ICPSR 2014

CAUSAL INFEERENCE IN THE SOCIAL SCIENCES

Course Syllabus

Instructors:
Dominik Hangartner—London School of Economics
Marco Steenbergen—University of Zurich

Teaching Fellow:
Giuseppe Pietratuono—Social Science Center Berlin

July 14–18, 2014
University of California–Berkeley

“Nothing can be more ludicrous than the sort of parodies on experimental reasoning which one is accustomed to meet with, not in popular discussion only, but in grave treatises, when the affairs of nations are the theme... How can such or such causes have contributed to the prosperity of one country, when another has prospered without them? Whoever makes use of an argument of this kind, not intending to deceive, should be sent back to learn the elements of some one of the more easy physical sciences.”

John Stuart Mill (1872)
Contact Information:
Dominik Hangartner, Department of Methodology, London School of Economics and Department of Political Science, University of Zurich. Email: d.hangartner@lse.ac.uk.
Marco Steenbergen, Department of Political Science, University of Zurich. Email: steenbergen@ipz.uzh.ch.
Giuseppe Pietrantuono, Department of Methodology, Social Science Center Berlin. Email: giuseppe.pietrantuono@uzh.ch.

Course Information:
Meeting times: July 14–18, 2013: 9am–5pm.

Course Description
This course provides an introduction to statistical methods used for causal inference in the social sciences. Using the potential outcomes framework of causality, we discuss designs and methods for data from randomized experiments and observational studies. In particular, designs and methods covered include randomization, matching, instrumental variables, difference-in-difference, synthetic control, and regression discontinuity. Examples are drawn from different social sciences.

Organization
Many problems of causal inference from observational studies revolve around the concept of confounders, i.e. true causes of an effect that may render the impact of a putative cause spurious. There are different ways of handling confounders, depending on whether they are observed or not. After proving a general introduction to causation and causal inference, we begin by considering research designs in which the confounders are unobserved but rendered impotent through randomization (day 1). Since randomization is not always feasible, we may have to rely on other methods of controlling for confounders. On day 2, we consider designs in which the confounders are observed and can be controlled statistically. On days 3-5, we focus on cases in which at least some of the confounders are unobserved.
Prerequisites

Prerequisites and assessment: Knowledge of multiple linear regression and some familiarity with generalised linear models, to the level of Freedman (2005, see below) or equivalent. Familiarity with notions of research design in the social sciences and the statistical computing language R, to the level of Fox (2002, see below).

If you need to review material on regression models, please consult this excellent textbook:


If you need to review some R basics, you may want to have a look at


Software

R will be used in lab sessions.

Materials

The main course texts will be:


Assessment

No formal assessment but formative feedback on daily homework.
Schedule

**Day 1**

**Causal Inference Using Potential Outcomes**

Today we will introduce the topic of causal inference. We will define causal effects based on the potential outcomes framework of Neyman and Rubin, encounter the fundamental problem of causal inference, and discuss confounding as what separates association from causation, and observational studies from randomized experiments. We introduce examples of well designed observational studies and discuss the foundations and limitations of statistical models.

Readings:


Further readings:


*For the truly dedicated:*


**Day 1**

**Randomized Experiments**

We review the logic of randomized experiments, a research design that is
widely believed to maximize internal validity and that is becoming ever more popular in the social sciences. We pay special attention to Fisher’s randomization inference, in which randomization is the “sole and reasoned basis for inference”. Lastly, we will meet the “Lady tasting tea”.

Readings:


Class 1:

- Re-analysis of Sesame Street experiment.

Day 2

Subclassification and Matching on Covariates

The advantage of randomized experiments is that potential confounders can be safely ignored since they will balanced, at least in expectations. But randomization is not always practical, nor is it always ethical. How can one ensure valid causal inference in a world without randomization? Today, we discuss designs which assume that selection into the treatment groups is based on observables. We start by considering two very intuitive methods, subclassification and exact matching techniques.

Readings:


Further readings:


Day 2

Matching and Weighting Using the Propensity Score

Next, we discuss matching techniques that are based the propensity score
and on Euclidean distance. We also consider some practical issues with matching such as matching with and without replacement, common support restrictions, and estimating standard errors. In addition, we consider weighting as a nifty alternative to matching. Lastly, we compare OLS regression with matching and weighting estimators.

Readings:


- Angrist & Pischke. Chapter 3.3.1-3.3.3, pp.69-91.


Further readings:


For the truly dedicated:


Class 2:


Day 3

**Difference-in-Differences**

Confounders cannot always be observed and if they cannot, then alternative research designs have to be found. One such alternative arises in the context of panel data or repeated cross-sections. Here one can take the difference between pre- and post-tests and then compare those across groups. To the extent that the differences in the confounders have remained constant over time, then this estimator can produce valid causal inferences.

Readings:
• Angrist & Pischke. Chapter 5, pp. 221-246.


Further readings:


Day 3

Synthetic Control Method
If you would like to use Difference-in-Difference but don’t have a good control unit: synthesize one. This estimator is a simple but potentially widely applicable generalization of the Difference-in-Differences estimator.

Readings:


Further readings:


Class 3:

• Replication of Card and Krueger (1994).

Day 4

Instrumental variables: LATE
Instrumental variable (IV) methods can be used to address unobserved confounders in the context of cross-sectional data. Today, we discuss the basic logic of IV-techniques, focusing in particular on the LATE estimator.

Readings:
Day 4
Instrumental Variables: LARF
In many settings, an instrument is only valid conditional on some covariates. As we will show, the LATE estimator is only able to identify average treatment effects conditional on covariates under very restrictive assumptions. Abadie’s LARF estimator develops an ingenious weighting scheme to get unbiased treatment effects under heterogeneity for the subsample of compliers only.

Readings:
• Angrist & Pischke. Chapter 4.5-4.7

Class 4:
• Re-assessing the effect of watching more Sesame Street on child cognitive development; program effect of watching TV vs. program effect of encouragement to watch. ITT, Wald, LATE, and LARF estimators.

Day 5
Regression Discontinuity Designs
RDDs arise when selection into the treatment group depends on a covariate score that creates some discontinuity in the probability of receiving the treatment. We discuss both sharp and fuzzy RDDs.

Readings:
• Angrist & Pischke. Chapter 6.

For the truly dedicated:


Day 5
Overview and review
Schematic overview of class: maximizing internal validity of local estimates and the price of sacrificing external validity. Q & A session.