# Using microsimulation to improve the quality of the official Austrian population projection

**Hannah Diethard[1], Pauline Pohl[1], Philip Slepecki[1], Martin Spielauer[2]**

[1]*Statistics Austria*

[2]*Austrian Institute of Economic Research (WIFO), Vienna, Austria*

**Keywords:** Dynamic microsimulation, Population projection, Cohort-component method

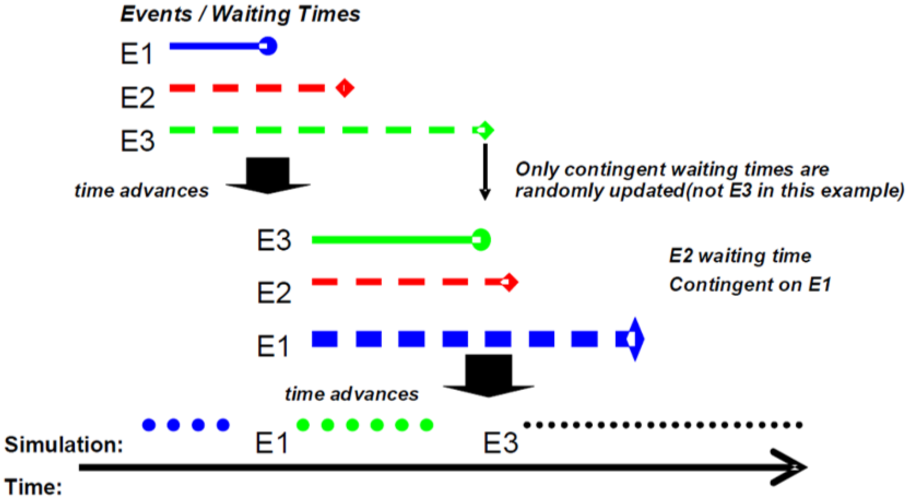
## 1. Introduction

Population projections in official statistics are generally produced using the cohort component method. It is computationally simple, does not require a broad range of input data, and is well-established among demographers. However, it cannot account for complex and dynamic demographic processes, model interactions, or produce detailed results for individual-level characteristics. To overcome these limitations, Statistics Austria has developed the dynamic microsimulation model STATSIM, which builds on the characteristics of individuals instead of entire cohorts and allows for the simulation of realistic life-courses. In the following, we present the model and compare its results with those of the cohort-component method and the observed data for Austria for the years 2013 to 2021. In the future, we plan to extend STATSIM to include additional characteristics for education and health.

## 2. Methods

The cohort component method uses demographic event rates to project fertility, mortality and migration patterns. The main drawback of this modelling approach is that it restricts the number of attributes and it cannot capture a broad range of demographic processes, particularly when interactions between variables or individuals are important (Burch, 2018) (Van Imhoff & Post, 1998). To overcome these disadvantages, Statistics Austria has implemented the microsimulation model STATSIM, in which event rates are replaced by waiting times, computed by inverse transform sampling. Similar to the event rates, the event waiting times depend on age, place of residence, and other person-specific characteristics. Following a competing risk approach, the event with the shortest waiting time for a person at a certain point in time is realized. After an event has been realized, the person's characteristics are updated and new waiting times are assigned. Figure 1 illustrates the waiting times for three events, E1, E2, and E3. Among these events, E1 (represented by the solid blue line) has the shortest waiting time at the beginning of the given time window. Consequently, event E1 is realized at the end of the assigned waiting time. Following the realization of this event, the individual's characteristics and waiting times for the events are updated. However, as shown in the figure, there is no change in the waiting time for event E3, as this event occurs at a specific time and is not influenced by other individual characteristics, such as their place of residence. An example for this type of event is the person’s birthday. Finally, the simulation ends for a person when they die, move abroad or reach the end of the projection horizon, after which the next person is simulated.

Figure 1: Evolution of a simulated life course



While the use of microsimulation models for the purpose of computing population projections is rare in official statistics, the method itself has been around for decades ( (Orcutt, 1957), (Orcutt, et al., 1961)) and it is used by various national statistical institutes (e.g. Demosim (Statistics Canada, 2022), MOSART (Andreassen, et al., 2020), DESTINIE (Blanchet, et al., 2011), MikroSim (Münnich, et al., 2021)).

STATSIM is a dynamic microsimulation that projects individual life paths over time, with a primary focus on projecting population dynamics through simulating demographic events (births, deaths, migration). The model operates in continuous time, allowing events to occur at any moment and then be aggregated to generate results for each projection year. We implement STATSIM using Modgen[1], a programming language developed by Statistics Canada.

Transitioning from the cohort-component method to microsimulation is a significant shift in methodology, requiring a deeper understanding of model construction and advanced statistical programming and data analysis skills. As an initial step towards enhancing the model, we implemented a more detailed and realistic emigration module that avoids the need for a wide array of additional model variables or data sources. Instead of relying on standard emigration rates, we estimate piecewise constant hazards for emigration across 17 different country groups.

## 2.1. Clustering of Countries and Estimation of Emigration Hazards

Emigration behaviour is influenced, among other things, by the migrant's country of birth, which, in turn, affects the reasons for their initial migration and associated costs (Gundel & Peters, 2008) (Jensen & Pedersen, 2007). In order to reflect this behaviour in the model, immigrants' origin countries are grouped into clusters based on similarities in duration of stay and sociodemographic traits. For this purpose, data on immigrants from countries with at least 300 persons immigrating to Austria from 2017-2019 was analysed so that the following factors could be considered in the cluster analysis: Education, employment, age, sex, length of stay, asylum applications and family characteristics. Hierarchical clustering determined the number of clusters, followed by k-means clustering to define 17 clusters across three regional groups (EU/EFTA countries and the United Kingdom, European countries outside the EU and the rest of the world). These clusters reflect patterns such as work-related migration (Eastern EU), education-related migration (Western EU), and asylum-seeking migrants (mostly Middle East,

---

[1] https://www.statcan.gc.ca/en/microsimulation/modgen/modgen

Africa). Some countries with few immigrants are manually assigned to clusters based on geographic proximity. For each cluster and sex, emigration hazards are estimated using piecewise constant hazard models.

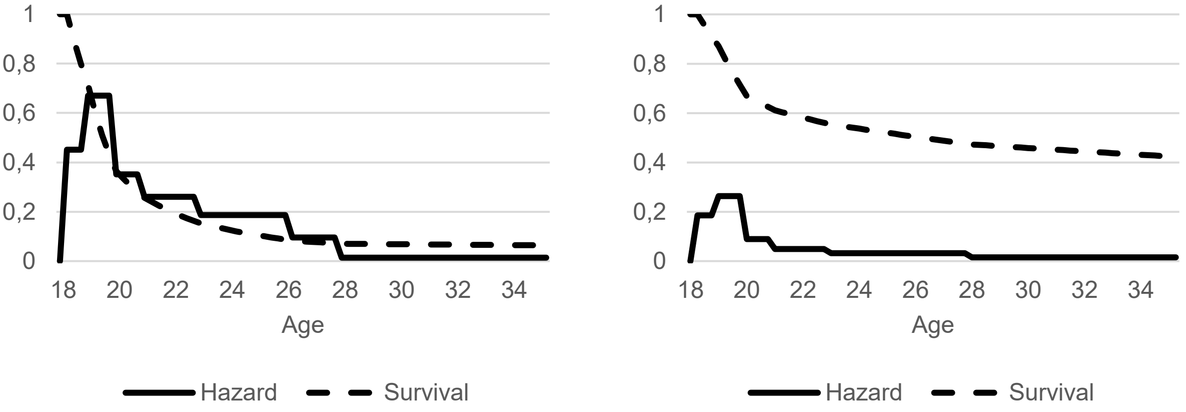
### 2.2. Data

As data sources, we use Austrian population data from Statistics Austria, which includes the population status as of 1 January of the base year, categorized by age, sex, federal province of residence, country of birth and length of stay in Austria. In addition, age- and sex-specific indicators for births, deaths and migration are derived from the Vital Statistics and Migration Statistics, supplemented by data from the Population statistics.

## 3. Results

Figure 2 plots the emigration hazards and the corresponding survival rates for a man who immigrates at the age of 18 and lives in Vienna, for two different country groups.
The left pane shows the results for a man born in a high-income EU member state in Northern or Western Europe, the right pane for a man born outside of Europe, in a country whose emigrants have a long duration of stay and a high number of asylum applications in Austria.The figure shows substantial differences in emigration behaviour, with immigrants from the Northern and Western EU member states experiencing much higher emigration hazards in the first 10 years following immigration. Using standard emigration rates, it would not be possible to capture these multifaceted patterns of emigration behaviour.

Figure 2: Emigration hazards and survival rates for a man who immigrates at age 18 and was born in a high-income EU member state in Northern/Western Europe (left pane) vs. a non-European country whose emigrants have a long duration of stay and a high number of asylum applications.



## 4. Performance of STASIM vs. the cohort-component method

In this section, we evaluate the results of a retrospective projection exercise, performed until 2021 with 2012 as its base year, using the parameters employed in Statistics Austria's 2013 population projection and observed immigration figures. Projections were computed using STATSIM as well as Statistics Austria's 2013 population projection model, which followed the CCM framework (Figures Figure **3** and Figure **4**). While the CCM projection does not capture the increased emigration flows after the years of high immigration (2015/2016), the microsimulation captures this pattern quite well. In STATSIM, the estimated emigration risks are used to determine that the immigrants with the shortest duration of stay are most likely to emigrate. Accordingly, the probability that those who immigrated to Austria in 2015/2016 are also the most likely to emigrate in subsequent years is higher.

Figure 3: Projected and observed annual emigration from Austria 2013-2021, based on the cohort-component method vs. STATSIM

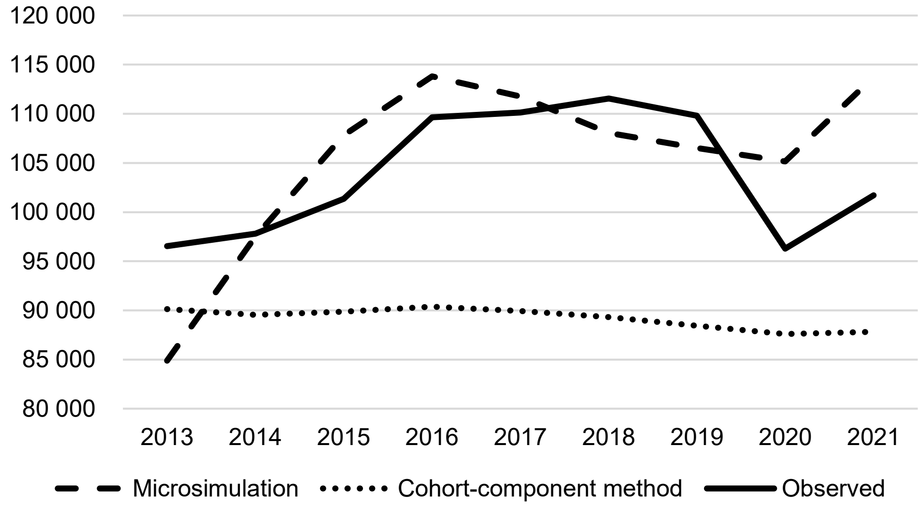
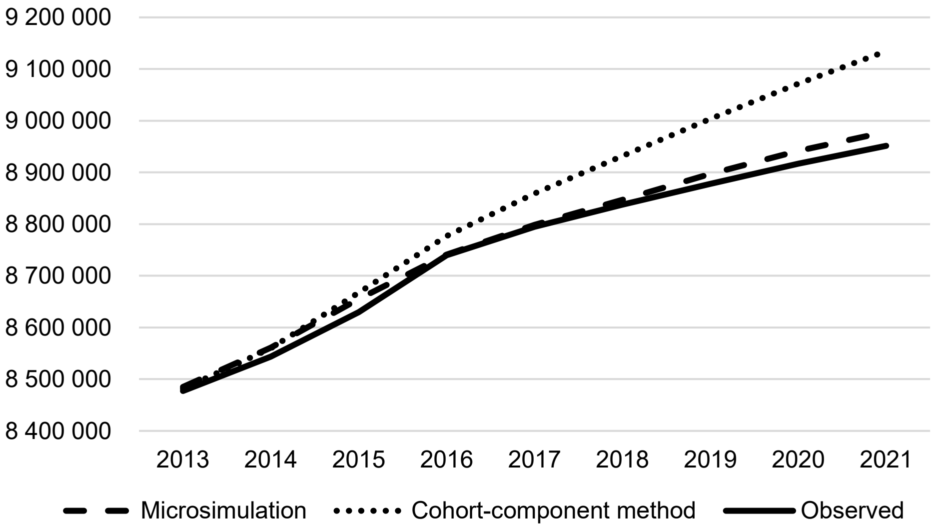


Figure 4: Projected and observed population of Austria 2013-2021, based on the cohort-component method vs. STATSIM



## 5. Extensions of the model

We are continuously improving the functionalities of STATSIM and extending its applications to different fields beyond population projections. Two additional modules are currently being developed: one for projecting cancer incidence, prevalence and mortality and another for school attendance projections.

For the projection of cancer incidence, prevalence and mortality, an additional simulation module is being developed in which cancer diagnoses and cancer deaths are simulated for 15 different groups of cancer localizations. For this purpose, as in the model of the population projection, event waiting times are calculated from the localization-specific incidence and mortality rates using the inverse transform sampling method. Cancer incidence, prevalence and mortality are then calculated from the sum of the diagnoses, the number of people living with a cancer diagnosis and the number of deaths within a projection year.

Moreover, an additional module is being developed to project the number of pupils, in which the educational pathways should be explicitly modelled. This means that, in addition to the demographic events (fertility, mortality, migration), for people between the ages of 5 and 30, the individual trajectories in the education system (school enrolment, changes between school levels, choice of school type, choice of school province, entries and exits in the education system) are simulated.

In the future, STATSIM can also be expanded so that demographic processes depend on individual-level education and employment characteristics; e.g. modelling women's fertility dependent on their education level and employment status.

## 6. Conclusion

The foreign-born population in Austria is a heterogeneous group and migrants differ in their demographic behaviour. In the case of emigration, country of birth and duration of residence are important determinants, alongside age and sex. Compared to previous population projections produced by Statistics Austria based on the cohort-component method, our microsimulation model can account for more complex and dynamic patterns in demographic processes and produce results for various individual-level characteristics. Using data for Austria and 2012 as the base year, we show that a microsimulation model that accounts for heterogeneity in migrants' emigration behaviour produces more accurate projections for the number of emigrants. The model is currently being supplemented by modules for cancer and school attendance projections and can therefore also provide improved results in these fields. In the coming years, we are also planning further extensions, such as one on employment status.

## References

Andreassen, L. et al., 2020. The dynamic cross-sectional microsimulation model MOSART. *International Journal of Microsimulatio,* 13(1), pp. 92-113.

Blanchet, D., Buffeteau, S. & Cren, E., 2011. Le modèle de microsimulation Destinie 2 : principales caractéristiques. *Economie et Statistique,* Volume 441-442, pp. 101-121.

Burch, T. K., 2018. The Cohort-Component Population Projection: A Strange Attractor for Demographers. In: *Model-Based Demography.* s.l.:Springer.

Gundel, S. & Peters, H., 2008. What determines the duration of stay of immigrants in Germany? Evidence from a longitudinal duration analysis. *International Journal of Social Economics,* 35(11), pp. 769-782.

Jensen, P. & Pedersen, P. J., 2007. To Stay or Not to Stay? Out-Migration of Immigrants from Denmark. *International Migration,* 45(87-113).

Münnich, R. et al., 2021. A Population Based Regional Dynamic Microsimulation of Germany: The MikroSim Model. *Methods, data, analyses : a journal for quantitative methods and survey methodology,* 15(2), pp. 241-264.

Orcutt, G. H., 1957. A New Type of Socio-Economic System. *The Review of Economics and Statistics,* 39(2), pp. 116-123.

Orcutt, G. H., Greenberger, M., Korbel, J. & Rivlin, A. M., 1961. *Microanalysis of Socio-economic Systems: A Simulation Study.* New York: Harper & Row.

Statistics Canada, 2022. *Projections of the Indigenous populations and households in Canada, 2016 to 2041: Overview of data sources, methods, assumptions and scenarios.* [Online] Available at: https://www150.statcan.gc.ca/n1/pub/17-20-0001/172000012021001-eng.htm

Van Imhoff, E. & Post, W., 1998. Microsimulation Methods for Population Projection. *Population: An English Selection,* Volume 10(1), pp. 97-138.