

Tobias Raabe

ECONOMIST · DATA SCIENTIST

✉ raabe@posteo.de | 🏠 tobiasraabe.github.io | 📄 tobiasraabe | 📄 tobiasraabe | 📄 tobiasraabe | 📄 Tobias Raabe

Education

University of Bonn

Bonn, Germany

M.Sc. IN ECONOMICS

Oct. 2015 - Nov. 2019

- The thesis "A unified estimation framework for some discrete choice dynamic programming models" develops a general framework to represent a certain class of economic models in Python which is called respy.

Skills

- Programming** Python, R, SQL, Julia, C, \LaTeX
- Programs** Stata, Octave, Matlab, Microsoft Office
- Languages** German (native speaker), English (fluent)

Experience

quantilope

Hamburg, Germany

DATA SCIENTIST

2020-

- Developing methods and tools to analyze data in marketing with a strong focus on automation.

Institute for Applied Microeconomics (IAME)

Bonn, Germany

RESEARCH ASSISTANT

2019-2020

- Maintained respy by redesigning the documentation, organizing contributions and helping contributors.
- Designed the computational backend and interface for gettsim, a simulator for the German tax and transfer system.
- Established software engineering best practices in several research groups. Best practices for designing applications, continuous integration (with Travis-CI, Appveyor, Azure Pipelines or Github Actions), static code analysis, and maintenance, release and code review manuals.

Institute of Labor Economics (IZA)

Bonn, Germany

RESEARCH ASSISTANT

2017-2019

- Co-authored the report "Early Identification of College Dropouts Using Machine-Learning" with I. Isphording sponsored by the German Ministry of Education and Research.

Institute for Applied Microeconomics (IAME)

Bonn, Germany

RESEARCH ASSISTANT

2015-2017

- For "What motivates effort? Evidence and expert forecasts" by Della Vigna and Pope (2018), I identified super-forecasters, researchers who predicted the outcomes of several behavioral experiments best, with machine learning methods.
- Some contributions to "The approximate solution of finite-horizon discrete-choice dynamic programming models" by Eisenhauer (2018).
- Developed ovmm, a tool to manage server instances for running behavioral experiments with oTree.

Projects

respy

CO-AUTHOR

2019-

- <https://github.com/OpenSourceEconomics/respy> and <https://respy.readthedocs.io>.
- respy is a framework for models to study the human capital accumulation process, labor market choices and outcomes of individuals.
- At the heart of respy are configuration files which enable researchers to compose models with various features.
- The model parameter are estimated with maximum likelihood or method of simulated moments and numerical optimization algorithms.
- Due to computational complexity the models are often programmed in Fortran. respy is developed in Python which has many advantages (accessibility, flexibility) and uses Numba to achieve Fortran- or C-like performance where it is necessary.
- respy leverages Dask to scale from single machines to HPC clusters.
- The package is used in publications and Master's theses.

sid

CO-AUTHOR

2020-

- <https://github.com/covid-19-impact-lab/sid> and <https://sid.readthedocs.io>.
- sid is a simulator for the spread of infectious diseases which combines features of the standard susceptible-exposed-infected-recovered (SEIR) model with an agent-based model.
- It allows to model explicit contacts between individuals based on social structures and mixing patterns (e.g. households, work places or assortative matching based on age, gender, and geography).
- Because of that, researchers are able to assess the impact of very targeted policies like school closures, county-level lockdowns, etc.

pytask

AUTHOR

2020-

- <https://github.com/pytask-dev/pytask> and <https://pytask-dev.readthedocs.io>.
- pytask is a build system based on the plugin framework pluggy which is used in pytest.
- Its purpose is to run workflows in research or data science projects from data preparation over analyses to the final reports.
- Its syntax is inspired by pytest and, thus, easy-to-learn for users who are not programmers by nature.

gettsim

SOFTWARE ENGINEER

2019-2020

- <https://github.com/iza-institute-of-labor-economics/gettsim> and <https://gettsim.readthedocs.io>.
- gettsim allows researchers and students to study the German tax and transfer system by analyzing the status-quo or the effect of reforms.
- I designed the computational backend of gettsim. The tax and transfer system consists of functions which compute quantities (e.g., child benefits) and which can request the results of other functions. The relationships between functions build a computational graph or DAG. The graph can be executed with pandas or Dask.

estimagic

CORE CONTRIBUTOR

2019-

- <https://github.com/OpenSourceEconomics/estimagic> and <https://estimagic.readthedocs.io>.
- estimagic is a package to for nonlinear optimization, numerical differentiation and statistical inference.
- I am involved in the design of the package and I implemented several optimization algorithms.

Publications

Early Identification of College Dropouts Using Machine-Learning: Conceptual Considerations and an Empirical Example

IZA Research Report No. 89

WITH INGO ISPHORDING

2019

- Can and should we use modern big data methods to predict which students will drop out of their studies? Such an early detection of at-risk groups of students would allow practitioners and policy-makers to timely intervene through targeted counter-measures as for instance mentoring, counseling, or institutional support. This report discusses the feasibility, benefits and potential hidden costs of such an approach which would allow to mitigate substantial costs of educational career frictions.
- The report uses a representative sample of students at German universities (NEPS). The data comprises socio-economics, demographic and psychological measures.
- The paper discusses the impact of various definitions of university dropout, metrics for classification problems and class imbalance, multiple machine learning models using scikit-learn and Keras, and cost-benefit discussions.