GESIS Spring Seminar 2024

Syllabus for course:

"Causal Machine Learning for Cross-sectional and Panel Data"

Lecturers:	Prof. Dr. Martin Spindler	Prof. Dr. Jannis Kück
Affiliation:	Universität Hamburg	Heinrich-Heine-Universität Düsseldorf / DICE
Email:	Martin.spindler@uni-hamburg.de	kueck@dice.hhu.de
Website:	https://www.bwl.uni-hamburg.de/statistik/team/spindler.html	https://www.dice.hhu.de/das-dice/team/professorinnen-und-professoren-1/kueck

Date: 11-15 March 2024
Time: Mo: 10:00-17:00 | Tu-Fr: 09:00-16:00
Venue: GESIS Cologne

About the Lecturers

Prof. Dr. Jannis Kück is Professor for Data Science in Economics at the Heinrich-Heine-Universität Düsseldorf. He holds an PhD in Statistics / Econometrics from the University of Hamburg. His fields of interest are high-dimensional Statistics, Machine Learning and in particular Causal Machine Learning. He has published in leading journals in his field, including Biometrika, Journal of Econometrics, and Journal of Business and Economic Statistics.

Prof. Dr. Martin Spindler is Professor for Statistics at the University of Hamburg. His fields of interest are Statistics and Econometrics and in particular Causal Machine Learning. He has published, amongst others, in Biometrika, Journal of Econometrics, and Journal of Machine Learning Research.

Course Description

Participants of this course will learn and apply recent Causal Machine Learning methods to analyze effects in either cross-sectional or panel data. Causal Machine Learning combines the field machine learning, which was developed for predictions and is based on correlation, and the field of causal inference. In this course we will focus on the so-called Double Machine Learning approach (DML) which allows for valid inference in high-dimensional settings.

This course will focus on tools that are easy to implement for practitioners in the R / Python and covers three blocks:

1. Basics of causal inference
2. Basics of machine learning
3. Double Machine Learning (DML) for cross-sectional and panel data (including difference-in-differences, instrumental variables and mediation analysis)

The final day will also give the opportunity to discuss own work and applications of the participants (if requested).

The course will be based on three pillars to teach the new methods: (i) lecture-based introduction of the theoretical concepts, (ii) getting to know the methods with hands-on examples / notebooks provided by the lecturer, (iii) supervised application to provided or own datasets.

Keywords

Causal machine learning, double machine learning, conditional average treatment effects

Course Prerequisites

- Basic understanding of probability theory (conditional expectations) and regression analysis (OLS)
- Basic understanding of causal research designs, in particular randomized experiments and observational designs that control for confounding factors
- Basic experience with the software R or Python
- (Not required, but an advantage:) Basic understanding of Machine Learning methods, in particular shrinkage methods (e.g., Lasso, Ridge) and tree-based methods (regression trees, random forest)

Target Group

Participants will find the course useful if:

- They are familiar with the basics of causal inference and regression analysis and are curious how machine learning methods could enter their empirical toolbox
- They work with experimental and observational data in social science or related fields.
- They want to learn and understand the new field of Causal Machine Learning, in particular Double Machine Learning

Course and Learning Objectives

By the end of the course participants will:

- Understand popular methods that are likely to appear in future studies they consume.
- Know in which settings and for which research questions the current state of Causal Machine Learning provides attractive alternatives to standard tools.
- Be able to apply Causal Machine Learning in basic settings.
- Have the background knowledge to learn about Causal Machine Learning methods for more complex settings that are not covered in the course.
- Understanding the Double Machine Learning approach

Organizational Structure of the Course

- During lab time, participants will apply the methods that were introduced in the morning session to synthetic and real datasets. A suggestive workflow for the analysis will be provided by the lecturer. Participants are encouraged to bring their own datasets if they come from a research design that is covered in this course.
- Lecturer will support work on the datasets and is available for questions. Further, he is available for individual consultations on participants' projects.

Software and Hardware Requirements

Course participants will need to bring a laptop with the latest versions of R (https://cran.r-project.org/) and RStudio (https://www.rstudio.com/) as well as Python installed. All three programs are freely available for download and use. Please install the DoubleML packages for both R and Python: https://docs.doubleml.org/stable/index.html

Participants will need to be able to download files from the internet (free Wifi is provided by GESIS) and have the rights to install packages on their laptops during the course.

Recommended Literature to Look at in Advance

Bach, P., Chernozhukov, V., Kurz, M. S., and Spindler, M. (2022), DoubleML - An Object-Oriented Implementation of Double Machine Learning in Python, Journal of Machine Learning Research, 23(53): 1-6, https://www.jmlr.org/papers/v23/21-0862.html.

Bach, P., Chernozhukov, V., Kurz, M. S., and Spindler, M. (2021), DoubleML - An Object-Oriented Implementation of Double Machine Learning in R, arXiv:2103.09603.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. and Robins, J. (2018), Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, 21: C1-C68, doi:10.1111/ectj.12097.

Course contents for certificates

Causal Machine Learning, Double Machine Learning, Cross-sectional and Panel Data

Day-to-day Schedule and Literature

Day	Topic(s)
1	**Introduction to Causal Machine Learning, Recap Causal Inference in Experimental Data, Recap Machine Learning**
	Literature: Chernozhukov et al. (2023+): Applied Causal Inference Powered by ML and AI. Book draft. Work in Progress.
2	**Statistical Inference on Predictive and Causal Effects in High-Dimensional Linear Regression Model and in Modern Nonlinear Regression Models**
	Literature: Chernozhukov et al. (2023+): Applied Causal Inference Powered by ML and AI. Book draft. Work in Progress.
3	**Instrumental Variables Estimation, Difference-in-Difference and Regression Discontinuity with Double ML**
	Chernozhukov et al. (2023+): Applied Causal Inference Powered by ML and AI. Book draft. Work in Progress.
4	**Heterogeneous Treatment Effects and Sensitivity Analysis**
	Literature: Chernozhukov et al. (2023+): Applied Causal Inference Powered by ML and AI. Book draft. Work in Progress.
5	**Panel Data, Case Studies / Presentation of own work**
	Literature: Chernozhukov et al. (2023+): Applied Causal Inference Powered by ML and AI. Book draft. Work in Progress.