The Springer Series on Demographic Methods and Population Analysis

Volume 43

Series Editor
Kenneth C. Land, Duke University

In recent decades, there has been a rapid development of demographic models and methods and an explosive growth in the range of applications of population analysis. This series seeks to provide a publication outlet both for high-quality textual and expository books on modern techniques of demographic analysis and for works that present exemplary applications of such techniques to various aspects of population analysis.

Topics appropriate for the series include:

- General demographic methods
- Techniques of standardization
- Life table models and methods
- Multistate and multiregional life tables, analyses and projections
- Demographic aspects of biostatistics and epidemiology
- Stable population theory and its extensions
- Methods of indirect estimation
- Stochastic population models
- Event history analysis, duration analysis, and hazard regression models
- Demographic projection methods and population forecasts
- Techniques of applied demographic analysis, regional and local population estimates and projections
- Methods of estimation and projection for business and health care applications
- Methods and estimates for unique populations such as schools and students

Volumes in the series are of interest to researchers, professionals, and students in demography, sociology, economics, statistics, geography and regional science, public health and health care management, epidemiology, biostatistics, actuarial science, business, and related fields.

More information about this series at http://www.springer.com/series/6449

Alain Bélanger • Patrick Sabourin

Microsimulation and Population Dynamics

An Introduction to Modgen 12

Springer

Alain Bélanger
Urbanisation Culture Société
Inst nat de la recherche scientifique
Urbanisation Culture Société
Montréal, Québec, Canada

Patrick Sabourin
Urbanisation Culture Société
Inst nat de la recherche scientifique
Urbanisation Culture Société
Montréal, Québec, Canada

Additional material to this book can be downloaded from http://extras.springer.com, http://www.statcan.gc.ca/.

ISSN 1389-6784 ISSN 2215-1990 (electronic)
The Springer Series on Demographic Methods and Population Analysis
ISBN 978-3-319-44662-2 ISBN 978-3-319-44663-9 (eBook)
DOI 10.1007/978-3-319-44663-9

Library of Congress Control Number: 2016948704

© Springer International Publishing Switzerland 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing Switzerland
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This new type of model consists of various sorts of interacting units which receive inputs and generate outputs. The outputs of each unit are, in part, functionally related to prior events and, in part, are the results of a series of random drawings from discrete probability distributions.

—Guy H. Orcutt (1957)

Foreword

This book is intended for anyone who needs a practical introduction to microsimulation in general and to the Modgen program in particular. We have selected Modgen, a Statistics Canada software package, as the platform for this initiation in microsimulation because of its flexibility, performance, accessibility and multifunctional capacity. Our view is that although Modgen is increasingly being used outside Statistics Canada itself, it is still under-exploited. Recent European microsimulation projects have invested significant resources in building their own microsimulation platforms, but these have not always proved to be as flexible or to perform as well. The underuse of Modgen is probably due mainly to the fact that the documentation produced by Statistics Canada is essentially aimed at those who are already experienced Modgen users. We believe that there is a need for a training manual designed for beginners, offering a simple but realistic microsimulation model. This book is intended to fill this gap. It provides a series of exercises which are specifically tailored to enable the beginner to build up his or her first dynamic microsimulation program step by step. To create a dynamic microsimulation model, a designer has to begin by identifying the events which will affect a population and which will drive its evolution over time. Whether the population is made up of businesses, households, people or other elements (or cases, to use the technical term), these individual elements all have to come into existence (or be born), evolve (grow older) and disappear (or die). Each particular population model, then, exists in a particular setting. In this book we use examples of human populations and build a model of demographic projection which simulates the evolution of the population of the different regions of Canada, based on certain demographic characteristics. Although we will be using some of the concepts and methods of demography, the non-demographer should find it relatively easy to practise the exercises and then to transpose his or her skills in using Modgen into the development of models in another expert field, such as economics, sociology, biology or elsewhere.

The idea for this book emerged from a seminar organised by the Population Change and Lifecourse Strategic Knowledge Cluster (PCLC Cluster). Because of a lack of documentation on Modgen, the Cluster decided to sponsor the production of a training manual to encourage Modgen users to learn how to use the software

program, and this task was given to us. Through the Cluster, this book has therefore received financial backing from the Canadian Social Sciences and Humanities Research Council (SSHRC). The authors would like to acknowledge the support of SSHRC and of its director, Professor Roderic Beaujot. We would also like to thank Samuel Vézina for proofreading the original manuscript and Julien De Gouffe from Statistics Canada for his precious technical support. The book was originally written in French and has been translated into English by Duncan Fulton, whom we also wish to thank for his work.

Of course, any errors or imprecisions that you may encounter in this book are entirely ours. We are always happy to receive suggestions and comments, and you can write to us at comments@microsimulationandpopulationdynamics.com.

Québec, Canada

Alain Bélanger
Patrick Sabourin

Installing Modgen 12

Modgen is a microsimulation package based on the C++ programming language. The Modgen software is designed to be used in conjunction with Microsoft Visual C++.

Fortunately, Visual C++ is now available at no cost, as Microsoft has released a free version of its Visual Studio suite (Community Edition). You must download and install Visual Studio Community Edition before installing Modgen.

You will then be able to download and install the Modgen 12 software which is also available for free on the Statistics Canada website.

It is important that you install Visual Studio before Modgen, since Modgen will introduce some of its components in the Visual Studio files. Both installations are straightforward and should not be problematic. Complete installation instructions, along with download links to Visual Studio and Modgen can be found on the book website: www.microsimulationandpopulationdynamics.com.

Modgen 12 is scheduled to be released sometime in the Winter of 2017. Should it still be unavailable by the time you buy this book, you can write to statcan.microsimulation-microsimulation.statcan@canada.ca and ask for an advanced copy of the program.

Users of previous versions of Modgen will notice that some substantial changes have been made in version 12. Most notably, the Modgen toolbar has been removed, and precompilation (more on that topic in the book!) is now performed automatically upon compilation with the Visual C++ compiler. The developer's guide and other files, previously accessible through the Modgen toolbar, are now only available through the start menu.

For up-to-date information on the book and on Modgen, visit http://www.microsimulationandpopulationdynamics.com/.

Contents

1 Creating a Basic Cohort Model 1
1.1 Using the New Model Wizard 2
1.2 The ModgenExample.mpp File (Appendix 1.1) 6
1.3 The PersonCore.mpp File (Appendix 1.2) 8
1.4 Compiling and Running the Simulation Program 16
1.5 Summary 21
Appendices 21
Appendix 1.1 ModgenExample.mpp 21
Appendix 1.2 PersonCore.mpp 24

2 The Life Table 29
2.1 Adding a Classification 30
2.2 Adding a Numerical Range 32
2.3 Adding an Event: The Birthday Event 33
2.4 Modifying the *Start* Function 35
2.5 Modifying the Mortality Event Function 37
2.6 Adding Values in the Parameters File 39
2.7 Generating New Tables: The Elements of the Life Table 41
2.7.1 Derived States in the Table 43
2.8 Modifying Parameters and Reorganising Tables with the User Interface 50
2.9 Summary 55
Appendices 55
Appendix 2.1 PersonCore.mpp 55
Appendix 2.2 Base(PersonCore).dat 61

3 The Multiple Increment-Decrement Life Table 63
3.1 Self-Scheduling Events: The *self_scheduling_int* Function 64
3.2 Creating a New Module for Interprovincial Migration 66
3.3 Assigning a Province of Birth Using a *cumrate* Parameter and the *Lookup* Function 71
3.4 Adding an Internal Migration Event 74

3.5 Modifying the Mortality Parameter 78
3.6 Multi-regional Tables 79
3.7 Manipulating Tables in Modgen (Advanced Topic) 83
3.8 Adding a Parameter File to a Scenario 86
3.9 Summary 88
Appendices 89
Appendix 3.1 PersonCore.mpp 89
Appendix 3.2 Migration.mpp 93
Appendix 3.3 Base(Migration).dat 97

4 Modelling Fertility 99
4.1 Adding Calendar Time 100
4.2 Creating a Fertility Module 103
4.3 Modifying the *PersonCore* Module and the *Start* Function 109
4.4 Bringing the Tables Together in a Results Module 114
4.5 Looking at the Results 117
4.6 Summary 120
Appendices 121
Appendix 4.1 PersonCore.mpp 121
Appendix 4.2 Fertility.mpp 125
Appendix 4.3 Tables.mpp 127
Appendix 4.4 ModgenExample.mpp 128

5 The Base Population 131
5.1 Preparing a Microdata File 132
5.2 Importing External Data into Modgen 133
5.3 Integrating Import Functions into the Main File 134
5.4 Using the Imported Data and Modifying the *Start* Function 135
5.5 Modifying Calendar Time 140
5.6 Modifying the Scenario and Running the Model 141
5.7 Summary 143
Appendices 144
Appendix 5.1 ModgenExample.mpp 144
Appendix 5.2 PersonCore.mpp 147
Appendix 5.3 Tables.mpp 151

6 International Migration 153
6.1 Creating an International Migration Module 154
6.2 The Emigration Event 157
6.3 Modifying the *Start* Function 158
6.4 Modifying the Simulation Loop to Include Immigration 161
6.5 Adjusting Weights 165
6.6 Generating Results 166

6.7 Summary 169
Appendices 170
Appendix 6.1 InternationalMigration.mpp 170
Appendix 6.2 ModgenExample.mpp 171
Appendix 6.3 PersonCore.mpp 175
Appendix 6.4 Tables.mpp 179

Conclusion 181

Appendix: Coding Standards 185

Glossary 187

References 189

Index 191

List of Figures

Fig. 1.1 Opening a new Visual C++ project 2
Fig. 1.2 Creating a new Modgen project 3
Fig. 1.3 Selecting the language for a model 5
Fig. 1.4 Visual Studio programming interface 5
Fig. 1.5 Compiler output messages 17
Fig. 1.6 Executable file in the project folder 18
Fig. 1.7 Modgen model user interface: opening the base scenario 19
Fig. 1.8 Table of results 19
Fig. 1.9 Modifying the set-up parameters 20
Fig. 1.10 Modifying the mortality parameter in the model 20

Fig. 2.1 Accessing the Modgen developer's guide 46
Fig. 2.2 The derived states section of the developer's guide 46
Fig. 2.3 Section of developer's guide explaining the use of the *duration* derived state 47
Fig. 2.4 Mortality parameters 50
Fig. 2.5 Scenario parameters 51
Fig. 2.6 Life expectancy at birth by sex 52
Fig. 2.7 Comparing methods for calculating life expectancy 52
Fig. 2.8 Survivors by age and sex 53
Fig. 2.9 Table properties 54
Fig. 2.10 Survivors, deaths and person-years lived by age 54

Fig. 3.1 Accessing the menu to add a new Modgen module 67
Fig. 3.2 Adding a Modgen module 67
Fig. 3.3 The empty migration module 68
Fig. 3.4 Menu for adding a new parameters file 68
Fig. 3.5 Creating a new parameters file 69
Fig. 3.6 New parameters file for the Migration module 69

Fig. 3.7 Adding a parameters file to a scenario 87
Fig. 3.8 User interface, multi-regional tables 87
Fig. 3.9 Multi-regional life expectancy table 88

Fig. 4.1 Demographic parameters according to event distribution 107
Fig. 4.2 Adding the fertility parameters file 118
Fig. 4.3 Table of births and age-specific fertility rates 118
Fig. 4.4 Comparison of simulated fertility rates and input fertility rates for the province of Québec (100,000 cases for the whole of Canada) 119
Fig. 4.5 Population, births, deaths, exits and entries, by province 120

Fig. 5.1 Format of data in the base population file 133
Fig. 5.2 Modifying the number of cases in the scenario 142
Fig. 5.3 Table of results 142

Fig. 6.1 Diagram of microsimulation algorithm 162
Fig. 6.2 Number of immigrants per five-year period 168
Fig. 6.3 Canadian population according to high (300,000), medium (250,000) and low (200,000) immigration scenarios 169

Introduction

IF WE ASK A DEMOGRAPHER to describe the future evolution of a given population, he or she will start by gathering two types of information. The first of these concerns the actual population numbers by age group and sex (and perhaps by region), which are the starting point for an elementary form of population forecast or projection. The demographer's next question will be about the frequencies with which the main demographic events occur – births, migrations and deaths. These frequencies will be used to calculate rates for each age and gender group. So our imaginary demographer will be building a model whose operationalisation requires population numbers for every group considered as "homogeneous" in some way (e.g. in terms of age or sex) and rates which represent the degree of risk (of mortality, birth, migration, etc.) that these groups are exposed to. The number of such events projected per year is obtained by multiplying the population numbers by the risk in question. Based on analysis of more or less recent trends or on comparison with trends observed in other regions or countries, hypotheses are formed about the likely evolution of these rates in the future.

From Multi-state Macro Models to Microsimulation

This projection methodology is determinist in nature. The uncertainty attached to the future evolution of the various rates is estimated by generating a consistent set of scenarios which incorporate different plausible hypotheses about future rates of fertility, migration and mortality. These scenarios allow us to generate a set of projections; as future events unfold, their actual values should fall within the limit values of these projections. A probabilistic component can also be added to the model by estimating, in addition to the central tendency, the variance of each parameter (e.g. based on a chronological series). In this way the same projection can be repeated several times, giving the rates a random value depending on the variance attributed to each of them.

Any model is necessarily an abstraction from reality, but these kinds of models, whether they are deterministic or probabilistic, have several major limitations. Firstly, all of them assume that each sex and age group is internally homogeneous, that is, that all the individuals in a group experience the same risk of undergoing a particular type of event (e.g. that all women aged 22 have the same risk of dying, migrating or giving birth). Such a hypothesis rarely if ever holds true in practice, and this is bound to result in a bias in the projected results. To illustrate this, suppose that a population is composed of two subgroups (e.g. with different languages or religions), which are different from each other in terms of their fertility (and, to keep things simple, solely in that respect). Assuming that fertility rates remain constant, the more fertile subgroup will gradually increase in its relative size within the population as a whole. But if the relative size of this more fertile group within the population increases, then the fertility of the population as a whole will automatically rise, even though there has been no change in the behaviour of the women of either of the sub-groups. This will mean that the further into the future the projections extend, the more this simple macro model will tend to under-estimate the number of future births.

Another limitation of traditional macro models is linked to the nature of the projection hypotheses. According to these, the probability of an event occurring in the future depends entirely on present conditions. To use the example of fertility again, the probability of a woman giving birth is assumed to depend solely on her age and the region where she is living at time t, without taking account of her country of birth or her parity (the number of children she has). But if fertility behaviour is partly a result of the social and cultural environment in which a person was brought up, then a woman born in a country with higher fertility will have, on average, higher fertility than one born in a region with lower fertility, irrespective of where she is living at time t. But a simple cohort-components model is unable to take these differences into account, because the fertility of both women are assumed to be the same as the fertility of all women of the same age in their region of residence.

We should point out that it is theoretically possible to make the cohort-components model more complex, so as to account for the underlying heterogeneity of populations (for example in terms of ethnic group or region of birth). But the difficulty here is more technical, and results from the fact that component-based projections usually depend on matrix computation. Without going into details, the key point here is that the size of the matrix increases exponentially with the number of variables being considered. So the model quickly becomes unworkable (Van Imhoff and Post 1998).

Microsimulation enables us to overcome these limitations by creating a randomised series of experiences for each individual in the projection and so building up a unique biography for each simulated case. Looking again at the fertility example, we can trace the stages in a simple microsimulation model. The first step is to compile a representative sample of the female population of the region being studied (e.g. using publicly available micro-level data from a census). For each woman in the sample and for each year in the simulation, a randomised experiment is then carried out whose probability of success is a function of the level of fertility at each age. To put this in concrete terms, a random value between 0 and 1 is compared to the probability of giving birth, for each woman in the sample. If the random value

is lower than the probability of giving birth, an individual aged 0 is added to the simulation. If the opposite is true, there is no birth and the model moves on to another event. The randomized experiment is repeated for every fertile age, varying the probability of giving birth. The result for the population as a whole is obtained by aggregating all the individual results, using appropriate weighting.

Access to increasingly powerful computers at ever lower cost, as well as improved availability of survey microdata, has meant that microsimulation has become increasingly widely used over recent decades. The development of programming languages and of software such as the Modgen program presented in this book has made it easier to create microsimulation models and has helped to increase the numbers of people using them. However, even with the help of Modgen, developing a dynamic microsimulation model is, initially, a potentially more demanding task than developing a cohort-component projection model or even a multiregional model. Considerable resources have to be devoted to programming, establishing parameters and validating the model.

But despite the effort needed to set them up, microsimulation models are nevertheless superior to macromodels. One of their main advantages, as argued by Orcutt (1957) in his original article, is theoretical in nature. The most powerful theoretical models for explaining human behaviour, such as decisions to have a child or to migrate, operate at the level of the individual. So it makes sense to simulate these behaviours at the individual level. The same argument explains another strength of microsimulation which we have already touched on. This concerns changes in the composition of the population which are hard to capture in macro models but which can be easily taken into account by a microsimulation model. Behaviours like those involving fertility and mobility can be made to vary according to a large number of relevant variables such as country of birth, level of education or religion. At the analytical level, microsimulation offers greater flexibility and is usually the only way to obtain consistent results, especially when a number of different characteristics have to be projected at the same time. For example, the microsimulation model developed at Statistics Canada to forecast the ethnocultural diversity of the Canadian population (Alain Bélanger and Caron Malenfant 2005) required simultaneous projection of several ethnic variables: immigration status and immigration period, membership of a visible minority group, religious denomination and mother tongue as well as age, sex and region of residence. Two other modelling options were available to the researchers developing this model, as alternatives to microsimulation – but one of them would have been impossible to implement, and the other would have yielded inconsistent results. If we had attempted to carry out this projection using a multistate model, the transition matrix would have had several hundred million cells, much more than the total number of inhabitants of Canada. This matrix would have been so big that it would have been impossible to create and manipulate. The other option would have been to carry out separate projections for each dimension of cultural diversity. Such projection models might have been manageable, but their results would not have been internally consistent. This is because there is no reason to assume that population projections based, for example, on immigrant status would arrive at the same totals as those based on visible minority group

membership or religious affiliation. Furthermore, this approach would not have allowed cross tabulation of all the projected variables, and this would necessarily have reduced the quality of the projection itself as well as the richness of the results produced. It is much easier to introduce a number of explanatory factors for each type of behaviour into a microsimulation model, because there is no need to increase the state space (the matrix) exponentially. A new explanatory factor (a state) can be simply added to the existing model without having to reorganise everything that has been done up to that point.

Conditional relationships (interactions) between explanatory variables are practically impossible to account for in a conventional projection model. For example, the relationship between the fertility of immigrant women, their ethnic origin and their religion can be modelled relatively easily in a microsimulation model, but significantly increases the number of cells in the transition matrix of a cohort component-model (it would be multiplied by the number of possible states in the immigrant status times the number of possible ethnic origins times the number of included religions). In the same way, modelling variables which vary over time is theoretically possible as an extension of multistate models, but in practice this would also lead to an exponential increase in the size of the transition matrix or to such complications that it is hardly ever attempted (Wolf 2001). The healthy immigrant effect-the effect of duration of residence of immigrants in the host country on their relative mortality risk-is an example of the implementation of a continuous variable in a projection model.

Finally, a multistate model does not allow for the modelling of possible interactions between the actors (in a model, individuals are referred to as "actors" or "cases"). There are numerous social processes in which the interactions between individuals are essential elements in understanding the phenomenon (and therefore in modelling it). For example, in family demography, modelling the formation of unions may require projecting the entire set of individuals in a population so as to enable a marriage market to be built up and interactions between the members of a couple to be generated. Modgen for instance enables us to build agent-based microsimulation models where individual agents may interact (by exchanging information, for instance), but these latter models are not dealt with in this introductory book. Our focus will be placed exclusively on the modelling of the life courses of non interacting individuals.

What Is Modgen?

Modgen is a framework for the development of microsimulation models. It includes a specialised microsimulation programming language, a runtime library and a suite of standalone applications which can be used with Modgen modules. Modgen was designed at Statistics Canada in 1994 to facilitate the programming and modification of microsimulation models. The aim was to develop a generic programming environment which would allow all kinds of microsimulation projects to be carried out.

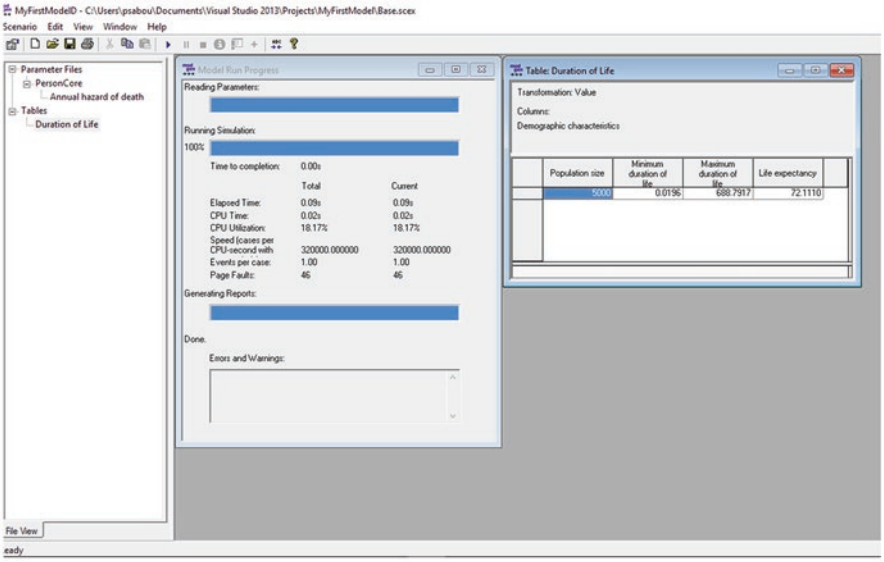

Fig. 1 The user interface of a Modgen model

The Modgen Language is a superset of the C++ programming language, so that Modgen functions are formed of C++ functions. The elements of the language specific to Modgen are processed by a precompiler which converts Modgen code into C++ code. The code is then processed by the C++ Visual Studio compiler to create a functional program. That is why Modgen 12 requires Visual Studio to be installed. There is no need to be an expert in C++ programming in order to use this book, but a basic knowledge of programming concepts will be very useful for understanding the code in the models. **The library** contains elements which are common to all the microsimulation models created using Modgen, such as the **user interface** (Fig. 1). Whereas the structure of the user interface is common to all Modgen models, its content is individualised for each one. It enables the user to run the model, consult and modify the parameters, control the simulations and manage the scenarios, as well as examine and export the tables generated by the model. With this interface it is possible to use projection models without necessarily knowing how to design them, in other words without needing to know the Modgen programming language. So a model can be distributed to external users, such as experts and decision makers, who can themselves set up their own scenarios and derive projections from them.

The Modgen Toolkit includes numerous optional applications, some installed with Modgen and others available online. The most commonly used tools are Modgen BioBrowser and the Modgen Web Interface. Microsimulation programs often require large databases to be created in which output information on the simulated individuals is stored; analysis of results then has to be done separately using external database software. With Modgen the simulation takes place at the individ-

ual level, but the tables generated by the model show aggregated results according to the designer's specifications. So Modgen generates tables in real time during the simulation. This means that there is no need for output databases, which can take up a large amount of time and processing power. But Modgen still maintains the option of observing individual life courses, thanks to the monitoring option provided by Modgen BioBrowser. BioBrowser allows a subset of the simulated population of individual life courses to be observed graphically, with its dimensions and variables specified by the user. Developers more interested by a life course perspective than by the aggregated projection results per se will find the BioBrowser very useful. The BioBrowser is not covered in this book, but information on this topic can be found in the Modgen developer's guide.

The Modgen Web Interface has been developed by the Canadian Partnership Against Cancer. It enables the user to operate a Modgen model from a website. This interface means that the user can see and modify the parameters, see the results in graphic or table form as well as compare the results of different scenarios.

Types of Models

Modgen has been designed to be very flexible, and the software allows for both static and dynamic[1] models, although dynamic models are the most commonly used. For dynamic models, Modgen adopts a competing-events approach which is general enough to permit the use of continuous-time, discrete-time, or combined models.

So-called "time-based" models, in which an entire population is simulated simultaneously so as to take into account the interactions between its members, are possible. These models are often known as multiagent or agent-based. So-called "case-based" models, which we will use for the examples in this book, are more usual. In these models the entire population is simulated, but one case at a time, with each single case representing an individual (or some other unit, a family or a firm for instance).

[1] Typically, static models use weighted aging techniques to display the population over the course of the simulation. The characteristics of the actors do not change, but their weighting in the population is modified. In practice these models use macrodemographic projections carried out separately, usually by an official body, to modify the weighting of individuals in each age group. Static models are often used to measure the anticipated effect of a change in public policy, for example a fiscal measure, on individuals in terms of winners and losers.

Dynamic models work by continuously updating each of the characteristics of each individual across time: this is the type of model described in this book. Individuals evolve over time, experiencing life cycle events such as unions and their breakdowns, births, migratory movements, educational advancements, entry into and exit from the labor market etc. up to the time of death or the horizon of the projection. Each of these events occurs with a probability varying according to the characteristics of the actor. A specific risk is therefore calculated for every event which each actor in the projection may encounter, based on their current characteristics.

Finally, whatever the type of model chosen, the simulated populations may be synthetic (determined by the designer) or based on empirical micro-data (e.g. census data). There are two main types of dynamic model: population models and cohort models. These two types use similar processes for aging and for modification of the characteristics of individuals, but differ in terms of the population used as the starting point for the simulation. A cohort model uses a cohort of individuals all born at the same time as its starting population. So the base population of a cohort model is a synthetic population. A population model instead takes as its starting point a representative sample of the projected population, generally derived from publicly available census data. The first four chapters of this book deal with developing a cohort model, and the two remaining chapters with a model of the Canadian population.

Structure of the Book

When a Modgen model is first created, the software automatically generates a default model that projects a population cohort subject to a fixed death rate. The program also calculates the life expectancy of this imaginary population and creates a table to present this result. We begin, in Chap. 1, by describing how to set up this automatically generated program and how to specify some of its parameters. We use this simple model to introduce certain fundamental ideas and to describe a number of essential functions for the development of all microsimulation models using Modgen.

Chapter 2 deals with the further development of the model automatically generated by the creation wizard. In this chapter we introduce the basic factors enabling some variation in the risk of death, namely age and sex. This produces a more realistic microsimulation model by cohort which simulates life expectancy and the other elements of the life table, using mortality data from the Canadian province of Ontario in 2006. The programming exercise developed throughout this chapter will teach the reader to add new states to a model, such as sex (male or female), and to create a new event function, such as the birthday. The latter will provide the age of the actor in completed years, thus enabling us to distinguish it from exact age, which is also used for the purposes of calculating the life table. Finally, the reader will learn how to modify an existing event function and how to add new dimensions to a results table.

Chapter 3 continues the development of this cohort model by integrating a regional dimension, thus transforming the life table created in Chap. 2 into a "multistate" table. This chapter's example will help the reader to learn how to create a new event function that is more complex than the birthday event. This event is part of a module on interregional mobility, where internal migration is modelled in two stages. First, the probability of an individual leaving his or her place of residence is compared to a random number. If the Monte Carlo experiment succeeds (if the random number is lower than the probability of leaving), the model will randomly

assign a new region of residence to the individual according to a distribution specific to each region of origin. Through reproducing the example in this chapter, the reader will have the opportunity to become familiar with a type of parameter (*cumrate*) and a function (*lookup*) which are specific to Modgen. These functions enable a new value to be randomly assigned to a state variable according to an arbitrary distribution matrix. In practice, this model will generate results which are equivalent to the results obtained by using a multiregional table.

In succeeding chapters, the cohort model will be gradually transformed into a population model. In Chaps. 4 and 5, we add a fertility module and build in a starting population. This makes it possible to carry out closed multi-regional projections i.e. without new entries resulting from international migration. International migration is included in the model in Chap. 6, making it possible to do "open" multiregional projections.

Data search and estimation of parameters are of course an important part of the creation of a microsimulation projection model, but this laborious task is not necessary in order to learn how to use Modgen. The data used in our examples are provided in Excel files which are available on the book website at http://www.microsimulationandpopulationdynamics.com/. The data is taken from widely available microdata files from the 2006 Canadian census and from vital statistics data available on the Statistics Canada website.

All the files used in this book are available on the Internet at the following address: http://www.microsimulationandpopulationdynamics.com/ or on Springer Extras Online. For each chapter, the Modgen code for the example (i.e. the complete Visual Studio solution) and the Excel files needed for the parameters are provided. The code for the relevant files for the first two exercises is given in full in the appendices to Chaps. 1, 2 and 3 to help the reader become familiar with the structure of the program. In succeeding chapters, the reader who is by then initiated will sometimes be required to look up the code directly in the solution files on the website.

***

The reader who works through all the examples in the book will learn how to create a dynamic projection model using relatively complex microsimulation tools and, what is more, a model that would have been difficult to implement using the traditional cohort-component method. By this time he or she will have enough working knowledge of Modgen to be able to develop his or her own model of dynamic microsimulation.

As a final note, the model developed in this book is a slimmed-down version of a projection model developed by Alain Bélanger and his team in the Dynamic Simulation Laboratory. This model enables a simulation of the Canadian population and its subregions based on a representative sample of the population, along a set of demographic and ethnocultural dimensions (immigration status, visible minority membership, mother tongue, home language, religion, etc.).