2023 DINA UPDATE FOR CANADA AND PROVINCES

SILAS XUEREB
MATTHEW FISHER-POST
FRANCOIS DELORME
CAMILLE LAJOIE

TECHNICAL NOTE N°2023/04

NOVEMBER 2023

WORLD INEQUALITY LAB

Distributional National Accounts Estimates for Canada and Provinces, 1920-2020

Silas Xuereb[1,2], Matthew Fisher-Post[2], François Delorme[2,3], and Camille Lajoie[4]

[1]University of Massachusetts
[2]World Inequality Lab (Paris School of Economics)
[3]University of Sherbrooke
[4]London School of Economics

This version: November 2023

The purpose of this technical note is to highlight the distributional national accounts series for Canada, including estimates through the year 2020. Where extensive previous work had been based solely on fiscal income data from the tax authorities, we have improved these estimates and cover all sources of national income, including income that is not reported on tax returns, as well as to cover post-tax incomes—and estimate distributions not only for all of Canada, but for its largest provinces.

1   Introduction

In this technical note, we provide a brief overview of the updated methods used to estimate distributional national accounts (DINA) for Canada for the period from 1982-2020. Previous estimates had relied on less detailed data and more simplified methods. This update revises previous estimates of the pre-tax national income distribution and provides new estimates of the post-tax national income, and post-tax disposable income distributions. We also estimate distributional national accounts for the four largest provinces, and the rest of Canada.

2   Data

The main data source for all our estimates is the Longitudinal Administrative Databank (LAD), which contains a 20% sample of tax microdata for every year from 1982 to 2020. We obtained custom tabulations of the LAD from Statistics Canada which provide average incomes for several types of income for each percentile (plus the top 0.1% and top 0.01%) of the national equal-split income distribution. The sorting variable is the LAD-defined market income plus employment insurance (EI) and Quebec/Canada Pension Plan (Q/CPP) income minus EI and Q/CPP contributions, designed to closely align with the DINA concept of pre-tax post-replacement fiscal income. Within each percentile, we observe several types of income, including employment income, several government transfers, capital gains, dividends, and self-employment income, as well as payroll and income taxes paid.

Summary statistics of income inequality based on the LAD (using individuals as the unit instead of equal-split adults) are directly available from Statistics Canada. The first step in converting this data into DINA is to convert the population included in the data to the population of all Canadian residents 20 years of age and older. The LAD includes people of all ages who filed their taxes plus a small number of nonfilers who have a social insurance number and some link to the tax system. Thus, we first remove taxfilers under the age of 20 from the database. We observe the location in the income distribution of taxfilers under the age of 25. Since 33.1% of taxfilers under the age of 25 are under the age of 20 (Robson and Schwartz, 2020), we remove 33.1% of the taxfilers under the age of 25 from the database.

Next, we need to add all nonfilers above the age of 20 so that all Canadian residents above the age of 20 are represented in the data. We compare the number of taxfilers in the data to population estimates in the World Inequality Database. The proportion of people aged 20 and over observed in the LAD rose during the 1980s and early 1990s as tax credits based on taxfiling were introduced. In 1982, 87.5% of the 20 plus population are observed, increasing to 96.1% in 1994. This proportion has remained relatively stable ever since, ranging from 95-97.3% each year until 2020, when it dipped slightly to 94.3%.

We impute nonfilers to the distribution based on the income distribution of nonfilers obtained by Robson and Schwartz (2020) through comparing 2015 tax and survey data. This distribution implies that about 13.5% of nonfilers are in the bottom decile of the distribution, with a relatively uni-

form distribution throughout the middle deciles, and only about 5% in the top decile. We assume that there are no nonfilers in the top centile of the distribution. Before 1994, when the proportion of unobserved adults is higher, we increase the share imputed at the bottom of the distribution, assuming that low-income individuals were less likely to file taxes when they had less of an incentive to do so.

3 Pre-tax national income

Next, we distribute all the national income not observed in the tax data among all 20 plus adults. The income concepts of the LAD correspond closely to the concept of fiscal income in DINA (which is roughly income that is observable in tax data; Blanchet et al., 2021). We calculate measures of fiscal labour and capital income. Following convention, we treat self-employment income as 70% labour income and 30% capital income. To move from the fiscal income distribution to the pre-tax national income distribution, we adjust total labour and capital income to match total labour and capital income from national accounts. For labour income, there is very little adjustment because almost all labour income is observable in tax data. On the other hand, on average only 29% of capital income is observed annually. Most remaining capital income is distributed according to the distribution of fiscal capital income. Pension accrual is the exception, which is distributed according to the fiscal labour income distribution. This creates the distribution of pre-tax national income, the benchmark DINA concept for earned income, that is, income before the operation of the tax and transfer system.

4 Post-tax disposable income

Starting from our estimates of the pre-tax national income distribution, we distribute all taxes and government transfers to individuals to obtain the post-tax disposable income distribution. Taxes fall into four main categories: personal income taxes, corporate income taxes, sales and other direct taxes, and property taxes. Personal income taxes are observed in the LAD and thus we deduct observed personal income taxes paid from pre-tax national income, scaling total PIT to match total PIT collected according to national accounts (about 95% of PIT is observed). Corporate income taxes are distributed according to the distribution of corporate operating surplus (dividends plus undistributed corporate profits). Property taxes should be distributed according to the distribution of real estate wealth. The 2019 Survey of Financial Security (SFS) public use microdata file (PUMF) was used to obtain the distribution of real estate wealth along the pre-tax income distribution. Since survey data often underestimates income and wealth at the top of the distribution, and the SFS PUMF is only available for 2019, we distribute 65% of property taxes according to the distribution of real estate wealth in 2019 and 35% according to the observed distribution of rental income. Sales and other direct taxes are distributed according to the distribution of disposable income (pre-tax income plus transfers less taxes and saving).

Government transfers to individuals are observed in the LAD. Specifically, we add the following observed transfers to pre-tax national income less taxes paid to obtain the post-tax disposable income distribution: old age security, guaranteed income supplement, family benefits, GST/HST credit, child tax benefits, working income tax benefit, social assistance, workers' compensation, other federal tax credits, and provincial transfers. The addition of these transfers and the subtraction of the four types of taxes listed above provides the post-tax disposable income distribution.

5   Post-tax national income

Since taxes raised are larger than spending on transfers to individuals, total post-tax disposable income is less than national income. To obtain the post-tax national income distribution, which allows income levels to be compared across countries, we distribute the rest of government spending to individuals as 'in-kind' transfers. It is very difficult to know who benefits from government consumption expenditure and so this distribution requires making significant assumptions. We follow the DINA Guidelines (Blanchet et al., 2021; Piketty, Saez, and Zucman, 2018) in making these assumptions.

Government consumption expenditure can be broken down into collective and individual government consumption expenditure. Individual government consumption expenditure includes health and education spending while collective consumption expenditure includes all other government spending, including spending on administration and defense. Given that Canada has relatively equal access to public healthcare services and primary and secondary education, we allocate government spending on these categories equally to all individuals. Since access to tertiary education in Canada is related to family income levels, an equal lump sum allocation does not make sense for this form of individual consumption expenditure. We allocate spending on tertiary education and collective consumption expenditure according to the post-tax disposable income distribution, following Piketty, Saez, and Zucman (2018) so that including these expenditures is neutral to the income distribution.

6   Re-ranking

Because we are working from tabulated data rather than raw microdata files, there is the possibility for re-ranking to occur as we move from the pre-tax national income distribution to the post-tax income distributions. Specifically, an individual who is in the top 0.01% of the pre-tax national income distribution who has a particularly high tax bill in a certain year may not be in the top 0.01% of the post-tax national income distribution. To estimate whether this re-ranking is a problem, we compute the ratio of the average after-tax income to the average market income for each income group in Statistics Canada Table 11-10-0055-01, where the after-tax income distribution is sorted by after-tax income and compare that to the ratio we observe in our tabulations where after-tax incomes are sorted by pre-tax fiscal income. We found that there was a small but consistent un-

derestimation of average after-tax incomes in the top decile and so applied an adjustment to the post-tax income distribution proportionate to this observed underestimation.

7   Distributional provincial accounts

One novel contribution of this work is that we present estimates of distributional national accounts at a subnational level. In particular, we estimate distributional national accounts at the provincial level for Ontario, Quebec, Alberta, and British Columbia, as well as the rest of Canada combined. The rest of Canada is combined to ensure a sufficient population size for distributional national accounts to be meaningful. We follow a very similar procedure as at the national level to construct distributional provincial accounts.

The custom tabulations from the LAD are also broken down by province (the three territories are combined), so we begin our provincial estimates from this same tax data. We obtain estimates of all provincial aggregates from 2007 onwards, and most provincial aggregates prior to 2007, from Canada's provincial and territorial economic accounts. We scale all provincial totals to the national aggregate so that the sum of the 11 provincial and territorial aggregates always sum to the national aggregate. For provincial aggregates not available in provincial and territorial economic accounts before 2007 (including property taxes and government saving), we use fiscal data from the Department of Finance to estimate trends in these aggregates in each province. Before creating the provincial pre-tax income distribution, the six smallest provinces and the territories are aggregated as 'Rest of Canada.'

8   Historical series

While the methods outlined above—relying first and foremost on custom, tabulated data from Statistics Canada's Longitudinal Administrative Databank—allow us to estimate detailed distributional national accounts for the years 1982-2020, there is also value in taking a long-run view of the history of economic inequality in Canada.

Saez and Veall (2005) estimated long-run Canadian inequality, starting in 1920, using only fiscal income data; their estimates are similar to pre-tax income, but include only the incomes tabulated by the tax authorities (and thus do not account for all of the income in the national accounts). We connect their historical, pre-1982 series to the current DINA series by matching their estimates of fiscal income inequality for the period of overlap (via LAD data). We make the (perhaps occasionally heroic) assumption that the ratio of fiscal income (aggregates and distribution) to national income (aggregates and distribution) remained constant over time. There is very little difference in the LAD era, but this may not have been equally true during the pre-1982 era. In any case, the long-run DINA series (1920-2020) adds an important perspective on more recent developments.

We will present results from analysis of this data in Xuereb, Fisher-Post, Delorme and Lajoie (2023, forthcoming).

References

Blanchet, Thomas et al. (2021). *Distributional National Accounts Guidelines*. en. Tech. rep., p. 186 (cit. on pp. 3, 4).

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman (May 2018). "Distributional National Accounts: Methods and Estimates for the United States". In: *The Quarterly Journal of Economics* 133.2, pp. 553–609. DOI: 10.1093/qje/qjx043. URL: https://doi.org/10.1093/qje/qjx043 (cit. on p. 4).

Robson, Jennifer and Saul Schwartz (Sept. 2020). "Who Doesnât File a Tax Return? A Portrait of Non-Filers". en. In: *Canadian Public Policy* 46.3, pp. 323–339. DOI: 10.3138/cpp.2019-063. URL: https://utpjournals.press/doi/10.3138/cpp.2019-063 (cit. on p. 2).

Saez, Emmanuel and Michael R Veall (May 2005). "The Evolution of High Incomes in Northern America: Lessons from Canadian Evidence". en. In: *American Economic Review* 95.3, pp. 831–849. DOI: 10.1257/0002828054201404. URL: https://pubs.aeaweb.org/doi/10.1257/0002828054201404 (cit. on p. 5).