The tale of the tails: Revisiting recent trends in Canadian after-tax income inequality using Census data

Marc Frenette, David A. Green, and Kevin Milligan[1]

Abstract

We present new evidence on levels and trends in after-tax income inequality in Canada between 1980 and 2000. We argue that existing data sources may miss changes in the tails of the income distribution, and that much of the changes in the income distribution have been in the tails. Our data are constructed from Census files, which are augmented with predicted taxes based on information available from administrative tax data. After validating our approach in predicting taxes on the Census files, we document differences in the levels and trends in after-tax inequality between the newly constructed data source and the more commonly used survey data. We find that after-tax inequality levels are substantially higher based on the new data, primarily because income levels are lower at the bottom than in survey data. The new data shows larger long-term increases in after-tax income inequality and far more variability over the economic cycle. This raises interesting questions about the role of the tax and transfer system in mitigating both trends and fluctuations in market income inequality.

Introduction

Equity goals represent an important target for government policy. From the direct effects of the tax and transfer system to the perhaps less direct effects of other policies such as trade, few policy decisions lack an equity aspect. Assuming that we care about the equity impacts of government policy, a complete assessment of their impacts requires accurate measures of inequality. Examining inequality movements is also useful for evaluating alternate theories of how economies function. For example, Beaudry and Green (2003) argue that movements in the wage structures in the US and Germany over the past 30 years fit with a model of technological change in which the rate of adoption of new technologies is endogenously determined according to movements in relative factor supplies. For both reasons, we need accurate measures of the degree of inequality in an economy and its movements over time.

In a recent article, Frenette et al. (2004) argue that the data source most widely used to characterize inequality in Canada – the combination of the Survey of Consumer

[1] Respectively,
Statistics Canada
Dept. of Economics, University of British Columbia and Research Associate, Institute for Fiscal Studies
Dept. of Economics, University of British Columbia and NBER

Finances (SCF) and the Survey of Labour and Income Dynamics (SLID)[2] – does not provide an accurate picture of either the level or trends in income inequality. In particular, comparisons with Census and tax data indicate that the SCF/SLID under-represents both very low and very high incomes. This implies both an under-estimation of the level of inequality and, potentially, misrepresentation of trends that are driven by movements in the tails of the distribution. Indeed, given the evidence in Saez and Veall (2005) showing that there have been important movements in inequality concentrated in the very top of the income distribution in Canada in the last few decades, a misrepresentation of trends seems likely.

Our objective in this paper is to provide more reliable measures of inequality for Canada for the period from 1980 through 2000. Our first step in this endeavour is to build a case for the claim that Census micro data provides the most reliable and complete data for analyzing income inequality in Canada. We argue that Census data is superior to SCF/SLID primarily because it has much better coverage. The SCF/SLID has an approximately 20% under-response rate relative to the Census and, following Frenette et al. (2004), we show that this leads to misrepresentation of the two tails of the pre-tax income distribution. In addition, the much larger sample size of the Census permits more reliable measures of percentiles in both tails of the distribution. Census data is also superior to using tax data because of its longer time span (lack of incentives to file mean that tax data has incomplete coverage before 1993) and its much greater list of covariates.

These advantages for the Census are partially countered by two key shortcomings. The first is that the Census is only available quinquennially, and therefore unable to capture higher frequency movements in inequality. We have no remedy for this problem, but note that quinquennial data is sufficient for studying longer term trends. In particular, the 1980, 1990 and 2000 Censuses were all taken roughly at the top of business cycles, allowing consistent comparisons across time.

The second shortcoming of the Census data is the lack of information on taxes. The income concept most closely related to family well-being is after-tax and transfer (disposable) income. The Canadian Census asks questions about transfers received but

[2] For example, they are used in Johnson (1995); Osberg (1997 and 2003); Rashid (1998); Wolfson and Murphy (1998).

not taxes and, as a result, researchers cannot construct disposable income using what is available in the Census dataset. Thus, the second part of our exercise in this paper, and perhaps our main contribution, is to impute taxes for families in each Census in the span from 1980 to 2000. Adding these imputed taxes to the Census micro data, we create what we call the Census after-tax (Census-AT) dataset.

Our imputation procedure matches administrative tax data with the Census, using observable characteristics common to both data sources. We employ a reduced form approach in which we first regress taxes paid on observable family characteristics using the tax data then use family characteristics recorded in the Censuses in combination with the estimated regression coefficients to form predicted taxes paid for each Census family. This approach is superior to one in which we try to use actual tax schedules to impute taxes paid for each Census family since it does not require us to calculate allowable deductions and it reflects actual patterns of take-up of those deductions. We perform a validation exercise that shows that our approach does a very good job of predicting the distribution of actual taxes paid.

Having selected a preferred dataset and adjusted it to allow examination of disposable income, in the final section of the paper we use the Census-AT data to reassess what we know about Canadian income inequality levels and trends. Throughout the paper, we compare the results from our enhanced Census data with those from the SCF/SLID, the source of received wisdom on income inequality, and show that there are substantial differences between the two. In particular, the Census-AT data reveals both fatter left and right tails of the income distribution. It also shows a different pattern over time (especially over the business cycle) and a difference in the differential between pre and post-tax and transfer inequality. These differences force a reconsideration of the level of income inequality (revising it upward), of the magnitude of its relationship with the economic cycle (revising it upward), and of the role of taxes and transfers in mitigating movements in inequality (revising it downward) relative to what has been documented in the past by several authors (e.g. Beach and Slotsve, 2004; Johnson, 1995; Osberg, 1997 and 2003; Rashid, 1998; Wolfson and Murphy, 1998). For example, we show that in 2000, the log of the ratio of the 95th to the 5th percentile of the disposable income distribution is .82 according to SCF/SLID. Using similar techniques and definitions, the same ratio is .95 according to Census-AT data (16% higher than in SCF/SLID). Furthermore, the ratio rises by 6.1% between 1980 and 2000 according to the Census-AT but only 2.4% according to SLID/SCF.

These results have potentially important ramifications for our notions of equity in the Canadian economy. They also point to the need for further research on the impact of Canada’s tax and transfer system on inequality. The Census-AT data tells a story in which market income inequality grew at a relatively constant rate in the 1980s and 1990s. In contrast, disposable income inequality was virtually unchanged in 1990 compared to 1980 but grew sharply between 1990 and 2000. The result is a substantial long-term increase in disposable income inequality. While definitive statements on changes in the impact of taxes and transfers are not possible from these comparisons since pre-tax and transfer income will partly reflect behavioural responses to the tax and transfer system, the differences in the movements of market and disposable income inequality in the 1990s compared to the 1980s is a smoking gun pointing to a potential weakening of the effectiveness of redistributive policies. One of our goals in this paper is to point out that smoking gun, opening an avenue for future research.

The paper most closely related to ours is Frenette et al. (2004). That paper provides a detailed comparison of the three main datasets available for studying Canadian income inequality: SCF/SLID, Census, and tax data. The emphasis in that paper is on pointing out that there are serious differences among the datasets, raising reason for concern. Frenette et al. (2004) do not, however, try to generate a preferred picture of inequality levels and trends. In this paper, we proceed to the next step, selecting, defending and enhancing a preferred dataset (Census-AT) and then using that data to establish basic facts about Canadian income inequality.

The paper proceeds in five sections. In the following section, we provide a discussion of the relative merits of the available data sources and some evidence on the differences in trends and levels of pre-tax income inequality as measured by the Censuses and the SCF/SLID data. Next, we describe our measurement choices in terms of income and inequality measures. We then provide an outline of our methodology for predicting taxes on the Census and validate the approach with actual tax data. The first set of income inequality estimates drawn from the new data source are presented in the next section, which is followed by a comparison of those results with estimates drawn from SCF-SLID. The study concludes with a summary in the last section.

Comparing and contrasting available data sources

Researchers interested in studying levels and trends in Canadian income inequality have three data sources at their disposal. The most commonly used is Statistics Canada's official source of income estimates, namely the Survey of Consumer Finances (SCF) until 1996, and the Survey of Labour of Income Dynamics (SLID) from 1996 onwards. The second is the Census of Population files, which are available quinquennially from 1970. The third available source is the T1 Family Files (T1FF), available from 1982 onwards. Each of these sources has their advantages and disadvantages, as illustrated by Table 1.

Table 1: Attributes of income data sources

	SCF-SLID	Census	T1FF
Continuity	Break in 1996	No break	Change in filing incentives in early 1990s
Coverage	80% response rate	Mandatory response	Almost all file since early 1990s
Sample size	Small	Large	Large
Socio-economic information	Many	Many	Few
Frequency	Yearly	Every 5 years	Yearly
Income variable	After-tax	Total income	After-tax

Census data, which is the source upon which we focus in this paper, has several appealing characteristics. First, it has no breaks over our period of interest. In contrast, the SCF was replaced by SLID in 1996. Although this did not affect average levels of income, it did affect incomes at the top and bottom of the distribution (Frenette et al., 2004), requiring some kind of adjustment at the time of the "seam". Data from T1FF are not suited to study incomes of families at the very bottom of the distribution in the 1980s since there were few (if any) financial incentives to file for people with no taxable income. This is because refundable tax credits were not as prevalent then and some forms of income were simply not reportable on the tax files even if they did file (e.g. social assistance income). This creates a break in the data between the 1980s and the 1990s.[3]

A second appealing feature of the Census is its coverage. Response to the Census is mandatory by law, and as such, coverage of the population is almost complete, with the exception of very specific groups (most notably, on-reserve aboriginals, individuals in collective dwellings, and the homeless). Response to the SCF/SLID is voluntary, and roughly 20% of selected households choose not to do so. This creates the potential for response bias that may be related to income. The SCF/SLID datasets include weights calculated so that key sample characteristics mimic those of the population as a whole, but income is not one of the characteristics. Thus, to the extent that response bias is related to income, even after controlling for observables that are directly addressed by the weights, the weighted income distribution obtained from the SCF/SLID may still not correspond to that for the whole population.[4] The population coverage on T1FF is quite good, but only after 1993 when the combination of incentives from child tax credits and GST rebates improved the filing incentives for very low income individuals.[5]

A third feature of the Census (like T1FF) is its very large sample size (20% of the population), allowing researchers to conduct more detailed analyses of income inequality. In particular, large samples are important for obtaining reliable measures of movements in extreme percentiles of the distribution. In contrast to the Census, SCF/SLID has approximately 30,000 to 35,000 observations, making both detailed decompositions and examinations of extreme tails of the income distribution more problematic.

---

[3] Frenette, Green, and Picot (2004) discuss these issues in more detail.

[4] Coinciding with the release of the 2003 SLID, Statistics Canada has retroactively adjusted survey weights to account for discrepancies with the distribution of *individual* earnings based on the T4 slips. Unfortunately, these adjustments only go back to 1990. Furthermore, calculations by the authors show that they do not fully account for discrepancies in *family* income, especially at the very bottom of the distribution. This is likely due to the fact that individuals with low earnings may be in low, middle, or high income families.

[5] The 1993 changes to federal child benefits combined the family allowance, refundable credit, and non-refundable credit into one package that required tax filing. The GST refundable credit which started paying in 1990 also required tax filing, and substantially expanded benefits from the earlier federal sales tax credit introduced in 1986.

A fourth advantage of the Census is that it contains detailed socio-economic information on its respondents. This is also true in the SCF/SLID, but not in tax data. In particular, education is missing from the tax files.

One drawback of the Census is that it is only available every 5 years, while the other two sources are available annually. Perhaps more important, however, is the omission of taxes paid on the Census (in contrast to SCF/SLID and tax data). This makes the measurement of after-tax income inequality more difficult with the Census. In the next section, we discuss the benefits of examining income net of taxes as a measure of a household budget constraint (i.e. the flow of resources available to the household prior to making consumption decisions).

Given the other advantages of Census data, adding tax information to this source seems a worthwhile exercise. The effort of doing so, however, only seems justified if the various problems with the SCF/SLID actually affect our depiction of the income distribution. Frenette et al. (2004) provide comparisons of pre-tax income data from the three major sources. Figures 1 and 2 are derived from the data in their Table 4, showing average total income (wages, salaries and other private income plus government transfers) for the bottom and top deciles using both the SCF/SLID and Census data.[6] We present the data in constant year 2000 dollars, with an adjustment for adult equivalents using the square root of the number of family members.

[6] Note that the SLID income levels in 2000 were actually derived by adding the growth rates between 1996 and 2000 from the SLID to the 1996 SCF values in order to obtain levels that are comparable to those from the earlier SCF data.

Figure 1: Average total income in the bottom decile

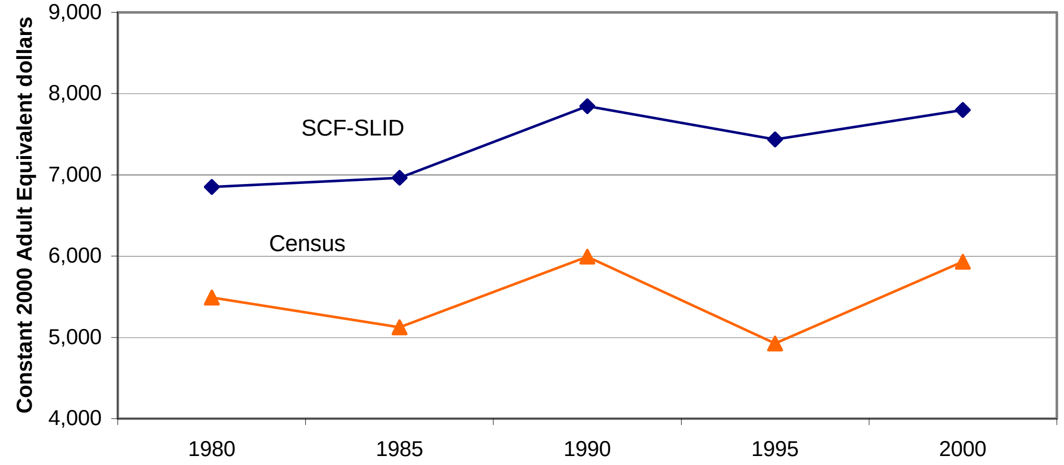
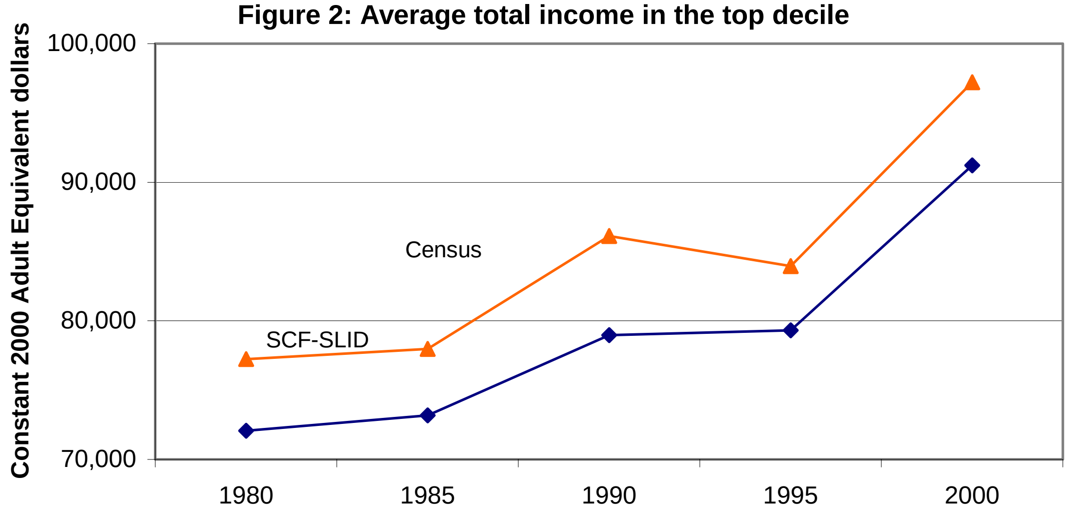


Figure 2: Average total income in the top decile



The Census and SCF/SLID data are dramatically different, with the Census recording lower levels of income in the bottom decile and higher levels in the top decile than the SCF/SLID. Thus, income inequality is higher in the Census in every year. For example, the ratio of the average income in the top decile to the average in the bottom decile was 16.4 in 2000 according to the Census but only 11.7 according to SCF/SLID.

The data also diverge substantially in the trends they register. In the bottom decile, Census data suggest more cyclical fluctuations in income. For example, incomes in the bottom decile of the Census fell by 7% and 18% in the downturns of the early 1980s and early 1990s, respectively. According to the SCF/SLID, incomes in the bottom decile rose by 2% in the early 1980s, and fell by only 5% in the early 1990s. In the recovery period of 1995 to 2000, incomes in the bottom decile rose by 20% in the Census, but only rose by 5% in SCF/SLID. In the top decile, the trends are largely similar, with the exception of the early 1990s. While Census data suggests a 3% decline in family income in the top decile between 1990 and 1995, SCF/SLID points to virtually no change at this time.

Table 1 offers several possible explanations for the differences across datasets but attempting to establish the exact source of the differences is beyond the scope of this study. Instead, we construct an alternative to the SCF/SLID for the purpose of studying after-tax income inequality. Specifically, we augment the Census data with predicted income taxes based on information available in the tax data.

Measurement of income and taxes

In this section, we motivate our choices for the measurement of income and taxes. As we discussed in the introduction, researchers may study inequality both out of direct interest in equity and as a form of evidence on how the economy functions. If our primary interest is the latter, we would focus on income measures that are closely related to factor prices and supplies: what we will call market incomes. Though even in that case, some discussion of taxes and transfers is necessary to make sure that movements in income that are generated from changes in government policy are not mistaken for the effects of, say, technological change. If, on the other hand, equity is our main focus then we would ultimately like a measure of well-being. In a world with homogeneous individuals and with all goods traded in perfectly competitive markets, differences in well-being are completely captured by differences in disposable income. However, in the real world, where there are market imperfections as well as heterogeneity in preferences and in the prices individuals face, income and well-being need not be nearly so directly related (Atkinson and Bourguignon, 2000). Even more generally, if we are ultimately interested in the justness of a society, focusing on income alone may take our attention away from elements of a society such as the existence of personal liberty and being treated as a social equal upon which no price can be placed (Sen, 2000). Nonetheless, income fills an important instrumental role in virtually any discussion of justice, implying that we have an interest in the financial resources available for households, even if we cannot claim they provide a direct representation of well-being. While an argument could be made that “available resources” should include the value of goods supplied by the public sector, they are traditionally measured by after-tax and transfer (disposable) income and we follow that tradition in this paper. Because of data constraints, we also do not consider the imputed income value of durables.

Given our desire to focus on disposable income, the obvious path to pursue is to start with income from the market, add government transfers, and then subtract taxes paid. The sticky question is which taxes to subtract. Subtracting income taxes is appropriate because they reduce the pre-consumption resources available to the household. We choose not to account for excise and sales taxes because they are a function of the consumption decisions of the household, raising complex endogeneity issues. While we agree that including these taxes would be interesting, we believe that it is important first to provide a consistent series on income available for households to use for consumption at the prices they face (which includes sales taxes) and which does not require behavioural assumptions with which the reader may not agree. Thus, we leave investigations of sales and excise taxes to future work.

The treatment of payroll taxes - particularly those that are (perhaps nominally) ear-marked for specific spending programmes - involves a different set of issues.[7] If the taxes are strongly linked to a particular benefit, and the benefit would have been purchased by the household in the absence of the government programme, then the payroll taxes can be thought of as a use of funds from the family budget.[8] The ‘purchase’ of the benefit just happens to be from the government rather than from a private provider. In this case, we should not account for payroll taxes given our interest in measuring the family’s pre-consumption and saving budget constraint. On the other hand, if the link between the payroll taxes and the benefits received is weak, the revenue collected should be thought of as reducing the size of the family’s pre-consumption budget. If so, we should subtract payroll taxes from the family’s income in order to arrive at after-tax

[7] See Bird and Tsiopoulos (1997) for a discussion of benefit taxation.

[8] More specifically, only the portion of the government programme that crowds out private spending should be accounted for. If the programme provides social insurance that would otherwise not be purchased by individuals then the counterfactual suggests that the household budget in the absence of the programme would not be reduced by spending on premiums.

income. The decision to include or exclude payroll taxes is therefore specific to each particular case, and necessarily in part subjective.

For Canada, there are three payroll taxes we consider: Canada/Quebec Pension Plan premiums, Employment Insurance premiums, and provincial health levies. For the first two cases, benefits are tied to earnings rather than directly to contributions. For example, the Canada Pension Plan is not a defined contribution plan, so the marginal dollar of contribution does not affect benefits. Moreover, there is a substantial intergenerational transfer component in the Canada and Quebec Pension Plan premiums beyond what is necessary to fund one's own benefits.[9] For Employment Insurance, total benefits and contributions are not closely linked in practice, in spite of a nominal legislative link.[10] Finally, health levies flow into general revenues and are not linked to the amount of health services received. Since we contend that the tax-benefit link is weak in all three cases, it would be appropriate to subtract payroll taxes to arrive at our desired household income measure. Because this decision is somewhat subjective, however, we provide separate results with and without payroll taxes.

Predicting income taxes on the Census

<u>Description</u>

Our goal is to impute income taxes paid for every family observed in a given Census. In a world with perfectly informed and rational agents, tax filers would minimize their tax burdens by claiming the optimal combination of income, deductions and credits. In such a world, we could predict income taxes by applying a 'calculation' approach (i.e. mathematically solving the filer's tax minimization problem). However, this requires detailed information on income sources, tax credits, and deductions, some of which is missing in the Census. Most importantly, the Census does not include questions on contributions to a Registered Retirement Savings Plan (RRSP) and Charitable Donations. Since the use of these tax measures are more common among middle and upper income families, their omission in a tax calculation approach would tend to overstate income

[9] See OSFI (2004), page 121. The internal rate of return for the CPP is calculated as 9.6% for the 1930 birth cohort and only 2.1% for the 1980 birth cohort, indicating substantial transfers across existing generations.

[10] In the 1980s, spending regularly exceeded contributions. In the last decade, the reverse has been true.

taxes in the upper portion of the income distribution by a considerable margin. This is in fact what we found in attempting to predict taxes in this manner.

In contrast, we adopt a more "reduced form" approach which essentially consists of defining homogenous groups of individuals based on a set of characteristics that are known to affect income taxes (e.g. income, family structure, age, province of residence) and obtaining average taxes actually paid by members of the group from tax data in the Census year of interest. Assuming we define the groups using characteristics that are also available in the Census, we can then assign the relevant average taxes paid to each member of the group in the Census. If the groupings are detailed enough, we expect income taxes to be about the same for most members in the group. This necessarily creates a tax measure that includes some degree of measurement error at the individual level (since all people in the group do not actually pay the exact same amount in taxes). However, within the group, there is no reason to believe the errors are systematic. We implement the approach using an initial regression of income taxes paid on a flexible function of observable characteristics. We then predict taxes paid for each person in the Census using the person's characteristics and the estimated regression coefficients.[11]

The regression approach has two main advantages over the calculation method. First, it has much less stringent data requirements. We do not need to know, for example, values of actual deductions and credits claimed. Instead we need to know only that people in a particular income class and with specific family characteristics pay a certain level of taxes, which will necessarily reflect whatever deductions they make. The second, related, advantage is that we obtain an estimated measure of what families in particular groups actually pay in taxes. This may differ systematically from their optimal tax bill to the extent that tax payers do not take full advantage of all available deductions and shelters. Since we are interested in the actual well-being of Canadian families, it is the actual tax bill that is relevant.

The tax data we use to predict income taxes is the T1 Family File (T1FF), which consists of T1 personal tax records with family level information added by Statistics Canada. The definition of the family on T1FF is the census, or nuclear family. For measuring economic well-being, a preferred concept is the economic family, which may

[11] The code used to predict taxes is available upon request from the authors.

include two or more census families, as long as there is a blood relationship between them (e.g. a brother living with his sister and her family). Since the Census includes identifiers for both types of families, we are able predict income taxes by using census family information, and then calculate after-tax income at the economic family level.

We predict taxes on the Census files for the years 1980, 1985, 1990, 1995, and 2000.[12,13] Unfortunately, T1FF is not available prior to 1982; however, this is not so limiting since the tax laws remained virtually unchanged between 1980 and 1982. Hence, we model income taxes in 1982 and use the estimated parameters to predict income taxes in 1980 on the Census.[14]

In each year, we predict federal and provincial taxes separately; that is, we estimate one federal tax model and ten provincial tax models.[15] Since the Quebec government does not provide provincial tax information to Canada Revenue Agency, actual Quebec taxes are not available. Since 1992, Statistics Canada has imputed Quebec taxes on the T1FF. For the years 1980, 1985, and 1990, we turn to the SCF to predict Quebec taxes.

Income taxes are collected from individuals, but require family level information for calculation. Thus, our strategy consists of estimating models on individuals, using individual and family level information as determining factors. All models are estimated on individuals who are at least 15 years old.[16] Predicted taxes are automatically set to zero for everyone else. In cases where negative federal or provincial taxes are predicted,

---

[12] The Census is actually conducted in May or June of the following year in each case, but the income collected refers to the previous year.

[13] Note that although there were few financial incentives for individuals with no taxable income to file taxes in the 1980s, this does not prevent us from using the tax files to model taxes paid since the models simply predict very low levels of taxes paid for these individuals by extrapolation.

[14] We also used the methodology to predict income taxes on the 1980 SCF data, and found that income inequality estimates based on predicted after-tax income lined up almost exactly with income inequality based on actual after-tax income. In other words, using 1982 tax information to predict 1980 income taxes yields accurate inequality measures.

[15] Note that payroll taxes are not included in our primary definition of taxes. Later in the paper, we will introduce a measure of after-tax income that incorporates payroll taxes.

[16] For tax purposes, December 31st is the reference date. However, the Census reference data is normally in May or June (of the following year). Since the version of the Census files we used for this study did not contain the exact date of birth, we randomly assigned individuals as either being the same age (in years), or being one year younger on December 31st of the previous year. To do so, we assigned individuals with a randomly chosen number between 0 and 1 from a uniform distribution, and assigned them as being one year younger if this number was less than or equal to n/365, where n=the number of days between the Census and December 31st of the previous year.

the values are set to zero. To reduce processing time on T1FF, a 20% random sample of census families is used in the estimations. Since the sample size is much smaller in the SCF, the full Quebec sample is used in the estimations.

The estimation approach consists of regressing income taxes on a set of determining factors by ordinary least squares (OLS). The most important factor in determining one's tax obligations is taxable income. Although this information is obviously available on the tax files, not all components are available on the Census files. We thus use a proxy for taxable income, which is defined as the sum of the following income sources:

- Wages and Salaries
- Other Employment Income
- Net Self-Employment Income
- Investment Income, Dividends
- Net Rental Income
- Alimony Received
- Private Pension Income
- Employment Insurance Benefits
- Other Income

The omitted components of taxable income include the Canada and Quebec Pension Plan Benefits, Old Age Security Income, and various deductions (e.g. RRSPs, Charitable Donations, Alimony Paid, Union dues, Child Care Expenses, Moving Expenses, Carrying Charges, Interest Expenses).[17] The Census either does not include these sources of income or deductions, or they are lumped together with other variables.[18,19]

The objective in specifying the models is to include variables that are expected to affect income taxes in as flexible a manner as possible. This flexibility is made possible by the large sample size available in the T1FF. A more flexible model in this context essentially corresponds to using smaller and more precisely defined groups, implying

[17] Social Assistance Income became taxable in the 1990s if a spouse had sufficiently high income (on the order of $50,000 in most years). Since eligibility for social assistance is normally based on family income, it is for the most part not actually taxed.

[18] The base deduction amount is not subtracted from the taxable income proxy. The intercept term should capture this since it applies equally to everyone. Also, all new alimony agreements after March 30th, 1997 are non-taxable and non-deductible. However, new arrangements can not be distinguished from previous ones in the Census. This only affects the predictions for the year 2000, and likely does not have a large impact since most existing alimony agreements in 2000 appear to be taxable (i.e. the aggregate amount of Alimony Received on the Census matches the amount on the T1FF quite closely).

[19] In the 1980s, certain child credits and family allowances were taxable. Since our model contains all the variables necessary to calculate the amount of these credits, their inclusion in the taxable income measure would not add to the fit of the model.

smaller measurement errors. The federal income tax model (denoted by FEDTAX) is shown below for person *i* at time *t*:

$$(1) FEDTAX_{it} = \alpha_t + \sum_{j=2}^{12} IR_{ijt} \cdot \beta_{jt} + \sum_{j=2}^{12} IR_{ijt} \cdot I_{it} \cdot \delta_{jt} + I_{it}^2 \cdot \phi_t + \sum_{j=2}^{12} SPIR_{ijt} \cdot \varphi_{jt} + \sum_{j=2}^{12} SPIR_{ijt} \cdot I_{it} \cdot \gamma_{jt} + SPI_{it}^2 \cdot \eta_t + CHIL_{it} \cdot \lambda_t + CHIL_{it}^2 \cdot \mu_t + SENIOR_{it} \cdot \nu_t + \sum_{k=2}^{10} PROV_{ikt} \cdot \pi_{kt} + \varepsilon_{it}$$

Most of the variables revolve around the taxable income proxy, which we shall refer to as "income" for simplicity. We capture the income dimension, in part, by using a set of 12 dummy variables, each corresponding to an income range for the individual (denoted by *IR*).[20] These are helpful in accounting for the non-linear nature of the taxation rules. To allow for heterogeneity within each range, we interact these dummy variables with income (denoted by *I*).[21]

Since many tax deductions are based on the couples' taxable income, we also include the same set of income variables for the spouse (denoted by the prefix *SP*).[22] Of course, income is set to zero if no spouse is present. The presence of children under the age of 18 is also used in calculating certain deductions. To capture this, we include variables indicating the number of children in the family (denoted by *CHIL*) and its squared value (to capture potential non-linearities).[23] An individual's tax obligations are also influenced by whether or not he or she is at least 65 years old, which we capture by including a 'senior citizen' dummy variable (denoted by *SENIOR*). Finally, to allow for

[20] The ranges include (in constant 2000 dollars): <=0 (omitted), 0-5K (i.e. 5,000), 5-10K, 10-20K, 20-30K, 30-40K, 40-50K, 50-60K, 60-100K, 100-150K, 150-250K, and >250K.

[21] Note that any predicted inequality measure that is sensitive to the tails of the distribution will be biased upwards when a local averaging technique is used for prediction. Within locally averaged groups, more tax dollars are taken from the bottom than in actual fact, while fewer tax dollars are taken from the top. This 'Reverse Robin Hood' effect (within locally averaged groups) should tend towards zero as the number of groups approach infinity. However, sensitivity tests suggested that increasing the number of groups had virtually no effect on predicted outcomes, but it increased the computational burden considerably.

[22] Prior to 1992, the definition of a "spouse" only included legally married spouses for tax purposes. Since then, the definition has also included common-law spouses.

[23] Note that we also restrict our samples to families with at least one individual who is 18 years old or above.

differences in the behaviour of taxfilers or in deductions across provinces, we've included provincial dummy variables (denoted by *PROV*).[24]

Our strategy for estimating provincial income taxes (denoted by *PROVTAX*) is identical to the federal model, except that the provincial dummy variables are necessarily excluded. The model is shown below for person *i* living in province *p* at time *t*:

$$(2) PROVTAX_{ipt} = \alpha_{pt} + \sum_{j=2}^{12} IR_{ijpt} \cdot \beta_{jpt} + \sum_{j=2}^{12} IR_{ijpt} \cdot I_{ipt} \cdot \delta_{jpt} + I_{ipt}^{2} \cdot \phi_{pt} + \sum_{j=2}^{12} SPIR_{ijpt} \cdot \varphi_{jpt} + \sum_{j=2}^{12} SPIR_{ijpt} \cdot SPI_{ipt} \cdot \gamma_{jpt} + SPI_{ipt}^{2} \cdot \eta_{pt} + CHIL_{ipt} \cdot \lambda_{pt} + CHIL_{ipt}^{2} \cdot \mu_{pt} + SENIOR_{ipt} \cdot \nu_{pt}$$

Once we have estimates of the parameters in the *FEDTAX* and *PROVTAX* regressions, we use these to predict tax values for individuals in the Census data sets based on their observable incomes and other characteristics. We then subtract imputed taxes from incomes and combine the resulting after-tax incomes for household members to arrive at our measure of family after-tax income.

<u>Validation</u>

Our next step is to assess the accuracy of our income tax prediction approach. To do so, we apply an internal validation technique, assessing how well the models predict income taxes on the tax data themselves. Note that we perform the assessment at the national level because we are only interested in national level income inequality in this study. The use of the prediction approach to study inequality among sub-groups of the population would require further assessment that is beyond the scope of this study.

We begin by predicting income taxes (federal and provincial combined) for each individual in the tax data based on the characteristics outlined above, and then aggregating income taxes to the census family level. In Table 2, we show income percentiles and measures of income inequality for the distribution of the actual after-tax income in T1FF 2000 (column 1) and the distribution of our predicted after-tax income (column 2). We also report the error as a percent of the actual amount for each of the percentiles. Because we are interested in comparing percentiles (and functions of the

[24] The omitted province is Ontario.

percentiles) across years, the percent error for the percentiles is the relevant error for our work. As a more stringent test, we also calculated the mean absolute error for the sample, finding an average error that is 5.0% of the actual amount.

Table 2: Percentiles and inequality indices based on actual and predicted after-tax income, T1FF 2000[a]

	Actual	Predicted	Error (%)[b]
P1	176	174	-1.3
P5	4,359	4,413	1.2
P10	8,301	8,240	-0.7
P25	14,672	14,042	-4.3
P50	23,681	23,592	-0.4
P75	35,564	35,645	0.2
P90	49,729	49,584	-0.3
P95	61,056	60,526	-0.9
P99	100,971	98,829	-2.1
Log (P95/P5)	1.1	1.1	-0.8
Log (P90/P10)	0.8	0.8	0.2
Decile ratio	17.9	17.5	-2.1
Gini	0.3644	0.3669	0.7

[a] The unit of analysis is the individual, but income is measured at the census family level and divided by the square root of the number of members in the family.
[b] Similar levels of error were obtained for the 1982, 1985, 1990, and 1995 T1FF files. These results are available from the authors upon request.

The results in Table 2 suggest that our fitting method does a good job, even though it is based on an incomplete measure of taxable income. In particular, actual and predicted incomes are almost identical throughout the distribution. Given this, it is not surprising that the various summary measures of inequality also show little difference. The specific summary measures we use here, and in the following sections are: the log of the ratio of the 95th to the 5th percentile, the log of the ratio of the 90th to the 10th percentiles, the ratio of the average income in the top decile to the average in the bottom decile, and the Gini coefficient. The log (P95/P5), log (P90/P10) and decile ratio measures are all sensitive to movements in the tails of the distribution, with log (P95/P5) being particularly so. The Gini, on the other hand, has a natural interpretation related to the Lorenz Curve for a distribution and is well know to be sensitive to changes in the middle of the distribution. All the measures indicate a strong similarity in the degree of inequality in the actual and predicted after-tax income distributions. Similar results were found using data from 1982, 1985, 1990, and 1995 and are available upon request.

After-tax income inequality levels and trends, 1980-2000

We now put the Census-AT files to their first use by documenting trends in after-tax income inequality over the period 1980 to 2000 in Table 3. For comparison, we also show inequality trends in market income (i.e. earnings, investment income, private pension income, and other non-transfer income) and total income (i.e. market income plus transfers). Later, we will discuss a fourth income concept: after-tax income, where taxes include income and payroll taxes.

While it is tempting to interpret differences in inequality in the pre- and post-tax distribution as the 'impact' of the fiscal system on household budgets, such an inference requires strong incidence assumptions.[25] This is so because we do not observe a true 'pre-tax' market outcome. Instead, we observe the market outcome in the presence of taxes. If the incidence of the income tax is not entirely upon the individuals paying it then the observed market wages and capital income receipts will reflect a premium to compensate for taxes paid. For example, if top executives are internationally mobile, their wages may be adjusted to offset differences in income tax between Canada and alternative countries of employment. In this case, the effect of the fiscal system on an executive would be smaller than would be measured by comparing the observed pre- and post-tax incomes. Similarly, the economic incidence of employer-paid payroll taxes and taxes on profit may be on workers, implying different wages if the taxes were not present.

To proceed, we assume the incidence of taxes is equivalent to the statutory burden, which has the virtue of transparency. Moreover, because our primary goal is to analyze after-tax income (which is independent of assumptions of incidence), the assumption is not critical to our primary conclusions. However, our incidence assumption should be kept in mind when comparing the market, total, and after-tax income results.

[25] See Fullerton and Metcalf (2002) for a complete discussion of tax incidence.

Table 3: Income inequality indices, Census-AT[a]

Year	Market income Log (P90/P10)	Market income Gini	Total income Log (P95/P5)	Total income Log (P90/P10)	Total income Gini	After income tax income Log (P95/P5)	After income tax income Log (P90/P10)	After income tax income Gini
1980	1.18	0.3923	0.98	0.72	0.3352	0.89	0.65	0.3083
1985	1.41	0.4157	1.02	0.75	0.3445	0.93	0.67	0.3140
1990	1.34	0.4142	1.00	0.73	0.3435	0.89	0.65	0.3070
1995	1.85	0.4458	1.08	0.79	0.3574	0.97	0.69	0.3194
2000	1.51	0.4387	1.07	0.76	0.3620	0.95	0.68	0.3219
% Growth								
1980-2000	28.4	11.8	9.1	6.0	8.0	6.1	4.0	4.4
1980-1990	13.7	5.6	2.2	1.8	2.5	-0.3	-0.4	-0.4
1990-2000	12.9	5.9	6.7	4.1	5.4	6.5	4.5	4.9
1980-1985	20.1	6.0	4.2	3.7	2.8	3.8	3.4	1.9
1985-1990	-5.3	-0.4	-1.9	-1.9	-0.3	-4.0	-3.7	-2.2
1990-1995	38.7	7.6	8.1	7.0	4.0	8.7	6.2	4.0
1995-2000	-18.6	-1.6	-1.3	-2.7	1.3	-2.0	-1.6	0.8

[a] The unit of analysis is the individual, but income is measured at the economic family level and divided by the square root of the number of members in the family.

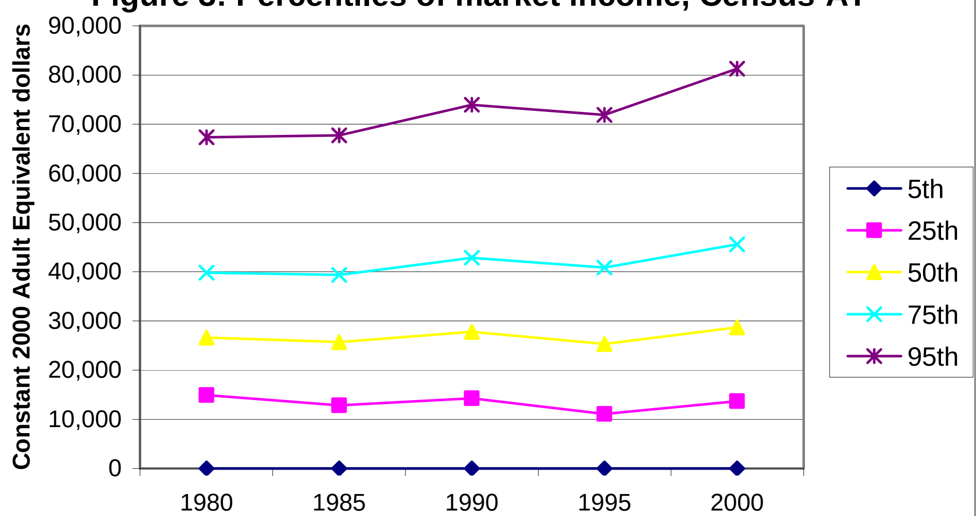
The first panel in Table 3 indicates that market income inequality has been rising over the entire period and at a similar pace in each decade. The upward trend results from the fact that market income inequality has risen during the recessions in the earlier part of each decade (as we would expect, since marginal workers are most likely to be laid off, and hiring is reduced substantially), but the decline in inequality in the recovery periods witnessed later in each decade was always much smaller. It is worth noting that while the 1980, 1990 and 2000 Census years are not perfectly comparable in terms of economic conditions, they do all roughly correspond to cyclical peaks and thus comparison across those years is useful for establishing longer term trends. The general rise in market income inequality around the developed world is well known and has spawned considerable debate over potential explanations, with the leading explanations including institutional change (e.g. DiNardo et al., 1996, for the US) and some combination of technical change and globalization (e.g. Acemoglou, 2002 for the US; Beaudry and Green, 2003 for the US and Germany). Piketty and Saez (2003) show that much of the increase is concentrated at the very top of the income distribution for the United States. The timing of the increase suggests a relationship with taxes, but further evidence on this question from Canada by Saez and Veall (2005) finds no obvious relationship to tax rates.

The second panel contains measures based on total income, which is constructed by adding transfer income to the market income measures described in the first column. The lower levels of all the inequality measures reflect the fact that transfers are primarily received by those in the lower part of the distribution. The over time patterns also imply that transfers tend to moderate cyclical fluctuations as the relative size of the increases in inequality across the recessionary periods (1980-85 and 1990-95) is smaller in total income compared to market income. This is expected since unemployed individuals usually qualify for some form of assistance but lose that eligibility when they return to the workforce, which can reduce the improvements in income resulting from the new job. However, there is an interesting difference in this pattern across decades. The log (P90/P10) measure indicates that the increases in inequality in the total income measure were larger in the 1990s than the 1980s even though the relative increases in inequality in market income were similar across the two decades. Examining the Gini, one finds that bringing in transfers cuts the growth in market income inequality in the 1980s by more than half, while in the 1990s it has virtually no impact on inequality growth.

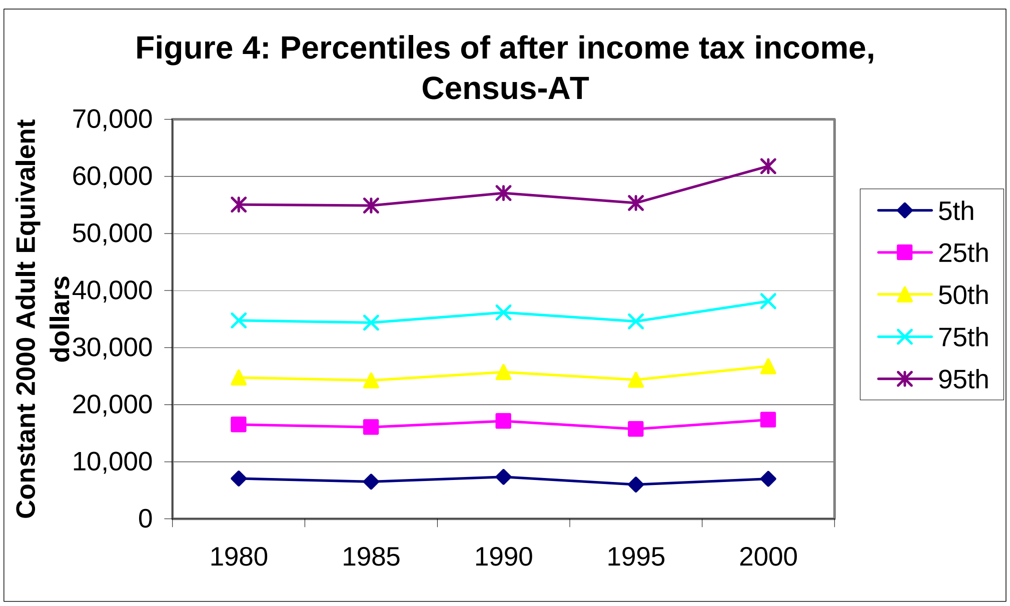
Levels of inequality are further reduced according to all measures when we shift to after-tax income in the third panel. Recall that the measure being examined in this panel consists of market plus transfer income minus income taxes. For all the measures of inequality, bringing in taxes actually makes the growth in inequality in the 1980s mildly negative. Thus, if we focus on the Gini, the initial substantial growth in inequality in market earnings is cut in half when we bring in transfers and then the remaining inequality growth is eliminated once we introduce the impact of taxes. The breakdown within the decade indicates that reductions in inequality in the first, recessionary, half of the decade are associated with transfers but not taxes. The impacts in the second half are the reverse: including transfers affects the growth in inequality witnessed in market inequality to a minor degree (as measured by the Gini) but once one then subtracts taxes, the inequality reduction becomes relatively substantial. This is all as one might predict. In a recessionary period, market income inequality increases as people in the lower part of the distribution lose jobs. This is offset by transfers, thus reducing inequality growth to some degree. In boom periods, market income inequality declines as low income people regain employment, but fewer transfers are required and so transfers have less impact in terms of reducing inequality. At the same time, incomes in higher parts of the distribution increase in boom times and taxes act to mitigate the associated inequality increases.

The first half of the 1990s also appears to fit with this predicted pattern. A substantial increase in market inequality is cut approximately in half when transfers are introduced but there is no further change when taxes are then subtracted. The last half of the 1990s, however, is a different story. Market income inequality falls as one would predict in a boom, but inequality in total (market plus transfer) income actually rises. Bringing in taxes mitigates this slightly but the net effect is an increase in after-tax inequality in spite of the fact that this is a boom period. Note that this does not imply that the tax and transfer system generated a level of inequality higher than what was witnessed in market income. (Comparisons across the 2000 row in the upper half of Table 3 show that is not true). However, as measured in terms of the Gini, changes in the tax and transfer systems in the latter half of the 1990s did lead to a growth in inequality in disposable income even though inequality in market income was decreasing. The more tail-sensitive log (P90/P10) measure indicates a decline in disposable income inequality in this period, but of much smaller magnitude than that witnessed in market income.

Figure 3: Percentiles of market income, Census-AT



We provide further detail on the movements in inequality over time in Figures 3 and 4. Figure 3 contains plots of the 5th, 10th, 50th, 90th and 95th percentiles of the real market income distribution as calculated from Census data. The 5th percentile of this distribution is always zero, but, fitting with our earlier description of movements in inequality across the cycle, the 10th percentile declined sharply in both the early 1980s and 1990s with recoveries in the second (boom) part of each decade. In both decades the recessionary declines in the 10th percentile were not fully offset by the increases in the ensuing recoveries. As a result, the 10th percentile dropped precipitously from $3,700 in 1980 to $2,000 in 2000. In contrast, both the 90th and 95th percentiles were essentially flat in the 1980s recession and slightly falling in the 1990s recession but experienced very substantial increases in the boom periods. The 90th percentile increases from $55,000 in 1980 to $65,000 in 2000. Thus, the substantial increase in inequality over the period described in Table 3 arose because of both increases at the top and declines at the bottom of the distribution.



Our numbers at the top of the distribution are corroborated by those derived from tax data by Saez and Veall (2005). Saez and Veall examine market income excluding capital gains using tax records over the period from 1920 to 2000. They find that the 95th percentile of their market income distribution was virtually unchanged between 1980 and

1990 but increased by 8.5% between 1990 and 2000. Our results for the 1990s are similar, showing an increase of 9.9% for the 1990s for the 95th percentile, but for the 1980s our data show a similar 9.8% increase for the 1980s. Saez and Veall (2005) show that these movements at the top of the distribution are dominated by increases in the share of total income going to the top 0.1% of earners and, thus, that even examining the 95th percentile tends to understate the extent of the increase in inequality. As a side point, Frenette et al. (2004) show that tax data indicate much larger increases in inequality than Census data in the 1990s in general. However, they find the major difference occurs at the bottom of the distribution and may be related to the way in which individuals are combined into families in tax data.

In Figure 4, we present plots of the same percentiles of the after tax and transfer income distribution. Note, first, that incomes at the bottom are higher and incomes at the top lower than what we see for market incomes in Figure 3. The tax and transfer system also dampened cyclical fluctuations in the lower end of the distribution. Further, once taxes and transfers are included, both the 5th and 10th percentiles of the distribution actually increased between 1980 and 2000. However, as described earlier, the impact of taxes and transfers is smaller in the 1990s and the net result is a decline in the 5th percentile from $7,400 in 1990 to $7,000 in 2000. Interestingly, this decline just exactly balances the increase at the 5th percentile over the 1980s, leaving the 5th percentile unchanged between 1980 and 2000. At the other end of the distribution, the tax and transfer system acted mainly to mitigate the growth in the top percentiles. This again fits with the discussion earlier in which we suggested that transfers would have little impact on movements at the top end but taxes would serve to reduce their boom period increases.[26]

Finally, we turn to results in which we also subtract payroll taxes (i.e. Canada/Quebec Pension Plan premiums, Employment Insurance premiums, and provincial health levies). As we argued earlier, while removing payroll taxes in our opinion provides a better measure of disposable income, this decision is more subjective. The results in Table 4 show our measures of inequality for the definition of income that

[26] In related work, we are investigating more carefully the roles played by taxes and transfers in income inequality over the 1980-2000 period.

subtracts payroll taxes as well as income taxes. The results change little from the set of results that did not exclude payroll taxes (i.e. after income tax income). Most log percentile values are slightly smaller when payroll taxes are factored in, suggesting that payroll taxes have a small dampening effect on these inequality measures. This is perhaps surprising, since the payroll taxes apply at a flat rate only up to a cap, meaning that lower income working individuals pay a higher percentage of their income in taxes. However, because the taxes are only levied on working individuals, these taxes miss all non-workers, who likely have lower total incomes than working individuals, and are thus more likely to be found in the left tail of the distribution (P5 or P10). The Gini coefficient, on the other hand, is more middle sensitive. It is thus not surprising that this measure is less affected by payroll taxes: workers in the middle part of the distribution pay roughly the same amounts in payroll taxes.

Table 4: Income inequality indices, Census-AT[a]

	After income tax income			After income and payroll tax income		
Year	Log (P95/P5)	Log (P90/P10)	Gini	Log (P95/P5)	Log (P90/P10)	Gini
1980	0.89	0.65	0.3083	0.89	0.65	0.3087
1985	0.93	0.67	0.3140	0.92	0.67	0.3139
1990	0.89	0.65	0.3070	0.88	0.64	0.3069
1995	0.97	0.69	0.3194	0.95	0.67	0.3177
2000	0.95	0.68	0.3219	0.93	0.66	0.3218
% Growth						
1980-2000	6.1	4.0	4.4	5.2	2.5	4.2
1980-1990	-0.3	-0.4	-0.4	-0.8	-0.9	-0.6
1990-2000	6.5	4.5	4.9	6.0	3.4	4.8
1980-1985	3.8	3.4	1.9	3.6	3.0	1.7
1985-1990	-4.0	-3.7	-2.2	-4.2	-3.8	-2.2
1990-1995	8.7	6.2	4.0	8.0	5.1	3.5
1995-2000	-2.0	-1.6	0.8	-1.8	-1.6	1.3

[a] The unit of analysis is the individual, but income is measured at the economic family level and divided by the square root of the number of members in the family.

Comparison of Census Patterns with SCF/SLID

As discussed earlier, most previous discussions of income inequality in Canada have been based on the SCF and SLID surveys. In the second section, we showed results indicating that Census and SCF/SLID generate very different levels and patterns in total income inequality. In this section, we create a more complete comparison, showing differences in all of our income concepts. To do this, we present Table 5, in which we replicate the measures in Table 3 using SCF/SLID data.

Table 5: Income inequality indices, SCF-SLID[a,b]

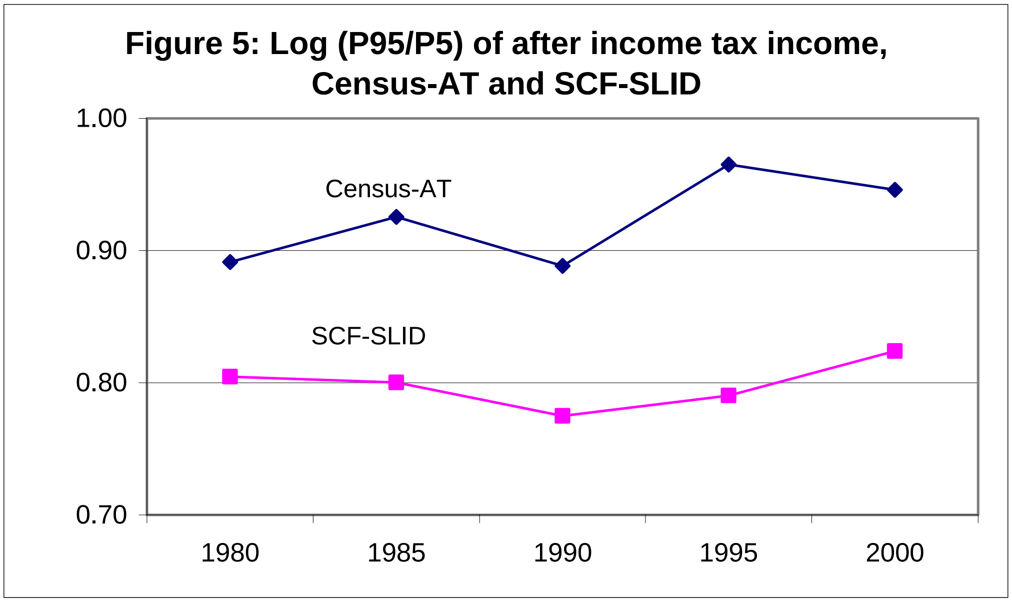
Year	Market income Log (P90/P10)	Market income Gini	Total income Log (P95/P5)	Total income Log (P90/P10)	Total income Gini	After income tax income Log (P95/P5)	After income tax income Log (P90/P10)	After income tax income Gini
1980	1.05	0.3688	0.89	0.69	0.3124	0.80	0.61	0.2848
1985	1.18	0.3938	0.90	0.69	0.3206	0.80	0.60	0.2884
1990	1.17	0.3947	0.90	0.69	0.3197	0.77	0.59	0.2805
1995	1.38	0.4205	0.91	0.70	0.3308	0.79	0.60	0.2877
2000[b]	1.22	0.4093	0.95	0.72	0.3377	0.82	0.61	0.2959
% Growth								
1980-2000	16.3	11.0	5.9	3.5	8.1	2.4	0.6	3.9
1980-1990	12.2	7.0	0.4	-0.7	2.3	-3.7	-3.7	-1.5
1990-2000	3.7	3.7	5.5	4.2	5.6	6.3	4.5	5.5
1980-1985	13.2	6.8	0.8	-0.5	2.6	-0.6	-0.6	1.2
1985-1990	-0.9	0.2	-0.4	-0.1	-0.3	-3.2	-3.2	-2.7
1990-1995	17.3	6.5	1.8	2.2	3.5	2.0	2.3	2.6
1995-2000	-11.6	-2.7	3.7	1.9	2.1	4.3	2.1	2.8

[a] The unit of analysis is the individual, but income is measured at the economic family level and divided by the square root of the number of members in the family.
[b] To partially account for the break that occurred in the series in 1996, the value in 2000 was derived by adding the growth in SLID between 1996 and 2000 to the value in the SCF in 1996.

The most notable difference between the Census based measures in Table 3 and the SCF/SLID based measures in Table 4 is their levels. The inequality measures are uniformly smaller when using the SCF/SLID. This is true for all three definitions of income, for all three measures of inequality, and at all five points in time. It is also interesting to compare these measures to inequality measures for the United States. Gottschalk and Smeeding (1997) present inequality measures for a variety of countries, including US results based on the CPS, which is directly analogous to the SCF for Canada. The income concept they use is disposable income, which is essentially the same as our disposable income concept. Based on their Table 3, the US disposable income Gini coefficient takes a value of .310 in 1980 and .352 in 1990. These are clearly larger than the Gini values based on the SCF for Canada and fit with common claims that inequality is both larger for the US at any point in time and grew faster in the 1980s. On the other hand, the Census-AT based Gini coefficient is almost identical to the US value in 1980. However, the Census-AT data again shows no growth in inequality over the 1980s, in strong contrast to the US, and, as a result, it indicates that inequality is lower than in the US by 1990. It would be interesting to investigate whether US data shows a similar difference between CPS and Census based measures of disposable income inequality as what we observe between the SCF and Census for Canada.

Differences in the time patterns of inequality generated from the two data sources is more nuanced. If we focus on inequality measures that emphasize movements in the tails of the distribution – the log (P90/P10) and log (P95/P5) measures in these tables – the Census indicates much more substantial increases in inequality in all three types of income across the whole 1980 to 2000 period. For example, Census-AT shows an increase in log (P95/P5) of 6.1% while the SCF/SLID shows only 2.4%. For market income using log (P90/P10), the difference is more pronounced, with a growth rate of 28.4% for the Census and only 16.3% for the SCF/SLID. In contrast, the growth rates in the Gini are more similar for all three income definitions. This is potentially consistent with differences in coverage of the population, with the SCF/SLID having much lower coverage. To the extent that lower coverage occurs systematically in the tails of the distribution, one might expect to see patterns such as this where the SCF/SLID tells a different inequality story when we focus on the tails but a similar story when we focus on the middle.

Regardless of the inequality measure we use, the SCF/SLID provides a different picture of both decadal and cyclical movements. For example, while the Census indicates a relatively even split of the overall increase in market income inequality between the two decades, the SCF/SLID implies that about two-thirds of the increase occurred in the 1980s. Again, this is related to what is happening in the tails since the inequality movements in the 1990s are even more heavily driven by what is happening in the tails in that decade than in the 1980s. Cyclically, market income inequality shows much stronger swings in Census data than in SCF/SLID data.



Our ultimate interest is in levels and movements in after-tax income inequality. In Figure 5, we plot the log of the 95th/5th percentile ratios for Census years based on both Census and SCF/SLID data.[27] Again, one can see the substantially higher level of inequality observed in Census data in all years. In addition, the Census based measure is clearly much more cyclical and shows a stronger trend increase in inequality. In the SCF/SLID, the tax and transfer system more than offsets increases in market income inequality in the 1980s. The 1990s witnesses a substantial increase in after tax income inequality but this places the final inequality level in 2000 only slightly above its 1980 level. In contrast, in the Census data, the tax and transfer system almost exactly offsets increases in market income inequality in the 1980s but leaves substantial growth in inequality in the 1990s. Interestingly, the two datasets point to very similar growth in inequality in the 1990s, with the difference in growth over the two decades from 1980 to 2000 arising because the SCF/SLID data implies a substantial drop in disposable income inequality in the 1980s that is not seen in the Census-AT data. Thus, both datasets point to a similar conclusion that the tax and transfer system ceased to fully offset increased market income inequality in the 1990s.

[27] Once again the SLID income levels in 2000 were derived by adding the growth rates between 1996 and 2000 from the SLID to the 1996 SCF values in order to obtain levels that are comparable to those from the earlier SCF data.

Conclusion

In this paper, we present results on measures of income inequality over the 1980s and 1990s for Canada using Census data. We argue that Census data is superior to the most commonly used data source, the combination of the Surveys of Consumer Finances and the Survey of Labour and Income Dynamics, because the latter has population coverage that is only about 80% of the Census. The main disadvantage of using the Censuses, apart from the fact they are collected only every 5 years, is that they do not include information on taxes paid, which are necessary for examinations of disposable income. Thus, a major part of our exercise is to impute taxes for families in each Census. We do this using a reduced form approach in which we estimate regressions of taxes paid on family characteristics using tax data and then use the estimated coefficients from those regressions to predict taxes paid in the Censuses. We demonstrate that this approach does a good job of predicting taxes paid using a validation exercise based on tax data.

Once we have the fitted taxes, we are in a position to generate a new series on inequality in after-tax income. We show that both the levels and time patterns of inequality depicted in this series is dramatically different from what is seen in the SCF/SLID data. Simple plots of percentiles of pre-tax income distributions constructed from the Census and the SCF/SLID indicate that the latter generates both substantially larger values of the lower percentiles of the distribution and substantially smaller values of the upper percentiles. Thus, it is not surprising that the levels of inequality as measured in the Census are substantially higher in all years and all income measures we examine. In addition, Census based inequality measures show much more cyclicality and much stronger increases in inequality over the time period we examine. The decadal patterns indicate that the 1980s can be characterized as a period in which strong increases in pre-tax and transfer income inequality were fully offset by the tax and transfer system. In contrast, equally strong pre-tax and transfer income inequality increases in the 1990s were not offset to nearly the same degree in the 1990s. This opens interesting questions about the impacts of changes in several parts of the social safety net in the 1990s.

References

Acemoglu, Daron (2002). "Technical change, inequality, and the labor market," *Journal of Economic Literature*, Vol. 40, No.1, pp. 7-72.

Atkinson, Anthony B. and Francois Bourguignon (2000). "Income Distribution and Economics," in Handbook of Income Distribution, Vol. 1, A.B. Atkinson and F. Bourguignon, eds., New York: North-Holland.

Beach, Charles .M. and George A. Slotsve (1994). "Polarization of earnings in the Canadian labour market," Bell Canada Paper on Economic and Public Policy.

Beaudry, Paul and David A. Green (2000). "Cohort patterns in Canadian earnings: Assessing the role of skill premia in inequality trends," *Canadian Journal of Economics*, Vol. 33, No. 4, pp. 907-936.

Beaudry, Paul and David A. Green (2003). "Wages and employment in the United States and Germany: What explains the differences?," *American Economic Review*, Vol. 93, No. 3, pp. 573-602.

Bird, Richard M and Thomas Tsiopoulos (1997). "User charges for public services: Potentials and problems," *Canadian Tax Journal*, Vol. 45, No. 1, pp. 25-86.

Dinardo, John, Nicole Fortin, and Thomas Lemieux (1996). "Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach," *Econometrica*, Vol. 64, No. 5, 1001-1044.

Frenette, Marc, David A. Green, and Garnett Picot (2004). "Rising income inequality in the 1990s: An exploration of three data sources," Statistics Canada, Analytical Studies Research Paper Series, No. 219.

Fullerton, Don and Gilbert E. Metcalf (2002). "Tax Incidence" in Handbook of Public Economics, Vol. 4., A. J. Auerbach and M. Feldstein, eds., Amsterdam: Elsevier.

Gottschalk, P. and T. Smeeding (1997). "Cross-National Income Inequality," Journal of Economic Literature, Vol. 35.

Johnson, Susan (1995). "More evidence on the effect of higher unemployment on the Canadian size distribution of income," *Canadian Public Policy*, Vol. 21, No. 4, pp. 423-428.

OSFI (2004). "Actuarial Report (21st) on the Canada Pension Plan as of December 31st, 2003," Office of the Superintendent of Financial Institutions, Ottawa.

Osberg, Lars (1997). "Economic growth, income distribution and economic welfare in Canada, 1975-1994," *North American Journal of Economics and Finance*, Vol. 8, No. 2, pp. 153-166.

Osberg, Lars (2003). “Long run trends in income inequality in the United States, UK, Sweden, Germany and Canada: A birth cohort view,” *Eastern Economic Journal*, Vol. 29, No. 1, pp. 121-141.

Piketty, Thomas and Emmanuel Saez (2003). “Income inequality in the United States, 1913-1998,” *Quarterly Journal of Economics,* Vol. 118, No. 1, pp. 1-39.

Rashid, Abdul (1998). “Family income inequality, 1970-1995,” *Perspectives on Labour and Income,* Vol. 10, No. 4, pp. 11-17.

Saez, Emmanuel and Michael R. Veall (2005). “The Evolution of High Incomes in Northern America: Lessons from the Canadian Evidence,” *American Economic Review*, Vol. 95, No. 3, pp. 831-849.

Sen, Amartya K. (2000). “Social Justice and Distribution of Income,” in Handbook of Income Distribution, Vol. 1, A.B. Atkinson and F. Bourguignon, eds., New York: North-Holland.

Wolfson, Michael and Brian Murphy (1998). “New views on inequality trends in Canada and the United States,” *Monthly Labor Review,* April, pp. 3-23.