Causal Inference

EUI Widening Programme - October 2024 - University of Sarajevo

Ken Stiller (ken.stiller@eui.eu)

This short course introduces causal inference in the social sciences. The objective is to learn how to make and evaluate causal claims. We start out with developing a framework of causality, learn about the fundamental problem of causal inference, and then learn to apply a range of basic causal inference approaches. To supplement and deepen the course materials, we occasionally use data from published papers to replicate and conduct our own analyses. The course is intended as an overview of the different methods ranging from experimental designs to quasi-experimental designs. Students will develop the skills to critique methods used in recent academic work and to begin to apply these methods in their own research.

The topics covered are extensive, so this course to be an introduction for students to then explore the different topics/applications on their own. You will receive guidance with respect to theory and how to implement different methods. All background readings, data and course material will be made available to you. Thus, this short course is an ideal starting point if you plan on using any of the introduced methods at a later point during your studies.

No knowledge of causal inference is required. Please bring your own laptop throughout the sessions. If possible, make sure R and R Studio are installed, more on this below - but we will take the time to install the software and introduce R to those who have not worked with it before.

Books & Readings

You are **not** requried to do any readings for this short course. However, I will provide a list of general and more specific readings and you are most welcome to use them throughout or after the course at your convenience. They will surely come in handy if you use any of the methods further down the road.

The main resource when it comes to the course contents is the recently released "Causal Inference: The Mixtape" by Scott Cunningham, which is available online for free. It provides a great resource for the theoretical underpinnings and implementation using statistical software. It is pretty accessible to newcomers and also really comes in handy if you want to refresh your knowledge on a specific topic.

Another, more econometric source is Angrist and Pischke (2014), "Mastering Metrics: The Path from Cause to Effect". This is the key introductory text that presents the essential tools of econometric research and demonstrates how to untangle cause and effect in human affairs. Most recently in 2021, one of its authors - Professor Joshua D. Angrist - along with Professor Guido Imbens (whose work we will be dealing with, too), won the Nobel Prize in Economics "for their methodological contributions to the analysis of causal relationships". If you want to read something more advanced on the same topics, see Angrist and Pischke's 2009 book "Mostly Harmless Econometrics"

I have listed additional readings for each bloc to allow participants to explore individual topics further as described in other textbooks and articles that implement the discussed research design on interesting research questions. Moreover, most blocs come with an empirical paper which provides an example of the method in use - I will make the data used in these papers accessible to you. All papers can be found online.

About R

Below you can find some useful information about R :

• To download R, go to: https://cran.rstudio.com/

• Many people like R-Studio as a way of managing your work in R. Like R, the basic version of R-Studio is free.

• You can download it here.

• This is entirely optional but if you want to improve your R skills, you will find a number of great tutorials here:
– Try R: http://tryr.codeschool.com/levels/1/challenges/1
– swirl: http://swirlstats.com
– Jared Knowles R bootcamp: https://www.jaredknowles.com/r-bootcamp/

Course Outline

Bloc 1: Introduction

Brief overview of the course. Setting up everything.

Bloc 2: R - The Basics

Using R in the social sciences - an introduction and refresher.

Bloc 3: Causality, Potential Outcomes Framework & Experiments

Discussion of what is causality in the social sciences, introduction to the potential outcomes framework and the ideal experiment. The second part will focus on randomization and why randomized experiments are the gold standard.

Readings:
• Angrist, Joshua D., and Jörn-Steffen Pischke. 2014. Mastering 'Metrics: The Path from Cause

to Effect. Princeton University Press: Chapter 1
• Angrist, Joshua and Jörn-Steffen Pischke. 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press: Chapter 2
• Cunningham, Scott. (2021). Causal Inference: The Mixtape: Chapter 4

Empirical Papers:
Lupu, N. (2013). Party brands and partisanship: Theory with evidence from a survey experiment in Argentina. American Journal of Political Science, 57(1), 49-64.
Hyde, S. D. (2007). The observer effect in international politics: Evidence from a natural experiment. World politics, 60(1), 37-63.

Bloc 4: Selection on Observables, Multiple Regression and Matching

Experiments/randomization is not always practical. Here we discuss designs that assume that selection into the treatment groups is based on observables and how we can use multiple regression to overcome endogeneity. We discuss regression and matching techniques.

Readings:
• Angrist, Joshua D., and Jörn-Steffen Pischke. 2014. Mastering 'Metrics: The Path from Cause to Effect. Princeton University Press: Chapter 2
• Angrist, Joshua and Jörn-Steffen Pischke. 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press: Chapter 3. Sections 3.1-3.3. For Matching see 3.3.1-3.3.3

Empirical Paper:
• Washington, E. L. (2008). Female socialization: how daughters affect their legislator fathers. American Economic Review, 98(1), 311-32.

Bloc 5: Panel Data and Fixed Effects Models

We will discuss what research designs we can use when we move from cross-sectional to panel data. These empirical strategies use data with a time or cohort/spatial dimension to control for unobserved, but time-specific or cohort-specific omitted variables.

Readings:
• Angrist, Joshua and Jörn-Steffen Pischke. 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press: Chapter 5
• Cunningham, Scott. (2021). Causal Inference: The Mixtape: Chapter 8

Empirical Paper:
• Gomez, B. T., Hansford, T. G., & Krause, G. A. (2007). The Republicans should pray for rain: Weather, turnout, and voting in US presidential elections. The Journal of Politics, 69(3), 649-663.

Bloc 6: Differences-In-Differences

Confounders cannot always be observed. However, many public policies are adopted at a certain point in time for a selected group which can be used to retrieve causal estimates. When there is data across time for both treatment and control groups, before and after treatment, a difference-in-difference approach can be implemented. This strategy addresses time-invariant confounders - but requires a careful consideration of assumptions.

Readings:
• Angrist, Joshua D., and Jörn-Steffen Pischke. 2014. Mastering 'Metrics: The Path from Cause to Effect. Princeton University Press: Chapter 5
• Angrist, Joshua and Jörn-Steffen Pischke. 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press: Chapter 5 - 5.1-5.3
• Cunningham, Scott. (2021). Causal Inference: The Mixtape: Chapter 9

Empirical Paper:
• Dinas, E., Matakos, K., Xefteris, D., & Hangartner, D. (2019). Waking up the golden dawn: does exposure to the refugee crisis increase support for extreme-right parties?. Political analysis, 27(2), 244-254.

Bloc 7: Instrumental Variables and 2SLS Estimation

If we discover a variable or phenomenon that is exogenously determined yet explains the variation in our explanatory variable of interest, we can use the method of instrumental variables (IV) to uncover the causal estimate of the explanatory variable on the outcome. A valid instrument induces changes in the explanatory variable but has no independent effect on the dependent variable. We will discuss the assumptions needed for a valid IV and appropriate model specifications.

Readings:
• Angrist, Joshua D., and Jörn-Steffen Pischke. 2014. Mastering 'Metrics: The Path from Cause to Effect. Princeton University Press: Chapter 3
• Angrist, Joshua and Jörn-Steffen Pischke. 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press: Chapter 4
• Cunningham, Scott. (2021). Causal Inference: The Mixtape: Chapter 7

Empirical Paper:
• Dinas, E. (2014). Does choice bring loyalty? Electoral participation and the development of party identification. American Journal of Political Science, 58(2), 449-465.

Bloc 8: Regression Discontinuity Designs

When selection into treatment group changes at an arbitrary cut-off for a continuous variable (i.e. a pass for a test score), a regression discontinuity design (RDD) can be implemented.

Readings:
• Angrist, Joshua D., and Jörn-Steffen Pischke. 2014. Mastering 'Metrics: The Path from Cause to Effect. Princeton University Press: Chapter 4

• Angrist, Joshua and Jörn-Steffen Pischke. 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press: Chapter 6

Empirical Paper:
• Meyersson, E. (2014). Islamic Rule and the Empowerment of the Poor and Pious. Econometrica, 82(1), 229-269.

Bloc 9: Fuzzy Regression Discontinuity Design

We will discuss more advanced topics around instrumental variables and regression discontinuity design. Specifically, we will discuss the fuzzy regression discontinuity design where the cut-off point increases the probability of receiving treatment.

Readings:
• Angrist, Joshua D., and Jörn-Steffen Pischke. 2014. Mastering 'Metrics: The Path from Cause to Effect. Princeton University Press: Chapter 4 (Fuzzy RD)
• Angrist, Joshua and Jörn-Steffen Pischke. 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press: Section 6.2.

Empirical Paper:
• Dinas, E., Riera, P., & Roussias, N. (2015). Staying in the first league: Parliamentary representation and the electoral success of small parties. Political Science Research and Methods, 3(2), 187-204.

Bloc 10: What to Make of All This? Limitations and Outlook

We will conclude by returning to the big picture. What can *you* do with all these methods, where are their limitations and can they be helpful even if your research question does not allow for an ideal causal inference design? In this session, we'll make space to discuss how these methods can be of use to your studies and research.