Different Strategies for Modelling and Simulation of the Impact of Migration on Regional Population Development

Matthias Obermair[1*], Martin Bicher[1,2], Felix Breitenecker[1]

[1]Institute of Analysis and Scientific Computing, TU Wien, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria; *matthiasobermair@gmail.com

[2]dwh simulation services, dwh GmbH, Neustiftgasse 57-59, 1070 Vienna, Austria

Abstract. Demographic developments are not only a point of interest on national but also on sub-national level where migration rates are typically much higher compared to international movements. This article presents several models for the simulation of the impact national and international migration has on regional population development. Hereby we contribute to the research field of demographic modelling as we demonstrate pros and cons of the different modelling strategies when trying to parametrize the models with real-world data. Since the influence of an individual's age and sex on its mobility is a common feature discussed in literature, the models developed mainly focus on these influences. While the first models use decoupled migration processes, one additional approach is designed to take into account an individual's wish of moving to a particular place and not being distributed to a random region. All models are further enhanced with a region-specific external migration tool.

Introduction

Demography plays a very important role in many aspects of governmental and economic planning. Many essential features of our society, such as the health-care system, need reliable data on population development, in this special case e.g. to guarantee an unfailing provision of medical care.

Tom Wilson[1] formulated three answers to the question, why in particular internal migration is worth a detailed examination. Firstly, for understanding the dynamics of a country's population geography, it is necessary to understand internal migration patterns. Secondly, to compute population estimates in years where no census takes place, reliable (internal) migration data are essential. Wilson thus argued that "internal migration is often the most important demographic variable shaping regional population age structures." Thirdly, multi-regional demographic prospects by single age-classes of course need precisely modelled migration profiles by single age-classes.

To embrace the influence of (internal) migration on regional population developments, this article further investigates its impact on a country's demography and how to model it in the most fitting way with respect to the data available for the regions in question. Therefore different model approaches will be presented and tested and their results compared. The models were implemented using the object-oriented programming language Python 3, altering a given basic model which will be presented in the next section. Among other advantages like free accessibility and an enormous pool of freely available Python packages, this language is capable of dealing with the high number of required agents.

1 The Generic Population Concept (GEPOC)

In 2015 a generic population model, capable of producing a valid virtual image of Austria's population and feasible prognoses, only using public accessible initial population data, was completed as part of the Decision Support for Health Policy and Planning (DEXHELPP) health-care research project.

Amongst other things, the GEPOC model was implemented as an agent-based model which was further developed as part of this work. It is defined by its initial setup and time-dynamics.

- Initial setup: Once a simulation start date is fixed, an agent-based model with $N + 1$ agents is initialized. N of them represent the inhabitants of Austria. Each of these agents can be imagined as representative of a real person and receives an individual birth-date and sex (male or female) as well as a unique ID. The empirical data used for the initial population is provided by the Austrian Bureau of Statistics[2]. The additional agent shall be referred to as government-agent, as it will play the role of the government.

- Time-dynamics: The model is updated in a-priori defined time-steps which are not necessarily equidistant. For every time-step all agents representing individuals are iterated in random order. For each one the model decides if they die, emigrate or (in case of a female agent) give birth to a new agent, using an event-based strategy. In case of death or emigration the agent is removed from the model and possibly later scheduled events are skipped. The birth event leads to the construction of a new agent with a birth-date according to the schedule. After the iteration is completed, the government-agent generates a given number of immigrants and adds them to the model as new agents.

After the systematic generation of results, they have been validated using prognosis data of the Austrian Bureau of Statistics[2]

2 Mathematical Models for Internal Migration

Henceforth (if not stated otherwise) migrant, immigrant and emigrant always refers to internal migration.

Various reasons and factors have to be considered when doing research on internal migration. For example Aude Bernard, Martin Bell and Elin Charles-Edwards concentrate on the life-course transitions that affect migration choices – especially of young adults – directly, calling them "proximate determinants". They state that economic, social and other rather general factors, shape people's plans for their lives and therefore

lead to migration age profiles. Age profile differences between countries occur through different timing of entry into education, labour market entry, partnership and childbearing[3]. The approach presented by them will be further explained in Section 2.1. Wilson and Bell state that multi-regional models with fixed migration rates tend towards dampening the net migration gains in fast growing regions and therefore causing convergent regional growth rates. The reason for this is that in such a fast growing region the potential number of emigrants is increasing at the same speed as the population, whereas the pool of possible immigrants cannot keep that pace. In addition, model approaches, where migration flows are influenced by both origin and destination, deliver better representations of migration behaviour compared to those depending merely on the original population size in the base year[4].

2.1 Migration age profile model (MAP)

Various scientists who have examined internal migration age profiles in different countries, have often found a high degree of regularity over space and time for migration age profiles, although huge variations in levels of migration occur[1]. Thus, they suggest the usage of generalised, age-dependent migration probabilities. The big advantage of such a MAP model is the small amount of additional input data for internal migration. It only consists of one matrix per gender, where migration flows between all regions (including the region of origin) for the required time are listed. A typical standardised migration age profile is presented by Aude Bernard, Martin Bell and Elin Charles-Edwards [3]. Migration propensities per age undergo a high variation, with young adults forming the most mobile group. Mobility peaks between twenty and thirty and declines steadily afterwards, with an eventual rise around retirement age and again shortly before death. Many authors consider this strong empirical regularity as almost universal, which is shown by the wide use of Tom Wilson's *"age schedule of migration"*[5].

A disadvantage of generalised migration age profiles is, according to David Plane, that fixed origin-destination migration rates cannot be justified on behavioural grounds, since true dynamics of regional demographic change are not reflected by them [6]. Figure 1 shows the existing regional differences in age-dependent migration intensities for Austria. Since this work concentrates especially on internal migration and the differences between the regions concerning a mi-

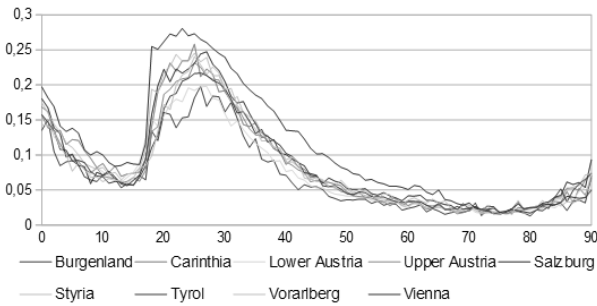gration age profile, this approach has been rejected and replaced by improved models with higher accuracy.

Figure 1: Differences in age-dependent migration probabilities between the federal states of Austria in 2017 according to Statistik Austria for ages 0–90[2]. Although a common tendency is visible, regional differences cannot be overseen. Especially Vienna shows higher internal migration rates due to its high population density.

2.2 Biregional model (BR)

With the MAP model handling only migration probabilities and not the whole migration process between regions, the first well-defined modelling approach for internal migration is the construction of a so-called biregional model, based on the one presented by Wilson and Bell[4]. The additional data needed for the internal migration flows consists of initial population as well as emigration and immigration data for every region partitioned into sex (s) and age (a). In fact, this approach is the combination of several biregional models instead of one whole multi-regional model. It is assumed that the country in question is divided into one region of interest and a combination of all other regions. Then emigration and immigration rates are calculated for this particular region. This process is repeated for every region in the simulation.

$$\text{emr}_{i,s,a} = \frac{\text{EM}_{i,s,a}(T)}{P_{i,s,a}(T)} \qquad (1)$$

$$\text{imr}_{j,s,a} = \frac{\text{IM}_{j,s,a}(T)}{\sum_i P_{i,s,a}(T)} \qquad (2)$$

The emigration rate corresponds to the total emigration per sex and age of the region divided by the total population of the region per sex and age at start time T of the simulation. The immigration rate (2) is calculated as the total number of immigrations per sex and age to a particular region from all other regions combined, divided by the total population per sex and age of the country at start time T.

For the implementation of the biregional model – since the GEPOC model is agent-based – it has to be evaluated for every agent if they emigrate. With that knowledge the calculation of the total amount of emigrants, which naturally has to be the same as the total amount of immigrants for internal migration, is possible. To ensure that the overall net internal migration sums to zero, it is necessary to rescale the calculated immigration rate. Now all emigrating agents can be distributed to their destination. This process is repeated for every time-step, leading to a very costly procedure. Therefore in the second approach the internal simulation was implemented via a migration pool model.

2.3 Migration pool model (MP)

Another model based on the paper of Wilson and Bell[4] is the so-called migration pool model. In theory it is executed in two steps: first the number of all emigrants from all regions is evaluated and the migrants concerned are put into a common "pool." Second they are divided into different regional destinations. While the emigration rate is calculated analogous to Equation (1) in the BR model, the difference to the BR model is that this division is not based on the population but on the total number of migrants.

To obtain the total number of immigrations to each region, immigration rates have to be calculated in the form of immigrants of one region separated by sex and age divided by total immigrants of the corresponding group

$$\text{imr}_{j,s,a} = \frac{\text{IM}_{j,s,a}(T)}{\sum_j \text{IM}_{j,s,a}(T)}. \qquad (3)$$

Equation (4) depicting the migration flow between two regions, shows the major advantage the MP model has compared to the BR model in an agent-based approach.

$$\text{M}_{i,j,s,a}(t) = \text{emr}_{i,s,a} \cdot P_{i,s,a}(t) \cdot \text{imr}_{j,s,a} \qquad (4)$$

Since the migration flow is calculated as the product of two fixed rates and the group-specific population of one region, rescaling is not necessary. This results in an agent's probability m to migrate from region *i* to region *j*

$$\text{m}_{i,j,s,a} = \text{emr}_{i,s,a} \cdot \text{imr}_{j,s,a}. \qquad (5)$$

Under these assumptions it is not necessary to actually create a pool of emigrants in an intermediate step. Therefore during the simulation the whole migration process of one individual can be handled in one quick step. Still, in the MP model immigration into regions is only dependent on the pool size and not the composition of the pool by region and origin [7].

2.4 Inter-regional migration model (IRM)

All models presented above have in common that emigration and immigration happen independently, thus there is no connection between origin and goal region. The fact that an individual emigrates from region *A* or *B* has no influence on their decision where to immigrate to afterwards. Although this has no effect on the number of migrants and the total population, from an agent-based point of view it should be considered. Since individuality is the unique feature of ABM, the agents in an internal migration model should decide where to immigrate based on their initial region. The inter-regional migration model reaches this goal by changing the internal migration process. While the decision to emigrate is still dependent on the emigration rate analogous to the BR and MP model, for immigration decisions age is not incorporated. Instead data which represents movement from every initial region to all goal regions is included. Thereby, while the agent's age is still important since it influences its emigration decision, the agent's destination is dependent on its origin.

Formalising this leads to an inter-regional migration rate for each sex from region *i* to region *j* which can be seen in Equation (6):

$$\text{imr}_{i,j,s} = \frac{\text{IM}_{i,j,s}(T)}{\sum_j \text{IM}_{i,j,s}(T)} \qquad (6)$$

Note that now the distribution and number of immigrants to one region depend on their origin instead of their age. The new migration rate between two regions thus is:

$$\text{m}_{i,j,s,a} = \text{emr}_{i,s,a} \cdot \text{imr}_{i,j,s} \qquad (7)$$

2.5 Enhancement with region-specific external migration (rem)

Until now all models have concentrated on internal migration only, thus external migration is handled for the conglomerate of all regions within the model. Emigration to a foreign country happens on the basis of a common probability for all agents, whereas immigrants from abroad are split proportionally according to the population of the regions.

With the additional region-specific external migration approach, every region has specifically adapted immigration and emigration rates for abroad. To implement this, foreign migration data for every region used has to be added, implying additional parametrisation data. Still, it has the big advantage of improving the simulation results significantly, which will be shown later on.

3 Results

The MP model has the highest accuracy of the standard models. The BR and MP models produce very similar results, with the MP approach being slightly more accurate, whereas the IRM model differs from them. The simulation results for the latter show the by far biggest deviation compared to the data. While for the first two approaches the relative difference overall constitutes a maximum deviation of 4.41% for the BR and 3.64% for MP model respectively (both Vienna, 2016), the difference for the IRM model reaches a deviation of up to 7.23% (Vienna, 2017). For all three approaches the growth of Vienna drags behind in favour of the other, less densely populated areas close to Vienna. With the MP model having a quicker computation time, it is satisfying to observe that it produces the best results of the basic implementations.
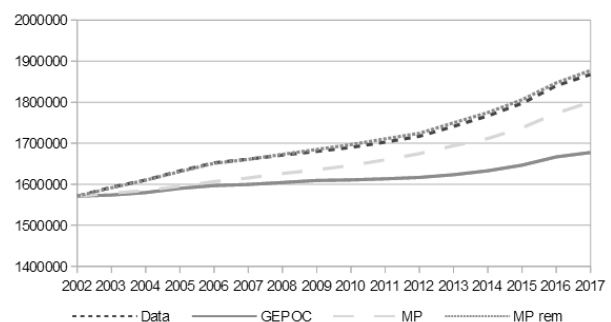

Figure 2: Comparison of the population of Vienna according to data, original GEPOC (without internal migration), the MP model and the MPrem.

Enhancement with region-specific external migration pays off. For all three upgrades a considerable improvement of simulation results can be ob-

served. Figure 2 exemplary shows this improvement for the MP model. While for Vienna the difference between data and outcome differ up to 81.952 people in 2017 for the MP approach, the upgraded MPrem reaches a maximum discrepancy of (minus) 9.802 for Vienna. Still, compared to the actual population of Vienna in 2017 of 1.867.582 inhabitants, the results of both approaches are rather accurate. The relative difference of the MP model upgraded with region-specific external migration now diminishes to a maximum of 1.99% (Upper Austria, 2017). The enhancement enables the successful transition of the (too many) people living in regions like Lower Austria, according to the standard MP model, to Vienna. The distribution of external immigrants changes from the allocations solely depended on population density in the standard models, to a distribution according to the real data. The magnitude of this change is the same for all approaches and for external emigration.

The IRMrem model has the highest accuracy of all models implemented. Of all modelling approaches compared, the IRMrem model reaches the most accurate results for the years 2002–2017. The highest deviation reaches an absolute number of 22.437 persons, compared to 29.173 in the second-best performing MPrem approach (both Upper Austria, 2017). Since the basic IRM model has the worst accuracy, the highly increased performance of the IRMrem model is very interesting. Apparently the internal immigration decision which is directly dependent on the origin of the agent adds up well with the region-specific external migration. This observation will be further investigated in the next paragraph. Considering the very good runtime performance of the IRMrem approach, this model is the best option to choose if the necessary data for the region-specific external migration is available.

Overall simulated migration rates are too low. A closer look at the internal migration data shows that – with very few exceptions – the number of migrants calculated tends to be lower than the actual ones. The rates are too low, especially for Vienna. This finding is in line with the results concerning the population figures. Again the BR and MP models produce similar results with the IRM model dragging behind. Altogether, the implementation of region-specific external migration has a positive impact on internal emigration rates, too. The similar behaviour patterns of the approaches

derive from the use of the same data and the identical data processing for the internal emigration process. The difference between the IRM model and the other two develops over time and is a consequence of the significantly lower internal immigrations happening right from the beginning.

The simulation forecasts produce very similar results. The population forecasts until 2030 produced by the different modelling approaches are nearly akin. Comparing them, the MP model predicts more people living in the city of Vienna, with accordingly less in the countryside. The IRM model mirrors the contrary development and the BR model is somewhere in between those two. These trends are confirmed when the internal migration rates of all simulation results are compared. Thus, the observations of the timespan 2002–2017 persist for the forecast, although, through the comparably simple extrapolation of the migration data by linear regression, the informative value of the predicted population development is questionable.

All models presented require demographic data available from most national (European) institutes for statistics or Eurostat, respectively[8]. The type and amount of parametrisation data required is listed in Table 1.

4 Conclusion

The main part of this article concentrated on the expansion of the existing demographic GEPOC model with internal migration to simulate regional population changes. All implemented models are constructed to handle an arbitrary amount of regions, as long as data for the migration flows between them is provided. The analysis and validation was undertaken for the federal states of Austria. In conclusion the inter-regional migration model with region-specific external migration turned out to be the best performing one, regarding both computation time and accuracy of the results.

Internal migration has a high impact on simulating regional population numbers accurately. As it has been shown in this work, regional demographic development cannot be simulated accordingly without region-specific external migration. Still the necessity of additional data for this enhancement has to be considered. The IRMrem approach needs fewest parametrisation data of all models with region-specific external

Model	Parameter	Datapoints
GEPOC	Emigration	3.584
	Immigration	3.584
	Birth	1.792
	Death	3.584
	TOTAL	12.544
BR & MP	GEPOC	12.544
	Internal Emigration	32.256
	Internal Immigration	32.256
	TOTAL	77.056
IRM	GEPOC	12.544
	Internal Emigration	32.256
	Internal Immigration	2.592
	TOTAL	47.392
BRrem & MPrem	Emigration	32.256
	Immigration	32.256
	Birth	1.792
	Death	3.584
	Internal Emigration	32.256
	Internal Immigration	32.256
	TOTAL	134.400
IRMrem	Emigration	32.256
	Immigration	32.256
	Birth	1.792
	Death	3.584
	Internal Emigration	32.256
	Internal Immigration	2.592
	TOTAL	104.736

Table 1: Parametrisation data necessary for the implemented models using a timespan of 16 years, 112 different age groups and nine regions.

migration. The generation of accurate population development forecasts is very difficult due to the unavailability of data concerning future migration trends.

The developed agent-based models can be further enhanced and used for many different applications. A possible usage lies in the analysis of probable developments of local labour markets. Various other possibilities lie in planning regional infrastructure like schools, retirement homes, hospitals and public transport.

References

[1] Wilson T. The impact of education-bound mobility on inter-regional migration age profiles in Australia. *Applied Spatial Analysis and Policy*. 2015; 8(4):371–391.

[2] Statistik Austria. STATcube. `http://statcube.at`. Accessed: 2018-07-18.

[3] Bernard A, Bell M, Charles-Edwards E. Life-Course Transitions and the Age Profile of Internal Migration. *Population and Development Review*. 2014; 40(2):213–239.

[4] Wilson T, Bell M. Comparative empirical evaluations of internal migration models in subnational population projections. *Journal of Population Research*. 2004; 21(2):127.

[5] de Jong PA, Brouwer AE, McCann P. Moving up and down the urban hierarchy: age-articulated interregional migration flows in the Netherlands. *The Annals of Regional Science*. 2016;57(1):145–164.

[6] Plane DA. Requiem for the Fixed-Transition-Probability Migrant. *Geographical Analysis*. 1993;25(3):211–223.

[7] Van der Gaag N, Van Imhoff E, Van Wissen L. Internal migration scenarios and regional population projections for the European Union. *International Journal of Population Geography*. 2000;6(1):1–19.

[8] Commission E. Eurostat. `https://ec.europa.eu/eurostat/data/database`. Accessed: 2018-09-18.

[9] Bicher M, Glock B, Miksch F, Schneckenreither G, Popper N. Definition, Validation and Comparison of Two Population Models for Austria. In: *Book of Abstracts, 4th International Conference on Business, Technology and Innovation*. 2015; .

[10] Ravenstein EG. The laws of migration. *Journal of the statistical society of London*. 1885;48(2):167–235.