Proceedings of Statistics Canada Symposium 2018
Combine to Conquer: Innovation in the Use of Multiple Sources of Data

Estimating annual wealth distributions within Canada's System of National Accounts

Margaret Wu, and Cilanne Boulet[1]

Abstract

In recent years, there has been growing interest in distributional measures of economic well-being. To address this need, Statistics Canada is building a series of annual tables integrating macro-level national accounts data with micro-level survey data on wealth, income and consumption. This product, the Distributions of Household Economic Accounts, adds a distributional component to Canada's macroeconomic accounts, thereby giving a more complete picture of the economic well-being of Canadian households within the national accounts framework.

This article describes the methodology used to build the wealth tables. In the past, Statistics Canada has conducted a survey of wealth only on an occasional basis. One of the major challenges of this project is to find a way to fill the relatively long gaps in the data between survey years. Modelling, calibration, benchmarking, and raking are combined to address the challenges of filling these gaps and ensuring consistency with macroeconomic accounts totals and the survey data.

Key Words: National accounts; Data integration; Calibration; Area-level model; Administrative data; Raking.

1. Introduction

In recent years there has been increasing focus on financial inequalities and financial stability, and correspondingly a greater interest in distributional measures of economic well-being. For Canadian households as a group, data on household wealth has been available for many years through the National Balance Sheet Accounts (NBSA) within the National Accounts. However they lack the distributional component that is important to understanding disparities and vulnerabilities among the Canadian population. The Distribution of Household Economic Accounts (DHEA) is a relatively new product in the System of National Accounts (SNA) that aims to fill this data gap. The DHEA integrates the NBSA with distributional information from the Survey of Financial Security (SFS), a household survey on wealth. The aim is to combine the strengths of both sources: the level of detail from the survey, with the coverage, concepts and international comparability of the NBSA (Statistics Canada 2018).

The DHEA wealth component consists of an annual series of tables on wealth distributed by household characteristics. A preliminary version was first released in December 2016, and regular releases have been ongoing since then. Currently, the tables cover the period from 2010 to 2017, and are based on the NBSA and the two SFS iterations falling within this time period. The wealth tables are available by the following household breakdowns: equivalized disposable income quintile, age range of the major income earner, household type, and province. Table-1-1 below is an example of a single year's table by age group. The total column is from the NBSA and remaining columns are based upon the micro-data source.

This paper will present the methodology of integrating the data sources to build the wealth tables. The methodology implemented follows the step-by-step approach recommended by the Organization for Economic Co-operation and Development (OECD) Expert Group on Disparities within the National Accounts (OECD 2013). Section 2 describes the micro-data and macro-data sources and the two major challenges faced in combining them, Sections 3 and 4 describe the current and past methodologies, and Section 5 presents a comparison of the methodologies.

[1]Margaret Wu, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa ON, Canada, K1A 0T6, (margaret.wu@canada.ca); Cilanne Boulet, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa ON, Canada, K1A 0T6, (cilanne.boulet@canada.ca)

Table 1-1
Sample wealth table for an individual year

	Age group of the major income earner					NBSA Total
	<35	35-44	45-54	55-64	65+	
Total Assets						
Total Financial Assets						
Life Insurance & Pensions						
Other Financial Assets						
Total Non-Financial Assets						
Real Estate						
Other Non-Financial Assets						
Total Liabilities						
Mortgage Liabilities						
Other Liabilities						
Wealth/Net Worth						

2. Data sources

2.1 National Balance Sheet Accounts (NBSA): macro-data source

The NBSA fall within the Canadian System of National Accounts, which is the product of a balancing framework in which data from various sources are combined and confronted. The NBSA consist of a balance sheet of the assets, liabilities and net worth at a given point in time, for each sector of the Canadian economy (household sector, corporate sector, government sector, etc.). This balance sheet is produced on a regular quarterly basis. To the degree possible, the NBSA follow international guidelines on SNA concepts and coverage for reasons of international comparability. The objective of the DHEA is to distribute the wealth totals of the household sector, which cover all Canadian households, among various household classifications of interest, while staying aligned with the SNA framework.

The NBSA at the end of the fourth quarter of each year from 2010 to 2017 are used to build the wealth tables.

2.2 The Survey of Financial Security: micro-data source

The SFS serves as the micro-data source that provides the distributional component. The SFS is a cross-sectional sample survey conducted by Statistics Canada that provides a comprehensive picture of the net worth of Canadians. It combines information collected through personal interviews on the value of all major financial and non-financial assets and on the money owing on mortgages, vehicles, credit cards, student loans and other debts; with additional information on income and pensions from administrative files and other sources through record linkage.

In the past, the SFS was conducted on an occasional basis, in 1999, 2005, and 2012, and starting with 2016, it will be undertaken triennially. The SFS covers the population living in the ten provinces of Canada and excludes certain groups which represent about 2% of the population (for instance, persons living on reserves or other Aboriginal settlements and chronic care patients living in hospitals or nursing homes).

The 2012 and 2016 SFS were used to build the wealth tables from 2010 to 2017 while the 1999 and 2005 SFS were used only for validation of the methods.

2.3 Challenges in integrating sources

One of the main challenges in integrating the sources is in aligning the survey micro-data and the NBSA macro-data so that they are coherent enough to be combined. The SNA and household surveys exist for different purposes and do not necessarily have the same coverage or concepts. For example, the SFS excludes the territories and the

institutionalized population but these are included in the NBSA. An example of where definitions differ is on credit card debt. In the SFS, the concept measured is the balance that is carried from one period to the next, while in the NBSA, the concept is the balance outstanding at a point in time. Integrating the two sources involves a careful examination of the differences and adjustments to align them as well as possible. To facilitate this, the wealth categories were aggregated from the NBSA to a level where there was more coherence with the SFS data in terms of the concepts and the Canada-wide totals.

The second main challenge is missing data, specifically in estimating wealth distributions in the years in which the SFS did not occur. This is addressed through modelling of wealth using auxiliary information as described in the next section. In addition, publication was limited to 2010 onwards during which gaps between survey years were shorter.

3. Methodology for modelling wealth

Two approaches have been used to model net worth and its components: one based upon calibration and another based upon area-level models; and the two approaches use different types of auxiliary data. The calibration approach will be used for the next release in March 2019. Previous releases in 2016 to 2018 were based either on the area-level approach or used a hybrid of both approaches together. More details on this methodology can be found in Statistics Canada (2018).

3.1 Current methodology: a calibration approach

The current methodology to combine the sources and estimate wealth distributions in both the years with and without survey data is based on calibration (Deville and Sarndal 1992), and the idea is similar to static aging. Calibration is a step of the usual survey weighting process and consists of adjusting the weights of the sampled units so that estimates from the survey coincide with known totals at the population level.

The first step of the process is the recalibration of the 2012 and 2016 SFS to population totals for both survey and non-survey years. The calibration totals are estimates of the sampled population based on Statistics Canada's projections of population counts. The totals used include counts of individuals by sex and age group categories, counts of households by household size, and counts of economic families by family size for select family sizes within provinces. These totals are very similar to the totals used in the production of SFS although some adjustments are made to better align with DHEA's released household categories. Recalibrating adjusts the weights of the sampled units so that estimates from the survey coincide with these population totals for non-survey years, in essence adjusting the survey weights to reflect demographic shifts.

In other words, the following sets of weights, $w_{i,y}$, are derived for all households i in the 2012 SFS, for each year $y = 2010, \ldots, 2017$ where $\mathbf{x}_i$ represents a vector of demographic characteristics for household i and $\mathbf{X}_y$ represents the population counts for year y that serve as calibration control totals. The same process is repeated for all households in the 2016 SFS using the same population counts $\mathbf{X}_y$.

$w_{i,2010}$ such that $\sum w_{i,2010}\mathbf{x}_i = \mathbf{X}_{2010}$
$w_{i,2011}$ such that $\sum w_{i,2011}\mathbf{x}_i = \mathbf{X}_{2011}$
…
$w_{i,2017}$ such that $\sum w_{i,2017}\mathbf{x}_i = \mathbf{X}_{2017}$

Once the weights of the SFS samples have been recalibrated to each year, the various components of wealth at the micro-data level are scaled to match the NBSA totals of each year. Putting all of this together, estimates for net worth and its components can be obtained for non-survey years using the scaled survey data and these adjusted weights.

Two series of estimates are obtained for the time period between 2010 and 2017, one using the 2012 SFS and the other the 2016 SFS. The series based on the 2012 SFS is used for estimates for 2010 to 2012. For 2013 to 2015, the series are combined by linearly interpolating between the series. For example, for 2013, the combined estimate is calculated as ¾ × estimate from the 2012 SFS recalibrated to 2013 + ¼ × estimate from the 2016 SFS recalibrated to 2013. The series based on the 2016 SFS is used for estimates for 2016 to 2017.

To mitigate turning points created when combining the two series of estimates, a three-point centered weighted moving average is applied to the combined series in 2012 and 2016. This slightly modifies the estimates derived for survey years. As a final step, raking is applied within each year to re-establish these relationships among the columns and rows that were perturbed during the smoothing step, while minimizing the change to individual cells of the table (Dagum and Cholette 2006).

It is worth noting that the calibration-based approach using a single SFS year only captures trends in wealth that are related to the demographic variables used, and it assumes that these relationships are stable over time. However, by using multiple survey years and combining them together in addition to incorporating the NBSA totals, this method should be able to capture trends more than just those related to demographics.

3.2 Past methodology: an area-level model approach

For the first releases of the DHEA, only one iteration of SFS was available and an area-level model approach was used to obtain wealth distributions in non-survey years instead. This approach was motivated by the area-level models that are a standard small area estimation technique (Rao and Molina 2015). It makes use of the knowledge that there exists a relationship between income and wealth and of the income tax data available from the Canada Revenue Agency.

In particular, Fay-Herriot models were used to estimate wealth using covariates from income tax returns, specifically the T1 Family File (T1FF). An example of one such model is as follows, for estimating net worth within the youngest age group:

$\hat{\theta}_i^{DIR} = \boldsymbol{z}_i^T \boldsymbol{\beta} + b_i v_i + e_i$ for areas $i = 1, \ldots, 49$ where

- $\hat{\theta}_i^{DIR}$ is the direct estimate of the logarithm of total net worth for the area for the youngest age group from the 2012 SFS
- $\boldsymbol{z_i}$ is the vector of income covariates from tax data
- b_i is a constant equal to 1
- $v_i \sim iid\ (0, \sigma_v^2)$ are area-specific random effects
- $e_i \sim iid\ (0, \psi_i)$ are sampling errors, with ψ_i estimated from the 2012 SFS
- the areas i are defined by crossing combined economic regions within the ten provinces with the household category of interest (such as age range)

The 2012 SFS data providing the direct estimates were first scaled at the micro-data level to match the NBSA 2012. The models were then fit on the scaled 2012 SFS data and the income tax data from 2012, and then applied to all years where the auxiliary tax data is available. This gave synthetic estimates $\boldsymbol{z}_i^T \widehat{\boldsymbol{\beta}}$ of the logarithm of total net worth for every area in non-survey years in addition to the survey year 2012. To obtain the distributions required for the wealth tables, which are at higher geographic levels, the estimates of total net worth were summed over the relevant domains, that is, as $\sum e^{\boldsymbol{z}_i^T \widehat{\boldsymbol{\beta}}}$. For example, the estimated total net worth for the youngest age group was the sum of the exponential of the synthetic estimates of log net worth, $e^{\boldsymbol{z}_i^T \widehat{\boldsymbol{\beta}}}$, over all geographic regions within the youngest age group.

To ensure consistency with the SFS, the estimated distributions were benchmarked to the distributions of the 2012 SFS. This step imposed the distributions of the SFS in 2012, while preserving the year-to-year movement in the estimated series as much as possible (Dagum and Cholette 2006). Benchmarking was also a way of dealing with the bias resulting from using a log model. As a final step, raking was applied within each year to re-establish these relationships among the columns and rows since each wealth component was modelled individually.

4. Comparison of approaches

In comparing the area-level model approach against the calibration approach, there are a number of advantages and disadvantages. The main advantage of the area-level model approach is that it can theoretically capture trends in wealth that are related to income. In fact, any other auxiliary data could be used in modelling, as long as it would be

available at the area level. On the other hand, incorporating new auxiliary variables into the calibration approach would require that they be available at the household level which is a more restrictive requirement.

However, there are practical disadvantages unique to the area-level model approach; namely, timeliness and the use of resources. The area-level model approach is much more time and resource intensive as each component must be modelled separately for each household category. This also makes it a more involved process to obtain wealth tables by additional household categorizations, while the calibration-based approach potentially allows for new distributions easily once recalibrated weights are produced. As well, area-level modelling requires timely auxiliary data; in the case of the income tax data, there is a lag of one or two years between when it is available and the release schedule of the DHEA; using this approach means that the last year(s) of the DHEA must be extrapolated in another way.

Another disadvantage of the area-level model approach is that it is not certain that an appropriate model can be found with the available covariates. This occurred with some of the smaller categories of assets and debts, and for these lines the calibration approach was used.

In terms of the wealth estimates, both approaches led to similar estimates on the 2010-2017 period once both the 2012 and 2016 SFS were available. This was true for most wealth categories, and where there were greater differences they were reduced once the SFS distributions were imposed through benchmarking. An example of the differences under the two approaches before benchmarking is shown in Table 4-1 below for net worth by age group. The differences compared to the SFS are comparable in size and direction; the sum of the absolute differences over the age ranges is in orange. Once the final steps of benchmarking and raking were conducted, the difference in the share held by each group was less than 1% in nearly all cells of the tables.

A last interesting note is that it was not clear that area-level modelling was better able to capture the effects of large events, specifically the 2008 recession. Given that the area models use income data, one might expect for example that the area models would show the wealth of the lowest income quintile being affected differently by the recession than the wealth of the highest income quintile. However the estimates were too noisy to distinguish any differences.

Starting with the next release in March 2019, the DHEA will be built using only the calibration-based approach. The practical disadvantages of the area-level model approach are one reason for the change particularly in light of the DHEA's tight production schedule. Another reason stems from the knowledge that wealth generally does not change rapidly. It is accumulated over a lifetime and with the SFS now occurring on a regular triennial basis, the gaps between survey years are not as long as before.

Nevertheless, it is worth nothing that area-level modelling might still be useful if tables are needed for the back period from 1999 to 2009 where the gaps between survey years are longer, or also in the situation where other household distributions are desired that are less related to the calibration variables, assuming that appropriate area-level modelling covariates would be available.

Table 4-1
Comparison of estimates from Survey of Financial Security (SFS), area-level model approach and calibration approach, with sum of absolute differences (orange)

		Distribution			Difference	
		SFS	Area model	Calibration	Area model	Calibration
2016	<35	5.6%	5.5%	5.9%	-0.1%	0.4%
	35-44	12.1%	11.6%	12.2%	-0.5%	0.1%
	45-54	21.6%	22.3%	23.5%	0.7%	1.9%
	55-64	30.9%	29.0%	27.9%	-1.9%	-3.0%
	65+	29.8%	31.6%	30.4%	1.7%	0.6%
	Total	100.0%	100.0%	100.0%	5.0%	6.0%
2005	<35	4.4%	6.1%	6.4%	1.8%	2.0%
	35-44	17.3%	14.5%	15.4%	-2.8%	-1.9%
	45-54	26.6%	27.8%	28.3%	1.2%	1.7%
	55-64	26.8%	25.7%	24.7%	-1.1%	-2.1%

	65+	25.0%	25.9%	25.2%	0.9%	0.3%
	Total	100.0%	100.0%	100.0%	7.7%	8.1%
1999	<35	7.0%	7.0%	7.4%	0.0%	0.3%
	35-44	18.8%	17.8%	18.8%	-1.0%	-0.1%
	45-54	25.8%	28.1%	27.4%	2.3%	1.6%
	55-64	22.7%	23.0%	21.5%	0.3%	-1.2%
	65+	25.6%	24.1%	25.0%	-1.5%	-0.6%
	Total	100.0%	100.0%	100.0%	5.2%	3.8%

References

Boulet, C., and M. Wu (2017), "Annual Household Distribution Tables: Wealth Tables Methodology", unpublished report, Ottawa, Canada: Statistics Canada.

Dagum, E. B., and P. A. Cholette (2006), *Benchmarking, Temporal Distribution and Reconciliation Methods for Time Series Data*, New York: Springer-Verlag.

Deville, J.-C., and C.-E. Sarndal (1992), "Calibration Estimators in Survey Sampling", *Journal of the American Statistical Association*, 87(418), pp. 376-382.

OECD (2013), *OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth*, Paris: OECD Publishing.

Rao, J. N. K., and A. Molina (2015), *Small Area Estimation, 2nd Edition*, Hoboken: Wiley.

Statistics Canada (2018), *Distributions of Household Economic Accounts, estimates of asset, liability and net worth distributions, 2010 to 2017, technical methodology and quality report*.