# Introduction to Causal Inference

## 12-13 June, Seminar Room Villa Sanfelice

***Organiser:*** *Fabrizio Bernardi*

**(10 credits)**

Please register with Martina.Selmi@eui.eu

## Motivation

> Whenever, looking at my watch, I see the hand has reached the figure X, I hear the bells beginning to ring in the church close by. But from the fact that the watch hands point to ten whenever the bells begin to ring, I have not the right to infer that the position of the hands of my watch is the cause of the vibration of the bells.
>
> Leo Tolstoy, *War and Peace*, trans. Constance Garnett (New York: Modern Library Classics, 2002), p. 939.

Do hospitals make people healthier? Is it a problem that more people die in hospitals than in bars? Does an additional year of schooling increase future earnings? Do parties that enter the parliament enjoy vote gains in subsequent elections? The answers to these questions (and many others which affect our daily life) involve the identification and measurement of causal links: an old problem in philosophy and statistics. To address this problem we either use experiments or try to mimic them by collecting information on potential factors that may affect both treatment assignment and potential outcomes. Customary ways of doing this in the past entailed the specification of sophisticated versions of multivariate regressions. However, it is by now well understood that causality can only be dealt with during the design, not during the estimation process. The goal of this workshop is to familiarize participants with the logic of casual inference, the underlying theory behind it and introduce research methods that help us approach experimental benchmarks with observational data. Hence, this will be a much applied course, which aims at providing participants with ideas for strong research designs in their own work.

## Summary of Content and Structure

This course will introduce participants into an authoritative framework of causal inference in social sciences. The objective is to learn how statistical methods can help us to draw causal claims about phenomena of interest. By the end of the course, students will be in position to:

1. critically read and evaluate statements about causal relationships based on some analysis of data;
2. apply a variety of design-based easy-to-implement methods that will help them draw causal inferences in their own research.

Either explicitly or implicitly, the goal of most empirical research is to interpret causally the co-occurrence of interesting phenomena. Addressing causality, however, has been notoriously difficult without the luxury of experimental data. This course will introduce you to methods that allow you to make convincing causal claims without working with experimental data. We will look at three such designs:

1. Instrumental Variables
2. Regression Discontinuity Design; and
3. Difference-in-Differences estimation.

All these designs will be presented through the potential outcomes framework, which constitutes the most widely acclaimed framework to discuss about causality.

You can only learn statistics by doing statistics. This is why this module includes a laboratory component, where you will learn to apply these techniques to the analysis of discipline specific data. Your coursework assignment will also be based on the need to actually implement these methods with real data. The data and all the instructions will be given to you. Thus, your goal will be to apply one of the methods you will have learnt on these data to answer questions about whether some variable X has an effect on some other variable Y .

## Educational Aims

This course aims to give students:

- an understanding of the methods of causal inference, using topics and examples from comparative politics, political economy and voting behaviour literature.
- a familiarity with software to implement the estimation based on these techniques. Most analyses will be based on STATA, a widely used package, which is very convenient for data management, statistical analyses and data visualisation. The equivalent ***R*** packages will be also discussed.

## Learning Outcomes

If you aspire to make causal claims in your own research or if you are interested in assessing other people's causal claims, you will find this course interesting and useful.

## Data

We will use various different data sets that will help us see how each of these designs works. All these data sets will become available to the participants in the first week of June. Participants want to apply any of these methods by using their own data are welcome to do so.

## Readings

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

Morgan Stephen L. and Christopher Winship. 2007. Counterfactuals and Causal Inference: Methods and Principles for Social Research, Cambridge: Cambridge University Press.

Design-specific articles and papers are shown below. More material and handouts will be distributed in the class.

## Timetable

**Day 1 (12/06/2014) – 09:30-12:30**

• Motivation, examples, discussion. We will see examples of the "fundamental problem of causal inference." Introduction to the potential outcomes framework. We will derive the causal quantities of interest (ATE, ATT, ATC). Reappraisal of the regression equation using the insights from the potential outcomes framework.

• Instrumental variables: Using a running example, we will derive the LATE, delve into identification, and we will introduce the Wald Estimator.

*Readings:*

•Angrist and Pischke: Ch. 1 & 4. Morgan and Winship Ch. 1, 2 & 7.

**Day 1 (12/06/2014) – 14:00–18:00**

• Instrumental Variables: Adding covariates, the 2SLS estimator. Relaxing the Constant Treatment Effects Assumption and the LARF Estimator.

• Regression Discontinuity Design: Intuition, Assumptions and Identification

*Readings*

• Angrist, Joshua, Guido Imbens, and Donal Rubin. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*, 91(434):444-55.

• Abadie, Alberto. 2003. Semiparametric Instrumental Variable Estimation of Treatment Response Models.*Journal of Econometrics*, 113: 231-63.

• Sovey, Allison & Donald Green. 2010. "Instrumental Variables Estimation in Political Science: A Readers' Guide." *American Journal of Political Science*, 55(1): 188-200.

• Imbens, Guido. 2014. Instrumental Variables: An Econometrician's Perspective. NBER Working Paper # 19983. [available Online here].

• Angrist & Pischke Ch. 6.

• Lee, David. "Randomized Experiments from Non-Random Selection in U.S. House Elections."

*Journal of Econometrics*, 142: 675-97.

**Day 2 (13/06/2014) – 10:00-13:00**

• Difference-In-Differences estimation: Motivation, examples, identification, estimation, Examples, what can go wrong and what to do about it.
• Hands-on practice in the lab. Introduction to the code and examples in STATA. We will see code for IVs, sharp and fuzzy RDDs and the difference-in-differences estimator.

*Readings*

• Imbens, Guido & Thomas Lemieux. 2007. Regression Discontinuity Designs: A Guide to Practice. NBER Working Paper, # 13039.

• Lee, David & Thomas Lemieux. Regression Discontinuity Designs in Economics, *Journal of Economic Literature*. 48:281-355.

• Hainmueller, Jens and Andrew Eggers. 2009. "MPs for Sale? Returns to Office in Postwar British Politics." *American Political Science Review*, 103(04): 513-33.

• Imbens, Guido & Karthik Kalyanaraman. 2010. Optimal Bandwidth Choice for the Regression Discontinuity Estimator. Cemmap Working Paper, # CWP05/10. [available Online here].

• Angrist & Pischke Ch. 5.

• Montalvo, José. 2011. "Voting after the Bombings: A Natural Experiment on the Effect of Terrorist Attacks on Democratic Elections." *Review of Economics and Statistics*, 93(4):1146-1154.

•Bechtel, Michael & Jens Hainmueller. 2011. "How Lasting is Voter Gratitude? An Analysis of the Short-and Long-Term Electoral Returns to Beneficial Policy." *American Journal of Political Science*, 55(4): 852-68.
• Calonico, Sebastian, Matias Cattaneo & Rocio Titiunik. 2014. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. Unpublished Manuscript. [Available Online here].

• McCrary, Justin. 2008. “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test.” *Journal of Econometrics*, 142(2): 698-714.

## Assessment

Assessment will be based on the implementation of one of the methods we will learn in this course. Finding your own case would be ideal. If this is not feasible, the task is going to be a 2,000 words essay, where you will be asked to replicate an existing paper using either of these methods. More details about these papers and the replication fies will be provided during the course.