EXPLORING MICROSIMULATION METHODS FOR ESTIMATING DEMOGRAPHIC CHARACTERISTICS

David Morávek

Abstract

The use of microsimulation methods in the field of population estimates and population projection is becoming increasingly popular with the development in the field of computer technology. Unlike conventional population projection methods, which are essentially based on aggregated data, these methods can produce richer output and more accurate results. These data are needed in policy planning not only at the national but also at the regional level. The aim of the paper is to estimate the expected number of inhabitants of Czechia according to age and sex and their other demographic characteristics – marital status and education. Furthermore, data on mortality and fertility are entered into the model to obtain the demographic structure of the population in the following years. The method is a Monte-Carlo simulation that assigns births and deaths to the population based on the age-sex and marital-education-specific fertility and mortality rates.

Key words: microdata, microsimulation, Monte Carlo, population census, Czechia

JEL Code: C53, C63

Introduction

Due to the demographic development in several developed countries, including Czechia, in recent decades there has been an increase in the proportion of the elderly as a result of the increasing life expectancy and the falling birth rate. It is necessary to deal with this issue, because the demographic development affects not only the economy of the given country, but also healthcare, social policy, housing policy, education, etc. In general, the aging of the population is a phenomenon that brings with it some obligations related to health and social care along with pension security. In order to propose a solution in this area, it is necessary to have an outline of future demographic development with more detailed results, not only about the number and age distribution of the population, but also about the development of the socio-demographic characteristics of the given population.

Traditional methods used to estimate future demographic development are most often based on the so-called cohort-component method (Van Imhoff & Post, 1998), which is referred to as deterministic approach. In recent years, together with the development of computer technology, new possibilities for estimating future demographic development are coming to the fore, which are essentially based on individual data. Their main idea is that they view the given population as heterogeneous, whereas traditional methods are based on a homogeneous population. An individual view of individual members of a given population and their characteristics is advantageous in the sense that it allows us to obtain a more detailed and at the same time more accurate outline of the expected demographic development, and thanks to this, it is possible to create more accurate models with subsequent recommendations for a given country's economy. Moreover, thanks to the maturity of today's computer technology, it is possible to perform a whole range of operations that were not possible before, and this facilitates the use of these methods not only in population projections.

The first microsimulation models date from the late 1950s and were introduced by author Guy H. Orcutt (Orcutt, 1957). Their importance lay mainly in the assessment of the effects of economic and financial policy (see e.g. Birkin, 2021). Later, these models began to be used in a number of other areas, including social sciences and demography (Lomax & Smith, 2017). Microsimulation models are essentially based on individual data, which is why they are sometimes also referred to as "individual-based state-transition" models. As Willekens and Hein (2014) state, these models create and describe individual life histories. Booth (2006) mentioned two important advantages of multistate modelling and its focus on transitions between states are the incorporation of behavioural feedbacks and the simulation of hypotheses for assessing the effect of behavioural change. Several successful applications of microsimulation models are known (see for example O'Donoghue, 2014). In Czechia for example, a dynamic microsimulation model is used to estimate the future development of the pension system (Deloitte, 2011).

1 Data and methods

The main data source for the microsimulation model is the 2011 Census, namely, it concerns micro data about individuals according to age, sex, marital status, education and other socio-economic characteristics. According to Morand et al. (2010), the starting population can be a cross-sectional population or a birth cohort. In the first case, a sample of the whole population at initial time is taken as a starting point and all individuals are simulated from this time to the end-ing date. In the second case, a sample of individuals belonging to the same birth cohorts are simulated from their birth to their death. Tanton (2014) noticed that the synthetic reconstruction method creates a synthetic list of individuals and households where the characteristics of the individuals, when aggregated, will match known aggregates from another data source in the area being estimated.

The model used for population projection is based on individual data. Moreover, data from vital statistics on births and deaths are used to calculate specific mortality and fertility rates. The microsimulation model used here is based on the demographyMicrosim model (Lomax & Smith, 2017), in which its basic input parameters have been modified.

The model works with the following input assumptions (Lomax & Smith, 2017):

- Only single births occur (i.e. we assume that multiple births are factored into the fertility rate).
- Newborn have an equally probable chance of being male or female.
- The marital and education of the newborn is the same as their mother's.
- Births occur before deaths – thus a newborn will survive if a parent dies within the same year.
- Migration is not taken into account.

The methodology is a Monte-Carlo simulation that assign births and deaths to the population based on the age-sex and marital-education fertility and mortality rates supplied. The simulation is discrete in that it operates in one-year intervals.

The model algorithm includes the following steps (Lomax & Smith, 2017):

1. Load the age-sex-marital-education-specific fertility and mortality rates.
2. Randomly assign births and deaths to members of the population in a manner that is consistent with the fertility and mortality rates.
3. Age the population by one year.
4. Insert newborns (aged zero) and remove the deceased from the population.
5. Repeat from step 2 until the target year is reached.

2 Results

The following chapter shows the results of population estimation by microsimulation model using Monte Carlo algorithm. Regarding the input data for the model, the official figures from the Census in 2011 were adjusted according to the input assumptions of the model. The initial population was modelled in time to 2021. The result is an estimate of the population by sex and age, and then by marital status and education.

2.1 Population by sex and age

In Table 1 is the starting population used for the microsimulation model based on data from the 2011 Census. In 2011, is the base population equal to 9 990 077 in total, of which 4 860 662 were men and 5 129 415 were women. Until 2021, a decrease in the number of inhabitants was observed.

Tab. 1: Total population by sex in 2011 and 2021 (est.), Czechia

Year	Total	Males	Females
2011	9 990 077	4 860 662	5 129 415
2021 (est.)	9 356 760	4 943 162	4 413 598

Source: Census 2011 data, author‘s calculations

In Table 2, the population is divided according to age groups 0–14, 15–64 and 65+. When comparing the years 2011 and 2021, the number of people over the age of 65 increased as a result of the ageing of the population. Conversely to that, the number of people aged 0–14 decreased, as did the number of people aged 15–64 representing the productive population.

Tab. 2: Total population by age in 2011 and 2021 (est.), Czechia

Year	0–14	15–64	65+
2011	1 502 470	6 885 003	1 602 604
2021 (est.)	745 758	5 553 248	3 057 754

Source: Census 2011 data, author‘s calculations

2.2 Population by marital status and education

In Table 3, the population is divided according to marital status, namely single, married, divorced and widowed. The structure of the population according to marital status has not changed over time. When comparing the years 2011 and 2021, there was an increase in the number of widowed persons. Otherwise, the number of persons decreased in other marital statuses, but singles decreased the most.

Tab. 3: Total population by marital status in 2011 and 2021 (est.), Czechia

Year	Single	Married	Divorced	Widowed
2011	3 973 658	4 255 958	1 024 304	736 157
2021 (est.)	3 429 097	4 110 764	1 018 348	798 551

Source: Census 2011 data, author‘s calculations

In Table 4, the population is divided according to levels of education, namely primary, lower and upper secondary, higher education and persons with no education. When comparing the years 2011 and 2021, the structure of persons according to education has not change over time. Due to the expected decrease in the number of inhabitants, the number of people according to educational levels also decreased at all levels.

Tab. 4: Total population by education in 2011 and 2021 (est.), Czechia

Year	No Education	Primary	Lower Secondary	Upper Secondary	Higher
2011	1 573 797	1 547 651	2 959 864	2 793 179	1 115 586
2021 (est.)	1 493 977	1 431 661	2 897 127	2 543 788	990 207

Source: Census 2011 data, author‘s calculations

Conclusion

The potential of using microsimulation methods in demography is beneficial given that it is possible to obtain more detailed results of the expected population development compared with conventional methods of population projections, which estimate the population by age and sex only. On the other hand, an advanced technical expertise in data preparation and data modelling is essential to ensure consistency and accuracy of estimated results.

Based on the microsimulation model, we obtained population estimates by age and sex, as well as by marital status and education. The input data were individual data from the 2011 Census. The 2011 baseline population was modelled using the microsimulation model demographyMicrosim, whose source code is publicly available in the R programming language by the authors (Lomax & Smith, 2017). The input assumptions of the model and the algorithm of the model were presented and discussed.

For the next steps, there is a need to keep the model assumptions more realistic as the projection omits crucial factors (most notably migration). In addition, different scenarios of expected demographic development should be tested and implemented it in the provided R code. Somehow, the population coverage in the input data should be increased (e.g. by data imputation methods) as some data were removed from the initial set because of their inconsistency with mortality and fertility data.

References

Birkin, M. (2021). Microsimulation. In *Urban Informatics* (pp. 845–864). Springer, Singapore.

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International journal of forecasting*, *22*(3), 547–581.

Deloitte. (2011). The summary based on the Final Project Report of the Dynamic Micro-simulation Model of the Czech Republic.

Lomax, N. M., & Smith, A. P. (2017). Microsimulation for demography. *Australian Population Studies, 1*(1), 73–85.

Morand, E., Toulemon, L., Pennec, S., Baggio, R., & Billari, F. (2010). Demographic modelling: the state of the art. *Paris, INED. FP7-244557 Projet SustainCity, 39*.

O'Donoghue, C. (Ed.). (2014). *Handbook of microsimulation modelling*. Emerald Group Publishing.

Orcutt, G. H. (1957). A new type of socio-economic system. *The review of economics and statistics*, 116–123.

Tanton, R. (2014). A review of spatial microsimulation methods. *International Journal of Microsimulation*, *7*(1), 4–25.

Van Imhoff, E., & Post, W. (1998). Microsimulation methods for population projection. *Population: An English Selection*, 97–138.

Willekens, F. (2014). Multistate analysis of life histories with R (p. 308). New York: *Springer*.

Contact

David Morávek

Prague University of Economics and Business

W. Churchill Sq. 1938/4, 130 67 Prague 3, Czech Republic

david.moravek@vse.cz