



SPS Seminar 2<sup>nd</sup> term 2019-2020

# The Statistics of Causal Inference

**Organiser: Professor Elias Dinas**

Contact: [Jennifer.Dari@eui.eu](mailto:Jennifer.Dari@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 causal inference, the underlying theory behind it and introduce research methods that help us approach experimental benchmarks with observational data. Particular emphasis is placed on how to use these methods in order to address historical questions. Why did Protestant countries flourish more than catholic ones? How did the European migrants integrate in the US throughout the 20th century? What is the effect of Nazi territorial control on patterns of resistance during the WWII? While discussing these and other applications, we will also shed light on how to think about archival work under the prism of causal inference. Hence, this will be a much applied course, which aims at providing participants with ideas for strong research designs in their own work.

## Content Summary and Learning Outcomes

The objective is to learn how statistical methods can help us to draw causal claims about phenomena of interest. During the five-days of the course, participants will be introduced into an authoritative framework of causal inference in social sciences, i.e. the potential outcomes framework. 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.
3. think about archival data under the logic of causal inference.

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

For every method, the following structure will be employed: first, a running example will provide the motivation and intuition. We will then proceed with the formal identification derivation and finally we will focus on estimation strategies and robustness checks. For each method there will be a hands-on lab section, where we will apply these methods with real data.

Depending on how much time we will have available in the course, we will also cover three additional topics:

1. Matching; and
2. Synthetic Control

Notwithstanding what we will cover, please do not hesitate to ask me about any of these topics. Hopefully, we will be able to at least touch upon them and provide some material that will be potentially useful via Piazza. I will also use Piazza to send you discussion on twitter or elsewhere, where there is currently taking place an important discussion about several issues regarding the strengths and weaknesses of the potential outcomes framework.

### Schedule

Wednesday	8 January	11:00-13:00	Seminar Room 2
Monday	13 January	09:00-11:00	Seminar Room 2
Monday	20 January	09:00-11:00	Seminar Room 2
Wednesday	22 January	11:00-13:00	Seminar Room 2
Monday	27 January	09:00-11:00	Seminar Room 2
Wednesday	29 January	11:00-13:00	Sala del Capitolo
Monday	3 February	09:00-11:00	Seminar Room 2
Wednesday	5 February	11:00-13:00	Seminar Room 2
Monday	10 February	09:00-11:00	Seminar Room 2
Monday	17 February	09:00-11:00	Seminar Room 2

### Textbook 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.

Angrist, Joshua D., and Jörn-Steffen Pischke. *Mastering'metrics: The path from cause to effect*. Princeton University Press, 2014.

Gerber, Alan S., and Donald P. Green. *Field experiments: Design, analysis, and interpretation*. WW Norton, 2012.

A very in-depth, path-breaking and by now legendary contribution in this field is Judea Pearl's *Causality*. Feel free to go through the book if you are particularly interested in the topic and you want to invest on this. Before doing so, it is worth knowing that Pearl uses a different notation, coming from computer science (hence the extensive use of  $\text{do}$  in the notation) which is not what we will be using in the course. Pearl is among other things famous for introducing the Directed Acyclic Graphs, aka DAGs. We will briefly introduce them and discuss one of the two criteria for identification, namely the back-door criterion. I wish we had more time to discuss the front-door criterion, which is linked to mediation analysis. If interested, Piazza or email me and I will send respective literature. Morgan and Winship use DAG's in their treatment.

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

## Timetable

### Wednesday — 08/01/2019

- Introduction to the Potential Outcomes Framework
  - 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.
  - How experiments solve the fundamental problem of causal inference. The logic of randomization and why it works. How do we know? Balance tests and other diagnostics.

Readings:

- Angrist and Pischke: Ch. 1 & 2. Morgan and Winship Ch. 1 & 2.

### Monday — 13/01/2019

- Randomization given by nature: Examples of Natural Experiments
- Instrumental Variables: Intuition, Identification & Estimation, Wald Estimator, 2SLS Estimator

Readings:

- Angrist and Pischke: Ch. 4. Morgan and Winship Ch. 7.
- Angrist, Joshua, Guido Imbens, and Donald 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, Joshua D. and Alan B. Krueger. 2001. Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments, *The Journal of Economic Perspectives*, 15(4): 69-85.
- Dunning, Thad. 2009. "Instrumental Variables." In Morlino et al. *International Encyclopedia of Political Science*, available online [here](#).

- Acemoglu, Johnson, & Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review*, 91(5): 1369-1401, available online [here](#).
- Lassen, David D. 2005. "The Effect of Information on Voter Turnout: Evidence from a Natural Experiment." *American Journal of Political Science*, 49(1): 103-18.
- Dunning, Thad. Natural experiments in the social sciences: a design-based approach. Cambridge University Press, 2012.

### Monday — 20/01/2019

- Lab Session: Instrumental Variables

### Wednesday — 22/01/2018

- Regression Discontinuity Design: Motivation, Identification, Estimation Strategies.
- Testing for Sorting, Robustness Checks, Examples, Applications.

#### Readings:

- Angrist & Pischke Ch. 6.
- Lee, David. "Randomized Experiments from Non-Random Selection in U.S. House Elections." *Journal of Econometrics*, 142: 675-97.
- 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.
- Dell, M. (2015). Trafficking networks and the Mexican drug war. *American Economic Review*, 105(6), 1738-79.
- Dell, M. (2010). The persistent effects of Peru's mining mita. *Econometrica*, 78(6), 1863-1903.
- Ferwerda, J., & Miller, N. L. (2014). Political devolution and resistance to foreign rule: A natural experiment. *American Political Science Review*, 108(3), 642-660.
- Eggers, Andrew, Anthony Fowler, Jens Hainmueller, Andrew B. Hall, and James Snyder Jr. 2015. "On The Validity Of The Regression Discontinuity Design For Estimating Electoral Effects: New Evidence From Over 40,000 Close Races." *American Journal of Political Science*, 59(1):259-74, available online [here](#).

### Monday — 27/01/2019

- The Fuzzy RD
- Extensions: Identification away from the Cut-off point & Randomization-Based Inference in the RD design.

#### Readings:

- Angrist, Joshua D., and Miikka Rokkanen. "Wanna get away? RD identification away from the cutoff." (2013).
- Hainmueller, Jens, Andrew B. Hall, and James M. Snyder Jr. "Assessing the external validity of election RD estimates: An investigation of the incumbency advantage." *The Journal of Politics* 77.3 (2015): 707-720.

- Cattaneo, Matias D., Brigham R. Frandsen, and Rocio Titiunik. "Randomization inference in the regression discontinuity design: An application to party advantages in the US Senate." *Journal of Causal Inference* 3.1 (2015): 1-24.

### **Wednesday — 29/01/2019**

- Lab Session: Regression Discontinuity Design

### **Monday — 03/02/2019**

- Difference-In-Differences: Motivation, examples, identification, estimation.
- Threads to validity and examples.

#### Readings:

- Angrist & Pischke Ch. 5.
- David, C., & Krueger Alan, B. (1994). Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4), 772-793.
- Fouka, Vasiliki. "How do immigrants respond to discrimination? The case of Germans in the US during World War I." *American Political Science Review* 113.2 (2019): 405-422.
- Fouka, Vasiliki. "Forthcoming." "Backlash: The Unintended Effects of Language Prohibition in US Schools after World War I." *Review of Economic Studies*.

### **Wednesday — 05/02/2019**

- Extensions: Triple-Differences, Adding Lags and leads, Fixed-Effects, and Interactions

#### Readings:

- Cantoni, D. (2015). The economic effects of the Protestant Reformation: testing the Weber hypothesis in the German lands. *Journal of the European Economic Association*, 13(4), 561-598.
- Acemoglu, D., Cantoni, D., Johnson, S., & Robinson, J. A. (2011). The consequences of radical reform: The French Revolution. *American Economic Review*, 101(7), 3286-3307.
- Autor, David H. "Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing." *Journal of labor economics* 21.1 (2003): 1-42.
- Just watch [these two videos](#).

### **Monday — 10/02/2019**

- Lab Session: Difference-In-Differences

### **Wednesday — 17/02/2019**

- Matching: Motivation, examples, identification, estimation. Genetic Matching and Entropy Balancing.
- Synthetic Control: Motivation, examples, identification, estimation.
- Quick Lab session

#### Readings:

- Diamond and Sekhon 2013, "Genetic Matching for Estimating Causal Effects: A General Multivariate

Matching Method for Achieving Balance in Observational Studies.” *Review of Economics & Statistics*.

- Rosenbaum, Paul R. 2002. *Observational Studies*. New York: Springer-Verlag 2nd edition.
- Hainmueller, Jens. "Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies." *Political Analysis* 20.1 (2012): 25-46.
- Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American economic review*, 113-132.
- Abadie, A., Diamond, A., & Hainmueller, J. (2012). Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program. *Journal of the American Statistical Association*.
- Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2), 495-510.

## Data & Software

We will use R and Stata to present examples from published and unpublished articles for each of the four methods.

## Assignments

Replicate one design-specific paper or use your own data to implement one of the methods we will learn. Deadline will be three weeks since the end of the course.