IMPRS
INTERNATIONAL MAX PLANCK RESEARCH SCHOOL
population. health. data science.

MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH

IDEM 185

Advanced Methods for Causal Inference

Course coordinator: Peter Eibich

Start date: 11 July 2022
End date: 15 July 2022

Location: Online Course. Link tba.

Instructors

- Peter Eibich
- Angelo Lorenti

Course description

Many research questions in demography, social sciences, economics, or epidemiology deal implicitly or explicitly with causal effects or causal mechanisms. For example:

- How does education affect the timing of fertility?
- How does health affect labor market participation?
- What is the impact of overweight and obesity on healthcare utilization?
- How does language proficiency affect the labour market outcomes of migrants?
- What is the impact of early life events on late-life mortality?

Unfortunately, conducting randomized experiments to study these questions is often infeasible, and we therefore rely on research designs and statistical and econometric methods to identify causal effects under certain assumptions.

In this course, we will review selected research designs and estimation methods that can be used to obtain causal estimates. Over the course of four days, we will review the theory behind these estimation methods and implement them in a series of practical exercises using real-world data.

Organization

The course will be offered online. On Monday through Thursday, there will be one lecture of about 90 minutes and one two-hour practical exercise / computer lab session. All lectures will be pre-recorded and made available on Thursday, July 7, four days before the course. Students are expected to watch the lecture carefully before the practical exercise / computer

lab session, which will be held live every day from 15:30-17:30 CEST (Central European Summer Time). During the daily practical exercise /computer lab sessions, students will gain hands-on experience. In each session, students and instructors will review code together, perform hands-on exercises, and evaluate these exercises together. This will allow students to develop intuition in using the concepts and deepen their understanding of the concepts.

On Friday, students who wish to obtain a certificate for participation in the course can take a short test (see below, "Examination"). We will also offer a limited number of virtual office hours with the instructors to provide students an opportunity to discuss their research projects with the course instructors.

Throughout the course week, an open forum (message board) will be available to promote and encourage dialogue between students and instructors. This forum will display questions and answers about the topics taught to facilitate the exchange of open questions and to raise and resolve any doubts.

In general, students should expect to spend about 6-8 hours per day on the course (lectures, practical exercise sessions, readings, presentations).

Detailed schedule

The course will cover the following topics in this order.

Monday

1. Advanced Instrumental Variables estimation
    a. Testing IV assumptions
    b. Linear and nonlinear IV estimation
    c. Characterizing the complier population

Tuesday

2. Regression Discontinuity Designs
    a. Sharp and Fuzzy RDDs
    b. Visual analysis
    c. Bandwidth choice and estimation

Wednesday

3. G-Methods: theoretical foundations
    a. The potential outcomes model
    b. Assumptions for conditional independence
    c. Intro to propensity scores

Thursday

4. G-Methods: practical implementation
    a. Inverse Probability of Treatment Weights (IPTW)
    b. Standardization
    c. Extensions to time-varying scenarios.

Course prerequisites

A working knowledge of STATA or R is recommended for the practical exercises. Students working with other statistical software packages can participate in the course, but support for the practical exercises will only be provided for STATA and R.

The course content will build on the topics covered in the 2021 PHDS Population Health course and assumes basic knowledge of linear and non-linear regression models as well as instrumental variables estimation.

In the week before the Causal Inference course (July 4-8), students who have not attended the 2021 PHDS Population Health course will have the opportunity to watch the pre-recorded lecture videos of the relevant parts of the PHDS Population Health Course.

Examination

Students who wish to obtain a certificate for participation in the course will have the opportunity to take a short online exam in multiple choice format anytime on Friday 15 July 2022.

General readings

Linear Regression Basics and Instrumental Variables:

> Angrist, J., Pischke, J.-S., 2009. Mostly harmless econometrics. Princeton University Press, Princeton and Oxford. (Chapters 3 & 4)

A broad overview of Regression Discontinuity Designs:

> Lee, D.S., Lemieux, T., 2010. Regression Discontinuity Designs in Economics. Journal of Economic Literature 48, 281–355. https://doi.org/10.1257/jel.48.2.281

G-Methods:

> Hernán M.A., Robins J.M. (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.
> The book is freely available online here:
> https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/
>
> Intro: Chapters 1,2,3
> IPW: Chapters: 2, 12
> g-formula: Chapters 2, 13
> Time Varying treatments (IPW + g-formula): 19, 20, 21

Further optional reading recommendations:

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Statistics in medicine, 34(28), 3661-3679.

Cole, S. R., & Frangakis, C. E. (2009). The consistency statement in causal inference: a definition or an assumption? Epidemiology, 20(1), 3-5.

Daniel, R. M., Cousens, S. N., De Stavola, B. L., Kenward, M. G., & Sterne, J. A. C. (2013). Methods for dealing with time-dependent confounding. Statistics in medicine, 32(9), 1584-1618.

De Stavola, B. L., Herle, M., & Pickles, A. (2022). Framing causal questions in life course epidemiology. Annual Review of Statistics and Its Application, 9, 223-248.

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. Annual review of sociology, 40, 31-53.

Holland (1986). Statistics and Causal Inference Journal of the American Statistical Association, Volume 81, Issue 396

Lundberg, I., Johnson, R., & Stewart, B. M. (2021). What is your estimand? Defining the target quantity connects statistical evidence to theory. American Sociological Review, 86(3), 532-565.

Naimi, A. I., Cole, S. R., & Kennedy, E. H. (2017). An introduction to g methods. International journal of epidemiology, 46(2), 756-762.

Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. Epidemiology, 11(5), 550-560.

Tuition

There is no tuition fee for this course.

Recruitment of students external to the IMPRS-PHDS network

Applicants should either be enrolled in a PhD program (those well on their way to completion will be favored) or have received their PhD. Applications from advanced master's students will also be considered. A maximum of 60 students will be admitted. The selection will be made by the MPIDR based on the applicants' scientific qualifications.

How to apply

Please apply online via https://survey.demogr.mpg.de/index.php/476348?lang=en and include as *single pdf file* (max. 10 MB), in English:

- A curriculum vitae, including a list of your scholarly publications (max. 2 pages).
- A one-page statement of your research and how it relates to the course. Please include a paragraph about how you meet the prerequisites for this course.

Application deadline is **17 June 2022**.

Applicants will be informed of their acceptance by 27 June 2022.

Applications submitted after the deadline will be considered only if space is available.

Email inquiries about the course and the application process should be sent to phds@demogr.mpg.de.