Unit 1: Introduction to microsimulation in official statistics

- ✓ History of microsimulation
  - ✓ Historic development of microsimulation
  - ✓ Different quotes what microsimulation is
  - ✓ Objectives of microsimulation and the 4M-Strategy

The birth of microsimulation

Guy Henderson Orcutt (1917-2006)
Source: Watts (1991).

Guy H. Orcutt (1957): A New Type of Socio-Economic System

Existing models of our socio-economic system have proved to be of rather limited predictive usefulness. This is particularly true with respect to predictions about the effects of alternative governmental actions and with respect to any predictions of a long-range character. It is even the case with respect to very short-run forecasting. In addition, it is recognized that current models of our socio-economic system have an unduly narrow reach in that they have little to say about such fundamental things as the size and location of the population of individuals, of households, or of firms.

Orcutt’s criticism

- Baseline problem: limited usefulness of macrosimulations for the prediction of effects within a society
  - Example: competing governmental actions with respect to some current socio-economic problems
- Macrosimulations: focus is limited on aggregates
  - Criticism: comprehensive solutions to socio-economic problems require the study of context-related distributions on small scales
- Small scales: individuals, households, neighborhoods, firms
  - Reason: behavior of decision-making units induces nonlinear dependencies and discontinuities within a society

Orcutt's criticism – Numerical example

Orcutt illustrated this by means of a simple numerical example:

- Let *U* be a population of *N = 100* individuals that are indexed by *i = 1;….;N*
- Each individual produces an output *Y* given an input *X*, where the individual realizations are denoted by the pair (y_i, x_i)
- Assume $y_i = 0$ when $x_i = 0$, and $y_i = 1$ when $x_i \in \{1,2\}$
- If $x_i = 1 \; \forall \; i \in U$, then $\sum_{i=1}^{N} y_i = 100$ and $\sum_{i=1}^{N} x_i = 100$
- However, if $x_i = 2$ for 50 individuals and $x_i = 0$ for the rest, then $\sum_{i=1}^{N} y_i = 50$ and $\sum_{i=1}^{N} x_i = 100$
- In both cases, the total of X is the same, but the total of Y depends on the distribution of X
- New type of model of the socio-economic system is needed

A new type of model

This new type of model consists of various sorts of interacting units which receive inputs and generate outputs. The outputs of each unit are, in part, functionally related to prior events and, in part, are the result of a series of random drawings from discrete probability distributions (Orcutt, 1957, p. 117).

Main components:

- Units: In general elemental decision-making entities (individuals, families, firms, ..)
- Input: Anything which enters into, acts upon, or is taken account of, by the unit
- Output: Anything which stems from, or is generated by, the unit
- Operating characteristics: Equation or tables which determine (the probability of) outputs

→ Need for a powerful giant, such as the IBM 704

IBM 704 (1954)

Source: Harding and Zaidi (2009).

What is a microsimulation? – Some quotes

Microsimulation(s)

- ... allow simulating the effects of a policy on a sample of economic agents (individuals, households, firms) at the individual level. (Bourguignon & Spadaro, 2006)
- ... is the simulation of dynamic feedback (in both directions) between individual states and states of the population as a whole or certain groups within a population. (Troitzsch et al., 1996)
- ... is a technique used to model complex real-life events by simulating the actions of, and/or the impact of policy change on the individual micro unit. (Li, 2011)
- ... is a tool to generate synthetic micro-unit based data, which can then be used to answer many\what-if" questions that, otherwise, cannot be answered. (Li & O'Donoghue, 2013)

What is a microsimulation? - Baseline

- Model conception that allows for the analysis of complex systems
- Already applied in various scientific fields, such as particle physics, cancer research, or transport research
- Microsimulation is the corresponding term commonly used within socio-economic research
- Basic procedure:
  1 Collect sample data on the population of interest
  2 Create a synthetic replica as base population
  3 Project the base population into future periods under scenarios
  4 Study the different outcomes on small scales
- Increasingly popular due to fast-growing computational power

Objectives of microsimulation

- Simulation of micro data to gain knowledge about a population
- Analysis of complex policy impacts on a population
  - Policies often based on individual characteristics
  - Interactions of different interventions
  - Effects can change over time
- Analysis of complex population developments
  - Interaction of individuals within households
  - Interdependency of socio-demographic and socio-economic events at micro levels
  - Broad focus on target values and distributions
- Analysis on spatial levels
  - Accounting for regional heterogeneity
  - Social structures as well as political impacts may vary between regional areas

Constructing a microsimulation model

4M-Strategy (Merz, 1993)

1. Microtheory
2. Microdata and adjustment of microdata
3. Microestimation, microeconometrics
4. Microsimulation

→ Microsimulation as the combination

Unit 1: Introduction to microsimulation in official statistics

- ✓ The development of microsimulation in Germany
  - ✓ History of microsimulation and microdata
  - ✓ SPES and SFB 3

Development of microsimulation models I

- Although the need for microsimulations has already been recognized in the 1950s, the scientific establishment took decades
- Time delay in development due to limited computational power
- Early 1970s: microsimulations exclusively used in the USA
- Application of various microsimulation models during the 1970s as result of technological innovations
- Origination of DYNASIM (Dynamic Simulation of Income Model) at Urban Institute in Washington D.C. in cooperation with Orcutt
- Adaption of DYNASIM methods for the CORSIM project at Cornell University

Development of microsimulation models II

- Main criticism in the 1970s: Heavy programming, computing, and data requirements
- Creation of simpler microsimulation models without temporal components in the 1970s
- Expansion of more complex dynamic microsimulation models as result of reduced computing costs in the late 1980s
- Today: more and more research projects using microsimulation models are launched

→ In the following, an exemplary overview is given

History of microsimulation and microdata

Socio-political decision-making and indicator system for the FRG (SPES)

1970

2000

First use of official microdata

SPES project overview

Project: Social policy decision-making and indicator system for Germany (SPES)

Duration: 1971-1979

Involved institutions:

- Goethe University Frankfurt
- Mannheim University
- Federal Statistical Office of Germany (DESTATIS) as data provider

Legal basis: Federal Statistical Act (StatGes) 1953

- First general definition of the basic principles of organizational law, procedural las and substantive law of statistic

SPES project content

Data:

- German Income & Expenditure Survey (EVS) 1962 and 1969
- Microcensus 1962 to 1969
- Microcensus Ad-hoc Modules 1971

Topics:

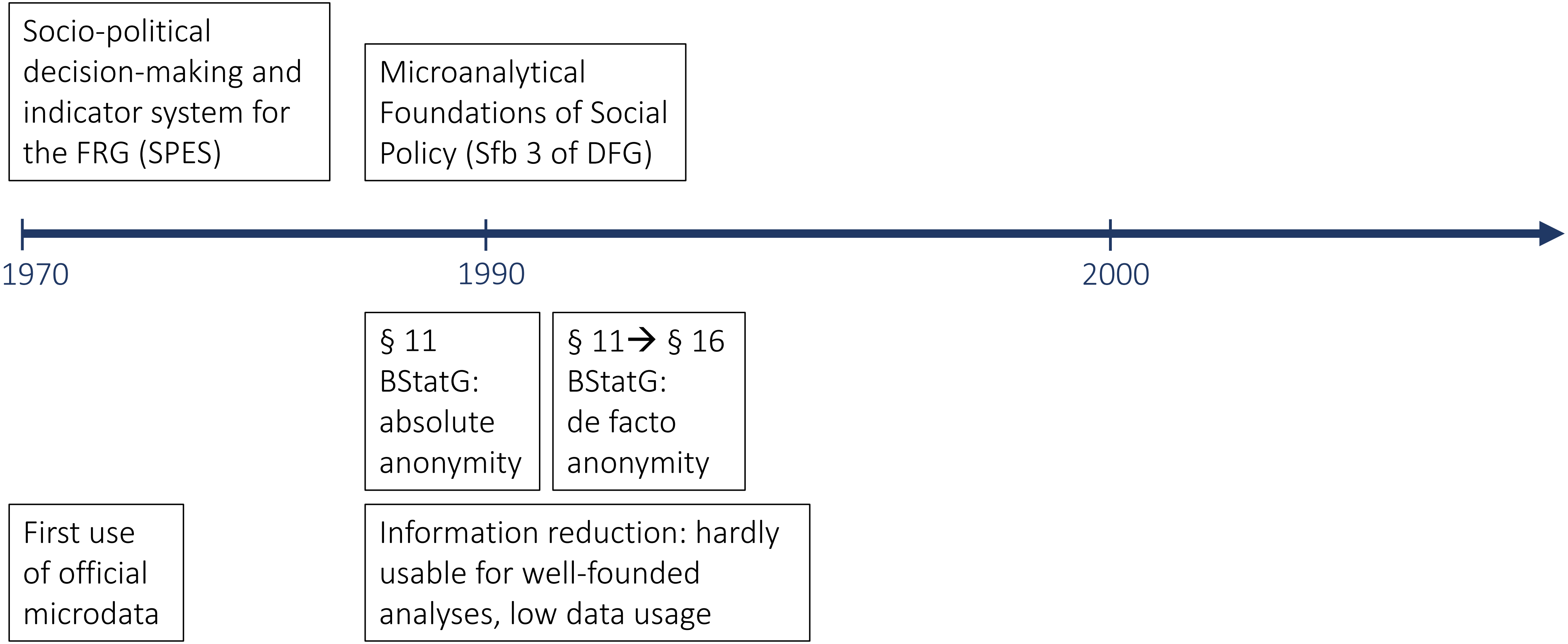
- Creation of a dynamic and static microsimulation model for Germany
- Focus on long-term social monitoring: demographic effects and professional activities

SPES project results

Main Results:

- First big data processing system in Germany
- First use if official statistics microdata for policy consulting
- Discussion within the Federal Statistical System of Germany as well as with academia and data protection officers (Zühlke et al. 2003)

History of microsimulation and microdata



SFB 3 project overview

Project: Collaborative Research Centre (SFB3) of the German Research Foundation: Microanalytical foundations of social policy

Duration: 1979-1992

Involved institutions:

- Goethe University Frankfurt
- Mannheim University
- Federal Statistical Office of Germany (DESTATIS) as data provider

Legal basis:

- StatGes is replaced by BStatG in 1980
- BStatG reform 1981: first transmission of microdata from Statistical Oces
- BStatG reform 1987: establishment of the research prerogative (“Wissenschaftsprivileg") for the use of de facto anonymized microdata (Merz et al. 2004)

SFB3 project content

Data:

- German Income & Expenditure Survey (EVS)
- Microcensus
- Income txt statistic (aggregated as well as microdata)
- Luxembourg income study
- SOEP, welfare surveys with subjective variables
- .....

Content overview:

- SPES project as basis for the SFB3
- Analysis of economic and social policies
- Broad variety of topics covering welfare analysis, social securities, health topics, micro as well as macro approaches, and many more

SFB3 project results

Main Results:

- Development of the Socio-Economic Panel (SOEP) for the investigation of individual dynamics
- First Scientific Use File of Microcensus 1989
- Establishment of the Research Data Centres (RDC) at the Statistical Offices
- New generation of experts in microanalysis: 7 habilitations and 61 dissertations within the project

→ SFB3 model as the first large-scale microsimulation model in Europe

History of microsimulation and microdata

Socio-political decision-making and indicator system for the FRG (SPES)

Microanalytical Foundations of Social Policy (Sfb 3 of DFG)

1970 1990 2000 2019

§ 11 BStatG: absolute anonymity

§ 11→ § 16 BStatG: de facto anonymity

First use of official microdata

Information reduction: hardly usable for well-founded analyses, low data usage

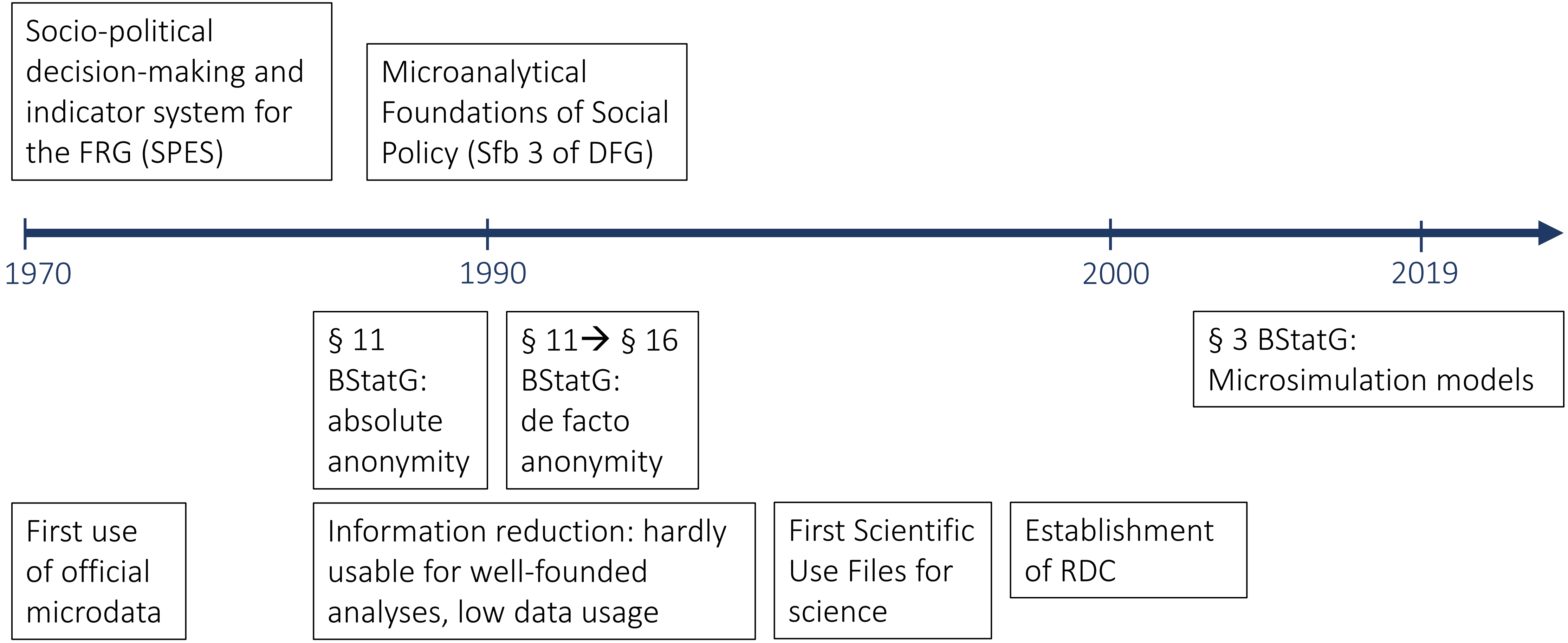
First Scientific Use Files for science

Establishment of RDC

Simulation models in tax policy

- Development of various static tax microsimulation models for the purposes of the federal ministry of finance (BMF) in the 1990s and 2000s
- Reform of the tax statistics act (StStatG) in 1996: first use of the administrative income tax data
- Two groups of microsimulation models:
  - based on SOEP data: e.g. GMOD (Wagenhals 2004), STSM (Jacobebbinghaus und Steiner 2003) und SimTrans (Kaltenborn 1998)
  - based on tax data: e.g. MICSIM, the microsimulation model of the BMF

History of microsimulation and microdata



Federal Statistical Act (BStatG) reform 2016

- Amendment of §3 (1) Nr. 6 BStatG in summer 2016 “development and application of microsimulation models as well as microeconometric analysis" as explicit task of the Federal Statistical Office of Germany
- Follow-up processing for federal purposes under the Federal Statistics Act
- Development of a competence center for microanalysis and -simulation within the Federal Statistical Office of Germany

History of microsimulation and microdata

Socio-political decision-making and indicator system for the FRG (SPES)

Microanalytical Foundations of Social Policy (Sfb 3 of DFG)

Multi-sectoral regional microsimulation model (MikroSim)

1970 — 1990 — 2000 — 2019

§ 11 BStatG: absolute anonymity

§ 11→ § 16 BStatG: de facto anonymity

§ 3 BStatG: Microsimulation models

First use of official microdata

Information reduction: hardly usable for well-founded analyses, low data usage

First Scientific Use Files for science

Establishment of RDC

Unit 1: Introduction to microsimulation in official statistics

- ✓ Static and dynamic microsimulations
  - ✓ The two concepts
  - ✓ Advantages and disadvantages
  - ✓ Typical examples

Basic definitions

Following Li & O'Donoghue (2013), there are two conceptions of microsimulation models: static and dynamic models.

Static modeling

- Focusses on the immediate distributional impact of changes on the population of interest
- No time dimension: population is studied for a single period given a proposed scenario with respect to some characteristic
- Example: How does the GINI coefficient for net income within the current population change given a new tax policy?
- Applications: EUROMOD, IZAΨMOD, STINMOD

Basic definitions

Dynamic modeling

- Focusses on micro-level dynamics arising from changes to the population of interest over time
- Micro-level dynamics: simulation units evolve in essential characteristics over simulated periods (e.g. income, family status)
- Allows for the study of long-term effects with respect to some assumed scenario
- Example: How does the care situation of elderly change over time in the light of demographic change?
- Applications: DYNACAN, DYNAMOD, MikroSim

Static modeling

According to Merz (1991), in a static microsimulation model, the following is considered:

- Certain behavioral relations and institutional conditions of a microdata base are systematically varied
- The microdata base usually comprises cross-sectional information
- It is assumed that the characteristics of the population of interest do not change rapidly

→ Suitable for short- and medium-range forecasts

Static ageing

According to Merz (1991) and Sutherland (2018), the static ageing process is realized by:

- Re-weighting the microunits using exogenously given aggregate (control) data of another time period
    - A microunit represents a certain number of other microunits in the population according to the weight
    - The relations among the variables of each microunit are generally maintained
    - Cross-section after simulation = Cross-section before simulation
- Uprating/deflating certain variables
    - This is usually done for metric variables (e.g. income components)
    - Data are directly altered

Dynamic microsimulation

- **Static:** evaluate the direct distributional impact on micro-units
- **Dynamic:** micro-units interact and evolve over time
- Simulation of micro-unit behavior over a temporal horizon
- Accounting for micro-level dependencies and complex interactions
- Long-term projections and time-dependent behavior simulations
- Note: primary objective is not to predict the future in the sense of point prediction (e.g. "In 2050, there will be ...")
- Focus is on sophisticated ceteris paribus analysis over time under an approximation to real-world complexity

Methodical differentiation

Dynamic microsimulation methods can be differentiated regarding the subsequent aspects (Li & O'Donoghue 2013):

- Base population
- Population of interest
  - cohort versus population model
- Population structure
  - open versus closed population
- Ageing method
  - static versus dynamic ageing
- Type of population ageing / time progression
  - continuous versus discrete stochastic process

Cohort and population models

Cohort models

- Model a single cohort over an extended time period
- Base population represents a relevant share of the real-world population given the research objectives
- Reduced computational requirements
- Model structure relatively inflexible

Population model

- Modeling of the entire population regardless of research objectives
- Base population represents a cross-sectional synthetic replica of the real-world population to be studied
- Higher flexibility: connecting points for further topics

Closed and open models

Closed model

- Closed: population evolution is exclusively driven by intrinsic events
- Example: evolution of Berlin only depends on Berlin citizens
- Population only changes through migration, birth and mortality
- Partner matching module required for demographic projection

Open model

- Population evolution may be also driven by external events that the researcher can implement artificially
- Flexible interfaces for adding new individuals
- Demographic projection includes synthetically generated partners
- Less computational requirements since no matching required

Discrete and continuous ageing

Discrete ageing ($\boldsymbol{S} \subset \mathbb{N}$)

- Simulation on discrete periods $\{0,1,2, \ldots.., S\}$, where *S* is the number of simulated periods (simulation horizon)
- Events are realized in each period *s* with associated probabilities
- Usually, ageing is with respect to annual or monthly intervals

Continuous ageing ($\boldsymbol{S} \subset \mathbb{R}_{\geq \mathbf{0}}$)

- Simulation on the continuous interval *[0; S],* where *S* is the upper boundary of the simulation horizon
- Events may be realized at any $s \leq S$
- Requires the estimation of "waiting times", hence the amount of time passed until a given event is realized

Unit 1: Introduction to microsimulation in official statistics

- ✓ Microsimulation as a policy tool
  - ✓ The MikroSim project (FOR 2559 „Multi-sectoral Regional Microsimulation Model”)

Motivation of MikroSim

- Increasing demand for statistical data and evaluation methods
- Demographic change, regional, supra-regional and international migratory movements

    → Challenges in the area of *Policy Oriented Research*

- Need for complex analyses of political measures and social constitutions
- Microsimulations as internationally established analytical instruments
    - Broad application e.g. in Canada, Australia, USA, Norway, Germany
- Microsimulations as a task of the Federal Statistical Office in the Federal Statistics Act (BStatG)

Base Data

- Requirements
  - Unit- and household-level data of the Germany population
  - Containing geographical information
  - Wide range of socio-demographic and -economic variables
- Problems
  - No appropriate unit-level data considering all requirements available
- Solution
  - Creation of a (partially) synthetic population

The MikroSim project

- Project: DFG Research Group FOR 2559: Multi-Sectoral Regional Microsimulation Model (MikroSim)
- Duration: Phase I 2018-2021; Phase II 2022-2024
- Collaborating institutions: Trier University, University of Duisburg-Essen, Federal Statistical Office of Germany (Destatis)

Aims of the MikroSim Project

Overarching topics (phase I & II)

- Development of a spatial microsimulation model for Germany
- Establishment of a microsimulation infrastructure for Germany
- Development of new microsimulation methods

Fields of application in phase I

- Elementary school demand
- Family and elderly care
- Labor market integration of migrants

Fields of application in phase II

- In- and Outpatient care
- Choice of means of transport
- Modeling housing
- Educational trajectories (M)

Base Data of the MikroSim project

- Basic component: synthetic population including several person- and household-level variables (based on anonymized population register data)
- Extension and calibration of the data up to the base year 2011
  - Implementation of unit-level information from the German Microcensus (GMC)
  - Finding an optimal combination of households to align to totals from the census 2011
- Geographical allocation to real buildings and addresses
  - Geographical information from the census grid cells (1000m × 1000m and 100m × 100m)
  - Allocation of individuals and households to addresses
- Methods
  - GREG calibration
  - Latent class analysis
  - (Multinomial) logit models
  - Combinatorial optimization: simulated annealing

Structure of the MikroSim Model