Unit 2: Hands-on microsimulation techniques

Hariolf Merkle

Content

- Introduction
  - Goals of unit 2
  - Data Access
  - Toolbox
  - Static and dynamic aging from practitioners' view
- Exercise with R- static microsimulation
- Exercise with R- dynamic microsimulation
- Revisiting MikroSim

Unit 2: Hands-on microsimulation techniques

- ✓Goals of unit 2
- ✓Data Access
- ✓Static and dynamic ageing from practitioners' view
- ✓Toolbox

What is the goal of this unit?

- Apply the theory of unit 1 to get even more in touch with microsimulation.
- Provide a toolbox for microsimulation methods.
- Improve the understanding of conducting microsimulation studies in R.
- Face and solve real world microsimulation issues.
- We focus on static and dynamic (discrete) ageing.
- We focus on dynamic discrete ageing since dynamic continuous ageing is a highly computer intensive task.
- We focus on the application of statistical microsimulation methods without dealing with the underlying theory in detail (unfortunately) due to time/scope constraints.

Preparation Steps

- Definition of the topic to be analyzed.
- Accessing useful input (micro-)data and other data sources.
- Choose an appropriate modeling strategy.
- Define Target and constraint variables.
- Loading input data.
- Subsetting to remove excess information.
- Re-categorising individual level variables: develops the R code to have pertinent categories in all variables (microdata and external sources/population totals).
- Matching individual and aggregate level data names: makes the data names correspond to avoid problems when executing the calibration.
- Perform microsimulation steps (uprating, reweighting, run set of modules).
- Validate your model based on your results.

(based on Lovelace & Dumont (2016), Ch. 4 and O'Donoghue & Loughrey (2014))

Where can I get suitable data?

- A crucial task in research projects is often the accessibility to suitable data.
- National statistical institutes provide many and relevant datasets to the scientific community:
  - The Research Data Centre (RDC) of the Federal Statistical Office and the Research Data Centre of the Statistical Offices of the Federal States: https://www.forschungsdatenzentrum.de/en
  - Data Access via Eurostat (https://ec.europa.eu/eurostat/web/microdata), e.g. European Community Household Panel, European Union Labour Force Survey, European Union Statistics on Income and Living Conditions (EU-SILC). EUROMOD is a EU-SILC-tailored static tax-benefit microsimulation model (https://euromod-web.jrc.ec.europa.eu/).
- Research Institutes operate RDCs, e.g. ifo (Economics & Business Data Center), IAB, Bundesbank or DIW.
- Usually, an application for the a specific research topic is required. Afterwards, a data usage agreement has to be signed including the chosen the way of data access (onsite or offsite).
- Different ways of data access: take into account the required software you want to use - usually, RDCs do only allow their preinstalled specific set of applications/software for onsite usage.
- Population totals/constraints can often be derived from publicly available data outside of a rdc, e.g. GENESIS database (https://www-genesis.destatis.de/genesis/online?operation=sprachwechsel&language=en).

Static Ageing Techniques in a Nutshell 1/3

- Evaluation of direct distributional impact(s) on micro-units by **reweighting** the micro units using exogenously given aggregate (control)
- variables.
- Recall unit 1: a microunit in a survey represents a certain number of other microunits in the population according to its weight.
- Example:

Microdata (Survey)

ID	age	married	...	weight
1	32	0	...	78.1
2	51	1	...	105.3
3	44	1	...	89.6
...	...	...	...	...
n	19	0	...	125.1

Microdata (Survey)

	married	not married
age 0-40	12 %	20 %
age 41+	33 %	35 %

Known Population Totals

	married	not married
age 0-40	15 %	19 %
age 41+	29 %	37 %

- The table calculated from survey data is not consistent with known totals, e.g. from administrative data or census data.

  –> calibration should be taken into account.
- If the totals for calibration are deduced from scenarios, e.g. future projections, we call the calibration process reweighting in the
- microsimulation context.
- Suitable methodology: Merz`s “Adjust”, raking/IPF, GREG calibration (cf. Dekkers (2015) for an overwiew of implementations).

Static Ageing Techniques in a Nutshell 2/3

- Weight adjustment (–> calibration) involves three steps (cf. Haziza and Beaumont, 2017):
  - 1) Obtain the design weights to account for the random sample selection. This is obtained by dividing one by the probability of a unit entering the sample e.g., a simple 1% random sample would lead to a weight of 100 for each unit.
    2) Adjust the design weights to compensate for unit non-response.
    3) Adjust the resulting weights such that the estimates agree with specific known totals in the population.
  - Recall: For microsimulations, micro units are reweighted using exogenously given aggregate (control) data
  - of another time period. Therefore, we extend the list above (usually data producers' duty) by:

    4) Adjust the provided weights such that the estimates agree with assumed/predicted totals in the future population.

    So we redo 3) with additional population totals or scenario-based totals.

How do we apply reweighting in case of population data for which weights are not provided?

–> Assign to each micro unit the weight "1" and proceed with 4).

(for more details on reweighting cf. Lovelace & Dumont (2016), Ch. 5-6)

Static Ageing Techniques in a Nutshell 3/3

- Evaluation of direct distributional impact(s) on micro-units by **uprating/deflating** certain variables:
  - Usually applied to monetary variables.
  - Goal: assure consistency with available projections or scenarios.
  - Example: the nominal mean labour income will increase by 17.8% until 2030 and unemployment benefits are 600 EUR per person.

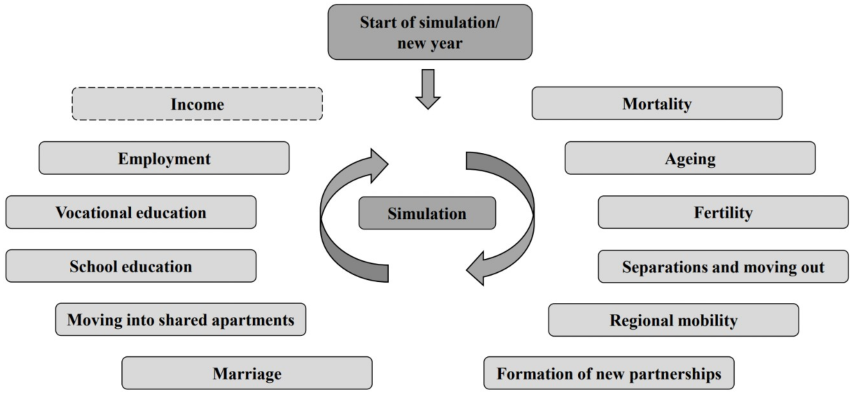
    –> Increase labour income by 17.8% and adjust income from unemployment benefits to 600 EUR per person.
    - Uprating or deflation can also be implemented using a regression model.
  - Important: order of uprating and reweighting. Applying both can be necessary.

Dynamic Ageing in a Nutshell

- Simulation of micro-unit behavior over a temporal horizon.

–> Update the micro unit data for each simulated period respectively (discrete model).

- Recall the module sequence of the MikroSim model from unit 1:



- How do the respective modules alter the data?
  - By a model, e.g. the transition propabilities from employment to unemployment (and vice versa) can be calculated using a simple logit model.
  - Adjustment to known totals can be implementend by adjusting the constant in a logit model.
  - By deductive logic, e.g. the activity status of employees reaching a certain age is automatically changed to “retired” or each year ever individual is one year older.
- Data consistency is crucial , defining editing rules can be necessary, e.g. if an age threshold implies “retired“ which might lead to a limited or zero labour income which might imply retirement benefits etc. ...

Dynamic Ageing Compared to Static Ageing

- Dynamic ageing vs static ageing (cf. O'Donoghue & Loureigh (2014), Dekkers (2015)):
  - Dynamic ageing produces intermediate results which allows the analysis of biographies or life cycles.
  - Dynamic ageing is suitable for multidimensional scenarios, e.g. a scenario incorporating changing retirement legislation with an increased retirement age of 70 years. No individuals with such characteristics might be in available in the dataset of a static microsimulation whereas dynamic ageing would create such individuals.
  - The broader range of possibilities for dynamic ageing comes at a much higher costs regarding programming, data processing (computation time) and data storage.
  - Dynamic continuous ageing is even more costly: with dynamic discrete ageing, all individuals are updated each period at once whereas with dynamic continuous ageing, all units are updated one by one. This is especially computationally intensive, if the units interact with each other, e.g. persons in a household, since the whole computation has to be halted to update only some units.

A Hint on Spatial Microsimulation

- Problem: we need microdata on a regional level that is not available to us or not available at all.
- Solution: create a synthetic data set with that regional level.
- Possible Solution: use survey data and reweigh the units iteratively for each region with respect to known regional totals.

Microdata (Survey)

ID	X	Y
1	A	0
2	B	1
3	C	1
...	...	...
235	A	0

Known Regional Totals

Region	n	Y=1
Region1	532	82
Region2	519	52
Region3	294	24
Region4	152	63
Region5	93	22
Total	1590	243

(Synthetic) Spatial Population

ID_new	ID_old	X	Y	region
1	235	A	0	region1
...	...	...	...	region1
532	2	B	1	region1
533	3	C	1	region2
...	...	...	...	...
1590	1	A	0	region5

- Include each unit in the survey data according to its respective weight into the synthetic population data.
- Here, the number of individuals in each region in the synthetic population should match the number of individuals from the known regionals totals, i.e. 532 individuals in region 1 with 82 individuals with Y=1.
- Further (optional) reading about spatial microsimulation can be found in Lovelace & Dumont (2016).
- Underlying assumptions: enough totals for each region are available to mimic regional distributions appropriately.

R-Toolbox for Microsimulation

The development of R packages is highly dynamic- no claim to completeness.

Reweighting (Static Ageing):

- `ipfp` or `mipfp`: can be used to calculate weights using iterative proportional fitting and other approaches.
- `survey`: a package for a wide range of survey methodology, several (re-)weighting procedures are implemented.
- `Ipfr`: IPU-algorithm for multilevel reweighting, e.g. household weights that also account for person-level characteristics.
- `simPop`: provides methods and frameworks for synthetic data generation, has also implemented the IPU-algorithm.
- `sampling`: another package for survey methodology, several (re-)weighting procedures are implemented.

Dynamic Ageing

- `MicSim`: entry-level toolkit allows performing continuous-time microsimulation for a wide range of life science (demography, social sciences, epidemiology).
- `microsimulation`: discrete event simulation using R and C++.

In General: you have to write code a lot yourself using methods provided by packages outside the microsimulation context!

Literature

- Dekkers, G. (2015). The simulation properties of microsimulation models with static and dynamic ageing–a brief guide into choosing one type of model over the other. International Journal of Microsimulation, 8(1), 97-109.
- Dumont, M., & Lovelace, R. (2016). Spatial Microsimulation with R. CRC Press.
- Haziza, D., & Beaumont, J. F. (2017). Construction of weights in surveys: A review. Statistical Science, 32(2), 206-226.
- O'Donoghue, C., & Loughrey, J. (2014). Nowcasting in microsimulation models: a methodological survey. Journal of Artificial Societies and Social Simulation, 17(4), 12.

Unit 2: Hands-on microsimulation techniques

- ✓ Exercise with R - static microsimulation
  - ✓ Exercise 1
  - ✓ Exercise 2
  - ✓ Exercise 3

Exercise 1

For this exercise, we use the R-Package `simPop` and the dataset `eusilcP` thereof. We also use the concept of the at-risk-of-poverty rate:

“The at-risk-of-poverty rate is the share of people with an equivalised disposable income (after social transfer) below the at-risk-of-poverty threshold, which is set at 60 % of the national median equivalised disposable income after social transfers. This indicator does not measure wealth or poverty, but low income in comparison to other residents in that country, which does not necessarily imply a low standard of living. “

(https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:At-risk-of-poverty_rate)

“The equivalised disposable income is the total income of a household, after tax and other deductions, that is available for spending or saving, divided by the number of household members converted into equalised adults; household members are equalised or made equivalent by weighting each according to their age, using the so-called modified OECD equivalence scale.“

(https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Equivalised_disposable_income)

a) Make yourself familiar with the dataset, i.e. get an overview of the variables.

b) Now, we want to analyse the at-risk-of-poverty rate (`arpr()` from `laeken`-Package) based on the variable `eqIncome` as target variable. Display your results for the whole population and on a regional level (federal states of Austria, variable `region`).

Exercise 2

A public discussion about inequality leads you to the idea of comparing the impact of different policies. Proceed based on exercise 1.

What if the government ...

a) ... increases the equivalised (household) disposable income (`eqIncome`) by 100 EUR per child (age < 14)? Recalculate the at-risk-of-poverty rate and compare it to the results of 1b).

b) ... is able to allocate a fixed amount of 3,725,700 EUR overall to retired individuals (`ecoStat == 5`). Test an allocation of your choice regarding household incomes (`eqIncome * eqsize`) and recalculate the at-risk-of-poverty rate based on `eqIncome`. Compare your results to 1b) and 2a).

c) Do we apply a cohort or population model?

d) What is the applied static ageing technique in this exercise called?

Exercise 3

You want to extend your analysis by taking population projections into account. It is projected that the share of part-time workers in 2028 will increase by 20 percent whereas the number of retired individuals will increase by 18 percent.

a) Apply static ageing using at least two different calibration methods, i.e. calculate new weights for `eusilcP`.

b) Calculate again the at-risk-of-poverty rate and compare the results to 1b). What developments can be observed?

c) Select one allocation from exercise 2b) and calculate the relative change of the at-risk-of-poverty rate in 2028.

d) What is the applied static ageing technique in this exercise called?

Unit 2: Hands-on microsimulation techniques

- ✓ Exercise with R - dynamic microsimulation
  - ✓ Exercise 4

Exercise 4

Please use again the dataset eusilcP provided by the R-package `simPop`. In this exercise, we set up a small dynamic microsimulation considering only the variables `age`, `gender`, `hid` and `ecoStat` in our dataset. Simulate 10 periods (years).

a) Create the following modules:

- death: according to this logit-model:

$$P(Y_i = 1) = \frac{1}{1 + \exp(6 - 0.03 \times age)}$$

- age: each year +1
- birth: according to this logit model but only for women between 18 and 40 :

$$P(Y_i = 1) = \frac{1}{1 + \exp(1.6 + 0.03 \times age)}$$

Exercise 4

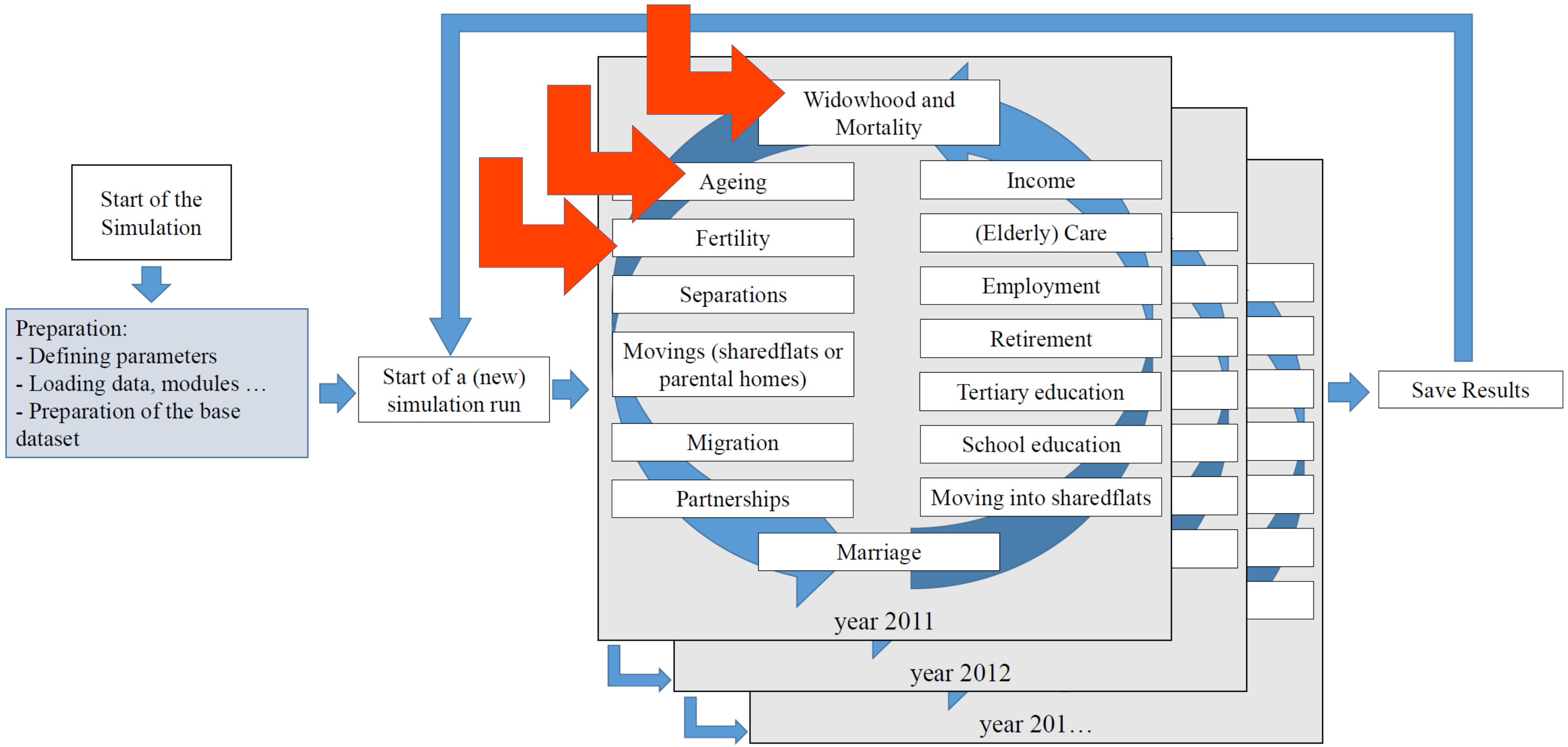
b) Run your microsimulation and calculate the average age and number of individuals of each period.

c) Take a look at the conditional distribution of `ecoStat` regarding `age`. What can you notice in particular?

Unit 2: MikroSim revisited

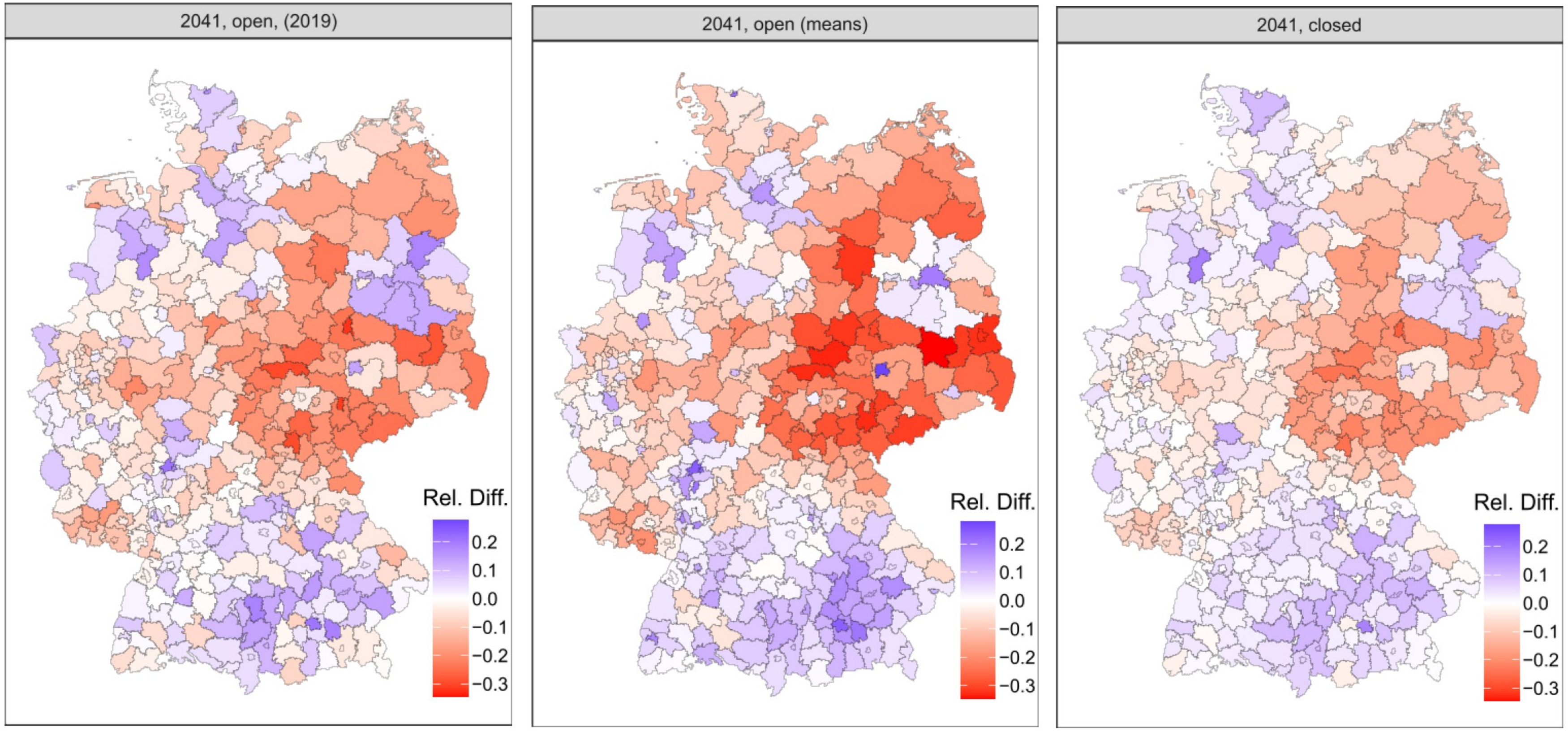
- ✓ The MikroSim project as an example of dynamic microsimulation
  - ✓ MikroSim applied

Structure of the MikroSim Model



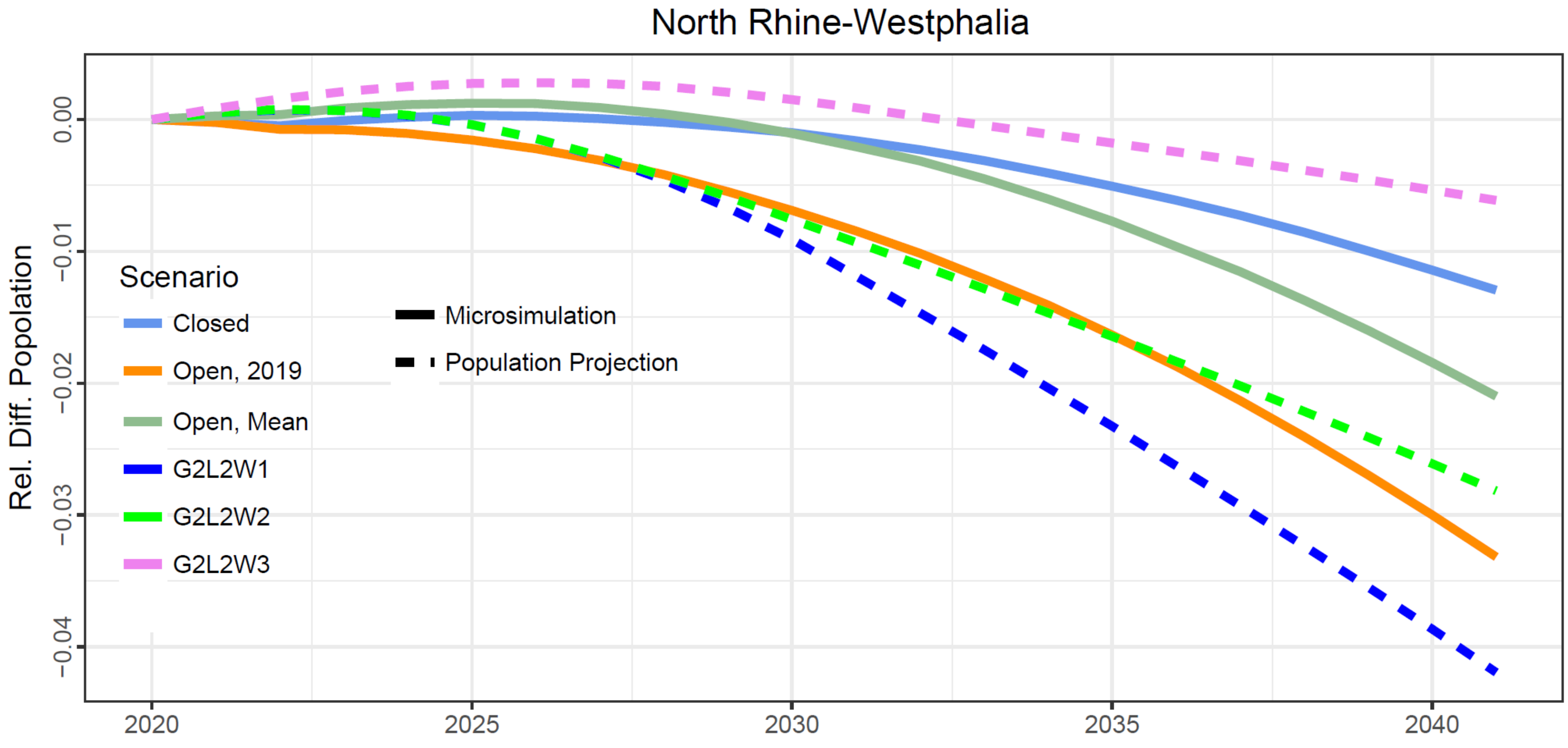
In our exercise, we created these modules accordingly.

Relative population difference 2041



MikroSim results of three different migration scenarios.

Relative Population Difference (2020): North Rhine-Westphalia



MikroSim results of differnt migration scenarios.

These slides were created as part of IPSDS.
IPSDS is sponsored by the German
Federal Ministry of Education and Research (BMBF) within the framework of the program „Aufstieg durch Bildung: offene Hochschulen“

Uploaded as part of the BERD Academy offers within BERD@NFDI,
funded by the DFG (460037581).