Chaire de recherche Industrielle Alliance
sur les enjeux économiques
des changements démographiques

CEDIA.ca

SIMUL

A DEMOGRAPHIC AND ECONOMIC MICROSIMULATION MODEL FOR QUEBEC

DOCUMENT TECHNIQUE
TECHNICAL DOCUMENT

Luc Bissonnette, David Boisclair, Nicholas-James Clavet,
Guy Lacroix, Steeve Marchand and Pierre-Carl Michaud

October 2016

Université Laval
Faculté des sciences sociales

HEC Montréal

ESG UQÀM

La Chaire de recherche Industrielle Alliance sur les enjeux économiques
des changements démographiques est une chaire multi-institutionnelle
qui s'appuie sur un partenariat avec les organisations suivantes :

- Centre interuniversitaire de recherche en analyse des organisations (CIRANO)
- iA Groupe financier
- Retraite Québec

SIMUL: A Demographic and Economic Microsimulation Model for Quebec

Technical Appendix[1]

Project team:

Luc Bissonnette, Université Laval
David Boisclair, HEC Montréal
Nicholas-James Clavet, Université Laval
Guy Lacroix, Université Laval, CIRANO
Steeve Marchand, Université Laval
Pierre-Carl Michaud, HEC Montréal, CIRANO, RAND

October 2016

[1]SIMUL is an ongoing project that started in 2010, under the supervision of Jean-Yves Duclos. We want to acknowledge the importance of his work the previous version and thank him for leaving his brainchild to our care. We also want to acknowledge the work of collaborators and graduate students who helped to improve the quality of the simulator. We are particularly grateful to Manuel Paradis and Laure Sébrier for their careful work with the model.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

SIMUL is a dynamic, reduced-form microsimulation model of the Quebec economy developed by Chair team members. It is based on a large cross-sectional sample of the Quebec population, various retrospective and longitudinal surveys as well as official projections regarding future population composition. The model provides a comprehensive view of the economic and socio-demographic evolution of the Quebec population until 2050.[1]

As with all microsimulation models, which contrast with macrosimulation (or simulation of aggregates), SIMUL allows for the analysis of actions and interactions at the individual level, and provides complete distributions – current and future – for variables of interest, such as earnings, retirement incomes, family composition, or receipt of government transfers. The approach also allows to compute macro-type results by aggregating individual-level ones. Each individual in SIMUL follows a life course that is characterized by social, demographic and economic events such as birth, death, migration, household formation and break-up, education, saving, employment and retirement and their associated types of income. The type and frequency of these events depend on the individual's and his household's characteristics. The model is said to be dynamic because it updates all of these individual and household characteristics over several periods of time.

The model's most attractive feature therefore lies in its dual "distributive" and "accounting" natures (for private and public finances in the latter case). Indeed, SIMUL can project public expenditures and revenue while accounting for the distribution of socio-economic variables within the population. Finally, SIMUL may be used to assess the impact on public and private finances, as well as on the current and future distribution of socio-economic variables, of an existing or prospective economic environment or of specific public policies.

Other microsimulation models have been built and used in Canada over time. The Social Policy Simulation Database and Model (SPSD/M) is a simulation-ready database offered by Statistics Canada that comes with its own "static accounting model", the SPSM. The latter "processes each individual

[1] The current version of SIMUL has been built for all of Canada and is fully functional for Canada as a whole. However, some elements such as public finances have not been implemented in provinces outside Quebec. For this reason, this document only discusses the case of Quebec – while sometimes referring to Canadian variables or aggregates, such as the Canada Pension Plan. In most cases, though, technical descriptions and modelling presented should be construed as applying to Canada, adapted as/where necessary.

and family on the SPSD/M, calculates taxes and transfers using legislated or proposed programs and algorithms, and reports on the results".[2] Another model built and used at Statistics Canada is the Demosim demographic model, which is used to generate population projections.

Perhaps the most recent and sophisticated model is LifePaths, which was built at Statistics Canada with support from various federal departments (mostly by that now known as Employment and Social Development Canada). Like SIMUL, LifePaths is a *dynamic* microsimulation model, meaning that individuals' behaviours and life events are updated over time in an economically consistent fashion. However, the LifePaths project and model are now discontinued; some researchers may keep using the archived version, but it is no longer updated or supported.[3] LifePaths and SIMUL are somewhat similar models (e.g., both are "open population", meaning that individuals have synthetic families that are not formally part of the model), but there are major differences: for instance, the former is in continuous time, while the latter offers a more detailed modelling of labour force participation and of provincial taxes and transfers. Another important difference is that SIMUL is based on the "real" population of the 2011 National Household Survey.

1.1 Model Structure

This version of SIMUL is written in FORTRAN. It comprises four modules interacting with each other. The first one is the initialization module, used to create the initial dataset, load data in the simulation and set the simulation parameters. This initial population has individual characteristics, such as gender or education, representative of the Quebec population. This module is presented in details in Chapter 3. The second module, on demography, deals with demographic transitions relevant to the simulations. This module is presented in Chapter 4). Other economic transitions are simulated using the third module, described in Chapter 5). Finally, SIMUL is linked to our taxes and transfers simulator, SIMTAX. This independent "fourth module" is only briefly introduced in this document, in Chapter 6, but detailed documentation is available elsewhere (see Marchand et al., 2016). SIMTAX may indeed be used with other microsimulation models, or independently with any microdata that provide income input to the simulator.

Figure 1.1 can provide a basic understanding of a simulation done with SIMUL. The figure illustrates what happens every year in the simulation after the initial dataset is loaded and the simulation begins. Colour on the chart indicates which module is responsible for which transitions, and the corresponding section(s) or chapter(s) in this document are provided next to the transition's description.

[2] Excerpt from Statistics Canada's website, available at http://www.statcan.gc.ca/eng/microsimulation/spsdm/spsdm.
[3] Information about LifePaths from Statistics Canada: http://www.statcan.gc.ca/eng/microsimulation/lifepaths/overview.

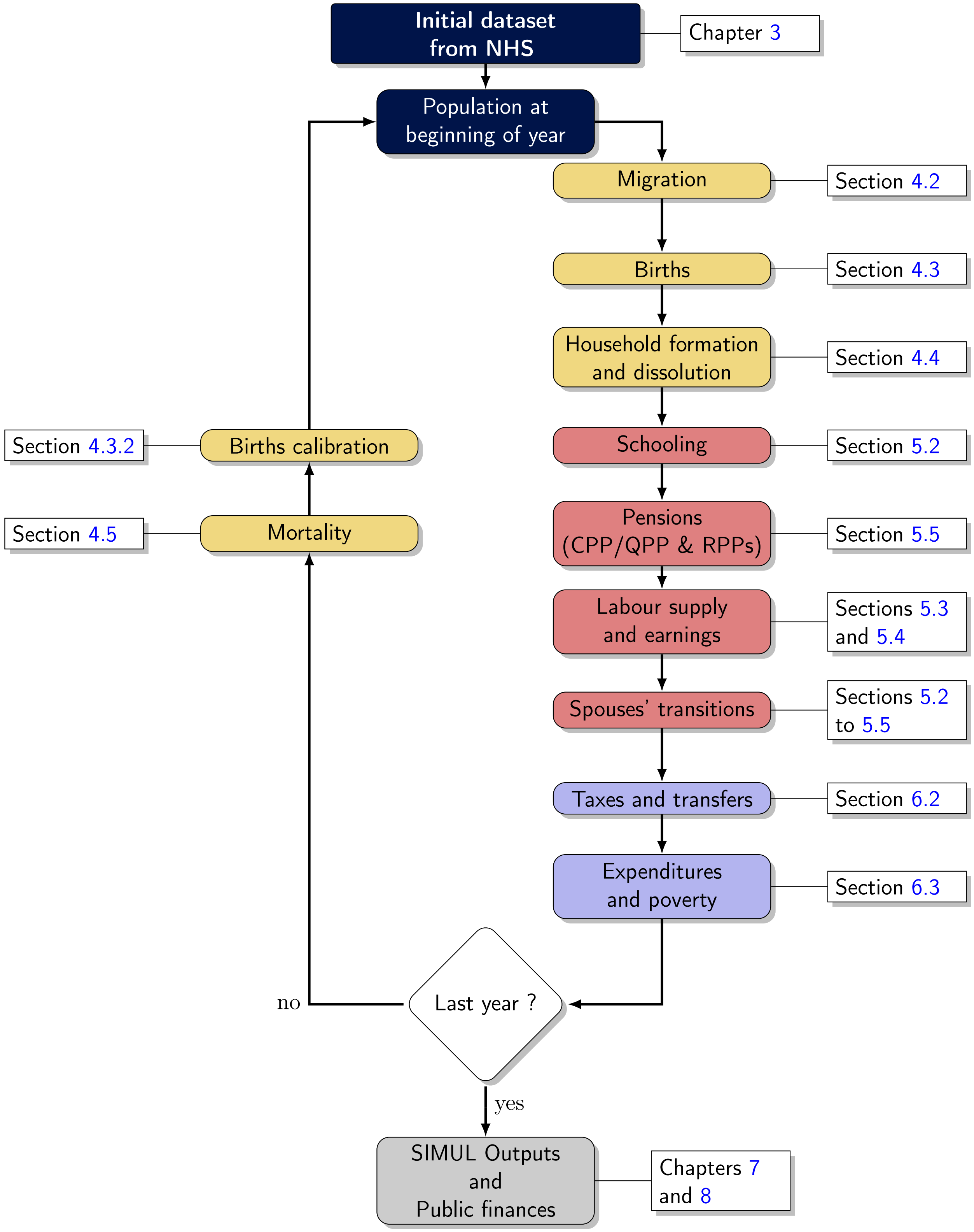

Figure 1.1: Overview of a simulation. Each step is coloured according to the relevant module: dark blue for the initialization module (Chapter 3), yellow for the demographic module (Chapter 4), red for the other transitions module (Chapter 5) and light blue for the taxes module, the SIMTAX simulator (Chapter 6). Gray refers to the outputs (Chapters 7 and 8).

Chapter 2

Data

2.1 Summary

This chapter presents the Statistics Canada data used in the construction and estimation of SIMUL. The 3 main databases are the *Longitudinal and International Study of Adults* (LISA), the *General Social Survey* (GSS), and the *National Household Survey* (NHS). The former two are used to estimate period transitions while the latter is used to set the initial population. The following chapters provide further details about complementary data sources.[4]

2.2 Longitudinal and International Study of Adults (LISA)

The Longitudinal and International Study of Adults (LISA) was conducted for the first time in 2012. The survey includes several self-reported socio-economic variables. This information is supplemented with administrative data on income derived from the Canadian Revenue Agency tax files, for 2012 but also for previous available years (1983-2011). These retrospective administrative data, along with the 2012 survey data, allow us to create a longitudinal database which we use in order to estimate SIMUL's dynamic income and employment models. A second survey wave is now available (for reference year 2014), but is not used in SIMUL because the 2012 wave was the only one available when SIMUL's models were estimated.

The survey covers the population living in the 10 provinces and excludes the 3 territories. It also excludes persons living on reserves and other Aboriginal settlements; official representatives of foreign countries and their families living in Canada; members of religious and other communal colonies;

[4]Some of the analyses presented in this document were conducted at the Quebec Interuniversity Centre for Social Statistics (QICSS) which is part of the Canadian Research Data Centre Network (CRDCN). The services and activities provided by the QICSS are made possible by the financial or in-kind support of the Social Sciences and Humanities Research Council (SSHRC), the Canadian Institutes of Health Research (CIHR), the Canada Foundation for Innovation (CFI), Statistics Canada, the Fonds de recherche du Québec - Société et culture (FRQSC) and the Quebec universities. The views expressed in this document are those of the authors, and not necessarily those of the CRDCN or its partners.

members of the Canadian Armed Forces stationed outside Canada; and persons living full-time in institutions.

2.2.1 Age and Birth Year

The age variable in our longitudinal data is constructed from the self-reported age in the 2012 survey. The birth year variable is simply 2012 minus this self-reported age.

2.2.2 Education

The education variable specifies the highest diploma completed and takes 4 possible values: no diploma, high school, college, or university. It is self-reported and derived from the 2012 survey. This could be a problem for those individuals who had not yet completed their schooling in 2012. However, we remove from the database the observations for the periods where individuals are considered students, since we use the LISA data only to estimate SIMULS's income and employment models. The "student" variable is constructed in our longitudinal data using the 2012 survey's self-declared year of graduation. The individual is considered a student until his or her graduation year, and a worker after that. If no year of graduation is specified, the individual is removed from the database completely.

2.2.3 Couple

LISA's administrative data indicate the individual's marital status for each year in the longitudinal data, because this information is required in the tax return. For each person-year, we give the variable "couple" a value of 1 if the person is married or in a common law union and of 0 otherwise.

2.2.4 Child Under 5 Years Old

This variable is constructed from the variable indicating the age of the youngest child in 2012. We first reconstruct the age of this youngest child for each year in the longitudinal data. Then, the variable is set to 1 for each year where this child's age is lower than 5, and to 0 otherwise.

2.2.5 Province

We construct the "province" variable based on the self-reported province of residence in the 2012 survey. It is assumed to be constant for all years. We build a categorical variable with 5 categories:

- Atlantic provinces (NB, NL, NS and PE);
- Quebec;
- Ontario;
- Prairies (AB, MB and SK);

- British Columbia.

2.2.6 Earnings

The earnings variable in LISA is obtained from administrative records. Because SIMUL uses real (2010) dollars, we convert earnings in LISA to 2010 dollars using the Bank of Canada Canada's Consumer Price Index (CPI) before using them. The variable includes total earnings declared on the T4 form; other employment incomes; net business incomes; net professional income; net commission income; net farming income; and net fishing income. Table 2.1 presents our earnings variable's deciles (i.e. the upper cutoffs for each category), for all years available in LISA.

Table 2.1: Earnings deciles in LISA (2010 dollars)

	Canada		**Quebec**	
Decile	**30 years**	**50 years**	**30 years**	**50 years**
1	0	0	0	0
2	6,525	10,694	4,271	7,870
3	14,094	22,053	13,171	19,434
4	22,607	31,724	19,629	28,668
5	30,717	39,770	27,009	36,577
6	37,378	49,700	34,091	45,179
7	44,053	61,959	38,973	54,641
8	52,721	75,865	47,962	67,324
9	65,364	10,1360	58,886	88,403

2.2.7 Employment

LISA does not include an employment status variable, so we define employment on the basis of earnings. Our employment variable is equal to 1 if an individual's earnings, in 2010 dollars, are $5,000 or more. Figure 2.1 shows LISA's employment rates by age, where all cohorts have been pooled. Employment rate for men is stable at 0.9 until age 55, while it peaks at 0.75 around age 50 for women.

2.2.8 CPP/QPP Income

The variable for Canada Pension Plan/Quebec Pension Plan (CPP/QPP) income in LISA is built from administrative records. We convert the values in 2010 dollars using the CPI. We also construct a dummy variable indicating whether the variable is greater than 0. Figure 2.2 presents the proportion of individuals who start receiving CPP or QPP benefits at each age, conditional on not having started to receive them in the previous year.

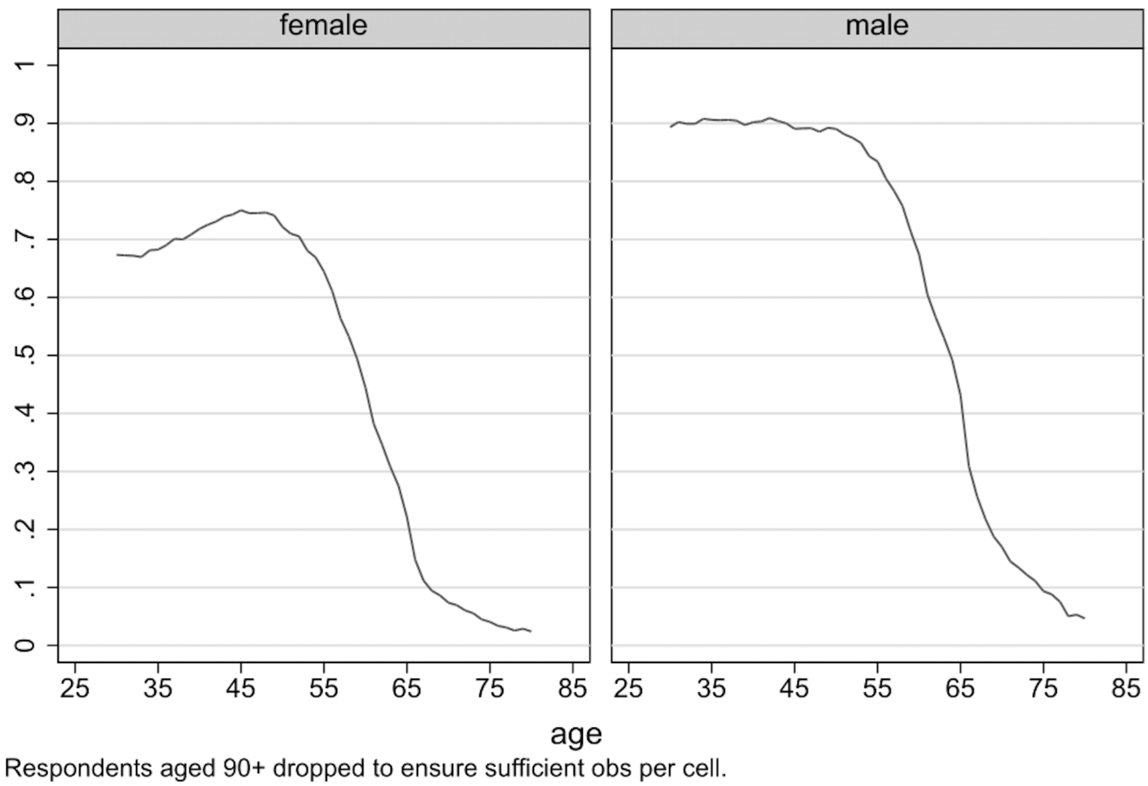

Figure 2.1: Employment rates in LISA, by age and gender, all cohorts pooled

2.2.9 Other Private Retirement Income

This variable is the sum of Supplemental Pension Plan and Registered Pension Plan (RPP) incomes, as provided in LISA's administrative records. This includes income received from Registered Retirement Income Plans (RRSPs) and Registered Retirement Income Funds (RRIFs), but not other investment income such as returns earned in a Tax-Free Savings Account – or TFSA – or in non-registered accounts. We convert the sum to 2010 dollars using the CPI. We also construct a dummy variable indicating whether the variable is greater than 0.

2.3 General Social Survey (GSS)

The *General Social Survey* contains retrospective questions regarding important life events. We rely on two waves of this survey (2006 and 2011) to estimate the probability of four events: child birth, schooling completion, couple formation, and couple separation.

From a modelling perspective, these events are considered interdependent their realization appear as explanatory variables for each other. For instance, those who obtain their final degree also have a greater probability of becoming parents.

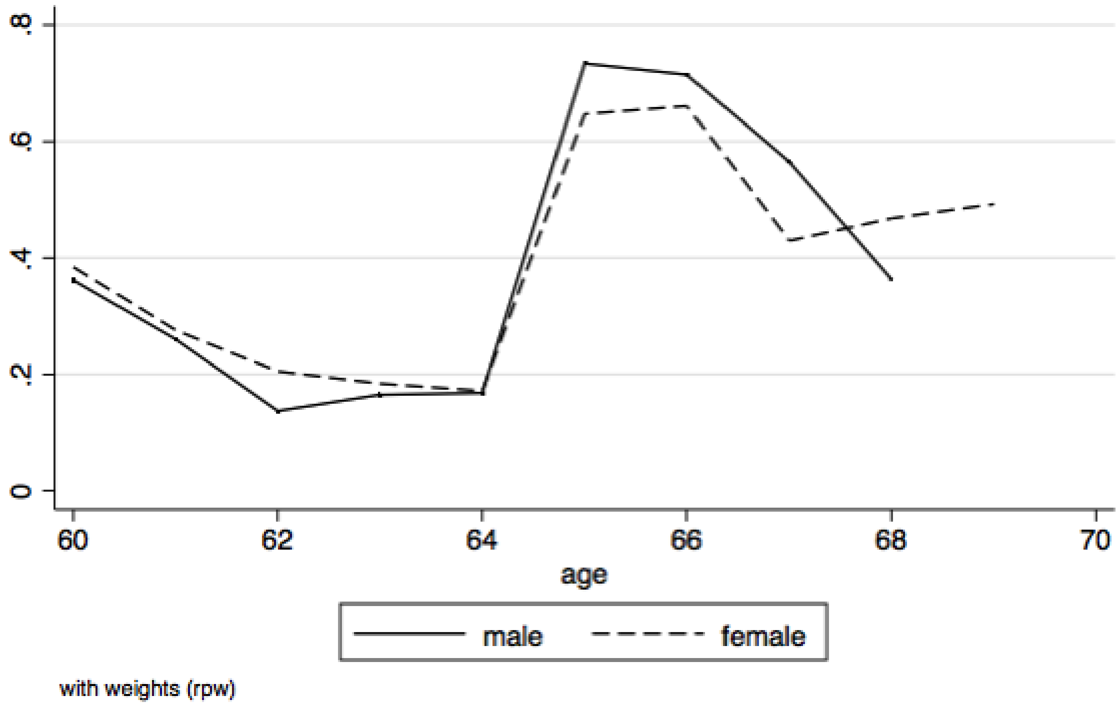

Figure 2.2: Proportion of respondents beginning to receive CPP/QPP benefits in LISA, by age, conditional on not receiving any in the previous year

The approach used for these transitions is to reconstruct the state of individuals at every age since they were 16. For instance, suppose we observe a woman aged 55 who answered all the questions concerning past unions and the age at which she had her children. From this information, we can recreate 40 time periods for which we can determine whether that woman was in a couple, formed a new union or separated or gave birth, as well as the age of her children. We can use this information to estimate the transition probabilities for the four aforementioned life events.

2.3.1 Children Transitions (Giving Birth)

The GSS contains information on the age of the respondent's children. This information is available in the variables *agechdc?* for 2006 and *agechd?c* for 2011, where ? is a token replaced by a digit corresponding to the child's position in the family. It is straightforward, based on the age of the children, to determine the years in which a mother gave birth. Based on this information, it is possible to determine whether a woman already had children when she gave birth. If she did, it is also possible to determine the age of the youngest among these previously born children.

SIMUL limits the number of children to 6 per family. Transitions were estimated accounting for that limitation.

2.3.2 Marital Status Transitions

Marital status transitions are a little more complicated and require some *ad hoc* assumptions. The GSS does not differentiate between civil union and marriage. In SIMUL, common law partners are also considered as being in a union. Information on all unions and their end is found in variables *age_ma*?, *agesema*? and *agedima*? in 2006, which were respectively renamed *age_ma?c*, *agesepm*? and *agedivm*? in 2011.

This information is used to create union spells and to determine years during which respondents were single. It is generally possible to determine if a respondent had divorced and formed a new union in the same year. We also use the information contained in the aforementioned variables to determine whether a common law union became a civil union or a marriage. In such cases, the couple is treated as having had a single spell.

We require respondents to be part of a couple in order to have children. This leads to some inconsistencies in the data. In such cases, the child is connected to the nearest union, assuming that this union took place within two years and giving priority to previous unions in cases where two spouses were at equal distance in time. If no such unions are present, we simply change the marital status for the period during which a woman gave birth.

The aforementioned variables also allowed us to determine how many unions respondents have had over time.

2.3.3 Education Transition

This variable is used to determine the age at which a respondent ended his schooling. This information is available in the variable *agecmplt*. If it has a value of 99.5, then the respondent is still studying. We assume that respondents are still students even if in practice, they might not be actively studying. The important variable here is whether someone is done with studying. SIMUL assumes that every respondent stops studying no later than age 35.

2.4 National Household Survey (NHS)

The *National Household Survey* (NHS) was implemented in 2011 as a complement to the Canadian Census, following the elimination on the census long form (B form) in 2010. Since the long form has been restored for the 2016 Census, the NHS is a one-wave survey. It was sent to almost a third of the Canadian population with a response rate of 68.6%. It contains an important amount of demographic, social and economic information. We use it to set the initial population of SIMUL.

The NHS comes in two versions: the Individuals file and the Hierarchical file. Most of the information comes from the Hierarchical file (age, gender, province of residence, etc). The missing part has been filled using the Individuals file. The latter was mainly used to obtain CPP/QPP benefits. A useful feature of the NHS is that most of its variables are similar to those found in LISA. Models estimated from LISA are therefore easily transferred to the initial population of SIMUL.

Most variable manipulations with the NHS are covered in Chapter 3.

Part I

Modules

Chapter 3

Initialization

3.1 Unit of Simulation

The unit of simulation in SIMUL is individuals, not families. However, each of these respondents has a *ghost family*, meaning that their spouse and children's characteristics are characteristics of this individual. Various individuals are not linked in the datasets.

In practice, this means that individuals will form unions and have children, but that their spouse and children are not accounted for when aggregating values for analysis. Family membres are simply treated as observable characteristics. These are useful when determining the probabilities of some transitions, or when computing a proportion of taxes and transfers. However, whenever there is a separation, the ghost spouse disappears. Similarly, when a new union is formed, spouse characteristics are randomly generated, as we do not seek to match two respondents from the dataset (see Section 4.4.1 for details on this).

3.2 Initial Dataset

The initial dataset for the simulation is derived from the NHS (2011) public use microdata files. The public files are available in two versions: the Individuals file, containing detailed information on individuals, and the Hierarchical file, containing information about households. Statistics Canada describes the latter as follows:

> This PUMF product provides access to non-aggregated data covering a sample of 1% of the Canadian households. It is a comprehensive social, demographic and economic database about Canada and its people and contains a wealth of characteristics on the population. The file enables the study of individuals in relation to their census families, economic families and households. The geographic identifiers have been restricted to the provinces, the three territories grouped into a region called Northern Canada, and selected metropoli-

tan areas (Toronto, Montréal, Vancouver, Edmonton and Calgary) to ensure respondents' anonymity.[5]

Given that household characteristics are of prime importance for demographic, economic and fiscal microsimulations, the initial population is mainly based on the Hierarchical file. The use of this dataset as a basis for the simulated population presents both advantages and disadvantages. As previously mentioned, the Hierarchical file is very rich in household characteristics. Table 3.1 shows the information available in the NHS that was imported in SIMUL. Some important characteristics, such as CPP or RPP benefits, were only available in the Individuals file and could not be used for our simulation. The major problem stems from the fact that we have to rely on the age categories rather than precise ages, which is not practical when studying retirement income (as explained in detail in Sections 5.5.1 and 5.5.2).

[5] See https://www12.statcan.gc.ca/nhs-enm/2011/dp-pd/pumf-fmgd/index-eng.cfm.

Table 3.1: Characteristics of the initial population and the source of information

Variable	**Source**	Variable name	**Note**
Household			
Province of residence	NHS (H)	*pr*	
Sample weight	NHS (H)	*weight*	
Number of children in household	NHS (H)	N.A.	Based on other household members.
Chidren's age	NHS (H)	*agegrp*	Age of other household members
			categorical in data, imputed within category.
Individual Respondent and Spouse			
Age	NHS (H)	*agegrp*	Categorical in the data; imputed within category.
Birth year	NHS (H)	N.A.	Follows from imputed age.
Gender	NHS (H)	*sex*	
Education	NHS (H)	*hdgree*	Recoded in a four-category scale.
Parents' education	NHS (H) and GSS	*hdgree*	Based on other members in household.
			Frequently missing; imputed with
			multinomial logit (using GSS) when unknown.
Done with schooling?	NHS (H)	*attsch*	Individuals over 35 y.o. cannot study anymore.
Worked in past year?	NHS (H)	*lftag*	
Migrant status	NHS (H)	*rimm*	
CPP/QPP benefits	NHS (I)	*cqppb*	Information on claiming decisions imputed.
			in initial dataset, benefits imputed in SIMUL.
RPP benefits	Imputed in SIMUL	N.A.	Claiming and benefits imputed in SIMUL.

Additional benefits of using the NHS include the possibility of using a unique source of information for migrants and non-migrants. Details on the elaboration of a migrants dataset is provided in Section 3.4.

Restricting our analysis to the province of Quebec, the initial population consists of 75,330 individuals entering the model in the first period. When sample weights provided by Statistics Canada are applied, the total size of the population is over 7.8 million individuals.

3.3 Details on Imputation

3.3.1 Age

The procedure used to impute initial age is the same as the one described in the COMPAS model's documentation (cf. Boisclair et al., 2016). The remainder of this section is an excerpt from this documentation describing the procedure. The difference in SIMUL is the survey used to build the initial population, which is the NHS instead of the CCHS.

> The public use CCHS classifies individuals into five-year age groups until 79 years of age and provides an open age category for individuals aged 80 years and over. We rely on public-use CCHS data to construct the initial database to ensure that we can initialize the model at will outside the walls of a Statistics Canada Research Data Centre (RDC). Using restricted-access data would mean that the user would be forced to run the model within a RDC. To overcome this problem, we attribute a precise age to each individual within his/her age group. In order to mirror the age distribution within an age group, we use the 2011 Census to calculate a probability for each precise age conditional to the membership of an age group. Thereafter we draw a random value from a uniform distribution to assign a specific age.

3.3.2 CPP/QPP Benefits

Some of the respondents had already claimed their CPP benefits when entering the survey. Unfortunately, this information is not available in the Hierarchical files, while it is available in the Individuals files. We must therefore rely on imputations to determine whether these respondents receive their benefits, and the amount that they receive.

To make matters worse, the age categories from the two files do not correspond. In order to impute CPP benefits, we use the imputed age variable and reconstruct the category presented in the Individual files.

Two additional problems arise when attempting to use this information in the model. First, given that we imputed age without conditioning on CPP/QPP benefits receipt, we cannot make sure that respondents of a given age are allocated a consistent value for CPP/QPP benefits. This is very important when considering retirement, as there are focal ages at which individuals are more likely to claim their benefits. Assigning an individual the wrong age could increase or decrease the real amount that he would receive. Second, we cannot disentangle where the CPP/QPP benefits comes from. It is therefore not possible to determine if part of the benefits comes from a survivor's pension or an disability program.

In order to have time-consistent imputation of CPP/QPP in the simulation, it was decided that every individual would receive CPP/QPP benefits according to simulations based on the stochastic process discussed in Section 5.5.1. If an individual is eligible to this pension, SIMUL generates an earnings history based using the labour and earnings model described in Sections 5.3 and 5.4. The resulting value is then used through the simulation.

3.3.3 Parents' Education

Parents' education is a variable that is used to predict a child's educational attainment. It is therefore important only for the younger respondents. Based on the family structure of the Hierarchical file, this information was often available in the NHS files. However, it turned out to be frequently missing, mostly for young adults not living in their parent's home. We did not drop the observations where this information was missing, and instead imputed the parents' education during the simulations if this information was needed. The imputation is based on a multinomial logit model[6] with 16 categories (both parents' education is estimated at once to control for the correlation) that was estimated with data from the GSS - in which information on the mother's and the father's education is available in variables *edum5* and *eduf5*, respectively. We recoded these variables in a four-category scale used throughout the model, in order to use the same data for estimating parents' and own education. Imputation is based simply on the child's birth year and province of residence.

The table of output for this regression is very large and is not reported here, but is available upon request.

3.3.4 Initial Marital Status

The transition models presented in Section 4.4 use the duration of the union and the number of unions as explanatory variables for the transitions models. This information is not available in the NHS. Imputation was used to determine how many unions the respondent lived in as well as the duration of the current marital status.

Even more important, the couple models rely on the number of previous unions. This was imputed based on observations from the GSS. We allow for a maximum of 2 unions per person over the life course. Hence, if we observe an individual entering the dataset as a single, there are three possible states: was never in couple, was once in a union before, was twice in a union before. An individual entering the model already in a couple can be in two states: either in a first or a second union.

Imputation of the initial state is based on a multinomial logit conditional on the marital status upon entry. As explained in appendix A.1.2, the multinomial logit for a case with two alternatives, which is the case if one is first observed in a union, is simply the logit model.

Estimated models are reported in Table 3.2.

[6] See Section A.1.2.

Table 3.2: Estimated parameters for the model of initial number of unions

	Initially single		Initially in couple
	In couple once ?	In couple twice ?	Second union ?
Birthyear	-0.021	0.003	0.034
Québec	0.242	0.455	0.204
British Columbia	0.281	0.510	0.344
Alberta	0.347	0.474	0.084
New Brunswick	0.397	0.477	-0.102
Prince Edward Island	0.253	0.519	-0.104
Manitoba	0.241	-0.428	-0.044
Saskatchewan	0.058	-0.390	-0.145
Nova Scotia	0.159	0.256	-0.008
Newfoundland and Labrador	-0.210	-0.931	-0.646
Splines for age 24 or less × male	0.597	0.904	0.271
Splines for age 24 to 35 × male	0.116	0.249	0.105
Splines for age 35 or more × male	-0.065	-0.043	0.031
Splines for age 24 or less × female	0.564	0.886	0.257
Splines for age 24 to 35 × female	0.163	0.253	0.126
Splines for age 35 or more × female	-0.050	-0.001	0.038
Constant	-15.468	-27.043	-10.887

Note: Ontario is the province of reference.

Note: Birthyears were normalized by subtracting the minimum value (1912).

The duration of the current union is estimated based on observations from the GSS. It is not necessary to impute the duration of marital status for those who were never in a union (as this is simply age, already controlled for) or for those who already dissolved two unions. We therefore need three equations for the imputations: for those currently in their first union, those separated from a first union, and those in a second union. Estimates for this model are reported in Table 3.3.

Table 3.3: Estimated parameters for the initial duration of marital status

	First union	Divorced once	Second union
Birthyear	-0.082	-0.005	-0.005
Québec	0.376	0.234	-0.203
British Columbia	-0.019	0.234	0.304
Alberta	0.618	0.387	0.463
New Brunswick	0.898	0.882	-0.341
Prince Edward Island	0.546	-0.438	0.682
Manitoba	0.713	0.538	-0.317
Saskatchewan	0.596	0.163	0.540
Nova Scotia	0.508	-0.053	-0.071
Newfoundland and Labrador	0.631	-0.143	-1.107
Splines for age 24 or less × male	0.348	0.267	0.210
Splines for age 24 to 35 × male	0.774	0.356	0.413
Splines for age 35 or more × male	0.994	0.369	0.774
Splines for age 24 or less × female	0.311	0.256	0.196
Splines for age 24 to 35 × female	0.644	0.313	0.337
Splines for age 35 or more × male	0.970	0.296	0.718
Constant	-1.759	-3.534	-2.547

Note: Ontario is the province of reference.
Note: Birthyears were normalized by subtracting the minimum value (1912).

Predicted value is rounded and censored such that marital status duration is an integer between 0 and the respondent's current age minus 18.

3.4 Initial Migrants Database

Our simulations allow for inflows of migrants, as described in detail in Section 4.2. We assume that the characteristics of migrants do not change over time. We therefore construct a dataset of migrants that will be added every year and re-weighted to match the assumed number of new migrants. This dataset contains all the respondents that reported arriving in the last 5 years.

As previously mentioned, the NHS is also a rich source of information regarding new migrants. It is therefore possible to extract information about migrants based on the variables defined above, but there is one challenge. As mentioned, age in the dataset is built from a categorical variable. Age at immigration (variable *ageimm*) is also a categorical variable. The procedure used to select who is a migrant is as follows. We first assign an individual an age based on the procedure described in

Section 3.3.1. Once this age is obtained, we determine if the age of a respondent falls in the age category defining age at immigration. Ideally, we would want to rely on the sample of migrants who arrived in the last 5 years. We therefore subtract 4 from the randomly assigned age and see if this value would be compatible with the immigration age defined in the *ageimm* categorical variable. This procedure allows us to create a sample of 2,123 individuals who will serve as new migrants in every year of the simulation, as described in detail in Section 4.2.

As previously mentioned, new migrants' families have characteristics sampled in the NHS. However, we enforce that new migrants do not already receive CPP/RPP benefits.

Moreover, in order to have a nationally representative sample of migrants, the selected population is reweighted to reflect the Canadian immigration rates observed in 2011.

3.5 A Note on Backward Imputation

It is to be noted that some of the transition probabilities in the following sections rely on information about an individual's history that was not available in the initial dataset. The simplest example is the computation of the value of CPP/QPP benefits received by an individual, as described in Section 5.5.1. This value must be computed based on the complete earnings history.

Whenever information about an individual's past is needed, we use the relevant SIMUL model and try to go back in time to impute the relevant missing value. For instance, when earnings history is needed for an individual claiming CPP benefits, the model goes back in time and generates an earnings history based on the same economic specification that is presented in Section 5.4. Specific additional hypotheses required for imputations are discussed in the relevant sections.

Chapter 4

Demographic Transitions

4.1 Summary

SIMUL's second module concerns the demographic transitions experienced every year. The module randomly generates births and is responsible for couple formations and dissolutions. As discussed in Section 2.3, these transitions are mostly estimated with the 2006 and 2011 waves of the GSS. The module also randomly determines which individuals die in a given year.

The demographic module also ensures that the population in the model remains representative of the Canadian population over time. Every year, the new births and new migrants are weighted to be consistent with official demographic projections made by Statistics Canada (2010), according to the medium growth scenario based on trends from 1981 through 2008. It is of note that differences between official Quebec projections (see Institut de la statistique du Québec, 2014) and population projected in this manner by SIMUL are very small.

4.2 Immigration and Emigration

Every year, immigrants are added to the data. Emigration is not modelled explicitly; rather, it was decided to add a net inflow of new migrants. Each province has an immigration target as a percentage of its current population. The migrants from the initial dataset described in Section 3.4 are added and proportionally reweighed to match this target. The underlying assumption is that the composition of this population of new migrants is constant over time.

Migration rates are taken from Statistics Canada's population projections by province. The rates used are those of the "medium" scenario.

4.3 Births

There are two different aspects of the births simulation. On the one hand, we want to determine who will be the parents of the newborn in the dataset. On the other hand, new observations must be added to the dataset while keeping the sum of their weights in line with the projections made by Statistics Canada. Each aspect is are covered in turn here.

4.3.1 Giving Birth

The probability of giving birth is estimated based on the GSS, as discussed in Section 2.3.1. This source of information contains retrospective data on the respondents. These respondents report the age of all their children, which allows to estimate a model based on actual children and circumvent the problem of mixed families where individuals live with their spouse's children or the problem caused by children who flew out of the coop.

Only couples can have children. We limit the number of children in a household to six. We use a different set of parameters for the probability of having these six children. We therefore estimate six different logit-type transition models (see Section A.1.1) depending on how many children are already in the household. As expected, a woman is less likely to give birth to a fifth child than to a second one, everything else equal. On the other hand, however, most women are more likely to give birth to a second child than to a first one.

If the main individual in the couple is a woman, then the child is added as a new observation to the dataset. Its parents' characteristics, such as province of residence or education attainment, are passed on to him, but that child will not keep a link to his parents later in the simulation. For instance, even if his parents die, a child never becomes an orphan and lives a simulated life of his own after the initial period. Should one of the parents still be a student, the education level of this parent is considered unknown and will be imputed using the model described above.

Note that there are only single births in the model.

4.3.2 Adding and Weighting New Births in the Population

From an aggregate perspective, there is a risk that our simulation would randomly generate too many or too few newborns in a given year, which would generate undesirable noise in the following time period. In a similar manner, a random gender imbalance could be simulated in the model, but is highly unlikely in reality. In order to enforce a population growth that is consistent with the official demographic forecasts, the newborn individuals who are added to the model are reweighed by gender based on Statistics Canada's population projections ("medium" scenario).

4.4 Household Formation and Dissolution

At each time period, singles can form new couples and existing couples may separate. Remember that partners are *ghost* individuals in the model. Hence, singles finding new partners do not have to find a

Table 4.1: Estimated parameters for the probability of giving birth

	Number of kids					
	0	1	2	3	4	5
Youngest child's age	0.000	-0.038	0.004	-0.004	0.078	0.053
High School	-0.129	0.036	-0.093	-0.079	-0.285	0.078
College	-0.282	-0.052	-0.026	-0.507	-0.086	0.090
University	-0.202	0.197	0.122	0.262	0.097	0.239
Done with schooling	0.446	0.088	0.074	0.192	0.412	-0.847
Québec	-0.186	-0.099	0.072	-0.261	-0.392	-0.969
British Columbia	-0.155	-0.053	0.172	-0.014	-0.013	-0.416
Alberta	0.029	0.025	0.218	-0.155	-0.072	-0.292
New Brunswick	-0.104	-0.021	-0.727	-0.105	0.227	-1.087
Prince Edward Island	0.179	0.080	0.143	-0.567	-0.349	0.000
Manitoba	-0.024	0.030	0.182	0.175	0.735	0.262
Saskatchewan	0.061	0.127	0.328	0.142	0.239	0.080
Nova Scotia	-0.093	-0.053	-0.025	-0.490	-0.316	-0.765
Newfoundland and Labrador	0.100	-0.334	-0.448	-0.730	-1.645	0.000
Dummy: age between 25 and 29	-0.066	-0.000	-0.517	-0.506	-1.355	-0.835
Dummy: age between 30 and 34	-0.102	-0.055	-1.051	-1.143	-2.484	-1.531
Dummy: age over 35	-0.982	-0.682	-2.000	-1.984	-3.608	-2.888
Constant	-1.845	-1.353	-1.657	-1.552	-1.095	-0.473

Note: Ontario and "less than high school" are the reference categories.

match within the dataset and the characteristics of the partner are simulated accordingly, as described below. Partners who separate simply disappear from the dataset. Children are assumed to stay with their mother; hence, a male individual's ghost children are also removed from the dataset.

The probabilities of forming a couple or separating are based on logit models (see Section A.1.1). We allow individuals to form up to two unions during their lifetime and to separate twice. Sequentially, separation comes first in a year. It is therefore possible that an individual will separate and form a new couple within the same year, but not the other way around (i.e. form a union and dissolve in the same year).

Table 4.2: Estimated parameters for the probability of forming or dissolving a union

	Form 1st	Dissolve1st	Form 2nd	Dissolve2nd
Birthyear	-0.009	0.029	0.038	0.002
Time since beginning of status	0.000	0.378	0.025	0.061
High School	-0.038	0.134	0.056	-0.153
College	0.026	0.144	0.106	-0.264
University	-0.050	0.475	0.110	-0.302
Done with schooling	0.752	-0.594	0.078	-0.168
Québec	0.145	0.159	0.083	0.120
British Columbia	0.058	0.208	0.047	0.122
Alberta	0.180	-0.165	-0.040	-0.240
New Brunswick	0.170	-0.313	-0.093	-0.223
Prince Edward Island	0.118	-0.070	-0.105	-1.052
Manitoba	0.098	-0.462	-0.135	-0.258
Saskatchewan	0.100	-0.225	-0.113	-0.580
Nova Scotia	0.050	-0.088	-0.044	0.165
Newfoundland and Labrador	0.064	-0.860	-0.061	0.211
Splines for age 24 or less × male	0.230	-0.206	0.169	0.048
Splines for age 24 to 35 × male	-0.153	-0.364	0.004	-0.038
Splines for age 35 or more × male	-0.107	-0.384	0.031	-0.058
Splines for age 24 or less × female	0.208	-0.187	0.164	0.076
Splines for age 24 to 35 × female	-0.053	-0.331	0.024	-0.098
Splines for age 35 or more × female	-0.147	-0.380	0.030	-0.048
Constant	-7.017	-1.001	-7.894	-4.129

Note: Ontario and "less than high school" are the reference categories.
Note: Birthyears were normalized by subtracting the minimum value (1912).

4.4.1 Generating New Spouses' Characteristics

Whenever a new couple is generated, the characteristics of the new spouse are randomly generated. Here is the procedure followed.

The age of the spouse is determined based on the distribution of spouses' ages difference observed in the NHS among respondents of the same age as the simulated individual. Spouses are always assumed to be of the opposite sex than the simulated individual; this can be assumed with minimal losses as the fiscal treatment of the *ghost* partner is not a function of his/her gender.

As for their education, in order to take into account the correlation between two spouses' educational attainment, we use the same model described for parents' education (see Section 3.3.3). While the use of this model enforces the consistency through time concerning education within the household, it has one drawback: it assumes that the members of the household have at least one child. As this will not be the necessarily the case when generating a new spouse, we assume that the new partner had a child in a couple in which the mother was aged 25.

In cases of a new spouse older than 65, a corresponding amount of CPP/QPP benefits is also determined based on the models above.

4.5 Mortality

Mortality is exogenous in the model and does not depend on individual characteristics beyond gender and age. At the end of each time period, a random draw determines whether simulated individuals, spouses and kids pass away. Mortality rates are taken from Statistics Canada's population projections. The rates correspond to the "medium" scenario. If a spouse passes away, information concerning the deceased partner is kept on file to compute survivor's benefits.

Chapter 5

Other Transitions

5.1 Summary

The third module concerns the other transitions, mostly of economic nature. Every year, each individual in the simulation is assigned transition probabilities for each characteristic that stochastically varies over time. The core transition in this module is the employment transition, i.e. movements in and out of the labour force. Other important transitions are the decision to end schooling and decisions to collect various types pension income, coming e.g. from the CPP/QPP and RPPs.

Every transition probability is a function of various individual characteristics such as age, gender, educational attainment, or the presence of a young child. The transition models are generally estimated based on the first wave of LISA (2012), as discussed in Section 2.2. The LISA dataset allows us to track individual income over time based on tax files, which in turn allows us to discuss employment dynamics or determine the amount of retirement income received by a respondent. It also allows the identification of cohort effects, as we observe individuals of various ages at different moments in time.

5.2 Schooling

Schooling is the only transition that is estimated using the GSS instead of LISA. The schooling model works in two steps. Starting at age 16, individuals have the possibility of terminating their schooling. This probability is modelled as a logit (see Section A.1.1). Estimated parameters are reported in Table 5.1.[7]

When an individual finishes schooling, her final level of schooling is randomly drawn based on a multinomial logit (see Section A.1.2). The end of schooling is a sinking state, as there are no possibilities to return to school in the model. While this assumption may seem strong, let us remember that the transition of interest is the probability of stopping studying for good. After 18 years old, if an individual

[7] Because there were enough observations available in the GSS, province dummies instead of region ones are used in estimating schooling variables.

Table 5.1: Estimated parameters for probability of ending schooling

Male	-0.061
Birthyear	-0.001
Is a father	-0.124
Is a mother	-0.434
Educ. of father: High school	-0.179
Educ. of father: College	-0.273
Educ. of father: University	-0.450
Educ. of mother: High school	-0.101
Educ. of mother: College	-0.204
Educ. of mother: University	-0.240
Newfoundland and Labrador	-0.028
Prince Edward Island	0.131
Nova Scotia	0.101
New Brunswick	0.118
Québec	-0.040
Manitoba	-0.013
Saskatchewan	0.064
Alberta	0.025
British Columbia	-0.046
Age is 18	-2.057
Age is 19	-1.581
Age is 20	-1.211
Age is 21	-0.933
Age is 22	-0.620
Age is 23	-0.542
Age is 24	-0.616
Age is 25	-0.616
Age is 26	-0.721
Age is 27	-0.683
Age is 28	-0.709
Age is 29	-0.739
Age is 30	-0.597
Age is 31	-0.437
Age is 32	-0.280
Age is 33	-0.155
Age is 34	0.170
Age is 35	0.790
Constant	-0.540

Note: Ontario is the province of reference.

Note: Birthyears were normalized by subtracting the minimum value (1912).

works and still has to get her final degree, she is assumed to hold a high school degree for the sake of determining her labour income and employment transition probabilities. Estimated parameters are reported in Table 5.2.

5.3 Employment

Employment is a central mechanism in the simulation model. It is also the most complex sub-model that is used in SIMUL.

Transitions in and out of the labour force are based on logit models. However, to allow for unobserved heterogeneity, we used a model with discrete types of workers. Specifically, there are two types of workers when it comes to entering the labour market: those with high probabilities to do so and those with low probabilities. This heterogeneity is captured by allowing the constant term in the model to be different for these two types of individuals. The same goes for the probability of leaving the labour market.

The model therefore uses four types of individuals:

1. Low probability of entering and exiting.
2. Low probability of entering, high probability of exiting.
3. High probability of entering, low probability of exiting.
4. High probability of entering and exiting.

In a nutshell, the simulation of employment requires three steps:

1. Those who did not receive their worker type are randomly assigned one, based on a multinomial logit (see Section A.1.2).
2. Those who did not work in the previous period may enter labour force, based on a logit model (see Section A.1.1).
3. Those who worked in the previous period may exit labour force, also based on a logit model (see Section A.1.1).

Estimation of this process is quite complex and is presented in detail in Section A.2. Tables 5.3 and 5.4 present the estimated parameters used to determine the worker type and to determine whether individuals enter or leave the labour force in a given year.

A final correction is made in order to have consistent employment rates. Given that we use tax data, our definition of work used with LISA is based on whether a respondent received labour income at any time during the year, they overestimate the proportion of workers employed at any moment. Two minor corrections are added to the employment status for these two definitions to be comparable.

First, respondents drawing a labour income inferior to 5,000$ are considered as not working anymore at the end of the year (see Section 5.4).

Second, respondents have an exogenous probability of 3.5% of losing their job at the end of a period, even if they worked during the year. In this case, the respondent keeps his labour income for that year,

Table 5.2: Estimated parameters for degree completion model

	Less than high school	College	University
Male	0.405	-0.273	-0.388
Birthyear	0.000	0.000	0.000
Is a father	0.511	0.003	-0.472
Is a mother	0.725	-0.039	-0.964
Educ. of father: High school	-0.466	-0.047	0.210
Educ. of father: College	-0.391	0.438	0.472
Educ. of father: University	-0.377	-0.183	0.951
Educ. of mother: High school	-0.459	-0.031	0.214
Educ. of mother: College	-0.204	0.296	0.495
Educ. of mother: University	-0.209	-0.211	0.803
Newfoundland and Labrador	0.383	0.359	0.248
Prince Edward Island	0.120	-0.172	0.124
Nova Scotia	0.338	0.087	0.310
New Brunswick	0.048	0.163	-0.130
Québec	0.503	0.389	0.840
Manitoba	0.226	0.012	-0.086
Saskatchewan	0.013	0.173	0.201
Alberta	-0.088	-0.043	-0.074
British Columbia	-0.314	0.106	-0.061
Age is 18	-5.603	1.770	1.722
Age is 19	-4.969	1.917	1.719
Age is 20	-5.691	2.007	1.856
Age is 21	-20.278	2.106	2.321
Age is 22	-6.228	2.069	2.677
Age is 23	-5.125	1.997	2.778
Age is 24	-20.354	1.996	2.781
Age is 25	-5.563	1.976	2.734
Age is 26	-20.527	1.967	2.724
Age is 27	-20.574	1.969	2.736
Age is 28	-6.302	1.950	2.733
Age is 29	-20.680	1.941	2.778
Age is 30	-20.713	1.942	2.775
Age is 31	-20.741	1.918	2.767
Age is 32	-6.328	1.907	2.797
Age is 33	-20.760	1.943	2.826
Age is 34	-5.222	1.921	2.805
Age is 35	-6.335	1.909	2.803
Constant	-0.501	-1.163	-2.005

Note: High school is the reference for the dependent variable.
Note: Ontario is the province of reference.
Note: Birthyears were normalized by subtracting the minimum value (1912).

Table 5.3: Estimated parameters to determine the probability of drawing unobserved worker type

	Entry/Exit		
	Low/Low	High/High	Low/High
Male	-0.876	-0.868	-1.579
Birthyear	-0.050	0.018	0.002
High School	-0.173	-0.586	-1.089
College	-0.049	-0.776	-1.398
University	0.036	-1.386	-1.639
Québec	0.257	0.039	-0.363
Ontario	0.249	0.061	0.309
Prairies	-0.055	-0.067	-0.059
British Columbia	-0.257	0.248	-0.148
Constant	2.205	-1.444	0.378

Note: Type "high entry/low exit" is the reference for the dependent variable.
Note: Maritimes and "less than high school" are the reference categories.
Note: Birthyears were normalized by subtracting the minimum value (1912).

Table 5.4: Estimated parameters to determine the probability of entering or leaving the labour force

	Enter	Leave
Male	0.548	-0.236
Birthyear	-0.004	-0.017
$age - 20$	0.031	-0.065
$\frac{(age-20)^2}{100}$	-0.163	0.133
Kids between 0 and 5	-0.045	-0.052
Woman with kids between 0 and 5	-0.470	0.631
Lagged log-income	0.000	0.000
Interaction Marital Status/Age	0.000	0.000
Couple	0.000	0.000
High School	0.268	-0.180
College	0.566	-0.378
University	0.789	-0.329
Québec	-0.202	-0.133
Ontario	0.123	-0.135
Prairies	0.166	-0.215
British Columbia	-0.145	0.215
Receives retirement income	-1.053	1.592
Worker type: low entry	-2.144	
Worker type: high entry	0.067	
Worker type: low exit		-1.740
Worker type: high exit		0.370

Note: Maritimes and "less than high school" are the reference categories.
Note: Birthyears were normalized by subtracting the minimum value (1912).

but is considered as not working anymore and must find a new job at the next period. The probability of such a shock was calibrated in order to match official employment rates.

At no point in time is leaving the labour force a sinking state, meaning that there is no retirement in the labour market sense. In theory, individuals may re-enter the labour market until they pass away, although very few individuals will actually do so.

5.4 Earnings

Conditional on working during a given year, individuals receive labour income (earnings). This income is based on a random-effects model taking into account the unobserved heterogeneity in worker types. Estimation of this model is described in Section A.2.2.

Earnings simulation involves two steps:

1. Those who did not receive a draw for their heterogeneity term are assigned one, drawn from the appropriate normal distribution. This term will remain constant through an individual's entire life.
2. Conditional on this effect, the log-income is generated based on a linear prediction, and is then converted to an income in level.

The model is estimated separately for men and women. Estimated parameters are reported in Table 5.5

5.5 Pension Claiming and Benefits

It is essential for SIMUL to allow for a detailed simulation of retirement and pension incomes. The next chapter will show that individuals in the models are entitled to receive pensions from the Old Age Security program. In this section, emphasis is on pensions from CPP/QPP and from private sources.

5.5.1 CPP/QPP Claiming and Benefits

Once eligible to claim CPP/QPP benefits (usually from age 60 on), a respondent can elect to claim this pension. Pension claiming is a sinking state and cannot be reversed. Every period, individuals eligible to receive CPP/QPP benefits have a probability of beginning to claim this pension. This probability is based on a logit model (see Section A.1.1) with dummies for every age when the respondent is eligible. This flexible specification allows for an increased probability of claiming at some ages. For instance, everything else kept constant, an individual has the highest probability to claim the benefits at age 65, conditional on not having claimed previously. Estimates for the parameters of this model are presented in table 5.6.

Once an individual claims his benefits, their value must be computed. In principle, individuals will have a work history based on the periods they lived in the model. However, this history may not

Table 5.5: Estimated parameters for the log-income model

	Women	Men
Birthyear	-0.005	-0.010
$age - 20$	0.039	0.059
$\frac{(age-20)^2}{100}$	-0.075	-0.131
Kids between 0 and 5	-0.175	0.003
Couple	-0.079	0.059
High School	0.333	0.294
College	0.575	0.442
University	0.870	0.658
Québec	0.149	0.133
Ontario	0.183	0.172
Prairies	0.173	0.204
British Columbia	0.118	0.163
Receives retirement income	0.000	0.000
Worker type: Low/Low	-0.044	0.039
Worker type: High/High	0.058	0.018
Worker type: Low/High	0.007	-0.003
σ_ν	0.272	0.177
σ_ε	0.301	0.187
Constant	9.460	10.075

Note: Maritimes, "less than high school", and "Worker type High/Low" are the reference categories.

Note: Birthyears were normalized by subtracting the minimum value (1912).

Table 5.6: Estimated parameters for the probability of first claiming CPP/QPP benefits

Birthyear	0.064	0.071
Couple	-0.159	0.302
High School	-0.006	-0.080
College	-0.276	-0.028
University	-0.011	-0.464
Québec	-0.252	-0.060
Ontario	-0.769	-0.628
Prairies	-0.317	-0.368
British Columbia	-0.165	-0.375
Age is 61	-0.012	-0.021
Age is 62	-0.755	-0.857
Age is 63	-0.606	-0.795
Age is 64	-0.813	-0.662
Age is 65	1.446	1.728
Age is 66	0.971	1.467
Age is 67	0.947	0.860
Age is 68	-1.088	1.588
Age is 69	0.688	0.763
Constant	-1.947	-2.753

Note: Maritimes and "less than high school" are the reference categories.
Note: Birthyears were censored at 1942 to reflect observed trends.
Note: Birthyears were normalized by subtracting the minimum value (1912).

be complete. In this case, the employment/income model described above is used to generate a full working history, and benefits are computed based on this history.

Note that when a respondent's spouse passes away, the respondent may be entitled to a survivor's benefits, which is computed based on the simulated values from the model.

5.5.2 RPP Claiming and Benefits

Additional income can come from a RPP (or a RRSP of RRIF, which are included in the RPP concept in SIMUL). Every individual aged 50 and over is eligible to receive RPP benefits.

The model is in three parts. First, individuals who do not receive RPP benefits decide whether to claim. Respondents who have not yet claimed RPP benefits can start doing so until they reach age 79. At 80, respondents who do not receive RPP benefits will never receive income from this source. The probability of first claiming RPP is based on a logit model (see Section A.1.1) with dummies at every age for reasons similar to what is described in the previous section. Estimates for the parameters of this model are presented in table 5.7.

In the first year when the individual claims benefits, an initial RPP outcome is simulated. We refer to this as starting benefits. This value is attributed based on a linear model predicting the log of this starting value. The simulation then transforms this log into the appropriate level variable. Estimates for the parameters of this model are presented in table 5.8.

In subsequent years, benefits are simulated based on starting benefits and on time-varying covariates. This is referred to as "current benefits". This model is also based on a linear model predicting the log value at every period. Estimates for the parameters of this model are presented in table 5.9.

RPP benefits for every respondent in the database are simulated independently from whether they were receiving such benefits in the initial NHS dataset. Because we had to impute the age of the individual in the initial dataset and because respondents are far more likely to start claiming at some ages than others (ages 65 and 70, for instance, are those with the highest hazard rates), using RPP information on individuals with randomly assigned age would bias upward the fraction of respondents who receive RPP benefits. The intuition for this is simple. Suppose that one observes respondents aged 64 and 65 in the initial dataset, but cannot tell who belongs to which age category. Moreover, assume, as is the case in SIMUL, that RPP benefits are not used to impute the correct age. One way to assign age randomly is simply to flip a coin. Note that about half the individuals really aged 65 will be assigned an age of 64. When running the simulation, they will go a second time through the claiming transition probabilities of those aged 65, meaning that a very large fraction of those without RPPs will now receive benefits. This means that for this age group, the fraction of RPP recipients will be much larger than what is expected.

To avoid this problem, respondents from the initial database are simply imputed a RPP claiming status and, if needed, "starting benefits" based on the aforementioned stochastic process. This initial simulation runs from the time the respondent is aged 50 until his current age.

Table 5.7: Estimated parameters for the probability of first claiming RPP benefits

Male	0.013
Birthyear	-0.025
Lagged log-income	0.171
Lagged employment status	-0.314
Couple	-0.045
High School	0.469
College	0.464
University	0.713
Québec	0.071
Ontario	-0.135
Prairies	0.005
British Columbia	-0.078
Age is 51	-0.811
Age is 52	-1.171
Age is 53	-0.652
Age is 54	-0.367
Age is 55	0.912
Age is 56	0.709
Age is 57	0.814
Age is 58	0.804
Age is 59	0.663
Age is 60	1.445
Age is 61	1.176
Age is 62	0.845
Age is 63	1.117
Age is 64	1.096
Age is 65	2.769
Age is 66	2.057
Age is 67	1.147
Age is 68	1.583
Age is 69	1.589
Age is 70	2.976
Age is 71	2.319
Age is 72	2.369
Age is 73	1.743
Age is 74	1.218
Age is 75	1.156
Age is 76	0.585
Age is 77	0.707
Age is 78	-0.298
Age is 79	1.214
Constant	-4.947

Note: Maritimes and "less than high school" are the reference categories.
Note: Birthyears were normalized by subtracting the minimum value (1912).

Table 5.8: Estimated parameters for the level of first RPP benefits

Male	0.399
Birthyear	0.013
$age - 20$	-0.044
Lagged log-income	0.064
Couple	-0.082
High School	0.210
College	0.316
University	0.679
Québec	0.063
Ontario	-0.037
Prairies	-0.037
British Columbia	0.016
Splines for age past 60	-0.000
Splines for age past 70	0.010
Receives CPP benefits	-2.406
Log of (CPP benefits + 1)	0.293
σ	1.161
Constant	9.261

Note: Maritimes and "less than high school" are the reference categories.
Note: Birthyears were normalized by subtracting the minimum value (1912).

Table 5.9: Estimated parameters for the level of current RPP benefits

Log of first benefits	0.916
Male	0.059
Birthyear	-0.008
$age - 20$	-0.002
Couple	-0.059
High School	0.101
College	0.117
University	0.177
Québec	0.008
Ontario	0.059
Prairies	0.050
British Columbia	0.005
Splines for age past 60	0.013
Splines for age past 70	0.007
σ	0.893
Constant	0.367

Note: Maritimes and "less than high school" are the reference categories.
Note: Birthyears were normalized by subtracting the minimum value (1912).

5.6 Other Income Sources

Other sources of income will be progressively added to SIMUL.

Chapter 6

Taxes, Transfers and Poverty

6.1 A Note on Constant Dollars

Models estimated with the LISA used values are in 2010 dollars. However, in order to use and account for the most recent version of the tax and transfer system, we convert the predicted values in 2015 dollars when determining net income. Taxes and transfers are hence computed using the 2015 values and results are reported in this currency.

This poses a challenge for elderly poverty analyses, as Old Age Security benefits are constant in real dollars (i.e. they are adjusted for inflation every year). In order to account for this, low-income thresholds are also adjusted for inflation and converted into 2015 dollars.

Hence, whenever monetary outputs are extracted from SIMUL, results should be interpreted as being reported in 2015 real dollars.

6.2 Household-Level Taxes and Transfers

Each year, total taxes and transfers paid and received by the simulated individuals are computed using SIMTAX, a fiscal simulator built and maintained by members of our research team and other colleagues (cf. Marchand et al., 2016). This allows us to determine various elements of interests, such as who will live with a low income, or how much tax a government can expect to collect. SIMTAX computes taxes and transfers at the household level.

SIMTAX calculates social transfers and personal income taxes in Canada as functions of market incomes and personal / family / social characteristics. SIMTAX lets the user calculate a reasonable budget constraint for people with and without children, for workers and for retired individuals. The simulator's most recent version, used with SIMUL, reflects the fiscal environment of 2015.

While the aim of this documentation is not to give a complete description of SIMTAX, Table 6.1, taken from Marchand et al. (2016), provides an extensive overview of the different taxes and transfers it covers.

Table 6.1: List of taxes and transfers included in SIMTAX

Federal income tax

Provincial income tax
Provincial tax reductions (NL, PE, NB, NS, ON, BC)
Surtaxes (PE, ON)
Health contribution (QC, ON)
Work income deduction (QC)

Guaranteed Income Supplement[a]
Old Age Security (including the clawback)
CPP and QPP contributions (contributions and NRTC)

Employment Insurance premiums (contributions and NRTC)
Québec Parental Insurance Plan (contributions and NRTC)

Social Assistance
- NL - Income Support Benefits
- PE - Social Assistance
- NS - Income Assistance
- NB - Social Assistance
- Québec - Aide financière de dernier recours (Aide sociale)
- Ontario Works
- Manitoba Employment income assistance
- SK - Social assistance
- AB - Income support
- BC - Employment and Assistance

Social Solidarity
- PE - Social Assistance - Disabled supplement
- NB - Social Assistance - Extended benefits
- Québec - Solidarité sociale
- Ontario Disability Support Program (ODSP)
- Manitoba Employment income assistance for disabled
- SK - Assured Income for Disability (SAID)
- AB - Income support - Barriers to full employment

Work Income Tax Benefit and Disability supplement

Goods and services tax/Harmonized sales tax (GST/HST) credit

Continued on next page

Universal Child Care Benefit
Canada Child Tax Benefit
National Child Benefit Supplement
Child Care deduction (CAN)
NRTC: Canada employment
NRTC: Disabled, Child, Pensions income, Age amount
NRTC: Senior supplementary amount (SK)
NRTC: Single (QC)
NRTC: Experienced worker (QC)
Newfoundland and Labrador Child Benefit
The Newfoundland and Labrador Harmonized Sales Tax Credit
Newfoundland and Labrador Seniors' Benefit
Prince Edward Island Sales Tax Credit
Nova Scotia Child Benefit
Nova Scotia Affordable Living Tax Credit
New Brunswick Child Tax Benefit
New Brunswick Work Income Supplement
New Brunswick Low-Income Seniors' Benefit
Québec work premium
Québec adapted work premium
Child Care Tax Credit (QC)
Child Assistance Program (QC)
Solidarity Tax Credit
Ontario Child Benefit
Ontario Trillium Benefit
Guaranteed Annual Income System (ON)
Manitoba 55 Plus
Manitoba Child Benefit
Saskatchewan Low-income Tax Credit
Saskatchewan Employment Supplement
Saskatchewan Seniors Income Plan

Continued on next page

Alberta Family Employment Tax Credit
Alberta Seniors Benefit
BC Family Bonus
BC Earned Income Benefit
BC Early Childhood Tax Benefit
BC Low-income Climate Action Tax Credit
BC Seniors' Supplement
Yukon Child Benefit
Northwest Territories Child Benefit
Nunavut Child Benefit

[a] Including the 2011 bonus.
Note: NRTC stands for *non-refundable tax credits* (see the complete official list here.)

Some of the transfers listed in the table are family-level transfers (e.g., the Universal Child Care Benefit). In such cases, when aggregating the national or provincial total cost of a transfer, only money given to women in the sample is accounted for in order to avoid double-counting.

6.3 Poverty

Microsimulation models like SIMUL allow for the analysis of projections and trends beyond means or medians. A relevant example is poverty analysis.[8] Once disposable income is computed by SIMUL, it is possible to determine whether a family's income is above a given low-income threshold.

SIMUL currently allows the computation of poverty based on the *Market Based Measure*. This threshold is of particular interest as it relies on a basket of goods whose price is specific to a given region. As pointed out by Bibi and Duclos (2010), this measure accounts for the fact that prices vary depending on region of residence. For instance, housing prices are much higher in Ontario and British Columbia than they are in Quebec. Using a common measure for poverty would lead to the conclusion that the population of Quebec is poorer than it really is, considering these prices. Here is how Statistics Canada describes this measure (cf. Statistics Canada, 2012).

> The Market Basket Measure (MBM) is based on the cost of a specific basket of goods and services representing a modest, basic standard of living. It includes the costs of food, clothing, footwear, transportation, shelter and other expenses for a reference family of two adults aged 25 to 49 and two children (aged 9 and 13). It provides thresholds for a finer geographic level than the low-income cut-off (LICO), allowing, for example, different costs for rural areas in the different provinces. These thresholds are compared to disposable

[8] The terms "poverty" and "low income" are used interchangeably in this section. Several agencies and analysts emphasize differences between the two concepts.

> income of families to determine low-income status. Disposable income is defined as the sum remaining after deducting the following from total family income: total income taxes paid; the personal portion of payroll taxes; other mandatory payroll deductions such as contributions to employer-sponsored pension plans, supplementary health plans, and union dues; child support and alimony payments made to another family; out-of-pocket spending on child care; and non-insured but medically prescribed health-related expenses such as dental and vision care, prescription drugs, and aids for persons with disabilities.

In the specific case of the province of Quebec, the MBM chosen is an average of the various thresholds for the province. The threshold is adapted for family size according to a standard formula:

$$MBM(size) = \frac{\sqrt{size}}{\sqrt{4}} MBM(4)$$

where the MBM for a family of four is the aforementioned known value.[9]

The current version of SIMUL relies on the 2010 value of *MBM(4)*. Hence, the threshold for a family of four is $30,552. We apply two corrections to this threshold. First, given that our fiscal simulator described above uses values for 2015, we inflate the threshold to make it comparable with the tax/transfer system. Second, following the recommendation by the *Centre d'étude sur la pauvreté et l'exclusion* (Fréchet et al., 2010), the MBM thresholds are increased by 7% to reflect the fact that some expenses made by individuals would be excluded from the amount used to determine disposable income (e.g., medical expenses).

[9] Use of the MBM may be problematic if the basket of goods consumed by a given family is very different from the representative basket used to determine the threshold. This may be a problem when studying poverty among the elderly, as we may not expect this group to consume the same basket of goods as a family of four with two working-age adults.

Part II

SIMUL's Output

Chapter 7

Standard SIMUL Outputs

7.1 Extracting Population Statistics

SIMUL can extract population statistics from the microsimulation. Once a variable is simulated through time, it is possible to synthesize the evolution of the variable by generating one of the following statistics:

- sum;
- mean;
- median.

The requested statistics can be obtained for all values of a given variable (e.g., obtaining average labour income for every age when individuals may be in the labour force) or conditional on certain characteristics (e.g., obtaining the average income only for women). Statistics are always extracted on a yearly basis. They can be weighted or not, depending on the context.

SIMUL relies on randoms draws to determine many individual outcomes. It possible to obtain a measure of the variation in relevant outputs due to this randomness by requesting many replications of the simulation. In this case, the returned statistics will be averaged and the output will consist in the mean value and the standard deviation of the simulated outcome.

While analyzing all the common output of SIMUL would be long and tedious, it is still interesting to discuss some of the more common variables that one may want to extract from the simulations and to explain how to extract these variables.

7.1.1 Requesting Statistics

Requesting statistics after a simulation is done through the `statslist.dat` file. The first line of the file must contain the number of statistics to be produced after the simulation. Each of these statistics is requested using the following format:

```
name statistics variable forvar step weight ifvar ifvarvalue
```

where

- `name` is a prefix that will be used once the variable is generated.
- `statistics` is the name of the statistics to be produced, as listed above.
- `variable` is the name of the variable to be generated (exhaustive admissible list follows).
- `forvar` is a variable that will be used to split the sample, as explained above (age, in the example). Use the value "year" to get the whole population.
- `step` is the step used for the `forvar` (e.g., at every age is a value of 1).
- `weight` is a logical value (either "F" for false or "T" for true) stating whether the information must be weighted.
- `ifvar` is a variable used to select a subsample ("none" if no variables are to be used).
- `ifvarvalue` is the value of the `ifvar` for which the subsample must be selected.

For instance, let us obtain the total population by age in the model. This is done with the following line.

```
pop sum wgt age 1 F none 0
```

In the output file, columns named `pop_sum_age_X` (with X taking the values of various ages) will be generated for every year in the simulation. They will contain the number of individuals at every age and for every year.

Note that in this case, the weight was summed directly (variable wgt) and hence, there was no need to weight the outcome. Not wanting to have the result for a subsample, we fixed the `ifvar` at `none`.

Suppose now that one wants to obtain the employment rate of women in the sample. The following line would generate this statistic.

```
workf mean work age 1 T male 0
```

This time, we obtain a weighted mean of the variable "work" for individuals for which the variable `male` is equal to 0.

7.1.2 Available Variables

Some variables were pre-selected and are available to compute statistics or to serve as for-variable or as if-variable. Table 7.1 shows a list of these variables. Extraction of other variables is feasible, but requires additional coding in the program itself.

Table 7.1: List of variables available to compute statistics

Variable	Individual identifier
age	Age
byear	Birthyear
male	Whether the respondent is a male
wgt	Sample weight
prov	Province of residence
year	Current year
work	Employment status (1 means employed)
nkids	Number of kids younger than 18
couple	Marital status (1 means in a couple)
single_male	Individual is a single male
single_female	Individual is a single female
schnotdone	Individual is still a student
lesshs	Individual's educational attainment is less than highs chool
hs	Individual's educational attainment is high school degree
college	Individual's educational attainment is college degree
university	Individual's educational attainment is university degree
poor	Individual is poor (see Section 6.3)
inc_earn	Earnings
inc_cpp	Income from CPP benefits
inc_spcpp	Income from spouse's CPP benefits
inc_fam_disp	Family disposable income
inc_oas	Income from Old Age Security
inc_gis	Income from Guaranteed Income Supplement
inc_rpp	Income from RPP benefits
inc_gross	Gross income
inc_surv	Income from survivor benefits
claim_cpp	Individual claimed CPP benefits
claim_rpp	Individual claimed RPP benefits
widow	Individual is a widow or a widower and is not in a new union
widow_male	Individual is a widower (male)
widow_female	Individual is a widow (female)
size	Household size
reprate	Retirement income replacement rate

7.2 Microdata

If the aggregated values of the SIMUL standard output are not sufficient, it is possible to extract the microdata from SIMUL. For instance, to study economic mobility, one must be able to reconstruct individuals' full history. This means that every record generated in the simulation will be written in a dataset. These microdata can be requested from the `setting` file used to set a simulation. In this case, the year interval for which the microdata are requested can be specified with the `step` option. If multiple replications are performed, each replication will generate its own microdata. This option should be used with care, as a complete simulation leads to a microdata file of about 3.5 Gb.

Chapter 8

Public Finances

SIMUL generates public finances aggregates that are calibrated to be consistent with those reported in the public accounts. At the time of writing this report, this part of the model is only functional for the expenditures, revenues and other aggregates of the Quebec government. This chapter details the variables created and used as well as the calibration process.

8.1 Budgetary Expenditures

Expenditures are broken down by government "mission", one of the ways by which they are reported in the Quebec public accounts. There are five missions, in addition to the debt service, each are the object of a sub-section below.

8.1.1 Health and Social Services

Health and social services expenditures (`hlth_total`) include the expenditures of the Ministère de la Santé et des Services sociaux, as well as those of the bodies and special funds related to this same "mission". They are split in two categories: programme expenditures (`hlth_prog`) and other expenditures (`hlth_other`). Programme expenditures are those related to health coming from the general government fund. Other expenditures are those from the Fonds de financement des établissement de santé et des services sociaux as well as others not coming from the general government fund.

Health programme expenditures are obtained by multiplying *per capita* expenses ($\texttt{hlth\_pc}_i$) from the Canadian Institute for Health Information (CIHI; see Canadian Institute for Health Information, 2015) for 20 age groups and 2 gender groups by the number of individuals projected for each age gender group ($\texttt{pop\_group}_i$). A structural growth factor for health expenditures (`htl_struc_rate` ; see Table 8.1 for value) is also added for the growth in expenditures that is unrelated to demography. Other expenditures vary according to real GDP growth ($\texttt{gdp\_rate}_t$). Programme expenditures and other expenditures are calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015) – see Gouvernement du Québec (2015). Table 8.1 shows the values in current (2014) and constant (2010) dollars.

Equations

$$\texttt{hlth\_prog}_t = \sum_{i=1}^{20} \texttt{pop\_group}_t^i \cdot \texttt{hlth\_pc}_{t-1}^i \cdot (1 + \texttt{htl\_struc\_rate})$$

$$\texttt{hlth\_other}_t = \texttt{hlth\_other}_{t-1} \cdot (1 + \texttt{gdp\_rate}_t)$$

$$\texttt{hlth\_total}_t = \texttt{hlth\_prog}_t + \texttt{hlth\_other}_t$$

Calibration

Table 8.1: Calibration of health and social services expenditures with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
hlth_prog	31,088	28,921
hlth_other	5,705	5,307
hlth_total	36,793	34,229
		Structural growth rate (in %)
htl_struc_rate		1.70

8.1.2 Education and Culture

Expenditures on education and culture (educ_total) comprise those of the Ministère de l'Éducation et de l'Enseignement supérieur[10]; those of the Ministère de la Culture et des Communications; those of the Ministère de l'Immigration, de la Diversité et de l'Inclusion; as well as those of the bodies and special funds related to this same mission. Education expenditures are split into two broad categories: programme expenditures in education (educ_prog) and other expenditures in education and culture (educ_other). Programme expenditures in education are then subdivided into three schooling levels: 1) preschool, primary and secondary (educ_scol_com); 2) college, or CEGEP in Quebec (educ_coll); and 3) university (educ_univ).

Programme expenditures in education are obtained by multiplying expenses per student (scol_com_ps, coll_ps, univ_ps; see Table 8.2 for values) by the projected number of students for each schooling level (stu_scol_com, stu_coll, stu_univ). Structural growth factors in education (struc_scol_com_rate, struc_coll_rate, struc_univ_rate, see Table 8.2) are also added to reflect expenditure growth that is unrelated to the number of students. Other expenditures in education and culture vary according to real GDP growth ($\texttt{gdp\_rate}_t$). Programme expenditures by schooling level and other expenditures are

[10]We use the names of ministries as they stand at the time of writing. These may differ from the way government activities were distributed in SIMUL's baseline year, but in every case public accounting ensures that there are no series breaks.

all calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.2 reports the values in current (2014) and constant (2010) dollars.

Equations

$$
\begin{aligned}
\texttt{educ\_scol\_com}_t &= \texttt{stu\_scol\_com}_t \cdot \texttt{scol\_com\_ps}_{t-1} \cdot (1 + \texttt{struc\_scol\_com\_rate}) \\
\texttt{educ\_coll}_t &= \texttt{stu\_coll}_t \cdot \texttt{coll\_ps}_{t-1} \cdot (1 + \texttt{struc\_coll\_rate}) \\
\texttt{educ\_univ}_t &= \texttt{stu\_univ}_t \cdot \texttt{univ\_ps}_{t-1} \cdot (1 + \texttt{struc\_univ\_rate}) \\
\texttt{educ\_prog}_t &= \texttt{educ\_scol\_com}_t + \texttt{educ\_coll}_t + \texttt{educ\_univ}_t \\
\texttt{educ\_other}_t &= \texttt{educ\_other}_{t-1} \cdot (1 + \texttt{gdp\_rate}_t) \\
\texttt{educ\_total}_t &= \texttt{educ\_prog}_t + \texttt{educ\_other}_t
\end{aligned}
$$

Calibration

Table 8.2: Calibration of education and culture expenditures with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
educ_prog	17,300	16,094
educ_scol_com	11,016	10,248
educ_coll	2,537	2,360
educ_univ	3,747	3,486
educ_other	3,605	3,354
educ_total	20,905	19,448
		Structural growth rate (in %)
struc_scol_com_rate		2.70
struc_coll_rate		0.77
struc_univ_rate		1.64

8.1.3 Support for Individuals and Families

Expenditures on support for individuals and families (welf_total) mostly comprise those of the Ministère du Travail, de l'Emploi et de la Solidarité sociale (except the sector of labor), of the Ministère de la Famille, of the Ministère de la Justice (access to justice sector) and of the bodies and special funds related to this same mission. They are subdivided into four categories:

1. welfare (`welf_socas`);
2. Quebec work premium (`welf_wp`);
3. child support (`welf_qccap`); and
4. other expenditures for support to individuals and families (`welf_other`).

The first three categories are three of the main support programmes and are modelled using SIMTAX. Various other support programmes are not yet modelled by SIMTAX and/or SIMUL. They are therefore included in "other expenditures on support for individuals and families". These vary according to real GDP growth ($\texttt{gdp\_rate}_t$). Programme expenditures computed by SIMTAX (`welf_socas`, `welf_wp`, `welf_qccap`) as well as the other expenditures (`welf_other`) are all calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.3 shows the values in current (2014) and constant (2010) dollars.

Equations

$$\texttt{welf\_other}_t = \texttt{welf\_other}_{t-1} \cdot (1 + \texttt{gdp\_rate}_t)$$

$$\texttt{welf\_total}_t = \texttt{welf\_socas}_t + \texttt{welf\_wp}_t + \texttt{welf\_qccap}_t + \texttt{welf\_other}_t$$

Calibration

Table 8.3: Calibration of expenditures on support for individuals and families with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
welf_socas	2,866	2,666
welf_wp	358	333
welf_qccap	2,206	2,052
welf_other	4,217	3,924
welf_total	9,647	8,975

8.1.4 Economy and Environment

Expenditures on economy and environment (`envr_prog`) include those of:

- the Ministère des Affaires municipales et de l'Occupation du territoire (except the sector of financial aid to municipalities and compensations in-lieu of taxes);
- the Ministère des Transports, de la Mobilité durable et de l'Électrification des transports;
- the Ministère de l'Agriculture, des Pêcheries et de l'Alimentation;

- the Ministère de l'Économie, de la Science et de l'Innovation;
- the Ministère de l'Énergie et des Ressources naturelles;
- the Ministère du Développement durable, de l'Environnement et de la Lutte contre les changements climatiques;
- the Ministère des Forêts, de la Faune et des Parcs;
- the Ministère du Tourisme;
- the Ministère des Relations internationales et de la Francophonie; and
- the bodies and special funds related to this same mission.

All these expenditures vary according to real GDP growth ($\texttt{gdp\_rate}_t$). They are jointly calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.4 reports the values in current (2014) and constant (2010) dollars.

Equation

$$\texttt{envr\_prog}_t = \texttt{envr\_prog}_{t-1} \cdot (1 + \texttt{gdp\_rate}_t)$$

Calibration

Table 8.4: Calibration of expenditures on economy and environment with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
envr_prog	11,458	10,659

8.1.5 Administration and Justice

Expenditures on administration and justice (just_prog) include those of:

- the Ministère des Affaires municipales et de l'Occupation du territoire (sector of financial aid to municipalities and compensations in-lieu of taxes);
- the Ministère de la Sécurité publique;
- the Ministère du Conseil exécutif;
- the Conseil du trésor;
- the Ministère de la Justice (except the access to justice sector);
- the Ministère des Finances (finance sector, except debt management);

- the Ministère du Travail, de l'Emploi et de la Solidarité sociale (labor sector);
- the National Assembly and its designated persons; and
- the bodies and special funds contributing to this same mission.

Revenues and expenditures linked to the application of any law under the responsibility of Revenu Québec also fall under this mission, except for the refundable tax credits that match the definition of a transfer funded by the tax regime. Expenditures on administration and justice vary according to real GDP growth ($\texttt{gdp\_rate}_t$). They are calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.5 reports the values in current (2014) and constant (2010) dollars.

Equation

$$\texttt{just\_prog}_t = \texttt{just\_prog}_{t-1} \cdot (1 + \texttt{gdp\_rate}_t)$$

Calibration

Table 8.5: Calibration of expenditures on administration and justice with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
just_prog	6,728	6,259

8.1.6 Debt Service

Expenditures on debt service (debt_srv) cover the activities of the Ministère des Finances (debt management sector). Debt service is obtained by multiplying the previous year's debt in constant dollars by the nominal interest rate. The nominal interest rate that allows us recover the debt service as published in the 2014-2015 public accounts (see Table 8.6) is 5.136%. Table 8.6 also shows the debt service's value in current (2014) and constant (2010) dollars.

Equations

$$\begin{aligned}
\texttt{debt_srv}_t &= \left(\texttt{total_debt}_{t-1} \frac{CPI_{t-1}}{CPI_{2010}} \cdot \texttt{nom_int_rate}_t\right) \frac{CPI_{2010}}{CPI_t} \\
&= \texttt{total_debt}_{t-1} \cdot \texttt{nom_int_rate}_t \cdot \frac{CPI_{t-1}}{CPI_t} \\
\texttt{debt_srv}_t &= \texttt{total_debt}_{t-1} \cdot \frac{\texttt{nom_int_rate}_t}{1 + \texttt{infla_rate}_t}
\end{aligned}$$

Calibration

Table 8.6: Calibration of debt service expenditures with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
debt_srv	10,270	9,554
		Nominal interest rate used (in %)
nom_int_rate		5.136

8.1.7 Total Expenditures

Total expenditures are the sum of the expenditures on all five government missions – health and social services, education and culture, support for individuals and families, economy and environment, and administration and justice – and for the debt service.

Equation

$$\texttt{tot\_spend}_t = \texttt{hlth\_total}_t + \texttt{educ\_total}_t + \texttt{welf\_total}_t + \texttt{envr\_prog}_t + \texttt{just\_prog}_t + \texttt{debt\_srv}_t$$

8.2 Generations Fund

The Generations Fund has two distinct values: a book value and a market value. The Fund's book value is that used in the Quebec government budget to calculate the gross debt's value. The Fund's market value is the amount of money the government could obtain if it sold all of the Fund's assets at market prices. This fund's revenue also has two values. Its book value is the value of the liquidities injected into the fund to which capitalized investment returns are added. Its market value is the sum of the book value and non-capitalized returns.

More precisely, the book value of the Fund's revenue (gen_rev_acc) equals the sum of investment returns (gen_rev_inv) and of revenue from other sources (gen_rev_other: specific tax on alcoholic beverages, hydraulic and mining royalties, unclaimed property, indexation of the price of heritage pool electricity, additional contribution from Hydro-Québec. Investment returns vary according to their previous year's book value and a rate of return (inv_rate). Revenue from other sources, on the other hand, vary according to real GDP growth ($\texttt{gdp\_rate}_t$). The Generations Fund's book value (gen_fun_acc) equals the sum of the previous year's value and the book value of the Fund's current year revenue.

As previously mentioned, the market value of the Fund's revenue (gen_rev_mark) equals the sum of book value and non-capitalized returns (gen_rev_notcap). Non-capitalized returns vary according the Fund's market value in the previous year and to a rate of return on non-capitalized investments (notcapinv_rate). The Fund's market value (gen_fun_mark) equals its previous year's market value plus the Fund's current year market revenue.

The values related to the Generations Fund are calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.7 reports the values in current (2014) and constant (2010) dollars.

Equations

$$
\begin{aligned}
\texttt{gen\_rev\_inv}_t &= \texttt{gen\_fun\_acc}_{t-1} \cdot \texttt{inv\_rate}_t \\
\texttt{gen\_rev\_other}_t &= \texttt{gen\_rev\_other}_{t-1} \cdot (1 + \texttt{gdp\_rate}_t) \\
\texttt{gen\_rev\_notcap}_t &= \texttt{gen\_fun\_mark}_{t-1} \cdot \texttt{notcapinv\_rate}_t \\
\\
\texttt{gen\_rev\_acc}_t &= \texttt{gen\_rev\_inv}_t + \texttt{gen\_rev\_other}_t \\
\texttt{gen\_rev\_mark}_t &= \texttt{gen\_rev\_acc}_t + \texttt{gen\_rev\_notcap}_t \\
\\
\texttt{gen\_fun\_acc}_t &= \texttt{gen\_fun\_acc}_{t-1} + \texttt{gen\_rev\_acc}_t \\
\texttt{gen\_fun\_mark}_t &= \texttt{gen\_fun\_mark}_{t-1} + \texttt{gen\_rev\_mark}_t
\end{aligned}
$$

Calibration

Table 8.7: Calibration of the Generations Fund with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
gen_rev_inv	302	281
gen_rev_other	977	909
gen_rev_notcap	604	480
gen_rev_acc	1,279	1,190
gen_rev_mark	1,883	1,670
gen_fun_acc	6,938	6,833
gen_fun_mark	8,182	7,612
		Real interest rates used (in %)
inv_rate		4.148
notcapinv_rate		2.727

8.3 Revenue

8.3.1 Personal Income Tax

Personal income tax revenue (`pers_inctax`) is computed based on SIMTAX. This revenue is calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.8 shows the values in current (2014) and constant (2010) dollars.

Calibration

Table 8.8: Calibration of personal income tax revenue with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
`pers_inctax`	27,415	25,504

8.3.2 Contributions to the Health Services Fund

Contributions to the Health Services Fund (`pers_fss`) vary according to real GDP growth ($\texttt{gdp\_rate}_t$). These revenues are calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.9 reports the values in current (2014) and constant (2010) dollars.

Equation

$$\texttt{pers\_fss}_t = \texttt{pers\_fss}_{t-1} \cdot (1 + \texttt{gdp\_rate}_t)$$

Calibration

Table 8.9: Calibration of contributions to the Health Services Fund with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
`pers_fss`	6,555	6,098

8.3.3 Corporate Tax

Corporate tax revenue varies according to real GDP growth ($\texttt{gdp\_rate}_t$). It is calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.10 shows the values in current (2014) and constant (2010) dollars.

Equation

$$\texttt{corp\_inctax}_t = \texttt{corp\_inctax}_{t-1} \cdot (1 + \texttt{gdp\_rate}_t)$$

Calibration

Table 8.10: Calibration of revenue from corporate tax with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
corp_inctax	5,823	5,417

8.3.4 Property Tax

Property tax revenue (prop_tax) varies according to real GDP growth ($\texttt{gdp\_rate}_t$). It is calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.11 reports the values in current (2014) and constant (2010) dollars.

Equation

$$\texttt{prop\_tax}_t = \texttt{prop\_tax}_{t-1} \cdot (1 + \texttt{gdp\_rate}_t)$$

Calibration

Table 8.11: Calibration of property tax revenue with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
prop_tax	1,942	1,807

8.3.5 Consumption Taxes

Revenue from consumption taxes (`cons_tax`) depend on two aggregates: gross revenue from consumption taxes (`gross_cons_tax`) and – refundable – solidarity tax credit (`soltaxcred`). The total value of this tax credit is subtracted from gross consumption tax revenue, since the main component of the credit is the Quebec Sales Tax credit.

The credit's total value is computed by SIMTAX. Gross consumption tax revenue, on the other hand, varies according to real GDP growth ($\texttt{gdp\_rate}_t$). The two aggregates are calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.12 shows the values in current (2014) and constant (2010) dollars.

Equation

$$\texttt{cons\_tax}_t = \texttt{gross\_cons\_tax}_{t-1} \cdot (1 + \texttt{gdp\_rate}_t) - \texttt{soltaxcred}_t$$

Calibration

Table 8.12: Calibration of consumption taxes revenue with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
`gross_cons_tax`	19,341	17,993
`soltaxcred`	1,784	1,660
`cons_tax`	17,557	16,333

8.3.6 Duties and Permits and Miscellaneous Revenue

Duties, permits and miscellaneous revenue vary according to real GDP growth ($\texttt{gdp\_rate}_t$). They are calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.13 reports the values in current (2014) and constant (2010) dollars.

Equation

$$\texttt{other\_tax}_t = \texttt{other\_tax}_{t-1} \cdot (1 + \texttt{gdp\_rate}_t)$$

Calibration

Table 8.13: Calibration of duties and permits and miscellaneous revenue with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
other_tax	11,491	10,690

8.3.7 Government Entreprises

Revenue from government enterprises varies according to real GDP growth ($\texttt{gdp\_rate}_t$). It is calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.14 shows the values in current (2014) and constant (2010) dollars.

Equation

$$\texttt{gov\_corp}_t = \texttt{gov\_corp}_{t-1} \cdot (1 + \texttt{gdp\_rate}_t)$$

Calibration

Table 8.14: Calibration of revenue from government enterprises with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
gov_corp	5,336	4,964

8.3.8 Federal Transfers

Federal transfers (fed_trans) are split in two categories: Canada Health Transfer and other federal tranfers. The Canada Health Transfer increase of 6% per year before 2017. From this year the growth of this transfer follows the GDP growth (real GDP growth plus inflation at 2%) with a lower growth limit of 3%. The other federal tranfers vary according to real GDP growth ($\texttt{gdp\_rate}_t$). They are calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.15 reports the values in current (2014) and constant (2010) dollars.

Equation

$$\texttt{fed\_trans}_t = \texttt{fed\_trans}_{t-1} \cdot (1 + \texttt{gdp\_rate}_t)$$

Calibration

Table 8.15: Calibration of revenue from federal transfers with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
`fed_trans`	18,539	17,247

8.3.9 Revenue of the Generations Fund

The Generations Fund's revenue corresponds to revenue on the Fund's books, as described in section 8.2.

8.3.10 Total Revenue

Total revenue (`total_rev`) is the sum of all Quebec government revenue: personal income tax, contributions to the Health Services Fund, corporate tax, consumption taxes, duties, permits and miscellaneous revenue, government enterprises, federal transfers, and revenue of the Generations Fund.

Equation

$$\begin{aligned}\texttt{total\_rev}_t &= \texttt{pers\_inctax}_t + \texttt{pers\_fss}_t + \texttt{corp\_inctax}_t + \texttt{prop\_tax}_t + \texttt{cons\_tax}_t \\ &\quad + \texttt{other\_tax}_t + \texttt{gov\_corp}_t + \texttt{gen\_rev}_t + \texttt{fed\_trans}_t\end{aligned}$$

8.4 Budget Balance

According to the Balanced Budget Act, the budget balance (`balance`) equals total revenue (`total_rev`) minus total expenditures (`total_spd`), minus contributions to the Generations Fund (`gen_spd`). The latter are only subtracted in computing the budget balance, in order to reproduce Quebec government accounting. They equal the Generations Fund's book revenue (`gen_rev_acc`).

Equations

$$\texttt{balance}_t = \texttt{total\_rev}_t - \texttt{total\_spd}_t - \texttt{gen\_spd}_t$$

8.5 Total Debt

Total debt (total_debt) represents the sum of the debt issued on financial markets and the net liabilities for the retirement plans and future benefits of government employees (Ministère des Finances, 2016). It is equal to the real value of previous year's total debt plus investments in non-financial (fixed) assets (real_estate_inv), plus other investments (other_inv), plus other factors (other), minus budget balance[11] (balance). Total debt is calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.16 shows the value in current (2014) and constant (2010) dollars.

Equations

$$\begin{aligned}
\texttt{total_debt}_t &= \left(\texttt{total_debt}_{t-1}\frac{CPI_{t-1}}{CPI_{2010}} + [\texttt{real_estate_inv}_t \right.\\
&\quad \left. + \texttt{other_inv}_t + \texttt{other}_t - \texttt{balance}]\frac{CPI_{2010}}{CPI_t}\right)\frac{CPI_{2010}}{CPI_t}\\
&= \texttt{total_debt}_{t-1}\frac{CPI_{t-1}}{CPI_t} + \texttt{real_estate_inv}_t + \texttt{other_inv}_t + \texttt{other}_t - \texttt{balance}_t\\
\texttt{total_debt}_t &= \frac{\texttt{total_debt}_{t-1}}{1+\texttt{infla_rate}_t} + \texttt{real_estate_inv}_t + \texttt{other_inv}_t + \texttt{other}_t - \texttt{balance}_t
\end{aligned}$$

Calibration

Table 8.16: Calibration of total debt, without the Generations Fund, with respect to year 2014-2015

Category	Target (2014) in M$	Target in constant 2010 M$
total_debt	203,466	189,286

8.6 Gross Debt

Gross debt equals total debt minus the Generations Fund's book value. It is calibrated in 2014 to match the values published in the 2014-2015 public accounts (2015). Table 8.17 reports the values in current (2014) and constant (2010) dollars.

[11] We use the reasonable assumption that any budget surplus goes to reducing the debt. By default, a deficit is added to the debt for the following year.

Equation

$$\texttt{gross\_debt}_t = \texttt{total\_debt}_t - \texttt{gen\_fun\_acc}_t$$

Calibration

Table 8.17: Calibration of gross debt with respect to year 2014-2015

Catégorie	Target (2014) in M$	Target in constant 2010 M$
`gross_debt`	197,807	184,021

Part III

Statistical Appendices

Appendix A

Econometric Models

A.1 Logit and multinomial logit models

Many of the simulated random variables in the model are discrete variables. Some are binary variables (e.g., claiming CPP) while other can take a larger set of outcome (e.g., last degree obtained). Most of the time, these variables are simulated using models from the logit family.

A.1.1 Logit model

Most of the transition models rely on the simple logit model. This model is appropriate when we want to simulate binary outcomes, such as the probability to finish schooling or to have a child.

We are interested in the probability that the outcome of a random variable, denoted y, will take the value of 1 conditional on a series of observed characeristics contained in the vector $\boldsymbol{x}$. A standard result in econometrics tells us that:

$$\Pr(y = 1 \mid \boldsymbol{x}) = \frac{\exp(\boldsymbol{x'\beta})}{1 + \exp(\boldsymbol{x'\beta})} \tag{A.1}$$

A.1.2 Multinomial logit model

The multinomal logit is an extension of the logit model to allow for multiple ouctomes. Suppose that there are K possibles outcomes. Then, the probability that $y = c$ is given by:

$$\Pr(y = c \mid \boldsymbol{x}) = \frac{\exp(\boldsymbol{x'_c\beta_c})}{\sum_{k=1}^{K} \exp(\boldsymbol{x'_k\beta_k})} \tag{A.2}$$

Most of the time, in or case, $\boldsymbol{x_1} = \boldsymbol{x_2} = \ldots = \boldsymbol{x_k}$, meaning that we normalise the vector $\boldsymbol{\beta_1}$ to be a vector of 0. In this case, notice that the simple logit is a special case of the multinomial logit where there are only two possible outcomes.

A.2 Employment and earnings models

A.2.1 Employment

We consider a case where there are C_0 class of agents when they are not on the market, and C_1 agents when they are. Note that this leaves us with $C = C_0 \times C_1$ classes.

The probability to enter the market, conditional on membership to class $c \in \{1, 2, \ldots, C\}$ is given by

$$p_0(\boldsymbol{x}_{i,t}, c; \boldsymbol{\beta}^{\mathbf{0}}, \alpha_0^c) = \Pr(d_{i,t+1} = 1 | d_{i,t} = 0, \boldsymbol{x}_{i,t}) = \Lambda\left(\boldsymbol{x}_{i,t}\boldsymbol{\beta}_0 + \alpha_0^c\right) \tag{A.3}$$

Similarly, the probability to transit out of the labour force is given by:

$$p_1(\boldsymbol{x}_{i,t}, c; \boldsymbol{\beta}^{\mathbf{0}}, \alpha_1^c) = \Pr(d_{i,t+1} = 0 | d_{i,t} = 1, \boldsymbol{x}_{i,t}) = \Lambda(\boldsymbol{x}_{i,t}\boldsymbol{\beta}_1 + \alpha_1^c) \tag{A.4}$$

Each class is therefore defined by a combination of α_0^c and α_1^c. The vector of these two values is denoted $\boldsymbol{\alpha^c}$. The vector of parameter with all these values is denoted $\boldsymbol{\alpha}$.[12] Moreover, we denote $\boldsymbol{x_i}$ the set containing the information of all $\boldsymbol{x}_{i,t}$. The same notation is used for the other characteristics or decision.

The EM algorithm relies on three objects:

- $\pi^c(\boldsymbol{z}_i; \boldsymbol{\gamma})$ Probability of belonging to class c, unconditional to the decision sequence (but conditional on observables this probability), defined with parameters $\boldsymbol{\gamma}$.
- $\Pr(\boldsymbol{d}_i \mid \boldsymbol{x}_i, c; \boldsymbol{\theta})$ Probability of realization of the decision sequence conditional on observables $\boldsymbol{x_i}$ and on the class to which a respondent belongs. The probability is defined by parameters $\boldsymbol{\theta}$. Note that:
$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \\ \boldsymbol{\alpha} \end{pmatrix} \tag{A.5}$$
- $\Pr(c \mid \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{d}_i)$ Probability of belonging to a class conditional on the observed sequence of decisions $\boldsymbol{d_i}$ and on observables characteristics $\boldsymbol{z_i}$ and $\boldsymbol{x}_i$.

The sequence goes as follows.

1. Draw random initial values for $\Pr^t(c \mid \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{d}_i)$ (or fix varying arbitrary ones) and assume equal $\pi^c(\boldsymbol{z}_i; \hat{\boldsymbol{\gamma}}^{\boldsymbol{t}})$ for all classes.
2. Find the optimal parameters vector $\hat{\boldsymbol{\theta}}^{t+1}$ maximizing the following weighted log-likelihood:
$$\hat{\boldsymbol{\theta}}^{t+1} = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{N} \sum_{c=1}^{C} \Pr^t(c \mid \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{d}_i) \ln \Pr(\boldsymbol{d}_i \mid \boldsymbol{x}_i, c; \boldsymbol{\theta}) \tag{A.6}$$
Note that this optimization can be achieved using the built-in Stata function `logit`, using importance weight given by $\Pr^t(c \mid \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{d}_i)$. Post estimation command `predict` allows us to automatically compute the predicted probabilities of the observed sequence.

[12] See Pacifico (2013) for details regarding the algorithm applied in the analysis of labour supply.

3. Update the conditional probability of belonging to each type. This is given by:

$$\Pr^{t+1}(c \mid \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{d}_i) = \frac{\pi_c(\boldsymbol{z}_i; \hat{\boldsymbol{\gamma}}^{\boldsymbol{t}}) \Pr(\boldsymbol{d}_i \mid \boldsymbol{x}_i, c; \hat{\boldsymbol{\theta}}^{t+1})}{\sum_c \pi_k(\boldsymbol{z}_i; \hat{\boldsymbol{\gamma}}^{\boldsymbol{t}}) \Pr(\boldsymbol{d}_i \mid \boldsymbol{x}_i, k; \hat{\boldsymbol{\theta}}^{t+1})} \tag{A.7}$$

4. Update the unconditional probability of belonging to each type.

   This requires first to estimate the vector parameter $\hat{\boldsymbol{\gamma}}^{\boldsymbol{t+1}}$, which solves the following equation.

$$\hat{\boldsymbol{\gamma}}^{t+1} = \underset{\gamma}{\operatorname{argmax}} \sum_{i=1}^{N} \sum_{c=1}^{C} \Pr^{t+1}(c \mid \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{d}_i) \ln \frac{\exp(\boldsymbol{z}_i \boldsymbol{\gamma}_{\boldsymbol{c}})}{\sum_k \exp(\boldsymbol{z}_i \boldsymbol{\gamma}_{\boldsymbol{k}})} \tag{A.8}$$

   Note that this optimization can be achieved using the built-in Stata function `mlogit`, using importance weight given by $P^{t+1}(c \mid \boldsymbol{x}_i, \boldsymbol{c}_i, \boldsymbol{d}_i)$.

   Then, under the assumptions presented above, it follows that:

$$\pi^c(\boldsymbol{z}_{\boldsymbol{i}}; \hat{\boldsymbol{\gamma}}^{\boldsymbol{t+1}}) = \frac{\exp(\boldsymbol{z}_i \hat{\boldsymbol{\gamma}}_{\boldsymbol{c}}^{\boldsymbol{t+1}})}{\sum_k \exp(\boldsymbol{z}_i \hat{\boldsymbol{\gamma}}_{\boldsymbol{k}}^{\boldsymbol{t+1}})} \tag{A.9}$$

   Again, note that the post estimation command `predict` gives us these weights automatically.

5. Estimated parameters $\hat{\boldsymbol{\theta}}_{\boldsymbol{EM}}$ and $\hat{\boldsymbol{\gamma}}_{\boldsymbol{EM}}$ are obtained when the sequence improves the log-likelihood by an arbitrarily small amount. If this condition is not fulfilled, go back to 2.

At each iteration of the sequence, estimations of the conditional and unconditional probabilities for each class are computed. Hence, we have *ex post* probabilities that respondents belong to each type of respondents.

A.2.2 Earnings

The potential log earnings equation for a member of a given class is given by

$$\log y_{i,t} = \boldsymbol{x}_{\boldsymbol{i},\boldsymbol{t}} \boldsymbol{\beta} + \delta^c + \nu_i + \epsilon_{i,t} \tag{A.10}$$

where ν_i is a random effect which is normally distributed with variance σ_ν^2 and $\epsilon_{i,t}$ is white noise with variance σ_ϵ^2.

When estimating the model, the type of each respondent is unobserved. However, when we simulate the earnings model each respondent will have drawn a type.

Suppose that we knew the type of the respondents. In this case, we could estimate a standard random-effects model by maximum likelihood. The likelihood would then be:

$$\begin{aligned} &L(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_\nu, \sigma_\epsilon \mid y_{i,t}, \boldsymbol{x}_{i,t}, \boldsymbol{z}_i, \boldsymbol{d}_i, c_i) \\ &= \int_{-\infty}^{\infty} \left\{ \prod_{t=1}^{T} \frac{1}{\sigma_\epsilon} \phi\left(\frac{\log y_{i,t} - \left(\boldsymbol{x}_{\boldsymbol{i},\boldsymbol{t}} \boldsymbol{\beta} + \delta^c + u\right)}{\sigma_\epsilon} \right) \right\} \frac{1}{\sigma_\nu} \phi\left(\frac{u}{\sigma_\nu}\right) \mathrm{d}u \end{aligned} \tag{A.11}$$

such that $\phi(x)$ is the standard normal density function.

The estimator that we use is simply a weighted version of this maximum likelihood estimator given by:

$$\begin{aligned}&\ln L(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_\nu, \sigma_\epsilon) \mid y_{i,t}, \boldsymbol{x}_{i,t}, \boldsymbol{z}_i, \boldsymbol{d}_i) \\ &\quad = \sum_{c=1}^{C} \Pr(c \mid \boldsymbol{x}_{i,t}, \boldsymbol{z}_i, \boldsymbol{d}_i) \ln L(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_\nu, \sigma_\epsilon \mid y_{i,t}, \boldsymbol{x}_{i,t}, \boldsymbol{z}_i, \boldsymbol{d}_i, c)\end{aligned} \tag{A.12}$$

such that the contribution to the likelihood of a respondent takes into account the *ex post* uncertainty concerning the class he belongs to. This estimation can be obtained directly in `Stata` by using the command `mixed`.

Bibliography

Bibi, S. and Duclos, J.-Y. (2010). Y a-t-il plus de pauvreté au Québec qu'au Canada ? *L'Actualité économique*, 86(3):277–318.

Boisclair, D., Côté-Sergent, A., Laliberté-Auger, F., and Marchand, S. (2016). COMPAS: A health microsimulation model for Québec and Canada. Technical document, *Industrial Alliance Research Chair on the Economics of Demographic Change.* http://www.cedia.ca/sites/cedia.ca/files/dt7_compas_anglais_sept2016.pdf.

Canadian Institute for Health Information (2015). National Health Expenditure Trends, 1975 to 2015.

Fréchet, G., Lanctôt, P., and Morin, A. (2010). Du revenu après impôt au revenu disponible aux fins de la mesure du panier de consommation (MPC). Centre d'étude sur la pauvreté et l'exclusion.

Gouvernement du Québec (2015). Public Accounts 2014-2015 – Volume 1.

Institut de la statistique du Québec (2014). Perspectives démographiques du Québec et des régions, 2011-2061.

Marchand, S., Bissonnette, L., Blanquaert, A., and Duclos, J.-Y. (2016). SimTax Documentation, Version 15.0. Technical document, *Industrial Alliance Research Chair on the Economics of Demographic Change.*

Ministère des Finances (2016). The Québec Economic Plan – Budget 2016-2017. Gouvernement du Québec, 17 March.

Pacifico, D. (2013). On the role of unobserved preference heterogeneity in discrete choice models of labour supply. *Empirical Economics*, 45(2):929–963.

Statistics Canada (2010). Population Projections for Canada, Provinces and Territories, 2009 to 2036. Demography Division. http://www.statcan.gc.ca/pub/91-520-x/91-520-x2010001-eng.pdf.

Statistics Canada (2012). Low Income Lines, 2010 to 2011. Income Research Paper Series. http://www.statcan.gc.ca/pub/75f0002m/75f0002m2012002-eng.pdf.