IMES DISCUSSION PAPER SERIES

Dimensions of Inequality in Canada

Matthew Brzozowski, Martin Gervais,
Paul Klein, and Michio Suzuki

Discussion Paper No. 2009-E-2

IMES

INSTITUTE FOR MONETARY AND ECONOMIC STUDIES

BANK OF JAPAN

2-1-1 NIHONBASHI-HONGOKUCHO
CHUO-KU, TOKYO 103-8660
JAPAN

IMES Discussion Paper Series 2009-E-2
January 2009

Dimensions of Inequality in Canada

Matthew Brzozowski*, Martin Gervais**, Paul Klein***, and Michio Suzuki****

Abstract

In this paper, we document some features of the distribution of income, consumption and wealth in Canada using survey data from many different sources. We find that wage and income inequality has increased substantially over the last 30 years, but that much of this rise was offset by the tax and transfer system. As a result, the rise in consumption inequality has been relatively mild. We also document that wealth inequality has remained fairly stable since 1999. A comparison of our results - obtained using confidential data - suggests that while some aspects of inequality are well captured by publicly available data, others paint a drastically different picture.

Keywords: Income Inequality; Consumption Inequality; Wealth Inequality

JEL classification: D12, D31

* Assistant Professor, Economics, Atkinson Faculty of Liberal and Professional Studies, York University (E-mail: brzozows@yorku.ca)
** Reader, Economics Division, University of Southampton (E-mail: gervais@soton.ac.uk)
*** Associate Professor, Department of Economics, University of Western Ontario (E-mail: pklein2@uwo.ca)
**** Economist, Institute for Monetary and Economic Studies, Bank of Japan (E-mail: michio.suzuki@boj.or.jp)

We would like to thank Jim Davies and David Green for their many comments and suggestions on this project, as well as participants at the Philadelphia conference on Heterogeneity in Macroeconomics and a seminar held at the Bank of Japan for useful comments. Gervais and Klein gratefully acknowledge financial support from the Social Sciences and Humanities Research Council of Canada. Views expressed in this paper are those of the authors and do not necessarily reflect the official views of the Bank of Japan.

1 Introduction

In this paper, we document some salient facts concerning the distribution of wages, hours worked, income, consumption and wealth in Canada. We use a variety of data sources to do so, as described in section 2. In general, these datasets have two access levels, one publicly available (Public-Use Files, or PUF) and one only available to Canadian researchers through Research Data Centres (RDC) administered by Statistics Canada. While our main conclusions in section 3 are drawn from RDC data, we discuss in section 4 the extent to which noise introduced by Statistics Canada to protect individuals' identity distorts various measures of inequality.

To put things into context, Table 1 displays aggregate statistics for the Canadian economy over the 1976 to 2005 period, the years covered by most of our micro data. Over those 30 years, the population of Canada grew at an average annual rate of 1.1 percent, from 23 to over 32 million. Meanwhile GDP grew at an average annual rate of 2.8 percent, so GDP per head grew at 1.7 percent per year. Both figures are slightly lower than in the U.S., where GDP and GDP per head grew at average rates of 3.0 and 2.0 percent per year, respectively. Similarly, GDP per hour only grew at an average annual rate of 1.2 percent over that period in Canada, compared to 1.5 percent in the U.S.

From 1976 to 2005, the Canadian economy experienced two recessions: one at the beginning of the 1980's and one at the beginning of the 1990's. Unlike the U.S., Canada did not experience a recession at the turn of the century. The importance of these episodes for various measures of economic inequality comes from the fact that recessions typically affect various segments of the population differently. Indeed, both recessions were accompanied by a rise in wage and income inequality, mainly because individuals and households at the bottom of the distribution were more likely to be affected by the downturn than those at the top of the distribution; this perspective is supported by the data.

Several tax reforms were carried out in Canada in the late 1980's and early 1990's. Overall, the picture that emerges is one of moderate and gradual reform rather than anything dramatic. There is not much evidence to support the view that changes in public policy were a major factor in causing changes in inequality. Perhaps the most

Table 1: Aggregate Statistics: 1976–2005

from to	1976 2005	1976 1979	1980 1984	1985 1989	1990 1994	1995 1999	2000 2005
Population	1.10	1.05	1.09	1.35	1.15	0.92	1.02
GDP	2.83	3.81	2.08	3.51	1.52	3.81	2.48
GDP Per Head	1.73	2.76	0.99	2.15	0.37	2.89	1.46
GDP per Hour	1.20	0.46	1.77	0.10	1.80	1.64	1.20
Employ't Rate	0.67	0.63	0.64	0.67	0.66	0.67	0.71

Note: The first four rows display the average growth rate of the variable. Employment rates are average employment rates for the 20–69 age group over the relevant period, from the April Labour Force Survey.

important reform was the replacement in 1991 of the manufacturers sales tax (MST) with the Goods and Services tax (GST), which is a federal value added tax. However, from a distributional point of view, the income tax reform of 1987 (implemented in 1988) is perhaps more interesting. As a result of this reform, households at all levels of income experienced somewhat lower taxes. As reported in Maslove (1988), the average effective income tax rate was reduced from 19.6 percent to 18.6 percent. Households in the top percentile experienced the largest reduction, from 33.1 percent to 30.6 percent. However, the 1987 reform resulted in an income tax schedule not very different from that in 1984. In fact, for most income levels (measured in constant dollars), average effective tax rates were slightly *higher* in 1988 than they were in 1984. For all households, it rose from 17.4 percent to 18.6 percent. The exception was the top percentile, which experienced a fall from 32.5 percent in 1984 to 30.6 percent in 1988, and the bottom two deciles which received larger tax refunds in 1988 than in 1984. Overall, the distributional impact of this tax reform was minimal.

Income inequality over the last 30 years or so has risen quite substantially in Canada. Wage inequality, as measured by the Gini coefficient, rose by almost 30 percent. Similarly, family earnings inequality increased by about 25 percent. Since 1980, while the Gini coefficient of before-tax total family income increased by 20 percent, that of after-tax income only increased by 13 percent, indicating that some of the changes in income inequality are absorbed by the tax and transfer system. Furthermore, the level of pre-tax income inequality is over 20 percent higher than that of post-tax income inequality on average. Similarly, disposable income inequality is close to 20 percent higher than consumption inequality, whose Gini coefficient increased by about 5 percent since the early 1980's.

Using income data from the Census, Frenette et al. (2007) find that the 1980's were characterized by a strong rise in before-tax income inequality, but that most of that rise was absorbed by the Canadian tax and transfer system, with the result that after-tax income inequality remained constant. They also report that while before-tax inequality also increased in the 1990's, this time the tax and transfer system did not offset the rise, and after-tax income inequality also rose in that decade, albeit not to the same extent. This broad picture emerges from our results as well.

Now the reason why Frenette et al. (2007) use Census data rather than survey data is mainly because they doubt the validity of results obtained though survey data. Indeed, Frenette et al. (2006) show that income inequality trends in survey data are inconsistent with both Census and tax data. However, perhaps because of these inconsistencies, Statistics Canada implemented a revision to the weights in survey data, mainly in order for the surveys to be consistent with information (from the Canadian Revenue Agency) on wages and salaries. The details of this revision and its impact on survey income data can be found in Lathe (2005). Although such a revision almost necessarily entails distorting other aspects of the data (such as employment to population ratios) and brings about a break in the series because the revision only goes back to 1990, it seems to have brought income inequality results closer to those that emerge from Census data. Indeed, a look at income inequality from Public-Use files, which for the most part still feature the pre-revision weights, confirms that the choice of Frenette et al. (2007) to go to Census data was warranted. Our results from Public-Use Files look much like theirs from survey data, and when we compare those results to the results emerging from the revised weights, we are inclined to support their view that the old survey weights led to a distorted depiction of the evolution of income inequality in Canada.[1]

While using Census data may have advantages over survey data, it also raises major concerns, of which Frenette et al. (2007) are of course well aware. The most

[1]Our results are not directly comparable, however, as they use a very different sample and a different way to equivalize earnings.

serious drawback of using Census data is that it contains no information on taxes. To remedy this problem, Frenette et al. (2007) impute taxes using administrative tax data. Essentially, they regress taxes paid on income and characteristics in tax data, and use it to predict taxes for each Census family. Another problem with Census data is that these data are only collected every 5 years, making it difficult to capture events that occur at higher frequencies, such as recessions. For these reasons, we prefer to use the survey data with the revised weights rather than the Census data.

2 Sources of Data

2.1 Datasets

As is typically the case, Canada does not have a single survey from which information on consumption, income and wealth is available. In fact, the best information on each of these variables has a separate source, and that source has changed recently, so 6 different datasets need to be explored. Furthermore, some of these datasets have two access levels, one publicly available and one only available to Canadian researchers within the confines of Research Data Centres (RDC) administered by Statistics Canada. Below we briefly outline our main data sources, with details contained in Appendix A.

The main survey for consumption in Canada currently is the *Survey of Households Spending* (SHS). This Survey started in 1997 and is available yearly until 2005, its latest release. The SHS replaced the *Survey of Family Expenditures* (FAMEX), which ran sporadically from 1969 until 1996. Both surveys contain data on consumption, income, and standard characteristics (other than education from 1997 to 2003) at the *household* level, or something close to a household (similar to consumer units in the Consumer Expenditure Survey for the U.S.). See Appendix A.1 for details.

The main survey for income currently is the *Survey of Labour and Income Dynamics* (SLID), which offers yearly data since 1993. This survey consists of overlapping samples interviewed in six consecutive years, with new waves introduced in 1993, 1996, 1999, 2002, and 2005. In any given year since 1996, then, there are two waves of data available. Although SLID contains a panel dimension, it replaced the *Survey* *of Consumer Finances* (SCF) as the main source of cross-sectional income data in 1996.[2] Through SCF and SLID data, a consistent dataset on income, hours worked, a measure of wage rates, with standard characteristics can be constructed from 1977 to 2005, with a two-year overlap between the two surveys.[3] See Appendix A.2 for details.

Wealth data is hard to come by in Canada. The SCF had a wealth supplement in 1977 and 1984, called the ADSCF, and the *Survey of Financial Securities* (SFS) was introduced in 1999 and ran again in 2005. Because of inconsistencies between the two Surveys, we only use the SFS in this paper.[4] The unit of analysis in the SFS is the *economic family*. This survey offers standard balance sheet data from which net worth can be constructed, as well as income and some characteristics of the economic family and the head of the family. See Appendix A.3 for details.

2.2 RDC Data versus Public-Use Files

While the data that we made available emanate from public-use files, the results presented in the next section come mainly from proprietary data which Statistics Canada occasionally grants access to through Research Data Centres (RDC) located at various institutions throughout the country. The main advantage of using RDC data is that these data are not modified in any way or, if they are, a flag identifying observations that have been modified is supplied in RDC data. In Section 4 we investigate the extent to which using public-use files can replicate results obtained from the original, RDC data.

The severity of distortions introduced in public-use data varies across surveys. At one extreme, public-use files for consumption data (FAMEX/SHS) are for our purposes identical to RDC data. At the other extreme the panel dimension of SLID data is only available through RDC.[5] But for income and wealth data, public-use files

[2] Because only the first wave of the survey is available from 1993 to 1995, the sample size of SLID is relatively small, and thus the SCF is deemed to contain better information for these years. Accordingly, we only use SLID data from 1996 on.

[3] SCF data was also collected in the early 1970's. We chose not to use years prior to 1976 for consistency reasons.

[4] Morissette et al. (2002) outline ways to make the ADSCF compatible with the SFS.

[5] The Cross-National Equivalent Files (CNEF) prepared by the Department of Policy Analysis and Management at Cornell University contain a version of the panel aspect of SLID. See

have some characteristics removed (SCF/SLID) and data have been shuffled (SLID) to preserve the identity of some of the respondents, mainly households at the top of the distribution. Finally, our public-use wealth data (SFS) suffers from the typical top-coding problem that work with many survey data must confront.

2.3 Survey Weights

The sampling frames of all the surveys used in this study are based on the Labour Force Survey (LFS), the equivalent of the Current Population Survey in the U.S.. The LFS is a monthly survey which covers the civilian, non-institutionalized population 15 years of age and over. The coverage rate of the LFS is over 98 percent of the Canadian population aged 15 or over.[6] Since all other surveys consist of a sub-sample of the LFS, their coverage rate is similar.

The LFS consists of a rotating panel in which households remain in the sample for 6 consecutive months. The total sample size of the LFS, which comprises 6 panels, has varied over the years but has been around 45,000 households since 1995. This sample is obviously chosen to be representative of the population. Initially, each individual in the sample is given a weight equal to the inverse of the probability of selection. These weights are then adjusted to account for non-response and to match census estimates for various age-sex groups by province and major sub-provincial areas. All other surveys use the LFS weights as their initial weights, to which adjustments are then made (more on this below).

Our income surveys are typically conducted as a supplement to the April (SCF) or January (SLID) LFS, much like the March Supplement to the CPS. The sub-sample of the SCF consists of 4 of the 6 rotating panels in the LFS at the time of the survey. Similarly, each SLID wave consists of 2 of the 6 rotating panels of the LFS, hence SLID has approximately the same number of observations as the SCF since 1996. Since both SCF and SLID consist of sub-samples of the LFS, the weights for these surveys are initially based on LFS weights. These weights are first adjusted for non-response in SCF/SLID (but not for LFS non-response, as this has already been taken

http://www.human.cornell.edu/che/PAM/Research/Centers-Programs/German-Panel/cnef.cfm.

[6]Although the survey is conducted nationwide, i.e. both in the provinces and the territories, territorial LFS results are not included in national estimates.

care of in the LFS weights), that is, the weight of non-respondents are distributed to respondents. The non-response adjusted weights are then calibrated to ensure that estimates on relevant characteristics of the population (age, sex, province, family size, and household size) match up with census data (or projections from a recent census).

Because of contradicting evidence on the degree and evolution of income inequality coming from different sources of data (tax data, census data, and survey data), as evidenced for instance by the work of Frenette et al. (2006), Statistics Canada recently revised their strategy to calibrate the survey weights. This new strategy includes wages and salary as one of the targets, in addition to standard characteristics. The targets themselves come from T4 slips (equivalent to W4 slips in the U.S.), which are employer remittances to Canada Revenue Agency (formerly Revenue Canada) for payroll tax purposes. This new strategy to impute weights is currently used for all surveys that we use in this study, and has been implemented retroactively to 1990 for SCF/SLID, and to 1996 for FAMEX/SHS. All SFS data uses this new methodology as well (see Lathe (2005) for more details).

The result of this new calibration strategy, documented in Rea and Greenlee (2005), is that more weight is given to observations with low/zero wages and salary as well as to observations with high wages and salary. The reason is that a disproportionate fraction of respondents are from the middle of the wages and salary distribution. Practically, as far as SCF data is concerned, this revision implies a higher level of inequality in wages and salary, which caries over to pre-tax income inequality and to a lesser extent to after-tax or disposable income inequality at the individual level. Ultimately, this revision had a similar impact on family/household level income inequality. This revision and its impact should therefore be kept in mind when interpreting the evolution of inequality over time in the rest of the paper.

While the new strategy is also used for consumption (FAMEX/SHS) and wealth (SHS) data, no impact can be detected either because all years of data have been revised (SFS) or because the survey was not conducted sufficiently frequently (FAMEX) to produce a noticeable break. It should be noted, however, that while the consumption surveys are also based on the LFS, 10 to 15 percent of SFS samples is drawn from geographic areas in which a large proportion of households had what was defined as “high-income”, while the rest of the sample is based on LFS sampling. In other words, much like the Survey of Consumer Finances in the U.S., the SFS over-samples the rich in order to improve wealth estimates.

2.4 Sample Selection and Statistics

The details of our samples are contained in Appendix B. Briefly, we restrict our samples to households whose head is at least 25 years or age and no older than 60. We also exclude observations with missing characteristics (gender, education, family type, marital status, and province of residence). For income data, we exclude any observation whose income was imputed, missing, or non-positive, as well as observations with a wage rate less than half the minimum wage. We also exclude observations with zero or missing hours worked. In order to exclude individuals with dubious hours worked, we discarded individuals whose wage is greater than 100 dollars *and* who report less than 100 hours worked. For consumption data, we exclude part-year households and households with zero food consumption.[7,8] In addition, since the territories were not covered by FAMEX surveys, we exclude households from the territories in SHS as well. Finally, for wealth data we exclude outliers and families with missing wealth information. Table 2 compares our samples along broad dimensions for three different years. Except for particularly low fractions of people with a university degree in FAMEX/SHS, our datasets are fairly consistent with one another.

3 Dimensions of Inequality

3.1 Aggregates

Figure 1 display aggregate income as reported in the System of National Accounts (SNA) as well as its implied counterparts constructed from survey data. The break in 1990 for SCF data is likely the result of the weight revision applied retroactively to 1990.

[7]A full-year household is defined as having at least one full-year member, i.e. a member who was in the household for the entire year; a part-year household is composed entirely of part-year members.

[8]For future versions, we will adopt the dollar-a-day criterion, whereby households with less than $350 of food expenditures will be discarded. Note that this is the case for only 73 observations in our entire data.

Table 2: Summary Statistics

	1986		1996			2005		
	FAMEX	SCF	FAMEX	SCF	SLID	SHS	SLID	SFS
Age	40.76	39.40	41.92	40.75	40.13	43.27	42.05	43.01
Family Size	3.10	2.96	2.87	2.85	2.77	2.85	2.54	2.64
% University	0.14	0.20	0.19	0.25	0.22	0.30	0.28	0.26
With Spouse	0.71	0.71	0.67	0.67	0.65	0.67	0.60	0.62
% Immigrant	0.19	0.19	0.21	0.18	0.16	n/a	0.16	0.20
Observations	7151	13850	7201	14580	12807	9035	10250	<3868

Notes. The 2005 SHS only provides categorized data on the age of the head of the household. We assign each household head the middle value of the age range to which the person belongs, and report the average over the assigned ages.

While the gap between SNA and SCF data increased over the SCF years, it is reassuring that it has been shrinking under SLID. In the last survey year available, as shown in the first panel of Table 3, average disposable income was over 90% of the SNA figure, either from SHS or SLID data.

At first glance, Figure 2 seems to suggest that the situation is not as bright for consumption, as the *nominal* gap between SNA and survey data has been rising moderately over the last 10 years. However, in real terms the gap has not been rising over the SHS years (since 1997), sitting between 80 and 82 percent (see second panel of Table 3). Furthermore, the ratio of non-durable consumption to income from SHS is consistently around 5 percentage points lower than the same ratio from SNA data, as shown in the third panel of Table 3 for 2005. Thus, the situation in Canada is not as dire as in the U.S., where the gap has been rising consistently over the last 25 years. For example, Slesnick (2001) reports that the ratio of per capita consumption in the Consumer Expenditures Survey (CEX) to the NIPA estimate of Personal Consumption Expenditures was 0.68 in 1980, 0.61 in 1990, and 0.56 in 1995. Similarly, Battistin (2003) finds that the ratio of non-durable expenditures in CEX to the NIPA estimate fell from around 80 percent in 1985 to around 65 percent in the late 1990's.[9]

[9]See Crossley and Pendakur (2006) for a discussion of the quality of expenditure survey data in Canada.

Figure 1: Mean Nominal Disposable Income

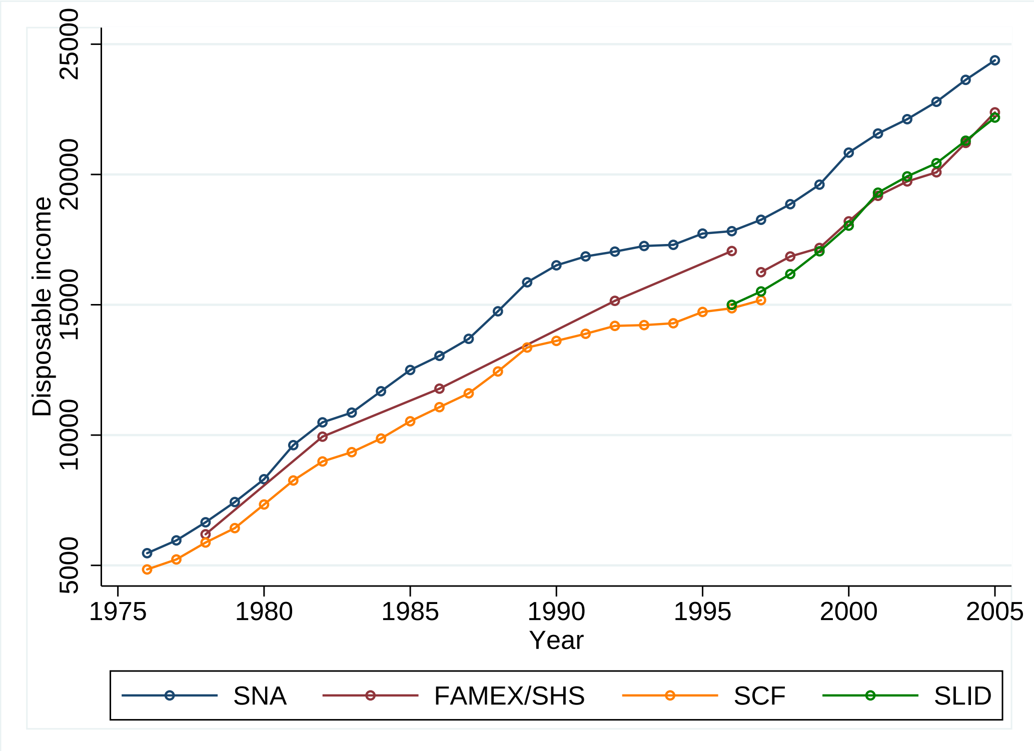
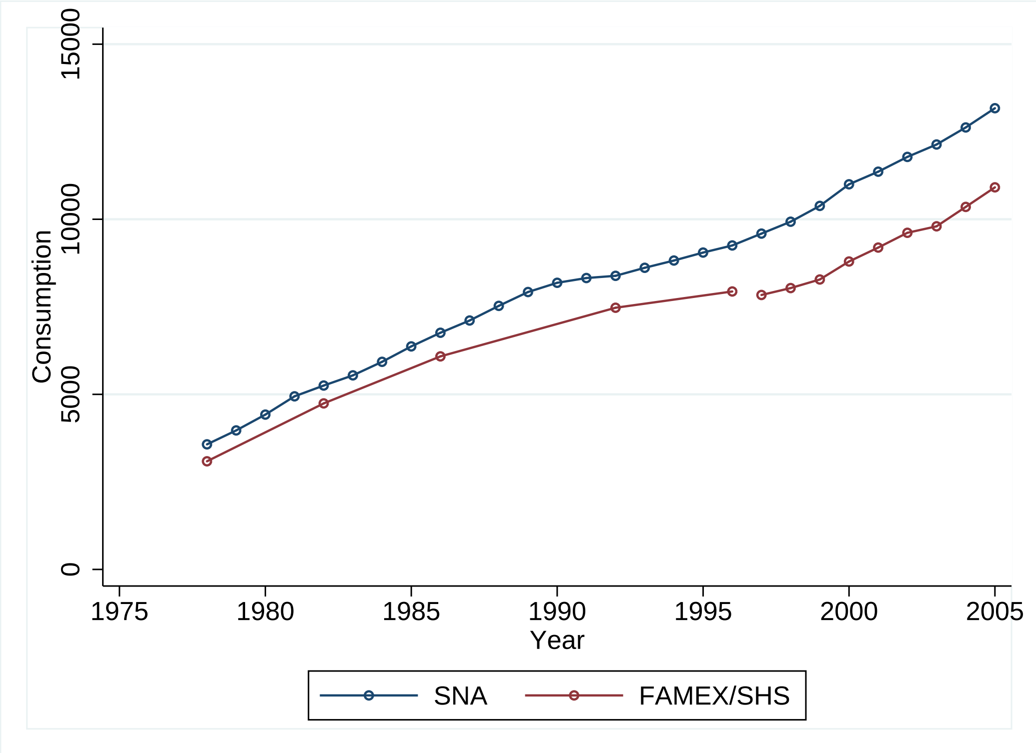


Figure 3 displays employment to population ratios at the aggregate level and from survey data over the sample period. It should be noted that the aggregate employment series also comes from a survey—the monthly Labour Force Survey (LFS). Also note that while most questions in our income surveys pertain to the year prior to the survey year, employment questions in SCF data pertain to the current month (typically April). By contrast, employment information in SLID is available for every month of the previous year. Furthermore, as mentioned in the previous section, both SCF and SLID surveys are conducted as a supplement to the LFS, that is, they consist of sub-samples of the LFS.[10] It is therefore surprising to see significant discrepancies between the two series. The discrepancies emanate from two distinct sources. First,

[10]Since the SCF is typically conducted as a supplement to the April LFS, both series in Figure 3 refer to the month of April. For 1976 and 1983, SCF data pertained to the month of May, and so we chose that month for the aggregate series as well for these two years. SLID offers monthly labor force data, so we also used the month of April for the years covered by SLID. However, to be consistent, our next version will use the January (of the interview year) employment status, which is compatible with the LFS survey.

Figure 2: Mean Nominal Non-Durable Consumption



the SCF had unusually small samples in 1977, 1979, 1981 and 1984, which explains the anomalies prior to 1990.[11] The post 1990 discrepancies can be explained by the 2003 historical revision of the weights discussed in the previous section. As a result of this revision, which was applied retroactively to 1990, higher weights were assigned to low/no earnings individuals as well as high earnings individuals and had the effect of increasing substantially the fraction of the population not in the labor force. This is reflected in Figure 3 as a low employment to population ratio.[12]

[11] Survey years 1977 and 1979, which contain data pertaining to 1976 and 1978, have other problems—in particular hours worked is missing—so we dropped these years for the rest of the analysis.

[12] This is confirmed by a similar figure (not shown) constructed using weights available prior to the revision, which shows essentially no discrepancy between the SCF and LFS numbers over the 1990's.

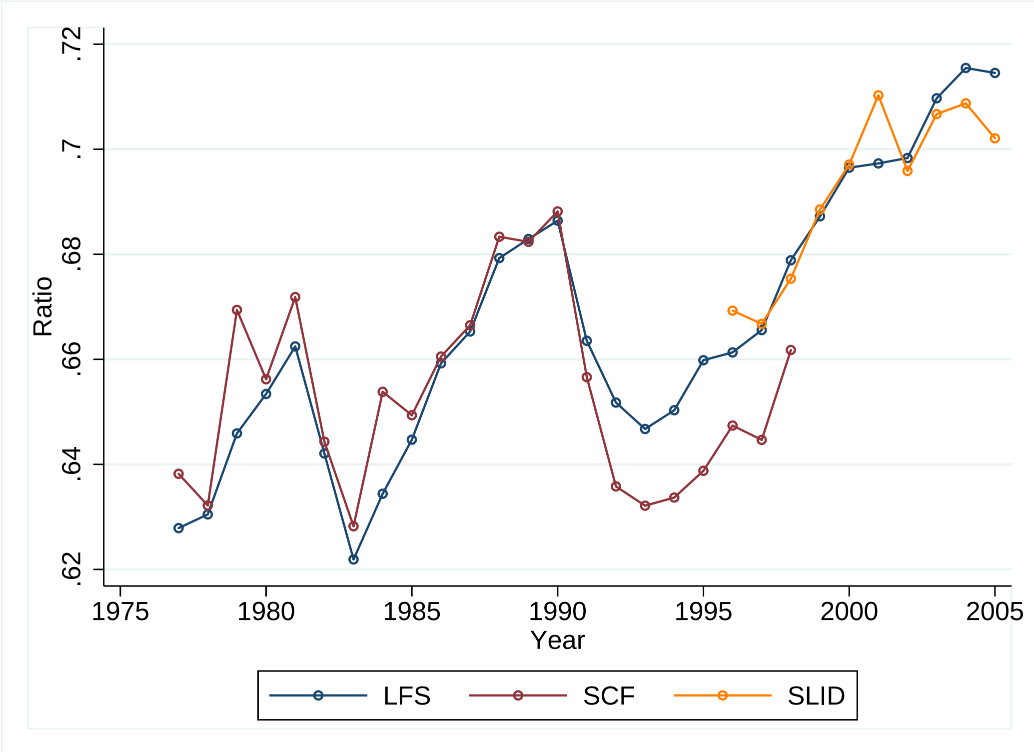
Table 3: Survey to Aggregate Ratios (%)

	1978	1986	1996	2005
Mean Disposable Income				
FAMEX/SHS	93.1	90.3	95.7	91.8
SCF	88.3	84.9	83.4	n/a
SLID	n/a	n/a	84.1	91.0
Mean Non-Durable Expenditures				
FAMEX/SHS	86.5	90.0	85.8	82.8
Ratio of Non-Durable Expenditures to Disposable Income				
FAMEX/SHS	49.9	51.6	46.5	48.8
SNA	53.7	51.8	51.9	54.0

3.2 Individual Level Inequality

Figure 4 displays four measures of wage inequality. The wage is defined as average hourly earnings, and earnings include wages and salaries as well as the labor share (62%) of income from self-employment. The first panel shows that wage inequality, as measured by the variance of log wages, increased substantially in the late 1970's and early 1980's, and more modestly in the early 1990's and around 2000. The fourth panel of that figure shows that a similar pattern emerges when wage inequality is measured by the Gini coefficient. Over the entire sample, the variance of the log has increased by 48 percent and the Gini coefficient increased by 26 percent. It is interesting to note that while high wage earners have been consistently gaining on the median over the sample period (upper right panel), low wage earners falling behind the median is the main reason behind the increase in wage inequality in the late 1970's and early 1980's (lower left panel). More precisely, the 50–10 percentile ratio increased by 13 percent from 1979 to 1983, but only by 4 percent from 1990 to 1992. This ratio has been essentially constant since 1992. On the other hand, the 90–50 percentile ratio only increased by 2 percent form 1979 to 1983 and actually *fell* by 2 percent from 1990 to 1992. However, the 90–50 ratio increased by about 15 percent from 1992 to 2005. Confining our attention to SLID data, the increase in the 90-

Figure 3: Employment to Population Ratio

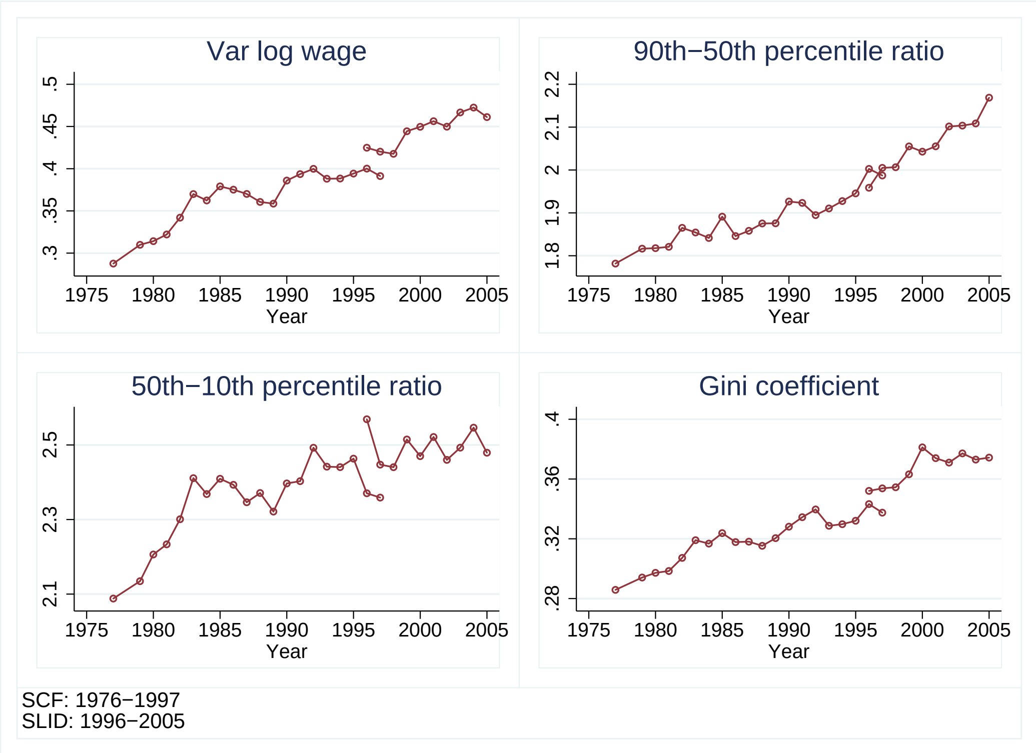


50 percentile ratio was 11 percent from 1996 to 2005.[13] Notice also that the Gini coefficient has remained roughly constant since 2000; as is well known, the degree of inequality as measured by the Gini coefficient is mainly driven by the middle of the distribution.

The bottom-right panel of Figure 5 shows that the above pattern of wage inequality is very similar to the pattern of residual wage inequality—inequality unexplained by observables—suggesting that most of the increase in wage inequality cannot be explained by observables. The top-left panel of Figure 5 shows the education premium, defined as the the average wage of males with at least a university certificate, a diploma, or a bachelor's degree, relative to the average wage of males without any such degrees. Interestingly, the education premium remained fairly constant from the beginning of the sample until the mid 1990's. Since then, the education premium rose

[13]This is consistent with Saez and Veall (2005), who report that the income share of the top 5 percent increased from around 22 percent in the mid 1990's to about 29 percent in 2000.

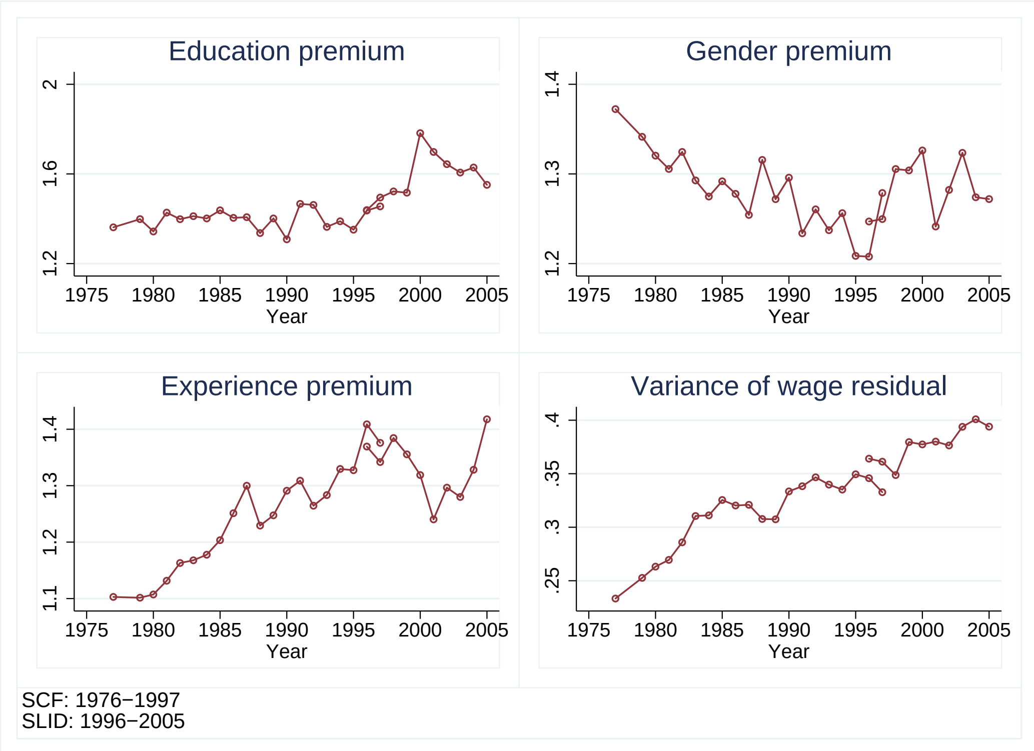
Figure 4: Basic Inequality in Wages



from an average of around 40 percent from the beginning of the sample until 1995 to just below 60 percent in 2005. Meanwhile, the fraction of the population with a university degree increased monotonically from 14 percent in 1977 to 28 percent in 2005. These patterns are quite different from those observed in the U.S., where the college premium *and* the supply of college graduates have been rising simultaneously since the late 1970's (Heathcote et al. (2008)).

The top-right panel of Figure 5 displays the gender premium, that is, the average wage of males relative to that of females. The gender premium narrowed considerably until the mid 1990's, from 37 percent in 1977 to just over 20 percent in 1995, but it has failed to decline in more recent years, averaging around 27 percent over the SLID years. As a result, the gender gap in Canada is now close to that in the U.S., but the drop in the gender premium since the late 1970's was more spectacular in the U.S., where the wage premium was at a much higher level (65 percent) than in Canada (37 percent) in the late 1970's (Heathcote et al. (2008)). The bottom-left

Figure 5: Wage Premia and Wage Residual



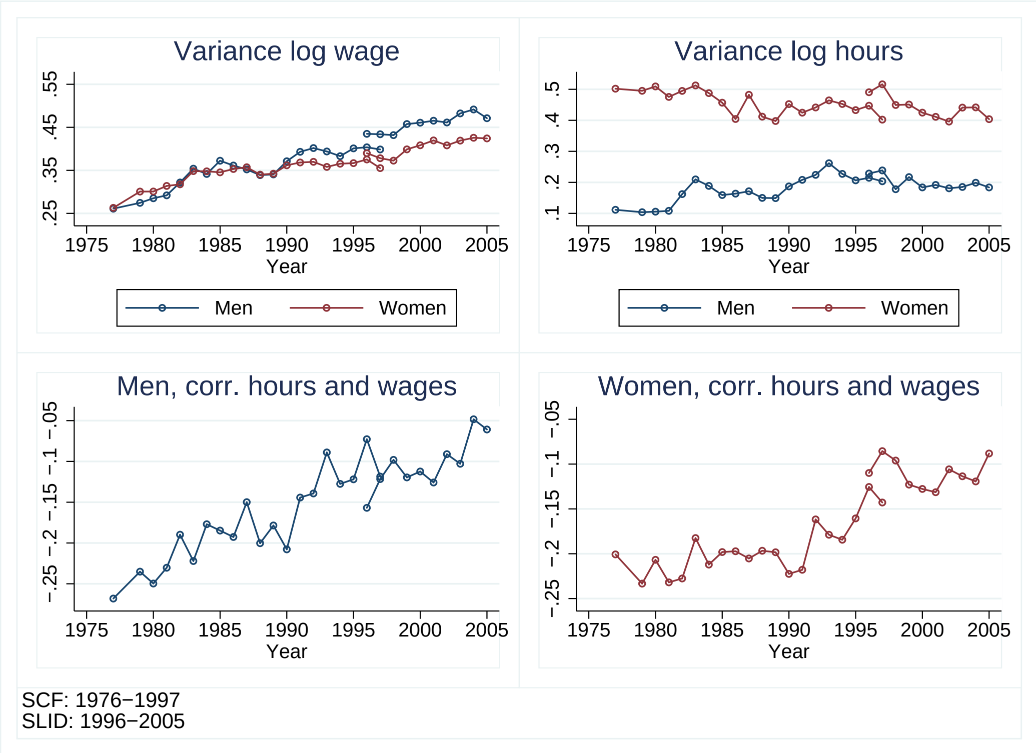
panel suggests that the reverse has happened to the experience premium, measured as the average wage of males aged 45–55 relative to those between 25 and 35 years of age. This premium increased until the mid 1990's and declined sharply during the dot-com years of the late 1990's. However, the experience premium came back up over the last 5 years of the sample.

The top-left panel of Figure 6 shows that, qualitatively, the evolution of wage inequality for both men and women mimics that of the entire population (top-left panel of Figure 4). The gap between male and female wage inequality, which opens up in the 1990's, may be an artifact of the retroactive revision of survey weights.[14] The last 10 years of SLID data suggests that the variance of log wages for men is around 13 percent higher than that of women.

Inequality of hours worked, as shown in the top-right panel of Figure 6, paints a

[14]Recall that as a result of the weight revision, more weight was given to observations with low or no income, which may have affected males and female wages differently.

Figure 6: Inequality of Labor Supply



drastically different picture for men and women. First, as one would expect, the level of hours inequality is much higher for women than for men throughout the sample. Second, inequality of hours worked for men is much more responsive to business cycle fluctuations than that of women: the variance of log hours almost doubled from 1981 to 1983 for men while it only increased by 7 percent for women; similarly, the variance of log hours from 1990 to 1993 increased by 40 percent for men and by 17 percent for women. It is also interesting to note that over the entire sample period, the variance of hours worked for women trended down, while that of men has been essentially stable since the recession of the early 1980's. Finally, the bottom panels of Figure 6 shows that the correlation between hours worked and wages has been rising throughout the sample period, approaching zero from below.

3.3 Family Level Inequality

This section pertains to inequality at the *economic family* level, where a family is defined as a group of individuals living at the same address and related by blood, marriage or adoption.[15]

The next two figures display various measures of family earnings inequality, where family earnings is the sum of the family members' wages and salaries plus the labor part (0.62) of self-employment income. Figure 7 shows that observables (family type, education, age and region of residence) do not explain much of the trend in earnings inequality, despite the fact that the variance explained by education has generally trended up over the sample period. On average, observables explain around 20 percent of the variance of equivalized earnings. All measures of inequality in Figure 8 show a large rise over the two recessions: the variance of log equivalized earnings increased by 25 percent from 1981 to 1983 and by 24 percent from 1990 to 1993. Both episodes were marked by the median gaining on the poor and the rich gaining on the median. The former is more pronounced than the latter for wages, but, interestingly, not for earnings. Notice also that inequality does not seem to revert back down to their pre-recession levels, except for the 90 to 50 percentile ratio after the first recession. Indeed, the level of the Gini kept increasing over the last 10 years of the sample. Over the entire sample, the Gini coefficient increased by about 25 percent and the variance of log earnings by 57 percent, the latter rise occurring entirely from 1977 to 1993.

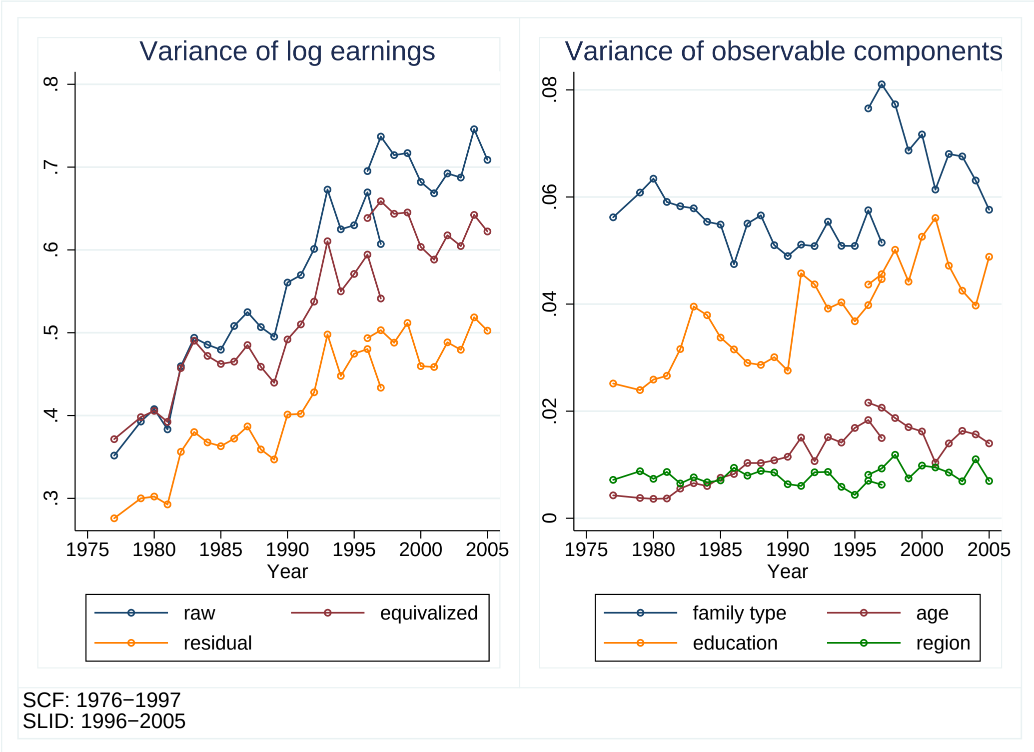
Figure 9 shows the variance of log income for many measures of income as well as wage and earnings of household heads. Notice that the distance between wage of head inequality (measure 1) and that of pre-government earnings of head (measure 2) has been rising over time. This is because the variance of log hours has been relatively stable and the correlation between wages and hours has been increasing over time.[16] Also note that the distance between measures 2 and 3 (pre-government earnings of the head and of the family), which was growing until the mid-1990's, suggests that there is substantial family insurance.

When comparing the variance of log disposable income and the variance of various

[15] Economic families may differ from households, defined as a person or a group of persons living at the same address.

[16] Note that more than 80% of the household heads are male in our benchmarks sample.

Figure 7: Earnings Inequality and its Decomposition

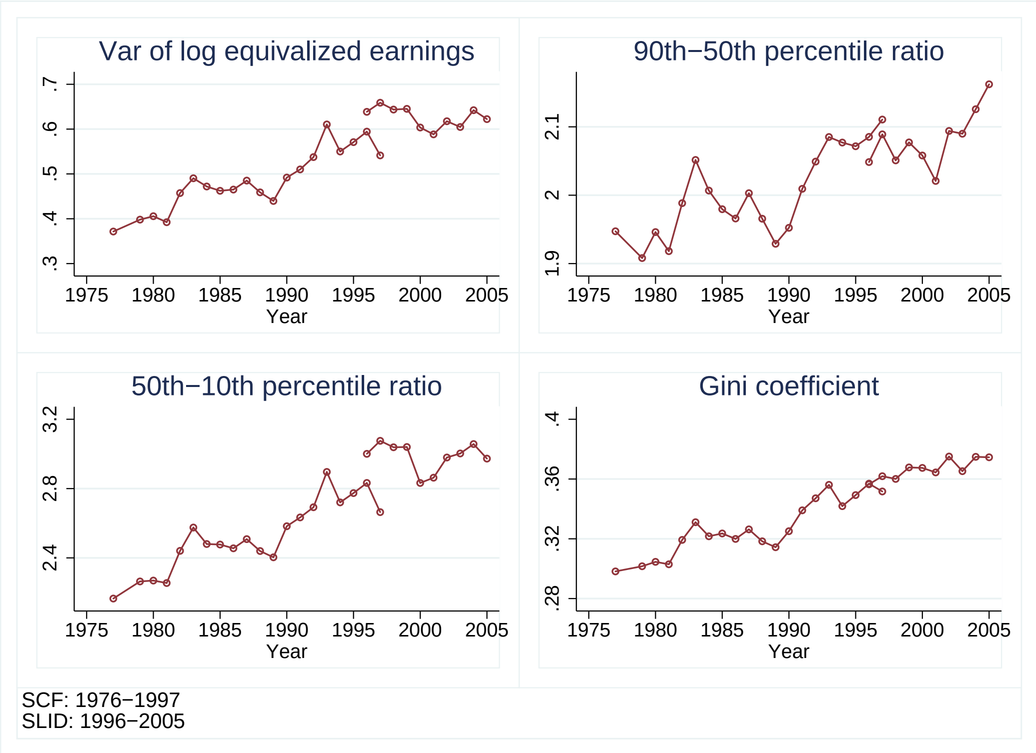


measures of pre-government earnings and income, a number of features stand out. First, disposable income exhibits much less inequality and the degree of inequality is much less variable than that of pre-government earnings/income. This suggests both that Canadian policy has been and remains redistributive, and that it smooths cyclical shocks to pre-tax income inequality.

On the other hand, there are noticeable changes over time. The first is that the correlation between pre-government inequality and post-government inequality seems to have disappeared. Post-government inequality exhibits very small cyclical fluctuations after 1996 and those movements that do occur are uncorrelated with the corresponding movements in pre-government earnings and income. Finally, while after-tax inequality remained fairly stable over the years covered by the SCF, it increased by about 10 percent over the SLID sample years.[17]

[17]This suggests that public policy has become gradually less redistributive in its effects, though not necessarily in its intent, since 1996.

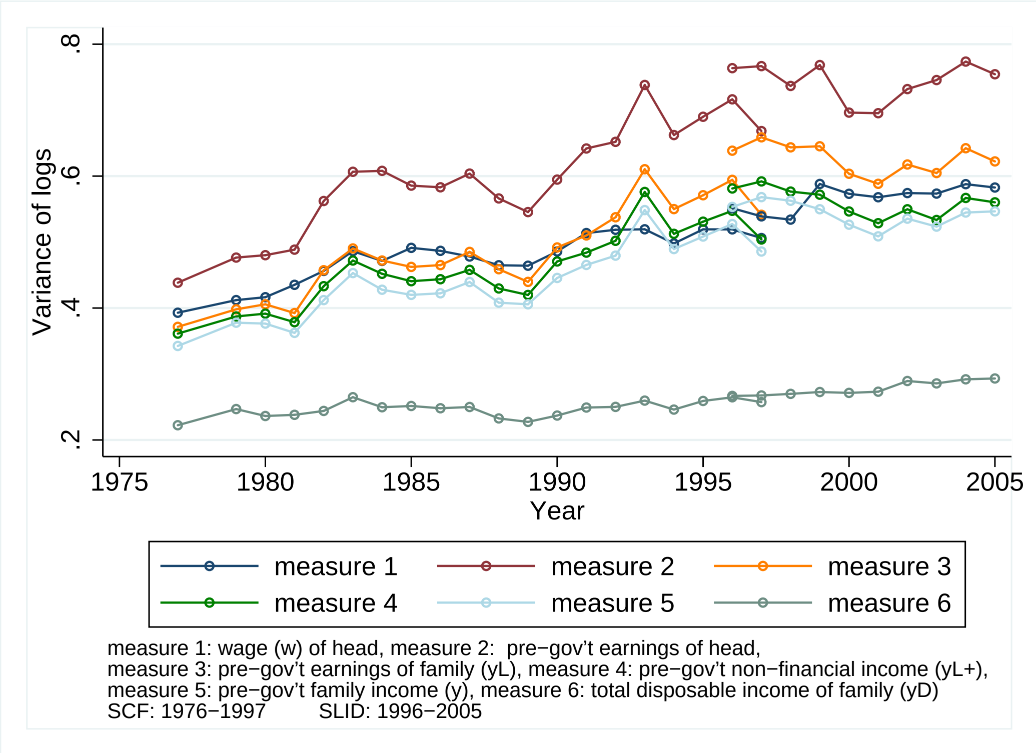
Figure 8: Basic Inequality of Equivalized Earnings



The scale in Figure 9 masks changes in disposable income inequality which are easier to discern in Table 4. This table is also meant to compare our results to those of Frenette et al. (2007), who use income data from the Census to characterize the evolution of income inequality over the 1980–2000 period. They conclude that most of the strong rise in before-tax income inequality over the first decade was absorbed by the tax and transfer system, leaving after-tax inequality unchanged. However, the equally strong increase in before-tax inequality in the 1990's was accompanied by an increase in after-tax inequality, albeit not of the same size. Our results in Table 4 reinforce that conclusion by reporting increases of much larger magnitude than theirs. As we will see in Section 4, the weight revision was instrumental in obtaining these results.[18]

[18]The noise introduced by Statistics Canada to protect individuals' identity could also play a role in explaining the difference between the results from of this Section relative to those from Section 4. However, results from Frenette et al. (2006) suggest that the weight revision was the main factor.

Figure 9: From Wages to Disposable Income



3.4 Consumption Inequality

We now turn our attention to consumption inequality, where consumption refers to non-durable consumption.

Figure 10 shows that consumption inequality exhibits a rising trend. In the "raw" data, where no adjustment has been made for family size, the variance of the log has almost doubled in 30 years. However, this is largely because of the rise in single-member households; the fraction of households that had a single member went up from 8 percent in 1969 to 21 percent in 2005. Once the data have been divided by the number of adult-equivalents (equivalized), the rise in inequality is less pronounced and only amounts to about 17 percent from 1969 to 2003.[19] It is also clear that the rise in consumption inequality is not driven by increased differences across groups

[19] The rise in inequality amounts to 10 percent if we consider the break in the data from FAMEX to SHS to be artificial, that is, adding the rise in inequality from 1969 to 1996 from FAMEX to the rise from 1997 to 2005 from SHS.

Table 4: Pre- and After-Tax Income Inequality

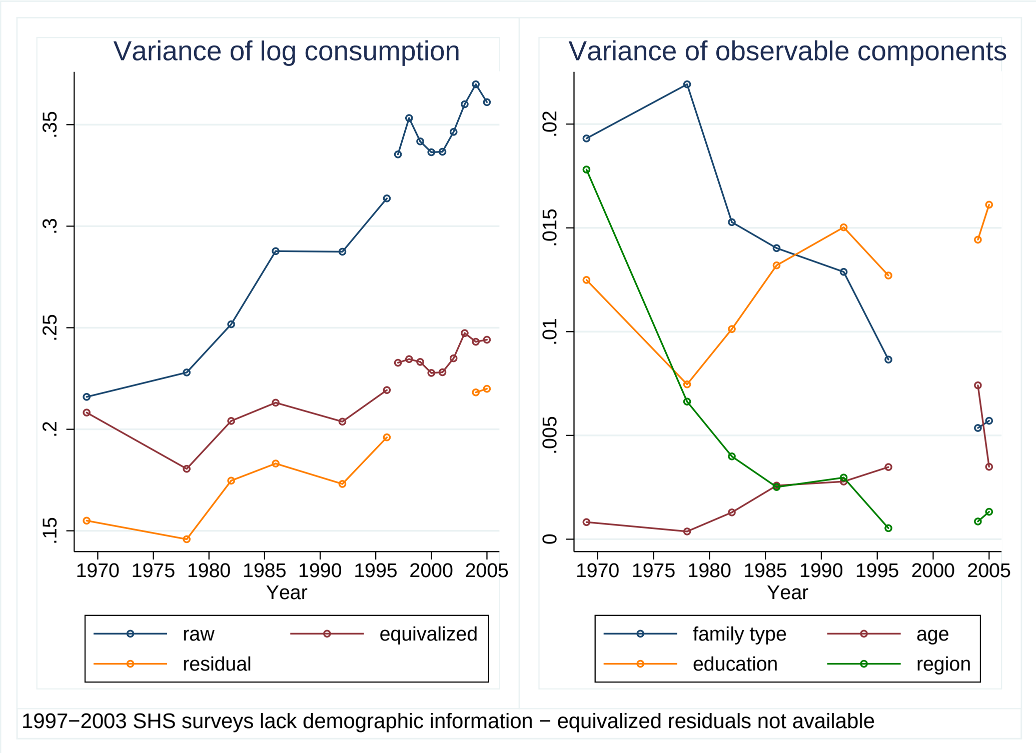
	Pre-Tax Income				After-Tax Income			
year	varlog	50/10	90/50	Gini	varlog	50/10	90/50	Gini
1980	0.3764	2.1443	1.9533	0.3039	0.2364	1.8319	1.8095	0.2635
1985	0.4200	2.3300	1.9655	0.3227	0.2515	1.8664	1.7929	0.2738
1990	0.4456	2.3731	1.9728	0.3209	0.2370	1.8618	1.7914	0.2620
1995	0.5084	2.5023	2.0487	0.3433	0.2592	1.8834	1.8374	0.2769
2000	0.5264	2.5653	2.0523	0.3635	0.2714	1.9151	1.8411	0.2932
2005	0.5466	2.6930	2.1237	0.3692	0.2932	1.9860	1.8927	0.3012
growth (%)								
1980–2000	35.23	13.07	5.72	18.48	13.70	-0.15	2.44	10.35
1980–1990	18.37	10.67	1.00	5.60	0.27	1.63	-1.00	-0.59
1990–2000	13.48	1.62	4.68	12.36	13.39	-1.79	3.47	10.97
1980–1985	11.56	8.66	0.63	6.19	6.41	1.88	-0.92	3.89
1985–1990	6.11	1.85	0.37	-0.56	-5.77	-0.25	-0.09	-4.31
1990–1995	14.09	5.45	3.85	6.98	9.34	1.16	2.57	5.71
1995–2000	-1.13	-3.92	0.81	5.23	3.86	-2.95	0.89	5.20
2000–2005	3.85	4.98	3.48	1.56	8.05	3.70	2.80	2.74

All changes that involve years over which SCF and SLID overlap are computed as the sum of the change in SCF until 1996 and the change in SLID from 1996.

distinguished by readily observable features such as family type, education, age or region of residence. In fact, the contribution to total variance of variance across family types and regions of residence has fallen noticeably since 1969.

Figure 11 juxtaposes the changes in consumption inequality as measured in four different ways with the corresponding changes in the inequality of disposable income. The main thing to take away from that figure is that consumption inequality is consistently lower than disposable income inequality. For the variance of the log, the difference is about 18 percent. Also, at least for the variance of the logs, the two tend to move up and down together. Finally, note that the Gini coefficient only rose by 3.7 percent over the entire sample, and by about 5 percent since the early 1980's.

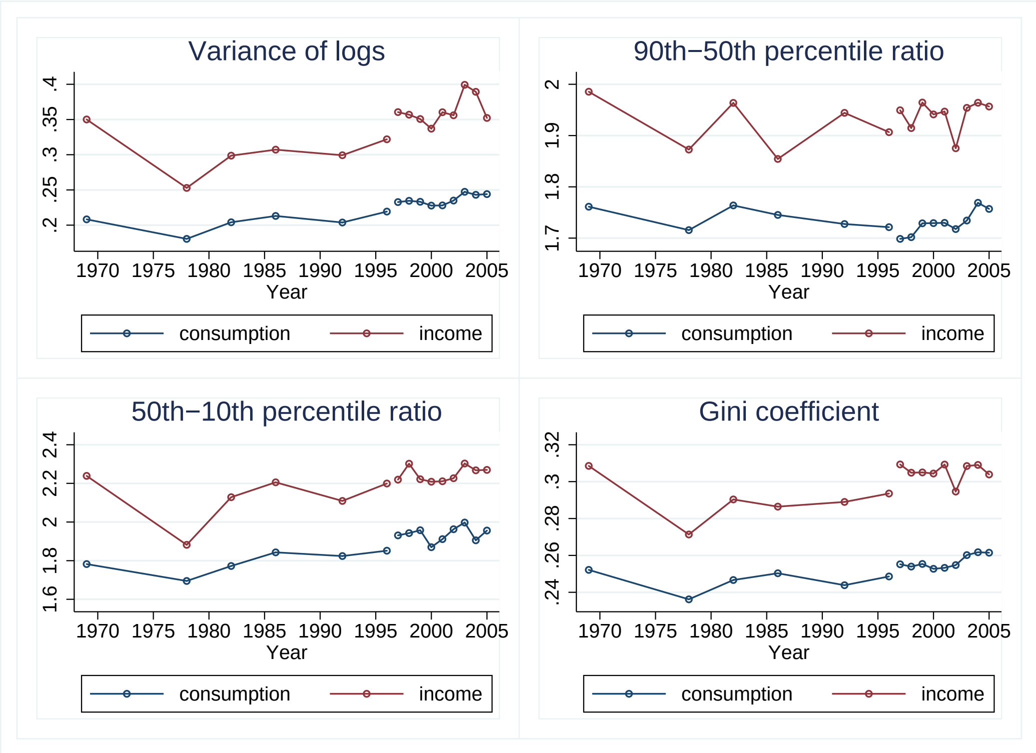
Figure 10: Consumption Inequality and its Decomposition



3.5 Inequality over the Life-Cycle

The evidence on life-cycle profiles of the variance of wages, earnings and consumption is reported in Figure 12, where we control for time effects and Figure 13, where we control for cohort effects. For wages, the picture is clear: variance increases monotonically over the life cycle. For raw earnings, the increase is monotonic when we control for cohort effects, but flat well into middle age if we control for time effects, the increase in variance coming only after age 50. For equivalized earnings, there is no monotonic increase whether we control for time or cohort effects, the pattern is instead J-shaped. Finally, the life-cycle profile of consumption variance again depends on whether we control for time effects or cohort effects. Controlling for time effects, consumption variance exhibits a cup shape; controlling for cohort effects, we see a monotonic increase. The behavior of earnings inequality over the life cycle is particularly interesting. The lack of a clear monotonic (let alone linear) pattern is in sharp contrast to the results constructed and presented in Storesletten

Figure 11: From Disposable Income to Consumption

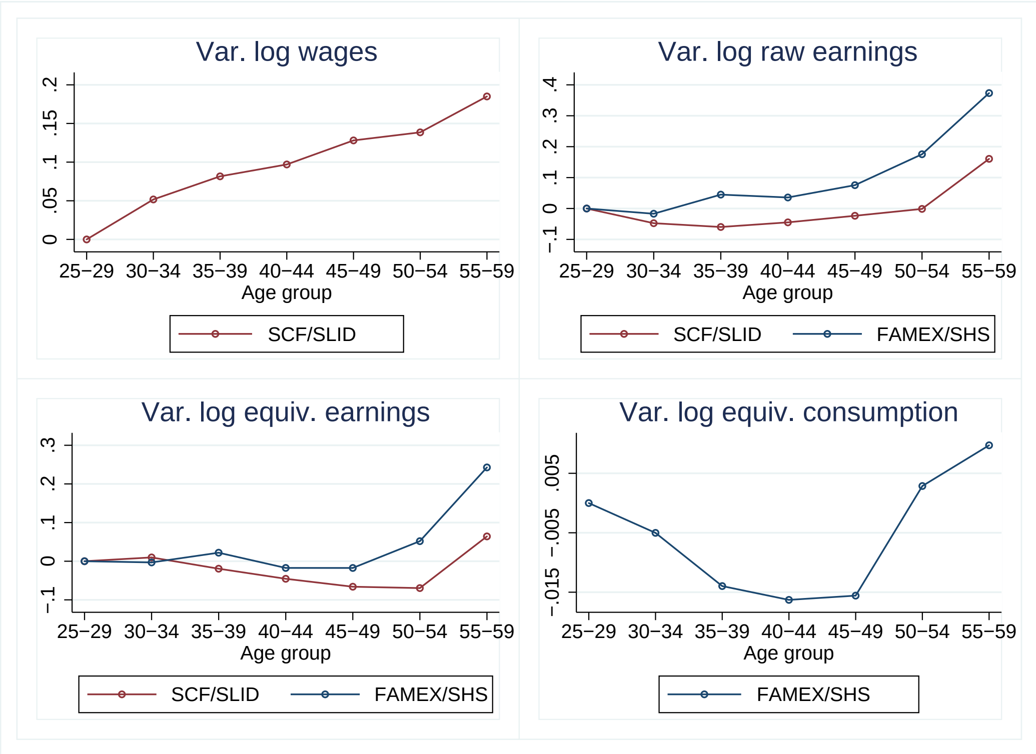


et al. (2004) for the U.S. This matters because it is precisely the linearity of the cross-sectional variance reported there that has motivated the profession to write down a unit root process as the main driving force for earnings; indeed we follow that approach in this paper as well. The facts reported in Figures 12 and 13 cast some doubt on the appropriateness of that approach, at least when applied to Canadian earnings, though it does seem to hold for wages.

3.6 Wealth Inequality

Morissette et al. (2002) document that wealth inequality went up in Canada form 1984 to 1999. The Gini coefficient for equivalized net wealth went up from 0.678 to 0.723, and this difference is statistically significant. We extend this analysis by considering the year 2005 as well. Note that our numbers are not directly comparable to theirs, chiefly because we confine our attention to working-age-headed households.

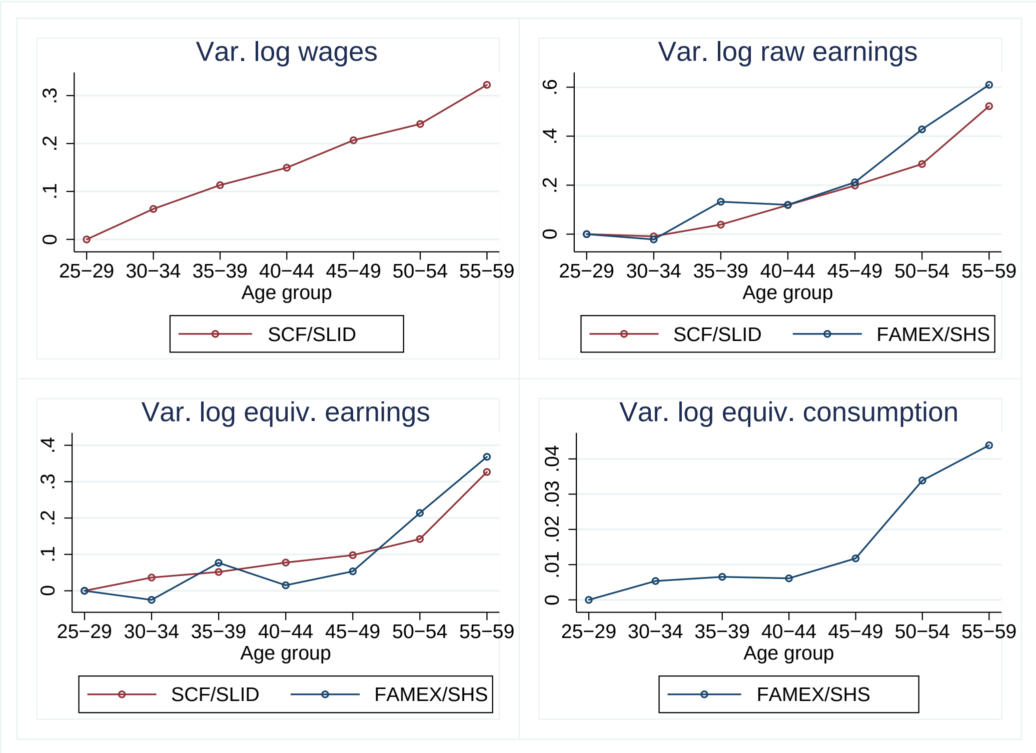
Figure 12: Inequality over the Life-Cycle (Time Effects)



We look at three measures of wealth which we call net financial wealth, net total wealth and net worth, respectively. Net total wealth is defined as net financial wealth plus net residential real estate and business equity. Net worth is simply total assets minus total liabilities.

Table 5 reports wealth to after-tax (disposable) income ratio and Gini coefficient for the income and wealth measures. For all three wealth measures, the wealth to income ratio increased between 1999 and 2005. Gini coefficients of the three wealth measures are about twice as large as those of disposable income. For all income and wealth measures, the Gini coefficient remained stable between 1999 and 2005. Table 6 reports correlation between income and wealth and the share of the top 1 and 5 percentile in the income and wealth distributions. For all wealth measures, correlation with disposable income has increased between 1999 and 2005. The share of the top 1 percentile in the wealth distribution is about three times as high as that for disposable income.

Figure 13: Inequality over the Life-Cycle (cohort Effects)



3.7 Wage/Income Processes

In this section, we examine (exogenous) idiosyncratic income risk that individuals and families face by estimating a stochastic process representing the dynamics of individual wages, equivalized family earnings and equivalized family disposable income. For this estimation, we use data from the SLID exclusively as only this survey provides panel data on income variables as well as individual and family characteristics. In order to extract the idiosyncratic component of a given variable, we first regress the variable of interest on observables and take residuals from the regression.[20] We control for age, gender, marital status, education, province of residence, immigration status and mother tongue in individual wages, while we control for head's characteristics listed above as well as family type and the number of earners for family earnings and disposable income. We model the dynamics of the residuals by the following

[20]We run the cross-sectional regressions year by year, allowing coefficients to vary over time.

Table 5: Disposable Income and Net Wealth

	1999	2005
Net financial wealth/disposable income	2.8841	3.2328
Net total wealth/disposable income	5.6318	6.8045
Net worth/disposable income	6.3066	7.4307
Gini index of disposable income	0.3355	0.3359
Gini index of net financial wealth	0.7054	0.6992
Gini index of net total wealth	0.6664	0.6636
Gini index of net worth	0.6603	0.6590

Notes. Net total wealth is net financial wealth plus net residential real estate and business equity. Net worth is net total wealth plus other net wealth including vehicles. (Source: SFS-RDC)

stochastic process:

$$\begin{aligned} u_{it} &= \alpha_{it} + \varepsilon_{it}, \text{ with } \varepsilon_{it} \sim N(0, \sigma^2_{\varepsilon,t}) \\ \alpha_{it} &= \alpha_{it-1} + \eta_{it}, \text{ with } \eta_{it} \sim N(0, \sigma^2_{\eta,t}), \end{aligned}$$

where u_{it} represents the residual earnings/income of person/family i at time t, α_{it} represents the permanent component of the residual earnings/income, ε_{it} represents the transitory component, and η_{it} represents innovation to the permanent component. We assume that η_{it} and ε_{it} are i.i.d., that $E(\varepsilon_{is}\eta_{it}) = 0$ for all s, t and that $E(\varepsilon_{is}\varepsilon_{it}) = E(\eta_{is}\eta_{it}) = 0$ for all $s \neq t$.

In Table 7 we present results from estimating processes for individual wages, equivalized family earnings and equivalized family disposable income. The estimation process is constructed so as to match the autocovariance function of log differences. The result is that a very high fraction of the overall cross-sectional variance and also a high fraction of the risk faced by households is accounted for by the permanent component as opposed to the transitory component. Inspecting the moment conditions, we see that the magnitude of the σ^2_ε estimates depends on the covariance between subsequent log differences. If subsequent log differences are not highly correlated, which they are not, we are led to conclude that transitory shocks account for a small fraction of the overall variance and risk. Indeed, given our estimates on wages, that

Table 6: Disposable Income and Net Wealth

	1999	2005
Correlation coefficient		
(Net financial wealth, disposable income)	0.3655	0.4377
(Net total wealth, disposable income)	0.3551	0.4146
(Net worth, disposable income)	0.3470	0.4255
Share of the top 1%		
Disposable income	0.0541	0.0583
Net financial wealth	0.1635	0.1449
Net total wealth	0.1665	0.1749
Net worth	0.1568	0.1668
Share of the top 5%		
Disposable income	0.1634	0.1719
Net financial wealth	0.4143	0.3926
Net total wealth	0.3702	0.3817
Net worth	0.3514	0.3639

Notes. Net total wealth is net financial wealth plus net residential real estate and business equity. Net worth is net total wealth plus other net wealth including vehicles. (Source: SFS-RDC)

fraction is below 3 percent for 60-year-olds; for family earnings and family disposable income, the corresponding number is below one percent. Of course, much higher numbers obtain for the young. [21]

In Table 7, we report estimates for σ_η^2 (permanent risk) and σ_ε^2 (transitory risk) for post-government disposable family income as well as those for family earnings. Government policies such as progressive taxation, social welfare programs, and employment insurance are designed to mitigate earnings risk that families face. In order to examine the extent to which government policies reduce earnings risk, we compare estimates for equivalized family earnings and disposable income. Our results suggest

[21] These fraction are likely to be much lower than we would get from an estimation designed either to match the age profile of cross-sectional variance (as in Storesletten et al. (2004)) or if we sought to match the autocovariance function of levels (as opposed to differences). Thus the conclusion that transitory shocks are unimportant should be taken with a very large grain of salt.

Table 7: Parameter estimates for the earnings process

	Individual wage		Equivalent family earnings		Equiv. family disposable income	
	σ^2_ε	σ^2_η	σ^2_ε	σ^2_η	σ^2_ε	σ^2_η
1994	0.0598	0.0588	0.0297	0.1051	0.0115	0.0288
1995	0.0535	0.0588	0.0298	0.1017	0.0085	0.0315
1996	0.0578	0.0592	0.0226	0.0942	0.0105	0.0344
1997	0.0632	0.0523	0.0238	0.0970	0.0080	0.0311
1998	0.0572	0.0538	0.0253	0.0876	0.0085	0.0314
1999	0.0733	0.0480	0.0200	0.0934	0.0097	0.0328
2000	0.0721	0.0485	0.0196	0.0865	0.0109	0.0300
2001	0.0639	0.0536	0.0275	0.0809	0.0114	0.0295
2002	0.0596	0.0566	0.0206	0.0805	0.0142	0.0317
2003	0.0579	0.0536	0.0240	0.0750	0.0127	0.0245
2004	0.0513	0.0583	0.0295	0.0715	0.0083	0.0288
2005	0.0513	0.0691	0.0295	0.0847	0.0083	0.0395
Mean	0.0609	0.0547	0.0248	0.0882	0.0104	0.0312

Notes. σ^2_ε and σ^2_ε are the variance of the transitory shock and that of the permanent shock, respectively. (Source: SLID-RDC)

that the tax and transfer system substantially reduces both permanent and transitory earnings risk, with the permanent and transitory risk reduced on average by 65 and 58 percent, respectively.

3.8 Log-Normality Tests

A highly desirable property of distributions is log normality. With that property, a distribution can be fully characterized by its first two moments, and two distributions are unambiguously comparable with respect to the degree of inequality. In particular, if earnings are log-normally distributed, it can never happen that the Gini coefficient goes up but the variance of log earnings goes down.

We investigate log-normality using two approaches. The first is formal tests, including the Kolmogorov-Smirnov and the Skewness-Kurtosis test. These tests tell us whether the deviations from normality that we observe are statistically significant. A further question, though, is whether they are sufficiently large to be important. To address that issue, we use Epanechnikov (1969)'s kernel density estimation in order to compare non-parameterically estimated densities with the corresponding normal densities.

The results from formal tests are on the whole negative. For earnings, we can always reject log-normality, no matter what the test, no matter what the year, no matter what the cohort and no matter what the (reasonable) significance level. The situation with consumption is rather different. As pointed out in Battistin et al. (2007), consumption is more log-normal than earnings. In fact, for observations in 1982, 1986 and 1992 and using the Kolmogorov-Smirnov test, we cannot reject, at the 5 percent level, log-normality of the cross-sectional distribution of consumption among two cohort groups: those born in the 1940's and those born in the 1950's. For other tests and other samples, however, we can reject normality.

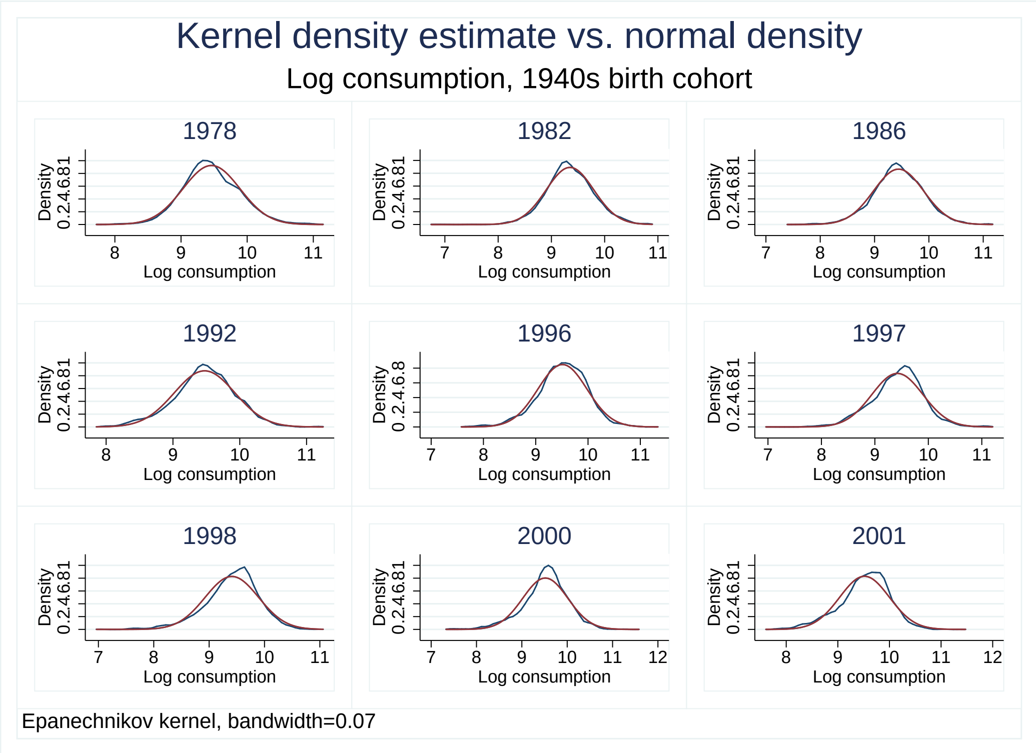
Turning our attention to the second approach, consider Figures 14–17. Comparing the non-parameterically estimated probability density functions with their normal distribution counterparts, we notice several things. The first is that it is very clear that log consumption is closer to normal than log earnings. Indeed, an inspection of the skewness and kurtosis of log consumption reveals that the distribution is very nearly symmetric and exhibits only a little more kurtosis than the normal distribution. The difference is barely visible to the naked eye. Thus log-normality is an excellent approximation of the distribution of consumption.

For earnings, things are quite different. The distributions are visibly skewed to the left (the mean is to the left of the mode) and strongly leptokurtic (the density has a sharper peak and fatter tails than the normal distribution). Log-normality is not a particularly good approximation of the distribution of earnings.

4 A View of Inequality from Public-Use Files

It may be of some interest to compare the conclusions about trends in inequality we draw from the ones that would be drawn from Public Use Files (PUF). PUF data

Figure 14: How log normal is consumption, part 1

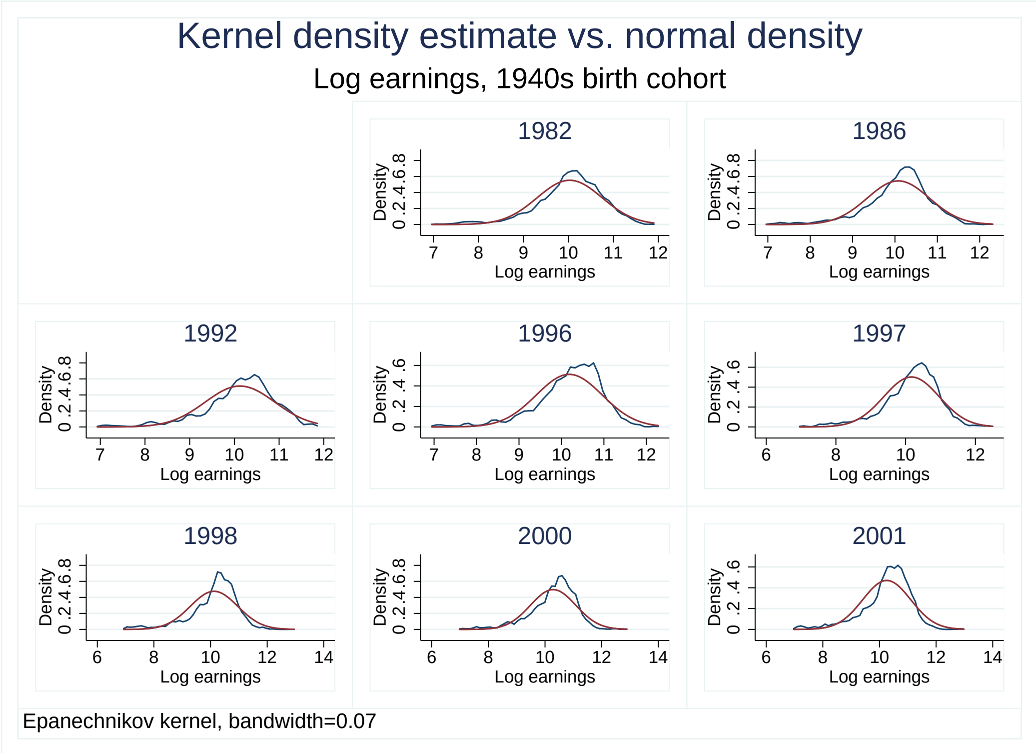


differ from RDC data in two important respects. First, Statistics Canada introduces noise in PUF data in order to preserve the identity of the respondents. Second, the revised weights are not currently available in PUF for the SCF sample. Since consumption data is only available through public files, we will briefly review how the main message from RDC data is altered when using public data for income and wealth.

4.1 Income Inequality

To see what Public-Use Files tell us about trends in income inequality, consider Table 8 and contrast it with the corresponding results from RDC data, displayed in Table 4. One striking result is that the level of inequality, however measured, is understated in Public-Use Files. In terms of changes, the difference is even more striking. For example, the Public-Use Files suggest that the level of inequality has

Figure 15: How log normal are earnings, part 1



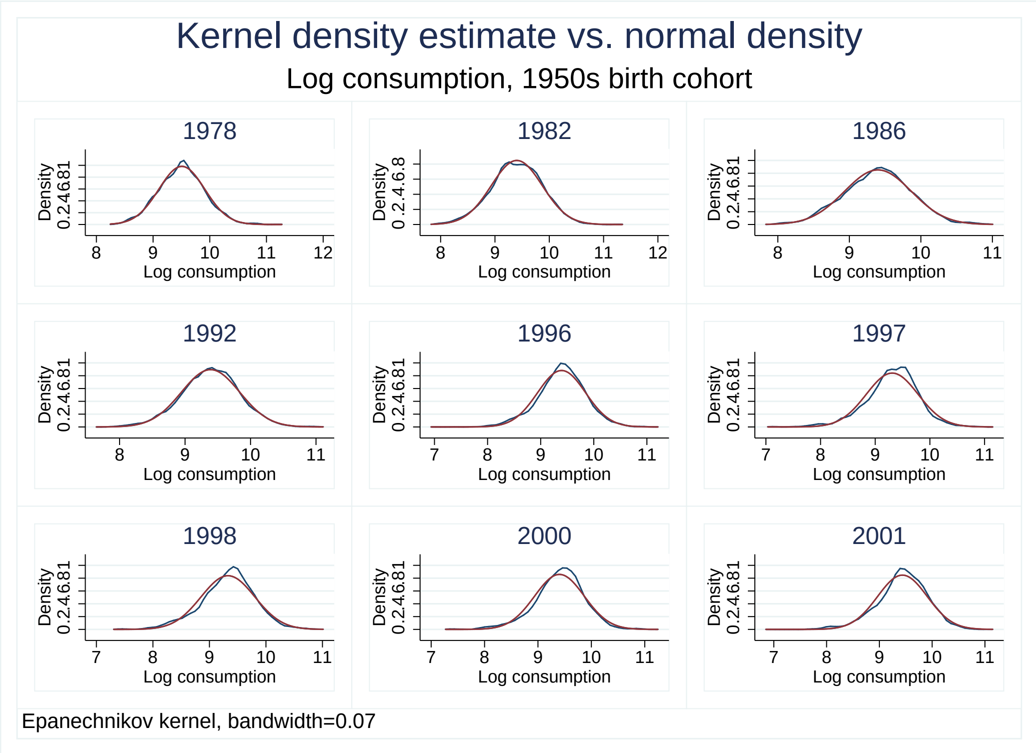
barely changed from 1986 to 2000. In particular, it completely misses the increase in inequality in the 1990's, which led Frenette et al. (2007) to abandon survey data in favor of Census data.[22]

4.2 Wealth Inequality

One aspect that public use files typically have a problem with is the very top of the distribution (because of top coding), and that is of course particularly important for the wealth distribution. Table 9 sheds some light on that. The Public-Use Files for the SFS suffer from top-coding problems where each observation above a certain threshold is given the value of that threshold. Since we have access to RDC data, in which no observation is top-coded, we can compare the accuracy of different statistical

[22]Some of the changes between 2000 and 2005 are due to the revision of survey weights as SLID data for 2000 feature old weights based on Census 1996, while the SLID data for 2005 feature the revised weights based on Census 2001.

Figure 16: How log normal is consumption, part 2



methods to impute top-coded observations in terms of various measures of inequality.

What we do is the following, taking net financial wealth as an example. First we consider the components of net financial wealth, like mutual fund holdings, mortgage debt etc. For each such component, we consider the top decile with the idea that that top decile has been drawn from a Pareto distribution. Excluding the top-coded observations, we then use the remaining observations in the top decile to estimate the single (remaining) unknown parameter of the Pareto distribution. This can be done either by least squares or maximum likelihood and we report the results from both approaches in Table 9, labeled Public 3 and Public 4 respectively. Having estimated the parameter, one can then mechanically compute the mean of the distribution and replace the top-coded values with that mean. As Table 9 reveals, both methods work pretty well. For the Gini coefficient of any of the wealth categories, the estimate is at most 1 percent away from the RDC value. However, The estimates are not as precise for the top 1%.

Figure 17: How log normal are earnings, part 2

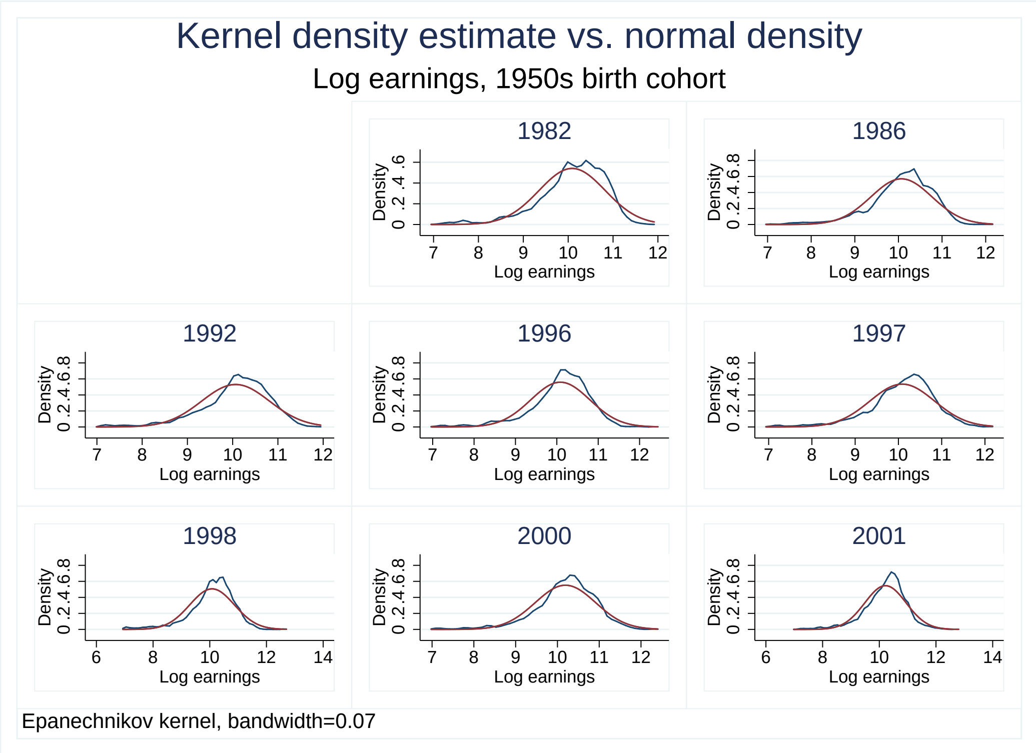


Table 8: Pre- and After-Tax Income Inequality form Public-Use Files

	Pre-Tax Income				After-Tax Income			
year	varlog	50/10	90/50	Gini	varlog	50/10	90/50	Gini
1980	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
1986	0.4126	2.3037	1.9627	0.3109	0.2430	1.8755	1.8078	0.2628
1990	0.4208	2.3573	1.9316	0.3120	0.2279	1.8669	1.7545	0.2554
1995	0.4508	2.4262	1.9832	0.3212	0.2382	1.8634	1.7867	0.2607
2000	0.4856	2.5177	2.0068	0.3320	0.2555	1.9074	1.8129	0.2697
2005	0.5166	2.6667	2.0859	0.3505	0.2737	1.9925	1.8592	0.2846
growth (%)								
1986–2000	6.51	-0.31	2.04	1.41	2.03	-2.66	0.59	-0.17
1986–1990	2.00	2.33	-1.59	0.35	-6.21	-0.46	-2.94	-2.83
1990–2000	4.29	-2.75	3.68	1.04	8.66	-2.20	3.61	2.75
1980–1985	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
1986–1990	2.00	2.33	-1.59	0.35	-6.21	-0.46	-2.94	-2.83
1990–1995	7.13	2.92	2.67	2.93	4.53	-0.19	1.84	2.09
1995–2000	-3.11	-5.71	1.00	-1.92	4.04	-2.01	1.76	0.64
2000–2005	6.38	5.92	3.94	5.57	7.13	4.46	2.55	5.54

All changes that involve years over which SCF and SLID overlap are computed as the sum of the change in SCF until 1996 and the change in SLID from 1996.

5 Concluding remarks

This paper characterizes various aspects of inequality in Canada over the last 30 years or so. All measures of income inequality have increased over these years. Most notable is the rise in wage and earnings inequality, which were partly offset by the tax and transfer system to produce a moderate rise in after-tax income. As a result, the rise in consumption inequality has been fairly mild, but still noticeable.

In light of the recent work of Frenette et al. (2006) and Frenette et al. (2007), our results suggest that the revision of the weights that Statistics Canada has recently implemented was valuable, in the sense that these revisions brought measures of inequality in Survey data much closer to the evidence emanating form Census data.

Table 9: Comparison of public-use and confidential SFS 1999.

	Public 1	Public 2	Public 3	Public 4	Confidential
Share of top 1%					
Disposable income	0.0481	0.0481	0.0482	0.0483	0.0541
Net financial wealth	0.1246	0.1431	0.1841	0.1966	0.1635
Net total wealth	0.1161	0.1424	0.1623	0.1831	0.1665
Net worth	0.1066	0.1323	0.1509	0.1683	0.1568
Gini index					
Disposable income	0.3252	0.3303	0.3319	0.3320	0.3355
Net financial wealth	0.6798	0.6926	0.7014	0.7032	0.7054
Net total wealth	0.6348	0.6536	0.6616	0.6685	0.6664
Net worth	0.6274	0.6456	0.6532	0.6591	0.6603

The column designated Public 1 reports estimates for the sample excluding observations with top-coded values. Public 2 reports the estimates for the sample including observations with top-coded values. Public 3 corrects for top-coding by OLS estimation suggested by David Domeij. Public 4 corrects for top-coding by maximum likelihood estimation. The estimation is conducted for each income and wealth component, since top-coding is done at the component level. For the estimation, the sample of families with the given variable above the 90th percentile and below the top-coding threshold is used. (Source: SFS-RDC and SFS-PUMF.)

In particular, our results tend to support the broad conclusion on the evolution of income inequality reached by Frenette et al. (2007).

Appendix A: Detailed Sources of Data

A.1 Consumption Data

The *Survey of Family Expenditures* (FAMEX) and the *Survey of Household Spending* (SHS) provide cross-sectional data on individual and household characteristics, income, and expenditure. For both surveys, income and expenditure data refer to a calendar year (called reference year). Data are collected through personal interviews conducted in the first quarter following a given reference year. FAMEX data are available for 1969, 1974, 1978, 1982, 1984, 1986, 1990, 1992, and 1996. But the surveys for 1974, 1984, and 1990 cover only urban areas and have a smaller sample (around 7,000 households instead of 14,000 or so for the years with national coverage). The SHS replaced the FAMEX in 1997 and has been conducted annually thereafter. Currently, the latest data are for 2005.

For 1969–1986, expenditure data are collected for *spending units*. Spending unit is defined as a group of persons dependent on a common or pooled income for the major items of expense and living in the same dwelling or one financially independent individual living alone. In 1990, the unit of analysis was replaced by *households*, which is defined as a person or a group of persons occupying one dwelling unit.

Disposable income after taxes and transfers (y^D) can be constructed for the whole sample period. However, pre-tax household earnings (y^L), which is defined as the sum of wages and salaries and the labor part of self-employment income, cannot be constructed in a consistent manner over time, since wages/salaries and self-employment income are not separately available over the 1997–2005 period. We use nondurable goods expenditure as our measure of consumption. Nondurable goods expenditure consists of the expenditure on food, alcohol and tobacco, personal care supplies and services, water, fuel, and electricity, household operation, public transportation, operation of automobiles and trucks, apparel, recreation, reading, and education, traveler accommodation, health care), and miscellaneous nondurable expenditures. We deflate income by region-specific CPI for all items. We deflate each expenditure component by the corresponding region-specific CPI.

Unfortunately, no information on education is available from 1997 to 2003. Thus, we cannot conduct analysis involving education level for those years.

A.2 Income Data

The *Survey of Consumer Finance* (SCF) and the *Survey of Labour and Income Dynamics* (SLID) provide cross-sectional data on individual and household characteristics as well as income components. Both surveys share the same sample design, as their sample consists of rotation groups from the *Labour Force Survey* (LFS) and includes roughly 15,000 households.[23] For both surveys, income data refer to a calendar year (called reference year). SCF data are collected through personal interviews conducted in April (so that individuals have their tax return in mind) following a given reference year. As for SLID, characteristics are collected in January following a given reference year, and income data is collected in May following a given reference year. SLID respondents are given the option to skip the May interview by giving Statistics Canada permission to access their T1 (income) tax information. Over 80% of respondents give their consent to the use of administrative records. The change in collection method may have had an impact on the comparability of data across the two surveys.[24]

Unlike SCF, SLID also has a panel dimension. Waves are introduced every three years, and each wave of respondents is interviewed for 6 years, with the first wave starting in 1993. The panels, which initially constitute a representative cross-sectional sample of the population, are thus as follows: Panel 1 from 1993 to 1998; Panel 2 from 1996 to 2001; Panel 3 from 1999 to 2004; Panel 4 from 2002 to 2005; Panel 5 from 2005. For each sampled household in SLID, up to 12 interviews are conducted

[23]The LFS is like the CPS, and the SCF is like the March supplement to the CPS. The SLID is also like a March supplement, only each sample is interviewed over the next 6 years.

[24]Statistics Canada reports that "[C]omparisons of figures produced from the SCF with other sources of data (Census of Population, Longitudinal Administrative Data, National Economic and Financial Accounts) reveal that certain income components, such as investment, self-employment earnings, social assistance payments and EI benefits, are under-reported in the SCF." Furthermore, "SLID's estimates of the number of income recipients, aggregate individual income and average family income are higher than the corresponding estimates from the SCF data." It should also be pointed out that the response rates, while relatively high in both the SCF and SLID, are higher in SLID than in SCF data: in the SCF it ranged from 78.1% in 1989 to 82.1% in 1995, while SLID's cross-sectional rate of response was 87.1% in 1996. The higher SLID response rate is primarily due to the use of administrative data from the tax files.

over a six-year period.

SCF data are available every other year from 1965 to 1971, and yearly thereafter until 1997. For all years, SCF data are available at the individual, census family and economic family level. In addition, household-level data are available for 1976-1997. We have SLID data from 1993 to 2005 at all levels. In the current paper, we use individual- and economic family-level data.

Though SCF data are available for odd years between 1965 and 1975, we do not use the data prior to 1976 for the following two reasons. First, some of the data necessary for our analysis are not available over these years. For example, data on labor force status and usual hours worked are not available for these years.[25] Second, SCF started adjusting non-responses by imputation in 1977, which is necessary to construct integrated survey weights at household, economic family, census family, and individual levels. (Non-responses were adjusted by re-weighting before 1977.) Though Statistics Canada dropped households with non-respondent members from the 1976 sample in order to apply integrated weights based on Census 1996, they have not applied the same adjustment to the data for 1965–1975. Without such adjustment, the data for 1965–1975 cannot receive revised weights.

With the detailed income data available in SCF and SLID, we are able to compute individual and family earnings (y^L), pre-government family non-financial income (y^{L+}), net financial income (y^A), pre-government family income (y), and post-government family income (y^D). In SCF, we compute annual hours worked by multiplying usual hours worked per week and weeks worked during the reference year. SLID provides data on total hours paid at all jobs during the reference year. We compute hourly wage (w) by dividing individual earnings by the person's annual hours worked. We deflate the income data by province-specific CPI.

SCF adopted a new set of questions on education, introduced by the Labour Force Survey in the 1989 reference year. The following three points are relevant to our analysis: 1. In 1976–1988, post-secondary education requires high-school graduation. In 1989–1997, any education that could be counted towards a degree, certificate, or diploma from an educational institution is taken as post-secondary education. For

[25] Note that LFS employment and unemployment estimates before and after 1976 are not directly comparable since the LFS questionnaire underwent significant changes in 1976.

example, trades programs offered through apprenticeship, vocational schools or private trade schools do not always require high school graduation. Such education is considered as post-secondary in 1989–1997 while only primary or secondary would have been recognized prior to 1989. 2. One cannot identify individuals whose educational attainment is exactly high school graduation in the SCF data prior to 1989, while it is possible for SCF 1989–1997 and for all SLID years. 3. Over the 1976–1988 period, one cannot distinguish between university (or college) certificate/diploma below bachelor's degree and university degree (bachelor's degree or above). For SCF 1989–1997 and for all SLID years, those two education levels can be identified separately. Due to the change described in the first point, one cannot construct the category of post-secondary education that is consistent over time. Due to the lack of detailed information on education before 1989, we use the following broad education categories in our analysis for consistency: 1. high school dropouts or graduates; 2. some secondary education below university; 3. university certificate/diploma/degree.

A.3 Wealth Data

The *Survey of Financial Securities* (SFS) provides various details of individual and family characteristics as well as income and wealth data. The data are available for 1999 and 2005. The SFS sample is drawn from two sources. The main sample is selected from the Labour Force Survey (LFS) sampling frame. The second portion of the sample, accounting for approximately 10 to 15 percent of the total sample, is drawn from geographic areas in which a large proportion of households have what is defined as "high-income". In other words, much like the Survey of Consumer Finances in the U.S., the SFS over-samples the rich in order to improve wealth estimates.

Demographic data are available at the individual level, while the questions pertaining to wealth are asked with reference to the economic family. With detailed data on income, assets and debts, we can construct a measure of disposable income (y^D), net financial wealth (a), net total wealth (a^+), and net worth for 1999 and 2005. The disposable income and the three wealth measures are deflated by province-specific CPI.

Appendix B: Sample Selection

The following tables document the various samples used in the text. Table 10 displays our FAMEX/SHS sample selection. Tables 11 and 12 present our samples for individual-level and family level data for the SCF, while Tables 13 and 14 are for SLID data. Finally, our sample selection for the SFS is displayed in Table 15.

Table 10: Sample Selection in FAMEX-SHS

	Observations deleted	Remaining observations
Original data set		225360
Observed in 1974, 1984, and 1990	15991	209369
Part-year household (SHS)	4603	204766
Aged less than 25 or more than 60	63474	141292
Missing main characteristics	1630	139662
Yukon, North-west Territories or Nunavut	3868	135794
Zero food consumption	17	135777

Main characteristics include education, region of residence, and family type. Yukon, North-west Territories and Nunavut are only covered in the SHS survey, not in FAMEX.

Table 11: Sample Selection in Individual-level SCF-RDC

	Observations deleted	Remaining observations
Original data set (1976-1997)		1965517
Observed in 1976 or 1978	79407	1886110
Aged less than 25 or more than 60	994857	891253
Income imputed	164839	726414
Missing main characteristics	0	726414
Missing earnings or hours worked	2880	723534
Wage less than half min wage	26821	696713
Zero hours worked	241878	454835
Wage greater than 100 and annual hours worked less than 100	775	454060
Non-positive earnings	0	454060

Main characteristics include gender, education, marital status, and province of residence.

Table 12: Sample Selection in Family-level SCF-RDC

	Observations deleted	Remaining observations
Original data set (1976-1997)		750396
Observed in 1976 or 1978	27077	723319
Aged less than 25 or more than 60	187911	535408
Income imputed	124512	410896
Missing main characteristics	0	410896
Missing income or head's hours worked	2274	408622
Head's wage less than half min wage	13823	394799
Zero hours worked by head	102511	292288
Head's wage greater than 100 and hours worked less than 100	292	291996
Non-positive income	59	291937

Main characteristics include head's education, province of residence, and family type.

Table 13: Sample Selection in Individual-level SLID-RDC

	Observations deleted	Remaining observations
Original data set (1993-2005)		875597
Aged less than 25 or more than 60	441891	433706
Income imputed	91145	342561
Missing main characteristics	16438	326123
Missing earnings or hours worked	17497	308626
Wage less than half min wage	16480	292146
Zero hours worked	46682	245464
Wage greater than 100 and annual hours worked less than 100	684	244780
Non-positive earnings	0	244780

Main characteristics include gender, education, marital status, and province of residence.

Table 14: Sample Selection in Family-level SLID-RDC

	Observations deleted	Remaining observations
Original data set (1993-2005)		319934
Aged less than 25 or more than 60	57278	262656
Income imputed	90882	171774
Missing main characteristics	7983	163791
Missing income or head's hours worked	7561	156230
Head's wage less than half min wage	7934	148296
Zero hours worked by head	19915	128381
Head's wage greater than 100 and hours worked less than 100	272	128109
Non-positive income	52	128057

Main characteristics include head's education, province of residence, and family type.

Table 15: Sample Selection in Family-level SFS-RDC

	Observations deleted	Remaining observations
Original data set (1999, 2005)		21215
Aged less than 25 or more than 60	5504	15711
Missing main characteristics	0	15711
Missing income or wealth	0	15711
Outliers in wealth	x	15711−x

We dropped outliers with respect to wealth. However, because the number of observations dropped violates the minimum threshold set by Statistics Canada for public release, we cannot report the exact number of observations affected by this criterion. Main characteristics include head's education, province of residence, and family type.

References

Battistin, E. (2003). Errors in survey reports of consumption expenditures. Institute for Fiscal Studies Working Paper W03/07.

Battistin, E., R. Blundell, and A. Lewbel (2007). Why is consumption more log normal than income? Gibrat's Law revisited. Institute for Fiscal Studies Working Paper 08/07.

Crossley, T. F. and K. Pendakur (2006). Consumption inequality in canada. In D. A. Green and J. R. Kesselman (Eds.), *Dimensions of Inequality in Canada*. Vancouver, Canada: University of British Columbia Press.

Epanechnikov, V. A. (1969). Nonparametric estimation of a multivariate probability density. *Theory of Probability and Its Applications 14*, 153–158.

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 G. Picot (2006). Rising income inequality in the 1990s: An exploration of three data sources. In D. A. Green and J. R. Kesselman (Eds.), *Dimensions of Inequality in Canada*. Vancouver, Canada: University of British Columbia Press.

Heathcote, J., K. Storesletten, and G. L. Violante (2008). The macroeconomic implications of rising wage inequality in the United States. Unpublished manuscript.

Lathe, H. (2005). Survey of labour and income dynamics: 2003 historical revision. Statistics Canada, Income Statistics Division, Income Research Paper Series.

Maslove, A. M. (1988). Distributional impacts of personal income tax reform. Discussion Paper, Institue for Research on Public Policy.

Morissette, R., X. Zhang, and M. Drolet (2002). The evolution of wealth inequality in Canada, 1984-1999. Statistics Canada Working Paper No. 187.

Rea, W. and E. Greenlee (2005). The effects of the revised estimation methodology on estimates from household expenditure surveys. Statistics Canada, Income Statistics Division, Household Expenditure Research Paper Series.

Saez, E. and M. R. Veall (2005). The evolution of high incomes in northern america: Lessons from Canadian evidence. *American Economic Review 95*, 831–849.

Slesnick, D. T. (2001). *Consumption and Social Welfare. Living Standards and Their Distribution in the United States.* Cambridge: Cambridge University Press.

Storesletten, K., C. I. Telmer, and A. Yaron (2004). Consumption and risk sharing over the life cycle. *Journal of Monetary Economics 51*(3), 609–633.