

# **A Population Based Regional Dynamic Microsimulation of Germany: The MikroSim Model**

Ralf Münnich, Rainer Schnell, Hanna Brenzel, Hanna Dieckmann,  
Sebastian Dräger, Jana Emmenegger, Philip Höcker, Johannes Kopp,  
Hariolf Merkle, Kristina Neufang, Monika Obersneider, Julian Reinhold,  
Jannik Schaller, Simon Schmaus, and Petra Stein

18.6.2020

Microsimulation models are widely used to evaluate the potential effects of different policies on social indicators. Most microsimulation models in use operate on a national level, disregarding regional variations. We describe the construction of a national microsimulation model for Germany, accounting for local variations in each of the more than 10,000 communities in Germany. The database used and the mechanisms implementing the population dynamics are described. Finally, the further development of the database and microsimulation programs are outlined, which will contribute towards a research lab that will be made available to the wider scientific community.

# 1 Introduction

What are the effects of changes in the demographic profile of the population on family formation processes? How does tax legislation impact on tax revenues? Do changes in women's employment alter the way in which the elderly are cared for? How do different qualifications and work biographies influence pensions? How will integration patterns develop in the future in the context of demographic transition regarding upcoming generations of migrants? These are some examples of problems studied using microsimulation techniques in past decades (Stein and Bekalarczyk, 2016, O'Donoghue and Dekkers, 2018, Schnell and Handke, 2020, or Zwick and Emmenegger, 2020).

However, due to limited data availability and high computational complexity, such simulations were mostly done at national levels of analysis. Given the fact that social and economic change is often different for some regions and subgroups of the population, the demand for detailed modeling is increasing. Therefore, highly detailed regional datasets covering the whole population are needed. A prime example is schooling: The demand for elementary schools varies at the local level. For secondary schools, increasing population heterogeneity requires complex school planning which accounts for the diversity of schooling demands. A second example is care for the elderly, where information regarding the distance between parents and their children, who are potential care providers, is essential for modeling the demand for care.

In the Netherlands or the Scandinavian countries, population covering register datasets could provide data for regional microsimulations. In Germany, however, hardly any datasets are available that foster regional microsimulation modeling. Either the sample size is too small for regional models or data protection regulations do not allow the use of low-level identifiers in population covering datasets. Therefore, modeling tasks requiring low-level information are challenging.

The MIKROSIM project described in this paper is aiming for a dynamic microsimulation

of Germany down to each municipality. The model is based on a highly detailed dataset build from many different sources, such as survey, administrative, and other data. Therefore, a series of different methods of data-integration and small area estimation have been used for building the dataset. Modeling the dynamic processes is achieved by estimating transition probabilities using different statistical methods. The MIKROSIM model consists of different modules, simulating births, deaths, marriage, education, and other dynamic processes.

This article provides an introduction to the design and implementation of the simulation processes of the model. At the end, we outline further steps in the development of the simulation program.

## 2 Base Data

A synthetic dataset is used as the base dataset for the microsimulation model MIKROSIM. In general, the purpose of synthetic datasets is to mimic a non-accessible or non-existing dataset so that the relevant characteristics of a synthetic dataset matches the characteristics of the underlying population as close as possible (cf. Münnich and Schürle, 2003; Münnich, Gabler, et al., 2012; Kolb, 2013; Alfons, Kraft, et al., 2011). The characteristics to be matched are distributional parameters, correlations, cluster effects, and totals.

The generated dataset is based on an anonymized national register of residents that has been used for methodological research for the census 2011. All 11,339 German municipalities are modeled.<sup>1</sup>

Since the register of residents contains only a few variables, additional data is generated. Since most data stem from German official statistics, data evaluation methods are performed within the statistical office to ensure confidentiality and privacy or, al-

---

<sup>1</sup>A previous project (REMIKIS) modeled the region Trier using a similar modeling strategy (Burgard et al., 2019).

ternatively, using scientific use files. No recode linkage between official microdata and the synthetic dataset or microdata from other sources takes place. Model coefficients and marginal information on distributions are provided to extend the dataset. For the data synthesis, cluster effects resulting from a positive correlation between household members are considered. The cluster effect is modeled by generating a latent household type. Depending on the household type, further characteristics of the household members are generated with multinomial logit models (Alfons, Filzmoser, et al., 2011; Kolb, 2013).

Finally, the synthetic population for each municipality is adjusted to published census totals. Using simulated annealing (cf. Laarhoven and Aarts, 1987; Huang and Williamson, 2001; Williamson, 2012; Tanton, 2014), households are selected randomly and entered or deleted sequentially to minimize the differences between the synthetic population and the known totals.

### **3 Construction Principles and Module Ordering**

The MIKROSIM model is designed as a closed population simulation model. Therefore, modules simulating the paths for individuals entering or leaving is key to obtaining realistic projections of the population. The modules providing these paths are mainly the modules MORTALITY, BIRTHS, and REGIONAL MOBILITY, described later. The simulation model is implemented in R.

Most of the modules in the MikroSim model are based on statistical models to estimate individual transition probabilities for the micro-units. For the simulation of state changes, the sequence in which the modules are processed is important.

The characteristics of the simulated population are updated once for each simulated year. Therefore, no information about the exact time of occurrence of an event within the simulated period is available. However, the occurrence of one event might determine other events (for example, a death triggers further changes). There are different strategies

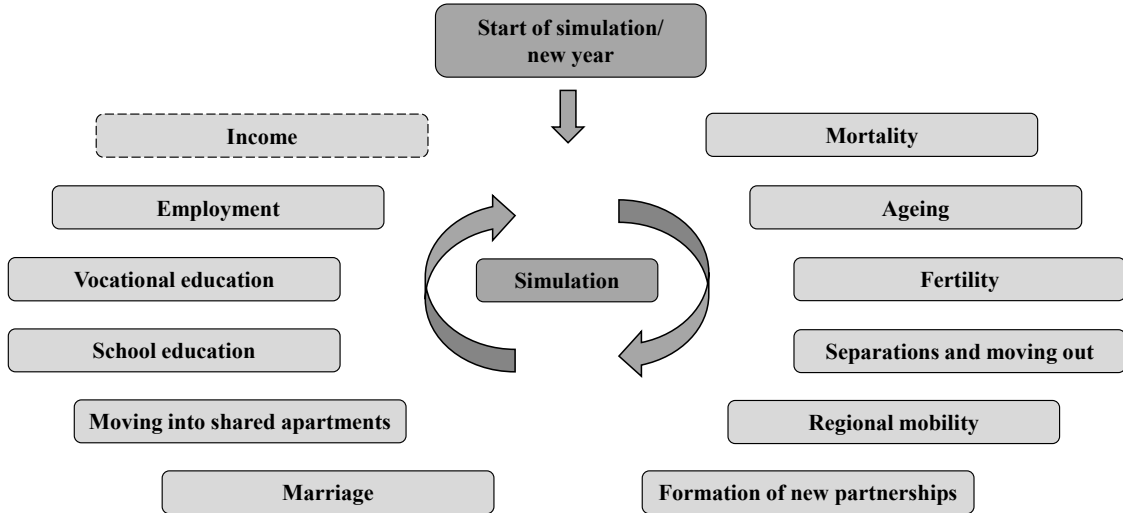


Figure 1: MikroSim Module Sequence

available to deal with such dependencies (van Imhoff and Post, 1998). In MikroSim, probabilities for many events are only estimated for those persons who are eligible for a change of state. The eligibility is modeled by the ordering of the modules. For example, as the event of birth is simulated prior to the event of a marriage, the probability of marriage can be conditioned on the birth event (Burgard et al., 2020). The sequence in which the modules are processed is shown in Figure 1.

The first module in the MikroSim model simulates widowhood for married persons not living in the same household with their spouse. Widowhood directly influences the possibility of the respective persons to enter the trailing modules such as separations or the formation of new partnerships. Updating this relationship status at the beginning can prevent an underestimation of widowhood within separated couples.

The MORTALITY module is placed prior to the AGING module since also newborns face a non-zero mortality risk. The event of death depends exclusively on age and gender.

Following the AGING module, births are simulated as the first way to add new individuals to the population. The position of the BIRTH module at this early stage of the simulation is required since birth decisions must precede the birth event. Thus, birth

probabilities are estimated mainly based on the characteristics in the previous period.

The LEAVING HOUSEHOLD module then simulates (1) relocations of adults from parental households, (2) shared apartments, and (3) dissolutions of households after separations. Since these relocation events often have a direct impact on regional mobility, migration across district borders is simulated subsequently.

New households are formed by creating new partnerships and shared apartments. Thus, persons can directly form new households within a simulation period after leaving a household or immigrating. Formal changes in the relationship status (divorces and marriages) are simulated separately.

Changes in school education, vocational education, and employment status are the final modules in the simulation process. An income module will be integrated soon.

The modules are based on a variety of different models, modeling methods, and datasets. Table 1 gives a brief overview, including the choice of independent variables.<sup>2</sup> The following sections describe the modules in more detail.

---

<sup>2</sup>A detailed explanation of the mechanisms sketched in Table 1 will be the subject of a different publication.

Table 1: Overview of the MikroSim Modeling approaches

Modules	Method	Y	X	Dataset
MORTALITY	RF	Death	age; sex	OS
	LR	Widowhood (not in cohabitation)	age; age <sup>2</sup> ; sex; working	GM
	D	Widowhood (in cohabitation)	death of the partner	
AGEING	D	Age	age <sub>t+1</sub> = age <sub>t</sub> + 1	GM
	LR	Birth (women between 15 and 49 years)	age; age <sup>2</sup> ; school education; vocational education; working; working (partner); marriage; age of children in household; Eastern Germany	
SEPARATIONS AND MOVING OUT	LR	Twin birth	age	GM
	LR	Leaving cohabitation/marriage (up to 60 years of age)	age; age difference; age of children in household; marriage; school education (both partners); vocational education (both partners); citizenship homogamy; working; Eastern Germany	GM
	LR	Leaving parental home (18 years and older)	age · sex; age <sup>2</sup> · sex; sex; school education; vocational education; marriage; citizenship; Eastern Germany	GM
REGIONAL MOBILITY (between districts)	LR	Leaving a shared apartment (no relationship within household; from 18 to 35 years)	age · sex; age <sup>2</sup> · sex; sex; school education; vocational education; marriage; citizenship; Eastern Germany	GM
	RF	Outflow	age; sex; marriage	OS
	RF	Inflow	age; sex; marriage	OS
FORMATION OF NEW PARTNERSHIPS	LR	Cohabitation (separate models for men and women living alone; from 18 years)	children in household; children under 6 in household; age; age <sup>2</sup> ; separated; widowed; school education; vocational education; working; Eastern Germany	SOEP
	M	Finding a partner	matching on age, education, citizenship	GM
	LR	Marriage (not married cohabitations)	age; age <sup>2</sup> (women); family status (both partners); ISCED-Level (both partners); working (both partners); children under 1 in household; Eastern Germany	SOEP
MOVING INTO SHARED APARTMENTS	LR	Moving into a shared apartment (one-person household; from 18 to 35 years)	age; age <sup>2</sup> ; sex	GM
	RF	Age of enrollment (5-year-olds)	federal states; sex	OS
SCHOOL EDUCATION	MR	No qualification; lower secondary; secondary; technical college level; university entrance level	age; sex; school education (parents); vocational education (parents); working (parents); citizenship (parents); single-parent household	NEPS
	MR	Enrolled; no vocational training; vocational training; tertiary	working (parents); vocational education (parents)	AIDA
EMPLOYMENT	MR	Working; unemployed; inactive (from 15 to 65 years)	age; age <sup>2</sup> ; sex; school education; vocational education; marriage; citizenship; federal states; working <sub>(t-1)</sub> ; children in household; children under 3 in household;	GM
	LR	Full-time employment	age; age <sup>2</sup> ; sex; school education; vocational education; marriage; citizenship; federal states; working <sub>(t-1)</sub> ; children in household; children under 3 in household;	GM
RF	Relative Frequencies	OS	Official Statistics	
LR	Logistic Regression	GM	German Microcensus	
D	Deterministic	SOEP	German Socio-Economic Panel	
M	Matching	AIDA	Growing up in Germany (Aufwachsen in Deutschland; Alltagswelten, AID:A)	
MR	Multinomial Regression	NEPS	National Education Panel Study	

## 4 Modules

### 4.1 Mortality

The MORTALITY module is the first step in the simulation process. The probabilities for death are assigned according to sex and age of the simulated person using the life tables published by the Federal Statistical Office of Germany (Statistisches Bundesamt, 2020a).

The event of death does not only affect the size of the population but also individual and household characteristics. For instance, a partner's family status has to be updated if the husband or wife dies and underage orphans living alone after a parent dies must be assigned to new households. In addition, widowhood for married persons who do not live in the same household cannot be updated deterministically and therefore has to be simulated. This model is based on the German Microcensus (Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, 2018a; 2018b; 2018c).

After removing the deceased individuals from the simulation and updating the family status, the age of all remaining units is increased by one year.

### 4.2 Fertility

The FERTILITY module in MikroSim simulates births in two steps.

In the first step, for all women of fertile age (15–49 years) a probability of giving birth is estimated. The model uses individual characteristics of the women as well as characteristics of other people living in the household, such as the employment status of a potentially existing partner or the age of the youngest child (see Table 1).<sup>3</sup>

In the second step, conditional on the event of a birth, twin births are simulated.<sup>4</sup> The estimation model includes only the age of the women. The sex of a simulated child is

---

<sup>3</sup>We plan to extend this model to include more independent variables such as nationality and regional information.

<sup>4</sup>The birth of triplets or more children is not simulated due to the small number of cases.



assigned in accordance with the known sex distribution of newborns. All other variables of a simulated person are initialized to reasonable values (such as age to zero and school or vocational qualifications to missing values).

Since the model is based on sample data lacking regional details, differences between observed and simulated birth rates may result for some districts. Therefore, the model for birth is calibrated to known birth rates of the German districts up to the last available data (Statistisches Bundesamt, 2019a).

### **4.3 Separations and Moving Out**

In the MikroSim model, the events of individuals leaving households are simulated by three mechanisms: (1) separating from a partner, (2) leaving of the parental household and (3) moving out of a shared apartment. Modeling these three transitions using survey data is quite challenging, as individuals are either not tracked over time at all (e.g., in the German Microcensus) or only for subgroups (e.g., in the German Socio-Economic Panel, Goebel et al., 2019). Since the person moving out is usually only observed before a change of residential status occurs, identifying the cause of a departure is difficult using available data.

The first mechanism simulates the separation from a partner as ending a cohabitation. Therefore, only persons living in a partnership with cohabitation are considered. The probabilities are estimated with a logit model based on longitudinal Microcensus data for the years 2012 to 2014 using information of the partners and their partnership (such as age difference, for details see Table 1). In the case of a separation, new households are formed. These new households are initially single or multiple-person households if children are present. Currently, children are assigned to the mother, but future versions of the simulation will include predictive models for assigning children to new households.

The second mechanism models leaving the parental home. Only persons who are

at least 18 years of age and still living in the parental household are considered. The probabilities are estimated with a logit model based on the same data as the mechanism described above. As predictors, age, the current level of education or vocational training and the relationship status are used (cf. Table 1). After leaving their parental homes, the individuals initially form new single-person households.

The third mechanism models moving out of shared apartments.<sup>5</sup> Probabilities are estimated using a similar model and the same data as for the other two mechanisms. The number of people moving out is used for modeling the number of persons moving into shared apartments in later iterations of the simulation.

To prevent overestimating the number of single person households, the moved-out persons can form new households by entering either new partnerships or shared apartments.

#### **4.4 Regional Mobility**

The module REGIONAL MOBILITY simulates between municipalities. Individuals leave the simulation population when moving out of the district and are added from a copy of the base population when moving into the district. This process is based on statistics produced by the Federal Statistical Office (Destatis, 2020). These statistics contain information on regional mobility broken down by age, sex, and relationship status. Subsequently, the probabilities for regional mobility are adjusted to known margins for each district level.

To prevent minors forming a household, regional mobility is additionally considered at the household level. Probabilities are estimated by Iterative Proportional Updating (Ye et al., 2009). Iteratively, the probabilities are adjusted to the frequencies of socio-demographic characteristics at the individual level in a randomized order and then scaled to a probability between zero and one (Stephensen, 2016). This is done using the base

---

<sup>5</sup>Shared apartments are defined by us as households containing at least two people aged between 18–35 without children or partnerships.

population to assign probabilities for leaving the district and on a copy of the base population to assign probabilities for moving into the district. Households that move into a district are selected from the copy of the population, which represents the remaining part of Germany. Newly arrived households are added to the base population. Outmoving households are removed from the population.

## 4.5 Formation of New Partnerships

The nuclear family (cohabitation of a mother, a father, and children) is still by far the most common family type in Western Europe. However, partners within the nuclear family increasingly remain unmarried (Schneider, 2015). Therefore, the simulation requires a module simulating the formation of new partnerships independent from the official status of a relationship (which is modeled in the MARRIAGE model).

A module for the creation of new partnerships performs two simulation tasks: The entry into the partner market and the matching of new couples considering covariates.<sup>6</sup> There are different modeling approaches in other simulation models (cf. Perese, 2002; Zinn, 2012). MIKROSIM uses stochastic matching, therefore allowing less favorable partner combinations (for example, in large age differences). The module consists of two-steps.

In the first step, for persons older 18 years not living in a partnership, a model based on German Socio-Economic Panel data (Goebel et al., 2019) estimates the probability of cohabitation with a partner (for details, see Table 1). Since the model is specified separately for men and women, the estimated propensities for a relationship might yield an imbalance of men and women available on the partner market. By only considering people in the same district, regional aspects of partner markets and the importance of

---

<sup>6</sup>The main mechanism is homogamy with respect to age, socio-economic status, and nationality (Klein, 2015).

spatial distance for partnerships are modeled.<sup>7</sup>

In the second step, the selected persons are matched.<sup>8</sup> The probability for cohabitation is estimated with a logit model using German Microcensus data. To account for potential age difference of the partners, spline functions are used in terms of generalized additive models (Wood, 2017). Therefore, also rare but possible partnerships (for example, with large age differences) are generated. The simulated imbalances of regional partner demands models the option of better available choices in asymmetric partner markets (cf. Klein, 2000).

## 4.6 Marriage

Within the MARRIAGE module, only single, widowed or divorced partners are eligible for marriage. The module starts with unmarried couples within existing households. Information on the household and individuals is used to predict the couple's probability to marry. The model is based on data from the German Socio-Economic Panel (for details of the model, cf. Table 1). Given the estimated transition probabilities, marriages are simulated by updating family status for the involved partners.

## 4.7 Moving into Shared Apartments

Since the number of apartment-sharing communities is likely to be high in large cities and university towns, neglecting this type of housing would lead to an unrealistically high number of single-person households. In the module, only people currently living alone between 18 and 35 years old are eligible to move into a shared apartment.

---

<sup>7</sup>People living in different districts are not matched in the module for two reasons. On the one hand considering all potential partners would result in high computational costs. On the other hand matched partners from different districts would again lead to regional mobility and subsequently distort the marginal distribution resulting from the REGIONAL MOBILITY module.

<sup>8</sup>Due to limited information on same-sex partnerships, the partnership module is restricted to heterosexual partnerships only.

The probability for moving into shared apartments is based on data from the German Microcensus and is estimated with a logit model including only age and gender as independent variables (cf. Table 1). The estimated probabilities for each district are calibrated via iterative proportional fitting so that the proportion of people living in a shared apartment remains the same after the first simulation period. The proportion of people moving into shared apartments is then left constant, such that a change of the total number of people living in shared apartments is a result of changing population structures.

To form new households, all relocating persons are randomly matched, such that the average household size of three persons is created while distributional size assumptions are approximately met.

## 4.8 School Education

The SCHOOL EDUCATION module consists of two sub-modules, (1) a SCHOOL ENROLLMENT module and (2) a module for assigning educational qualifications.

The ENROLLMENT module assigns the time of enrollment based on relative frequencies from official data (Statistisches Bundesamt, 2020b). A child is either enrolled early (at age 5), regularly (at age 6) or late (at age 7).<sup>9</sup> The probabilities are available by federal state and sex for different types of primary schools.<sup>10</sup>

The SCHOOL EDUCATION module assigns the duration of school attendance and the resulting certificate. Grade levels of students are promoted yearly. To take repetitions of classes into account, a pragmatic approach was chosen: The duration of primary

---

<sup>9</sup>In Germany, there are different key dates for school enrollment depending on the federal state. In MIKROSIM we do not take them into account, since we do not model birthdays and we use a yearly time framework to update characteristics.

<sup>10</sup>An initial attempt to estimate the probabilities by a model using the National Education Panel Study (NEPS) (Blossfeld, Roßbach, and Maurice, 2011) was discarded due lack of predictive power: The model including education, partner education, age, age of partner, work, marital status, and body length at birth yielded a McFaddens Pseudo- $R^2$  of 0.03.

schooling is estimated for each child. Based on the time of enrollment and the estimated duration, the age at which the students will complete the fourth grade is calculated. At this age, the child will be promoted to the fifth grade. Before this, they are simply in primary school.

Children at the beginning of the simulation must be assigned a grade level. They are assigned a grade level according to their age, not considering any covariates such as repeated classes or late enrollments.

The NEPS (Starting Cohort 2) is used to estimate the probabilities for the duration of primary schooling. The current module promotes people yearly until they leave school. A model predicting re-attendance of a class from fifth grade onward will be implemented in later stages of the project.

Since school qualifications in Germany can be obtained after grade 9, degrees are tentatively assigned at this grade. As soon as a qualification is assigned, the school career is continued depending on this qualification. The probabilities for grades are estimated using a multinomial logit model based on NEPS (Starting Cohort 4) data (see Table 1 for details).

## 4.9 Vocational Training

In the VOCATIONAL TRAINING module, people who left school are assigned a job qualification. Possible values are no vocational training, vocational training degree, or university degree.<sup>11</sup> All model estimates are based on the dataset “Growing up in Germany” (AID:A II) produced by the German Youth Institute.

The time until graduation (1–6 years) of people with an assigned Bachelor’s degree is taken from official statistics (Statistisches Bundesamt, 2019b). Persons not obtaining a degree receive a vocational training which after 3 years ends with a professional

---

<sup>11</sup>Degrees from universities of applied sciences and PhDs are not modeled because the proportions in the population are small and are often coded as university degree in survey data.

qualification. Future versions of the module will use distributions from official data for the duration of training.

At the start of the simulation, people already attending a university need an estimate of the duration of their attendance, which is estimated using a regression model based on a student survey data (Georg, Ramm, and Bundesministerium für Bildung und Forschung, 2016). Age is used to predict the stage of the academic career of the student.

## **4.10 Employment**

The EMPLOYMENT module consists of three models: First, the employment status is assigned to the individuals in the base dataset. Second, full-time and part-time employment are estimated for the working population. Hereby, only individuals between the age of 15 to 66 years are considered. Third, the simulated objects are sent into retirement.

The required transition probabilities are estimated by a multinomial logistic model using German Microcensus data. The model uses individual-level variables (such as age, gender, education, and employment status last year), household characteristics (children and type of relationship), and considers the regional variation in labor market status to predict employment status.

For the working subpopulation the probabilities for part-time or full-time employment is estimated using the same predictors (see Table 1).

In the module, currently all people passing the age-threshold of 67 get retired. In a later version we will account for early retirement.

## **4.11 Income**

A module of central importance is the INCOME module. This module provides necessary information for policy analyses, such as tax or family policy reforms. Moreover, income

is an important explanatory variable for other models, such as fertility decisions, internal migration, or education opportunities of children.

The INCOME module is based on the Taxpayer Panel, an administrative data source covering the entire population of taxpayers in Germany from 2001 to 2014. Since the tax data does not contain detailed socio-demographic variables, the dataset will be enhanced by Microcensus data using statistical matching methods (predictive mean matching and nearest-neighbor random hotdeck).

The purpose of the module is the regional prediction of income and its changes over time. Estimates of individual incomes based on a mixed model will be calibrated with published regional data, such as income, poverty and inequality indicators. To model changes, a two step approach is used. First, the probability of a change in income is predicted, then the differences to the previous year is modeled.

## **5 Application Modules**

The core simulation model can be extended easily. Currently, two extensions are implemented: (1) elderly care in the family and (2) the labor market outcomes of migrants in Germany. Other modules will be added in the future. We describe the current non-core modules briefly.

### **5.1 Elderly Care**

A central subject matter problem in the MikroSim project is the increase of informal care for the elderly depending on demographic changes and new family structures. The current aim of the module is the study of the development of intra-family care. Therefore, household structures have to be updated. For example, children and grandchildren can be potential informal care providers for the elderly. Since the same function can be performed by family members outside the household, a mechanism to add these external



families to a household is needed. The details of this mechanism are the subject of ongoing research.

In addition to this modeling of care supply, the demand for care has to be simulated. Therefore, a model for the degree and duration of care required is needed. Since the degrees of care have to be consistent with other variables in the model, additional constraints have to be fulfilled.

## **5.2 Labor Market Outcomes for Migrants**

The second subject-matter topic is the development of labor market qualifications in the migrant population in Germany. The aim of the module is to supplement the projection of labor market integration with a regional perspective since both the allocation of migrants and the labor market outcomes of individuals differ regionally.

Membership to different migrant groups (e.g., EU/non-EU) and migrant generations (i.e. born elsewhere/second-generation) are modeled regionally. A specific citizenship model within the module generates naturalization probabilities for all migrants in the simulation. The models specified in Table 1 will be estimated for the migrant population.

To estimate the integration of migrants in the labor market, the `EMPLOYMENT` module and its first sub-module concerning the labor market status of the simulated individuals are of interest.

The relative labor market positioning of migrants in comparison to the majority population is predicted dependent on the assignment of outcomes in the `SCHOOL` and `VOCATIONAL EDUCATION` modules run previously. The `INCOME` module will allow the study of income disparities and their potential change over time for the migrant population. Through the planned `CITIZENSHIP` module and the extensions in the `EMPLOYMENT` module, we can then estimate different scenarios of regional labor market integration for different ethnic minorities.

## **6 Achievements, current work, and further development**

After the first 18 months of work, the datasets required for estimation have been obtained, harmonized, documented and been used to estimate parameters for cross-sectional characteristics and transition probabilities. The base dataset has been updated to new margins and enhanced by survey data estimates.

The basic structure of the simulation model was planned and implemented as separate modules as shown in Figure 1. For each module, the processes required for updating the model were specified and estimated using available data. Resulting estimates were calibrated to known totals. The program code has been documented and tested. Currently, first test runs of the schooling simulation are in progress.

Within the next six months, the refinements described above will be implemented. After that, sensitivity studies and policy scenarios will be run and analyzed. We intend to publish first simulation results 30 months after the project has started. The availability of further data will furnish additional calibration methods. The modules will be fine-tuned further, leading to improved reproduction of known regional patterns and rare subgroups. To support continuous data updates while preserving model reproducibility, a data versioning system will be implemented in the simulation environment.

During the next funding period, additional modules will be implemented to test scenarios for policy studies in housing, health service research and urban travel demand. We intend to open the simulation model for other research groups by building a research data center, operating on similar principles as other data research centers in Germany.

## **Acknowledgement**

The research unit FOR 2559 MikroSim is funded by the German Research Foundation.

The principal investigators are Ralf Münnich (Speaker), Rainer Schnell (Co-Speaker),

Johannes Kopp and Petra Stein. The research unit cooperates with the German Federal Statistical Office, with Hanna Brenzel as primary contact, succeeding Markus Zwick. The paper was initiated by the principal investigators and drafted by those members of the research unit who are responsible for the modules described. The latest version of the paper was rewritten by Rainer Schnell.

## References

- Alfons, Andreas, Peter Filzmoser, Beat Hulliger, Jan-Philipp Kolb, Stefan Kraft, Ralf Münnich, and Matthias Templ (2011). *Synthetic Data Generation of SILC Data*. Research Project Report 6.2. Advanced Methodology for European Laeken Indicators.
- Alfons, Andreas, Stefan Kraft, Matthias Templ, and Peter Filzmoser (2011). “Simulation of Close-to-reality Population Data for Household Surveys with Application to EU-SILC”. In: *Statistical Methods & Applications* 20.3, pp. 383–407.
- Blossfeld, Hans-Peter, Hans-Günther Roßbach, and Jutta von Maurice, eds. (2011). *Zeitschrift für Erziehungswissenschaft Special Issue 14: Education as a Lifelong Process – The German National Educational Panel Study (NEPS)*.
- Burgard, Jan Pablo, Joscha Krause, Hariolf Merkle, Ralf Münnich, and Simon Schmaus (2019). “Conducting a Dynamic Microsimulation for Care Research: Data Generation, Transition Probabilities and Sensitivity Analysis”. In: *Workshop on Stochastic Models, Statistics and their Application: Dresden, Germany, March 2019*. Ed. by Ansgar Steland, Ewaryst Rafajłowicz, and Ostap Okhrin. Cham: Springer, pp. 269–290.
- Burgard, Jan Pablo, Joscha Krause, Hariolf Merkle, Ralf Münnich, and Simon Schmaus (2020). “Dynamische Mikrosimulationen zur Analyse und Planung regionaler Versorgungsstrukturen in der Pflege”. In: *Mikrosimulationen*. Ed. by Marc Hannappel and Johannes Kopp. Wiesbaden: Springer VS, pp. 283–313.

- Destatis (2020). *Bevölkerung: Wanderungen*. URL: [https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Wanderungen/\\_inhalt.html](https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Wanderungen/_inhalt.html).
- Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder (2018a). *Mikrozensus 2012, SUF, Version 0*. DOI: 10.21242/12211.2012.00.00.3.1.0.
- Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder (2018b). *Mikrozensus 2013, SUF, Version 0*. DOI: 10.21242/12211.2013.00.00.3.1.0.
- Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder (2018c). *Mikrozensus 2014, SUF, Version 0*. DOI: 10.21242/12211.2014.00.00.3.1.0.
- Georg, Werner, Michael Ramm, and Bundesministerium für Bildung und Forschung (2016). *Learning Conditions and Student Orientations 2012/13*. DOI: 10.4232/1.12510.
- German Youth Institute (2014). *Growing Up in Germany: Everyday Life's World (AID:A II)*. München. URL: <https://surveys.dji.de/index.php?m=msw,0&sID=107> (visited on 04/07/2020).
- Goebel, Jan, Markus M. Grabka, Stefan Liebig, Martin Kroh, David Richter, Carsten Schröder, and Jürgen Schupp (2019). “The German Socio-Economic Panel (SOEP)”. In: *Jahrbücher für Nationalökonomie und Statistik* 239.2, pp. 345–360. DOI: 10.1515/jbnst-2018-0022.
- Huang, Zengyi and Paul Williamson (2001). *A Comparison of Synthetic Reconstruction and Combinatorial Optimisation Approaches to the Creation of Small-Area Microdata*. Working Paper 2001/2. Liverpool: Department of Geography, University of Liverpool.
- Klein, Thomas (2000). “Partnerwahl zwischen sozialstrukturellen Vorgaben und individueller Entscheidungsautonomie”. In: *Zeitschrift für Soziologie der Erziehung und Sozialisation* 20.3, pp. 229–243.
- Klein, Thomas (2015). “Partnerwahl”. In: *Handbuch Familiensoziologie*. Ed. by Paul B. Hill and Johannes Kopp. Wiesbaden: Springer, pp. 321–343.

- Kolb, Jan-Philipp (2013). “Methoden zur Erzeugung synthetischer Simulationsgesamtheiten”. PhD thesis. Trier: University of Trier.
- Laarhoven, Peter J. M. and Emile H. L. Aarts (1987). *Simulated Annealing: Theory and Applications*. Dordrecht: Springer.
- Münnich, Ralf, Siegfried Gabler, Matthias Ganninger, Jan Pablo Burgard, and Jan-Philipp Kolb (2012). *Stichprobenoptimierung und Schätzung im Zensus 2011*. Vol. 21. Statistik und Wissenschaft. Wiesbaden: Statistisches Bundesamt.
- Münnich, Ralf and Josef Schürle (2003). *On the Simulation of Complex Universes in the Case of Applying the German Microcensus*. DACSEIS Research Paper Series 4. Eberhard Karls University of Tübingen.
- O’Donoghue, Cathal and Gijbrecht Dekkers (2018). “Increasing the Impact of Dynamic Microsimulation Modelling”. In: *The International Journal of Microsimulation* 11.1, pp. 61–96.
- Perese, Kevin (2002). *Mate Matching for Microsimulation Models*. Technical Report 3. Washington: Congressional Budget Office.
- Schneider, Norbert F. (2015). “Familie in Westeuropa: Von der Institution zu Lebensform”. In: *Handbuch Familiensoziologie*. Ed. by Paul B. Hill and Johannes Kopp. Wiesbaden: Springer, pp. 21–54.
- Schnell, Rainer and Thomas Handke (2020). “Neuere bevölkerungsbezogene Mikrosimulationen in Großbritannien und Deutschland”. In: *Mikrosimulationen*. Ed. by Marc Hannappel and Johannes Kopp. Wiesbaden: Springer, pp. 35–56. DOI: 10.1007/978-3-658-23702-8\_3.
- Statistisches Bundesamt (2019a). *Bevölkerung: Eheschließungen, Geborene und Gestorbene 2018 nach Kreisen*. URL: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Eheschliessungen-Ehescheidungen-Lebenspartnerschaften/>

- Publikationen/Downloads-Eheschliessungen/eheschliessungen-geborene-gestorbene-5126001187004.html.
- Statistisches Bundesamt (2019b). *Bildung und Kultur: Prüfungen an Hochschulen*. URL: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Publikationen/Downloads-Hochschulen/pruefungen-hochschulen-2110420187004.pdf> (visited on 01/21/2020).
- Statistisches Bundesamt (2020a). *Bevölkerung: Sterbefälle und Lebenserwartung*. URL: [https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Sterbefaelle-Lebenserwartung/\\_inhalt.html](https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Sterbefaelle-Lebenserwartung/_inhalt.html).
- Statistisches Bundesamt (2020b). *Schulanfänger: Bundesländer, Schuljahr, Geschlecht, Einschulungsart, Schulart*. URL: <https://www-genesis.destatis.de/genesis/online?operation=table&code=21111-0010> (visited on 04/07/2020).
- Stein, Petra and Dawid Bekalarczyk (2016). “Zur Prognose beruflicher Positionierung von Migranten der dritten Generation”. In: *Empirische Prognoseverfahren in den Sozialwissenschaften*. Ed. by Reinhard Bachleitner, Martin Weichbold, and Markus Pausch. Wiesbaden: Springer, pp. 223–257.
- Stephensen, Peter (2016). “Logit Scaling: A General Method for Alignment in Microsimulation Models”. In: *International Journal of Microsimulation* 9.3, pp. 86–102.
- Tanton, Robert (2014). “A Review of Spatial Microsimulation Methods”. In: *International Journal of Microsimulation* 7.1, pp. 4–25.
- Van Imhoff, Evert and Wendy Post (1998). “Microsimulation Methods for Population Projection”. In: *Population: An English Selection* 10.1, pp. 97–138.
- Williamson, Paul (2012). “An Evaluation of Two Synthetic Small-Area Microdata Simulation Methodologies: Synthetic Reconstruction and Combinatorial Optimisation”. In: *Spatial Microsimulation: A Reference Guide for Users*. Ed. by Robert Tanton and Kimberley L. Edwards. Dordrecht: Springer, pp. 19–47.

- Wood, Simon N. (2017). *Generalized Additive Models: An Introduction with R*. 2nd ed. Boca Raton: CRC Press.
- Ye, Xin, Karthik Konduri, Ram M. Pendyala, Bhargava Sana, and Paul Waddell (2009). “A Methodology to Match Distributions of Both Household and Person Attributes in the Generation of Synthetic Populations”. In: *88th Annual Meeting of the Transportation Research Board*.
- Zinn, Sabine (2012). “A Mate-Matching Algorithm for Continuous-Time Microsimulation Models”. In: *International Journal of Microsimulation* 5.1, pp. 31–51.
- Zwick, Markus and Jana Emmenegger (2020). “Mikrosimulation und Gesellschaftspolitik – ein kurzer historischer Abriss”. In: *Mikrosimulationen*. Ed. by Marc Hannappel and Johannes Kopp. Wiesbaden: Springer, pp. 17–34. DOI: 10.1007/978-3-658-23702-8\_2.