



Econometrics IV: Machine Learning

Course instructor:

Prof. Dr. Helmut Farbmacher
Room: 2412 (TUM School of Management, Arcisstr. 21)
+49 (0)89 289 28566
farbmacher@tum.de
Office hours: by arrangement

Timetable: 1st part in person: August 1, 2022 (9am to 5pm, room 2544)
2nd part in person: September 13-14, 2022 (9am to 5pm, room 2544)

Prerequisites: Ideally, Econometrics I, II and III (TUM) or equivalently solid introductory courses in econometrics. Preferably some basic knowledge of R and Python. Participants should bring their own laptop with R and/or Python installed. The target audience are PhD students.

Grading: Successful participation; details below

Registration: Until July 31, 2022, via Moodle.

Course description:

The course is part of a series of econometrics courses at TUM School of Management that also comprises “Econometrics I: Research Design and Estimation Methods” by Prof. Dr. Hanna Hottenrott, “Econometrics II: Causal Inference” by Prof. Dr. Joachim Henkel and “Econometrics III: Advanced Econometrics” by me. Econometrics IV will be a block lecture but conceptualized as a seminar based on student presentations. The course covers a selection of state-of-the-art methods in econometrics and machine learning. It aims to provide students with a sound understanding of the methods discussed, such that they are able to do research using modern econometric techniques, as well as critically assess existing studies.

In particular, the course will likely cover the following topics:

- Regression Shrinkage Methods (Ridge, Lasso, Elastic Net)
- Decision Trees, Random/Causal Forests
- Advanced Identification Strategies (e.g., Double Machine Learning)
- Introduction to Neural Networks

1st part of the course (in person):

In the first meeting, we will briefly discuss the econometric methods (including some applications to illustrate them). In this meeting I will also assign a (replication) project to each student, which (s)he will present at the second part of the course. Students will then apply these methods and will replicate recent research papers in economics. They are supposed to work on the replication project before the second part of the course begins.

2nd part of the course (in person):

The second part of the course will be similar to a reading course in which we will discuss the (replication) projects. All participants are expected to read the papers before the meetings. The presentation (roughly 30 minutes) together with a short report that summarizes the assigned paper (roughly 5 pages w/o figures, tables and references) will be relevant for the grading.

Recommended textbooks:

- Goodfellow Ian, Bengio Yoshua and Courville Aaron. *Deep Learning*, MIT Press, [available here](#)
- Bishop Christopher. *Pattern Recognition and Machine Learning*, Springer, [available here](#)
- Hansen Bruce. *Econometrics*, [available here](#)
- Hastie Trevor, Tibshirani Robert and Friedman Jerome. *The Elements of Statistical Learning*, Springer, [available here](#)
- James Gareth, Witten Daniela, Hastie Trevor and Tibshirani Robert. *An Introduction to Statistical Learning with Applications in R*, Springer, [available here](#)

Papers you definitely should read (potential project applications):

- Angrist and Frandsen (2022): Machine Labor, *Journal of Labor Economics*, 40(S1), S97–S140.
- Bach *et al.* (2022): DoubleML - An Object-Oriented Implementation of Double Machine Learning in Python, *Journal of Machine Learning Research* 23(53), 1-6.
- Borgschulte and Vogler (2020): Did the ACA Medicaid Expansion Save Lives?, *Journal of Health Economics*, 72, 102333.
- Brot-Goldberg *et al.* (2017): What does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics, *Quarterly Journal of Economics*, 132(3), 1261–1318.
- Buchner, Wasem and Schillo (2017): Regression Trees Identify Relevant Interactions: Can this Improve the Predictive Performance of Risk Adjustment?, *Health Economics*, 26, 74–85.
- Everding and Marcus (2020): The Effect of Unemployment on the Smoking Behavior of Couples, *Health Economics*, 154–170.
- Farbmacher, Guber, Klaassen (2021): Instrument Validity Tests with Causal Forests, *Journal of Business and Economic Statistics*, forthcoming.
- Farbmacher, Löw, Spindler (2021): An Explainable Attention Network for Fraud Detection in Claims Management, *Journal of Econometrics*, forthcoming.
- McGuire, Zink and Rose (2021): Improving the Performance of Risk Adjustment Systems, *American Journal of Health Economics*, 7(4).
- Rose (2016): A Machine Learning Framework for Plan Payment Risk Adjustment, *Health Services Research*, 51(6), 2358–2374.
- Rose, Bergquist and Layton (2017): Computational Health Economics for Identification of Unprofitable Health Care Enrollees, *Biostatistics*, 18(4), 682–694.
- Tibshirani (1996): Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58(1), 267–288.
- Windmeijer, Farbmacher, Davies, Davey Smith (2019): On the Use of the Lasso for Instrumental Variables Estimation with Some Invalid Instruments, *Journal of the American Statistical Association*, 114(527), 1339–1350.
- Zou (2006): The Adaptive Lasso and Its Oracle Properties, *Journal of the American Statistical Association* 101(476), 1418-1429.

Papers you could read if you have plenty of time:

- Athey and Imbens (2019): Machine Learning Methods Economists Should Know About, *Annual Review of Economics*, 11, 685–725.
- Chernozhukov *et al.* (2018): Double/debiased Machine Learning for Treatment and Structural Parameters, *Econometrics Journal*, 21, C1–C68.



- Fu (1998): Penalized Regressions: the Bridge versus the Lasso, *Journal of Computational and Graphical Statistics*, 7, 397–416.
- Giannone, Lenza and Primiceri (2021): Economic Predictions with Big Data: The Illusion of Sparsity, *Econometrica*, 89(5), 2409–2437.
- Mullainathan and Spiess (2017): Machine Learning: An Applied Econometric Approach, *Journal of Economic Perspectives*, 31(2), 87–106.
- Su, Shi and Phillips (2016): Identifying Latent Structures in Panel Data, *Econometrica*, 84(6), 2215–2264.
- Varian (2014): Big Data: New Tricks for Econometrics, *Journal of Economic Perspectives*, 28(2), 3–28.
- Wüthrich and Zhu (2021): Omitted Variable Bias of Lasso-based Inference Methods: A Finite Sample Analysis, *Review of Economics and Statistics*, forthcoming.