cfm>, (accessed June 21, 2013).

Statistics Canada. 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(3): 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.

ⁱ Prior to 2006, income data was self-reported.

ⁱⁱ 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.

ⁱⁱⁱ It should be noted that measurement error is not a problem unique to survey data. While earlier studies comparing administrative to survey data more or less treated administrative data as 'error free', more recent studies allow for and compare measurement error in both sources of data. See Abowd and Stinson (forthcoming) for details.

^{iv} There have been technical papers both in Canada (e.g. Abraham et al. (2001); Michaud et al. (1995)), 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 public policy.

^v Net under-coverage was 2.67 percent in the 2006 Census (Statistics Canada 2010).

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

^{vii} Fortin et al. (2012), which provides a current overview of the Canadian inequality literature, suggest that income inequality has increased in Canada since the early 1990s.

^{viii} 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.

^{ix} This approach was also favoured by Frenette, Green and Picot (2006), Frenette, Green and Milligan (2007), Frenette, Green and Milligan (2009), and Heisz (2007).

^x Although it is not the focus of this paper, switching to a response method that reduces heaping could matter for researchers interested in studying discontinuities in social programs that are based on individual income. In particular, it could reduce the amount of heaping-induced bias (e.g. Barreca, Lindo and Waddell (2011)). Note that we are not claiming that tax data are immune to measurement error. There are many instances where we could expect to see more measurement error in administrative/tax data. For instance, there is evidence that administrative/tax data can be less accurate than survey data for charitable donations (Rehavi 2010), and for outlawed behaviours (Blank, Charles and Sallee 2009).

^{xi} The mean incomes for the top and bottom deciles are the averages of the bottom two and top two income vingtiles, respectively.

^{xii} Using Labour Force Survey data, Fortin et al. (2012) find that wages increased more at the top of the distribution than at the bottom over the 2000 to 2005 period.

^{xiii} 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 focusing on individuals where all economic family members agreed on the consent decision. These findings are available upon request.

^{xiv} Using German data, Sakshaug and Kreuter (2011) found that foreign citizens were less prone to provide consent than their German counterpart.

^{xv} 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. Abraham et al. (2001), however, found that more educated individuals were less likely to grant permission to access their tax data.

^{xvi} Murphy and Welch (1990) suggest that the age-earnings profile is better captured by a quartic rather than a quadratic function.

^{xvii} However, in order to minimize the potential impact of choosing this particular functional form on the estimated coefficient of interest, we also ran regressions relying on binary righthand-side variables. In particular, we transformed our continuous age variable into a series of dummy variables covering 5-year intervals. These results are very similar to the ones presented here.

^{xviii} 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.

^{xix} In addition to the original source (Firpo, Fortin and Lemieux 2009), Fortin, Lemieux and Firpo (2011) provides an excellent introduction to the RIF methodology.

^{xx} Since the RIF estimation requires the inclusion of a constant term in the estimated equation, we compare the OLS and RIF results based on Specification (4) of Table 4, 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 \$3,759 when not.

^{xxi} We thank an anonymous referee for this suggestion. One crucial assumption in doing this exercise is that if we could force non-sharers to reveal their true tax-information income, the expectation of the OLS parameter estimator would be the same whether we regress the true tax-information income using the sharer or the non-sharer samples. This means that the relation between tax-information income and both the observables and unobservables characteristics should be the same for both groups. Finally, such an approach also assumes that non-sharers do not report differently when given the option to share than when not given the option.

^{xxii} Since the family-adjusted total income regression does not include a constant term, the residuals average for the sharers is not equal to 0 (although it is close to). We therefore have to look at the difference in average residuals.

^{xxiii} Having said this, the non-sharers' estimates are, for the most part, slightly smaller than those of sharers which could be taken as evidence of some negative selection into the non-sharer's group.

^{xxiv} 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. We also verified whether the observed differences in estimates vary across quantiles. The differences in estimates are relatively stable across quantiles with the exception of immigrant status, where the difference seems larger at the bottom of the wage distribution. These results are available upon request.

^{xxv} Alternatively, if there was no self-selection in the consent decision, the only confounding factor would be measurement error differences across data sources; recall that income was 100 percent self-reported in 2001, but the number dropped to 15 percent in 2006.

^{xxvi} To be able to more thoroughly understand the impact of the consent decision and of measurement error would require that we be able to link, ex-post, the Census data with the tax (TIFF) data, something that is presently not possible.

^{xxvii} While many Canadians with zero income have significant incentives to file tax forms (e.g. those with children and the elderly), many unattached individuals have little incentive. With small incentives, they might still be likely to file, but might be more likely to file late - well after the Census forms are sent out in May. Boame (2009) does find that singles - a group of individuals more likely to be unattached - had a lower filing compliance rate than married and common-law individuals. Given that unattached individuals are also much more likely to be in the non-sharers' group, we investigated whether they were driving our findings by excluding them when estimating the family-adjusted regressions. The coefficient estimates are very close whether unattached individuals are included or not in our regressions, and this is especially true for the *Sharer* coefficient estimate. We have also looked at the vintiles where unattached individuals were excluded and we still see a large difference at the bottom of the distribution, and this holds true for all four income measures. Results are available upon request.

Table 1: Summary Statistics, Means (15 years and up)

	Sharers	Non-Sharers	Overall Population
<i>A. Gender</i>			
Female	.516	.510	.515
<i>B. Age</i>			
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
<i>C. Educational Attainment</i>			
Less than High School	.225	.254	.237
High School	.256	.257	.256
Post Secondary	.332	.318	.326
Bachelor's and up	.187	.171	.182
<i>D. Immigrant Status</i>			
Yes	.226	.291	.237
<i>E. Visible Minority Status</i>			
Caucasian	.830	.761	.814
Aboriginal	.027	.039	.030
Non-Caucasian, Non-Aboriginal	.143	.200	.156
<i>F. family Size</i>			
1	.175	.226	.170
2	.318	.300	.304
3	.184	.165	.183
4	.196	.174	.201
5+	.128	.134	.142
<i>G. Province/Territories</i>			
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.

Source: Canadian Census 2006 (long-form).

Table 2: Consent Decision, Linear Probability Model Results

	(1)	(2)	(3)
Female	0.004*** (0.0004)	0.004*** (0.0007)	0.003** (0.0007)
Visible Minority	-0.040*** (0.0008)	-0.040*** (0.0008)	-0.037*** (0.0008)
Aboriginal	-0.073*** (0.0011)	-0.065*** (0.0011)	-0.061*** (0.0011)
Immigrant	-0.042*** (0.0012)	-0.044*** (0.0012)	-0.041*** (0.0012)
Years Since Migration	5.91e-04*** (7.6e-05)	6.95e-04*** (7.7e-05)	8.01e-04*** (7.7e-05)
(Years Since Migration) ²	-4.59e-06*** (1.2e-06)	-6.21e-06*** (1.2e-06)	-6.62e-06*** (1.2e-06)
No Diploma		-0.031*** (0.0008)	-0.033*** (0.0008)
Bachelor		0.013*** (0.0010)	0.012*** (0.0010)
Master's		0.018*** (0.0014)	0.017*** (0.0014)
<i>PhD</i>		0.019*** (0.0026)	0.017*** (0.0026)
Rural			-0.008*** (0.0005)
Constant	-0.062*** (0.0061)	0.057*** (0.0063)	0.059*** (0.0063)
Quartic Age Controls	Yes	Yes	Yes
Female × Education Dummies	No	Yes	Yes
Province Fixed Effects	No	No	Yes
<i>R</i> ²	0.02	0.02	0.02
N	5,119,850	5,119,850	5,119,850

Notes: The dependent variable equals one if the individual gave consent to Statistics Canada for them to access their tax files, and zero otherwise. The omitted educational-attainment category is High-School Graduation Certificate. The educational attainment dummies used in the regressions but not presented in the table are: Some Post-Secondary, Certificate Above Bachelor Level, and Degree in Medicine, Dentistry, Veterinary Medicine or Optometry. 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%.
Source: Canadian Census 2006 (long-form).

Table 3: Total Income (Individual Level), OLS Results

	(1)	(2)	(3)	(4)
Sharer	5,277*** (49.1)	4,424*** (46.0)	3,840*** (44.1)	3,977*** (43.9)
Female		-13,948*** (34.3)	-12,271*** (54.3)	-12,256*** (54.0)
Visible Minority		-3,451*** (66.2)	-3,918*** (63.2)	-4,826*** (63.6)
Aboriginal		-10,134*** (66.5)	-6,313*** (63.4)	-6,512*** (65.5)
Immigrant		-13,848*** (93.1)	-18,822*** (93.3)	-19,686*** (93.4)
Years Since Migration		714.7*** (6.17)	908.9*** (6.13)	871.5*** (6.13)
(Years Since Migration) ²		-7.011*** (0.099)	-9.224*** (0.097)	-8.932*** (0.097)
No Diploma			-7,726*** (58.7)	-7,052*** (58.3)
Bachelor			23,241*** (132)	23,035*** (131)
Master's			31,549*** (241)	31,395*** (239)
PhD			43,036*** (473)	43,080*** (472)
Rural				-2,818*** (40.1)
Constant	28,553*** (44.8)	-70,897*** (368)	-20,218*** (395)	-17,175*** (395)
Quartic Age Controls	No	Yes	Yes	Yes
Female × Education Dummies	No	No	Yes	Yes
Province Fixed Effects	No	No	No	Yes
R ²	0.00	0.16	0.24	0.25
N	4,740,330	4,740,330	4,740,330	4,740,330

Notes: Individual total income is the dependent variable. The omitted educational-attainment category is High-School Graduation Certificate. The educational attainment dummies used in the regressions but not presented in the table are: Some Post-Secondary, Certificate Above Bachelor Level, and Degree in Medicine, Dentistry, Veterinary Medicine or Optometry. 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%.

Source: Canadian Census 2006 (long-form).

Table 4: Total Income (Economic Family Level), OLS Results

	(1)	(2)	(3)	(4)
Sharer (adj)	5,443*** (68.2)	4,203*** (64.9)	3,624*** (62.3)	3,759*** (61.9)
Female (adj)		-3,186*** (89.8)	-3,739*** (138)	-3,809*** (136)
Visible Minority (adj)		-4,643*** (95.1)	-4,764*** (91.4)	-5,429*** (91.8)
Aboriginal (adj)		-13,197*** (88.3)	-7,999*** (81.3)	-8,264*** (86.5)
Immigrant (adj)		-14,815*** (125)	-20,111*** (130)	-21,225*** (131)
Years Since Migration (adj)		742.5*** (8.69)	956.7*** (8.87)	895.6*** (8.85)
(Years Since Migration) ² (adj)		-6.840*** (0.15)	-9.631*** (0.15)	-9.217*** (0.15)
No Diploma (adj)			-8,463*** (112)	-7,272*** (111)
Bachelor (adj)			23,983*** (203)	23,756*** (201)
Master's (adj)			30,665*** (355)	30,542*** (352)
PhD (adj)			39,011*** (644)	39,436*** (641)
Rural (adj)				-2,616*** (56.0)
Constant (adj)	27,493*** (62.6)	-56,753*** (892)	2,140** (885)	4,284*** (877)
Quartic Age Controls (adj)	No	Yes	Yes	Yes
Female × Education Dummies (adj)	No	No	Yes	Yes
Province Fixed Effects (adj)	No	No	No	Yes
R ²	0.59	0.63	0.67	0.68
N	2,490,890	2,490,890	2,490,890	2,490,890

Notes: Adjusted family income is the dependent variable. The omitted educational-attainment category is High-School Graduation Certificate. The educational attainment dummies used in the regressions but not presented in the table are: Some Post-Secondary, Certificate Above Bachelor Level, and Degree in Medicine, Dentistry, Veterinary Medicine or Optometry. 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%.

Source: Canadian Census 2006 (long-form).

Table 5: Log Weekly Wages (Full-Time Workers), OLS Results

	Sharers	Non-Sharers	Difference	All Full-Time Workers
Age	.081*** (.0002)	.080*** (.0006)	.001** (.0007)	.081*** (.0002)
Age ²	-.001*** (2.99e-06)	-.001*** (7.41e-06)	-6.10e-06 (7.95e-06)	-.001*** (2.77e-06)
Female	-.305*** (.0008)	-.263*** (.0021)	-.042*** (.0022)	-.299*** (.0008)
Visible Minority	-.187*** (.0016)	-.120*** (.0034)	-.067*** (.0038)	-.175*** (.0014)
Aboriginal	-.133*** (.0026)	-.145*** (.0057)	.011* (.0063)	-.136*** (.0024)
Immigrant	-.099*** (.0014)	-.075*** (.0031)	-.025*** (.0034)	-.096*** (.0013)
No Diploma	-.140*** (.0015)	-.117*** (.0036)	-.023*** (.0038)	-.137*** (.0013)
Some Post-Secondary	.146*** (.0011)	.145*** (.0027)	.001 (.0029)	.146*** (.0010)
Bachelor and Up	.459*** (.0012)	.437*** (.0030)	.022*** (.0032)	.456*** (.0011)
Constant	4.80*** (.0047)	4.79*** (.0117)	.010 (.0130)	4.80*** (.0044)
Province Fixed Effects	Yes	Yes		Yes
R ²	.27	.23		.26
N	2,009,140	329,160		2,338,305

Notes: The log weekly wages of full-time workers is the dependent variable. 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%.

Source: Canadian Census 2006 (long-form).

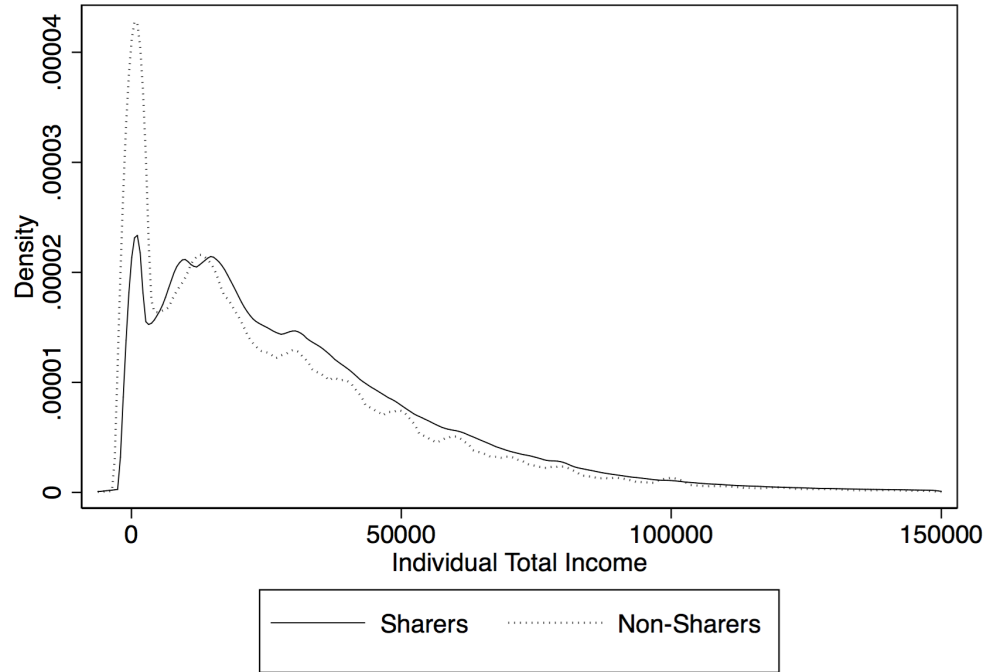


Figure 1: Individual Total Income Densities of Sharers and Non-Sharers

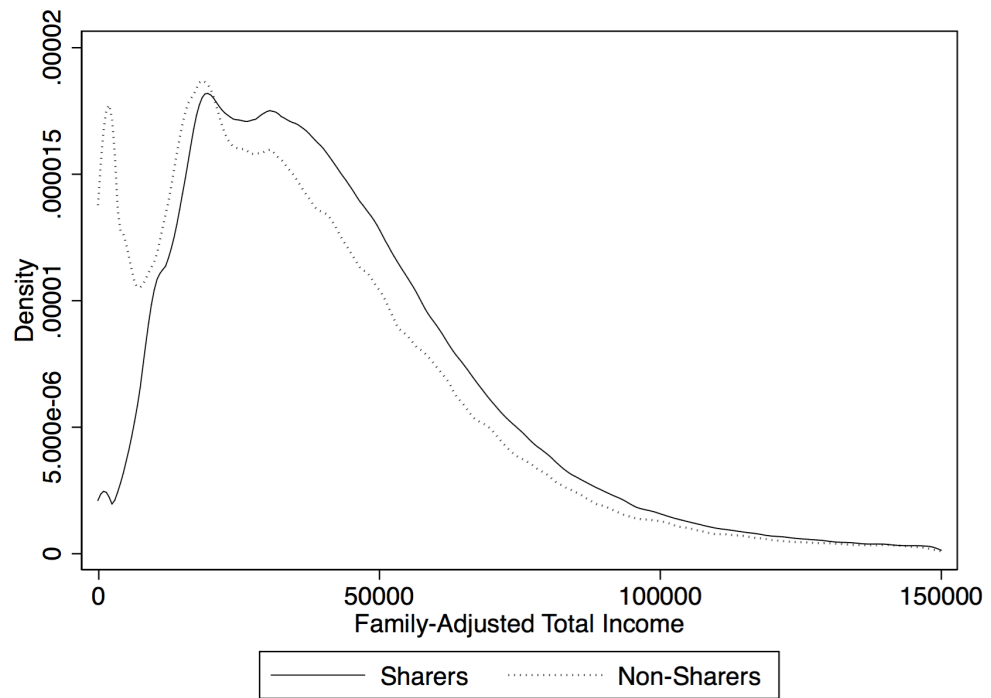


Figure 2: Family-Adjusted Total Income Densities of Sharers and Non-Sharers



Figure 3: CDFs of Family-Adjusted Incomes of Sharers and Non-Sharers

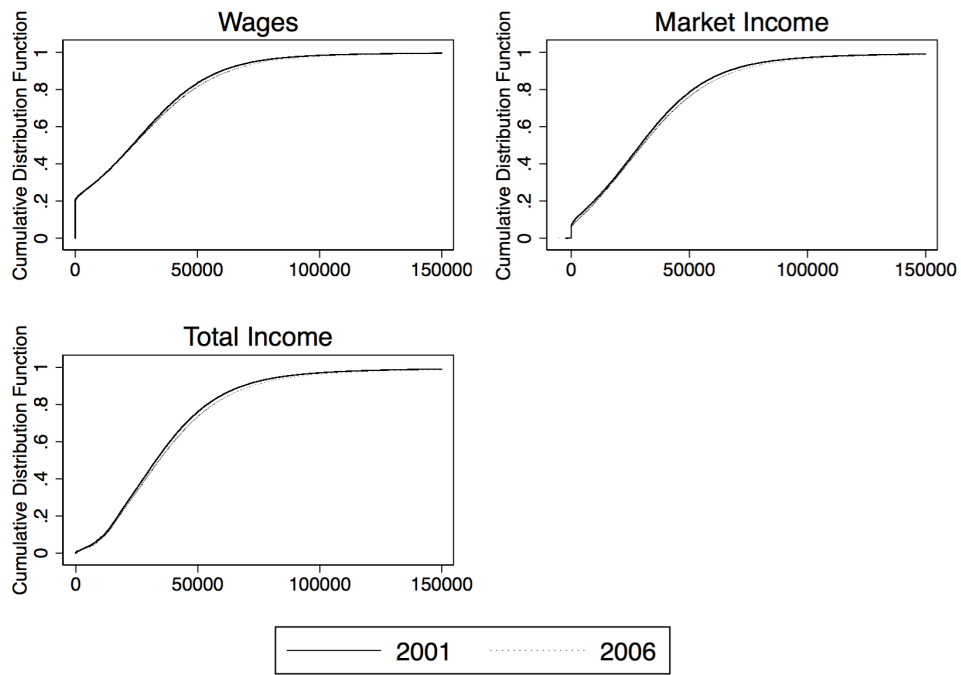


Figure 4: Evolution of Family-Adjusted Incomes of Sharers and Non-Sharers (\$2000)

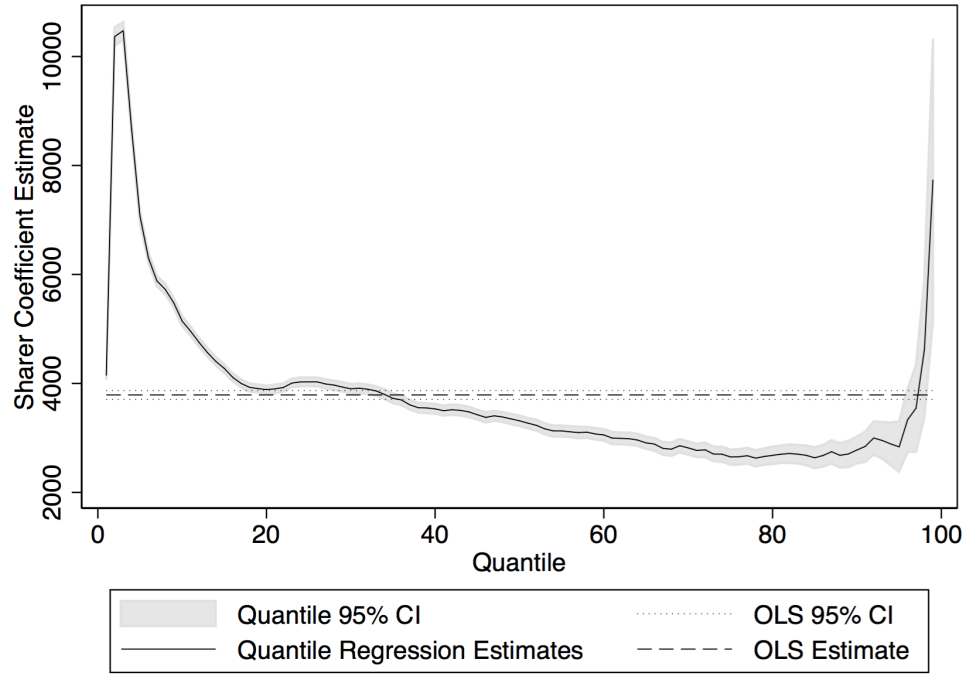


Figure 5: Unconditional Quantile Regression Estimates (Family-Adjusted Total Income)

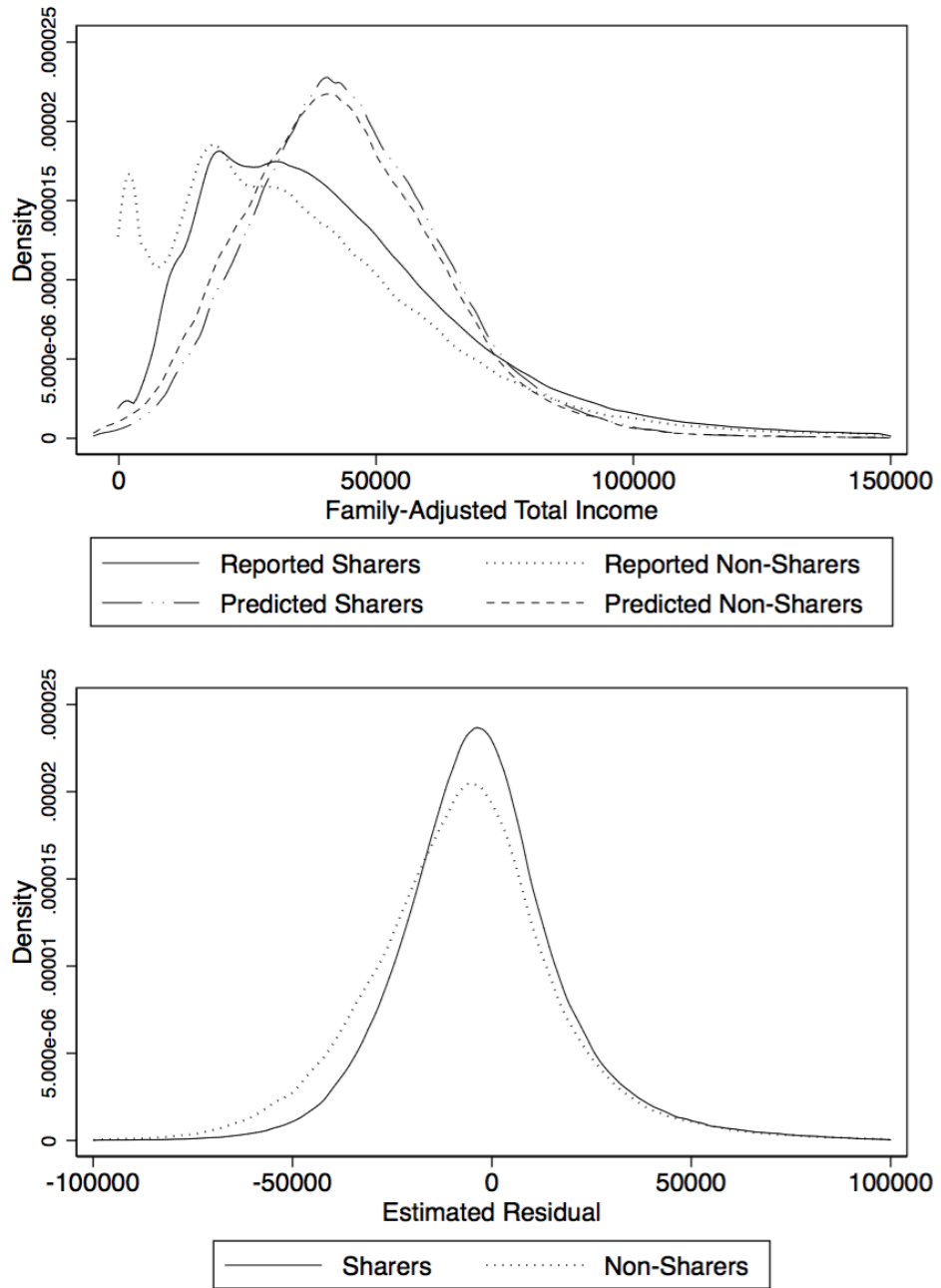


Figure 6: Reported vs. Predicted Family-Adjusted Total Income Distributions and Associated Residual Distribution