

MikroSim – Sektorenübergreifendes kleinräumiges Mikrosimulationsmodell

Daten für Mikrosimulationen

Prof. Dr. Rainer Schnell

Lehrstuhl für Empirische Sozialforschung

23. November 2018



MikroSim
DFG FOR 2559



Gefördert durch



Ausgangsdatenbestand und Simulationsinfrastruktur

Spezifikation des Simulationsmodells

- Welche Prozesse sollen abgebildet werden?
- „Dynamik“ des Modells

Basisdatensatz

- Basiert auf der Zensus-Simulationsgesamtheit aus dem Zensusprojekt 2011
- Erweiterung des Ausgangsdatenbestands
 - Bevölkerungsdaten der amtlichen Statistik (z. B. Bevölkerungsvorausberechnung bzw. -fortschreibung)
 - Prozess-produzierte Daten (z. B. Geburten- und Sterbestatistik)
 - Surveys des Statistischen Bundesamts (z. B. Einkommens- und Verbrauchsstichprobe und Mikrozensus)

Ausgangsdatenbestand und Simulationsinfrastruktur

Erweiterung der Datenbasis zu Schulen

- Suche und Bewertung von Surveydaten
 - National Educational Panel Study (NEPS)
 - Programme for International Student Assessment (PISA)
 - Internationale Grundschul-Lese-Untersuchung (IGLU)/Progress in International Reading Literacy Study (PIRLS)
 - Trends in International Mathematics and Science Study (TIMSS)

Evaluation von Flüchtlingsdaten

- Identifizierung und Bewertung verfügbarer Daten
- Sehr unklare und heterogene Datenlage
 - IAB-BAMF-SOEP-Befragung von Geflüchteten
 - Ausgewählte Migrantengruppen in Deutschland (RAM)
 - IAB-SOEP Migrationsstichprobe

Integration weiterer Mikrodatenbestände in den Basisdatensatz

Evaluierung und Anpassung der verfügbaren Datenbestände

- Datensätze des Statistischen Bundesamts
 - Demographische Kernbereiche (z. B. Statistik der Eheschließungen, Wanderungsstatistik, Ausländerstatistik)
 - Gesundheit und Pflege (z. B. Fallpauschalenbezogene Krankenhausstatistik, Statistik der schwerbehinderten Menschen)
 - Asyl- und Flüchtlingsdaten (Asylanträge, Auszahlung von Leistungen)

Integration weiterer Mikrodatenbestände in den Basisdatensatz

- Sozialwissenschaftliche Querschnittsdaten:
 - Viktimisierungssurvey des BKA
 - Media-Analyse
 - European Values Study
 - World Values Survey
 - Studie zur Gesundheit Erwachsener in Deutschland (DEGS)
 - Gesundheit in Deutschland (GEDA)
 - Programme for International Student Assessment (PISA)
 - European Social Survey (ESS)
 - Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS)
 - Soll und Haben
 - Surveys des DJI, z. B. Aufwachsen in Deutschland: Alltagswelten (AID:A)
- Anpassung der ggf. abweichenden Variablen und Variablenausprägungen

Integration weiterer Mikrodatenbestände in den Basisdatensatz

Modellierung auf Basis der Erhebungsdaten

- Modellierung der zu identifizierenden zentralen Prozesse des Simulationsmodells durch statistische Modelle (z. B. Regressionsverfahren, Klassifikationsbäume)
- Kreuzvalidierung zwischen Datensätzen mit den gleichen Variablen
- Sensitivitätsanalysen der geschätzten Modelle

Dokumentation

- Dokumentation der Datenstruktur sowie Schätz- und Prädiktionsmethoden
- Nutzung von MediaWiki
 - Versionierung und Bearbeitungshistorie
 - Intuitive Bedienung und damit erleichterte Kollaboration
 - Nutzungsfreundlicher als Datenbankalternativen

Modellierung der Zustandsänderungen

Identifikation geeigneter Datensätze zur Schätzung der Übergangswahrscheinlichkeiten

- Amtliche Daten (z. B. Mikrozensus-Panel)
- Sozialwissenschaftliche Panelsurveys:
 - Generations and Gender Survey (GGS)
 - NAKO Gesundheitsstudie
 - Sozio-oekonomisches Panel (SOEP)
 - Survey of Health, Ageing and Retirement in Europe (SHARE)
 - Panel Arbeitsmarkt und soziale Sicherung (PASS)
 - Panel Analysis of Intimate Relationships and Family Dynamics (pairfam)
 - Deutscher Alterssurvey (DEAS)
 - Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB)

Modellierung der Zustandsänderungen

- Dokumentation in MediaWiki
 - Datenquellen
 - Variablen
 - Recodierungsanweisungen

Pooling der Surveyschätzer

- Kombination verschiedener Schätzungen inkl. notwendiger Adjustierung auf Populationsrandverteilungen
- Evaluation verschiedener statistischer Pooling-Ansätze
 - Meta-Analyse
 - Zusammenfassung der Datensätze zu einem gemeinsamen Datensatz
 - Bayesianische Mehrebenenregression mit Poststratifizierung

Modellierung der Zustandsänderungen

Externe Schocks

- Plötzliche Ereignisse wie z. B. Klima- und Technikkatastrophen, Flüchtlingsbewegungen
- Recherche und Überlegungen zur experimentellen Modellierung (Bundesamt für Bevölkerungsschutz und Katastrophenhilfe)
- Zunächst: Implementierung entsprechender Routinen für die räumliche Verteilung und Modifikation der Übergangswahrscheinlichkeiten

Schlussbetrachtung

- Kombination von 36 sozialwissenschaftlichen Datensätzen und amtlicher Statistik
- Aufbau und Erweiterung des Basisdatensatzes
- Schätzung der Übergangswahrscheinlichkeiten mittels Längsschnittdatensätzen
- Davon ausgehend theoriegeleitete Modellierung der inhaltlichen Szenarien (Grundschulbedarf, Pflege, Migration, externe Schocks)

Literatur

- Ballas, D., Rossiter, D., Thomas, B., Clarke, G. & Dorling, D. (2005). *Geography Matters: Simulating the Local Impacts of National Social Policies*. York: Joseph Rowntree Foundation.
- de Kramer, R. M. (2016). *Evaluation of Cross-Survey Research Methods for the Estimation of Low-Incidence Populations*. (Diss., Boston College, Lynch School of Education).
- Lovelace, R. & Dumont, M. (2016). *Spatial Microsimulation with R*. Boca Raton: CRC Press.
- O'Donoghue, C. (Hrsg.). (2014). *Handbook of Microsimulation Modelling*. Bingley: Emerald.
- Rahman, A. & Harding, A. (2017). *Small Area Estimation and Microsimulation Modeling*. Boca Raton: CRC Press.