# Causal Inference for the Social Sciences II

2025 ICPSR Summer Program

Ann Arbor, Michigan

## 1 Basic Information


| | | |
|---|---|---|
| Instructors: | Sebastian Calonico | Nicolas Idrobo |
| | Assistant Professor | Ph.D. Candidate |
| | UC Davis | University of Pennsylvania |
| | scalonico@ucdavis.edu | idrobo@sas.upenn.edu |


| | |
|---|---|
| Schedule: | Monday July 14 through Friday August 1, 2025 |
| | Main lecture: 2:00 – 5:00pm every day |

## 2 Overview

The goal of this class is to provide a thorough introduction to recent methodological developments to study causal inference in non-experimental settings, including reweighting methods, regression discontinuity designs, difference-in-differences, synthetic controls and causal machine learning. The main focus will be on methodology and empirical practice, with an emphasis of the key steps to follow when implementing the methods.

It is assumed that participants have elementary working knowledge of statistics, econometrics and policy evaluation. It would be useful, but not required, if participants were familiar with basic results from the literature on program evaluation and treatment effects at the level of Wooldridge (2010). This course is nonetheless meant to be self-contained and hence most underlying statistical concepts and results are introduced and explained in class.

## 3 Reading Materials

We will not follow any specific books. For general background, we recommend Imbens and Rubin (2015), Angrist and Pischke (2009), Rosenbaum (2023), and Hernán MA (2020). We also provide specific background references under each topic below.

## 4 Assignments

Assignments are due each Friday, at the beginning of class.

## 5 Computing and Software

To participate in the hands-on exercises, you are strongly encouraged to bring a laptop computer with the most recent version of `Stata` or `R`/`RStudio` installed, as we will use them to analyze several empirical illustrations. Please make sure you have `Stata` (or `R`) and the modules/commands listed below installed and fully functional in your personal computer before the course begins. Datasets, do-files and `R` files will be provided in advance.

The following Stata/R packages will be used:

**RDD**

- rdrobust: RD inference employing local polynomial and partitioning methods. Calonico, Cattaneo and Titiunik (2014, 2015*b*).
- rddensity: Manipulation testing for RD designs. Cattaneo, Jansson and Ma (2018).
- rdlocrand: RD inference employing randomization inference methods. Cattaneo, Titiunik and Vazquez-Bare (2016).
- rdmulti: Analysis of RD designs with multiple cutoffs or scores. Cattaneo, Titiunik and Vazquez-Bare (2020).

**DiD**

- R: dplyr, fixest, did, drdid, bacondecomp.
- Stata: csdid, drdid, reghdfe, bacondecomp, eventstudyinteract, did_multiplegt, did_imputation.

**Machine Learning**

- R: ISLR2
- Stata:

### 5.1 Software installation instructions

If you will be using R, we recommend also installing RStudio. To install these programs, follow the instructions below:

1. Download the most recent version of R and install it in your computer. R is free and available for Windows, Mac, and Linux operating systems. Download the version of R compatible with your operating system from CRAN. If you are running Windows or MacOS, you should choose one of the precompiled binary distributions (i.e., ready-to-run applications) that are linked at the top of the page, *not* the source codes.
2. Download and install R Studio. This is an integrated development environment (IDE) that includes a console, a syntax-highlighting editor, and other tools for plotting, history, debugging and workspace management for R use. RStudio is free and available for Windows, Mac, and Linux platforms.

If you are using Stata, you will have to obtain a license from here. Stata is licensed through StataCorp and is frequently offered at a significant discount through academic institutions to their employees and students. Seminar participants who are not yet ready to purchase Stata could take advantage of StataCorp's 30-day software return policy and obtain Stata 18 on a trial basis in the weeks immediately preceding this course. Use this link for 30-day trial copy: http://www.stata.com/customer-service/evaluate-stata/

## 6 Outline, Schedule & Background References

This section gives an overview of the topics covered and optional readings.

### WEEK 1

#### Day 1: Introduction to Causal Inference

Potential outcomes framework. Randomized experiments: identification, estimation and inference. Non-experimental approaches to causal evidence. Introduction to Directed Acyclic Graphs (DAGs). Instrumental Variables.

Assigned Reading: Abadie and Cattaneo (2018) (Ch. 1-3), Hernán MA (2020) (Ch. 1-2, 6).

Background References: Imbens and Rubin (2015), Holland (1986), Rosenbaum (2010), Rosenbaum (2023).

#### Day 2: Unconfoundedness

Identification of causal effects by Conditional Independence. Regression Adjustments. Matching Estimators. Doubly robust approaches.

Assigned Reading: Abadie and Cattaneo (2018) (Ch. 4), Hernán MA (2020) (Ch. 3).

Background References: Imbens and Rubin (2015), Holland (1986), Rosenbaum (2010).

#### Days 3-4: RD Designs - Foundations

Introduction to RD designs. Graphical presentation of RD results. Conventional continuity-based estimation and inference methods based on local polynomials. Bandwidth selection. Falsification methods. The role of covariates.

Assigned Reading: Cattaneo, Idrobo and Titiunik (2020), Sections 1 and 2.

Background References: Hahn, Todd and van der Klaauw (2001), Lee and Lemieux (2010), Lee (2008), Thistlethwaite and Campbell (1960). Calonico, Cattaneo, Farrell and Titiunik (2019), Cattaneo, Jansson and Ma (2020, 2018), McCrary (2008).

Robust bias-corrected local polynomial methods for continuity-based analysis.

Assigned Reading: Cattaneo, Idrobo and Titiunik (2020), Sections 1-4.

Background References: Calonico, Cattaneo and Titiunik (2015*a*), Card et al. (2015), Keele and Titiunik (2015).

## WEEK 2

### Day 5: RD Designs - Extensions

Fuzzy RDD. Local randomization. Discrete running variables. Multicutoff and multiscore RDD.

Assigned Readings: Cattaneo, Idrobo and Titiunik (2023), Sections 2 to 5.

Background references: Keele and Titiunik (2015), Cattaneo, Titiunik and Vazquez-Bare (2017), Cattaneo, Keele, Titiunik and Vazquez-Bare (2016), Cattaneo, Keele, Titiunik and Vazquez-Bare (2021).

### Days 6-9: Difference-in-Differences

Classical two-periods/two-groups DiD setup with heterogeneous treatment effects. The parallel trends assumptions. DiD with multiple periods and simultaneous treatment adoption. Sensitivity analysis and robustness checks. Staggered treatment adoption designs: problems with two-way fixed effects estimators and possible solutions. Synthetic controls.

Assigned Reading: Abadie and Cattaneo (2018), de Chaisemartin and D'Haultfœuille (2022), Roth et al. (2023), Steigerwald, Vazquez-Bare and Maier (2021).

Background References: Abadie and Gardeazabal (2003), Abadie, Diamond and Hainmueller (2010), Athey and Imbens (2017), Sun and Abraham (2021), de Chaisemartin and D'Haultfœuille (2020), Tazhitdinova and Vazquez-Bare (2023), Abadie (2021).

## WEEK 3

### Days 10-14: Machine Learning Methods for Causal Inference

Supervised and unsupervised approaches. Average and heterogeneous causal effects. Regression trees and random forest. Double Selection to Double ML. Average treatment effect estimation: AIPW-Double ML. Heterogeneous effects with inference. Policy learning.

Assigned Reading: James et al. (2023), Mullainathan and Spiess (2017), Athey and Imbens (2017), Athey and Imbens (2019), Colin Cameron's notes (UC Davis).

Background References: Hastie, Tibshirani and Friedman (2009), James et al. (2014).

## References


Abadie, Alberto. 2021. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature* 59(2):391–425.

Abadie, Alberto, Alexis Diamond and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105(490):493–505.

Abadie, Alberto and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93(1):113–132.

Abadie, Alberto and Matias D Cattaneo. 2018. "Econometric methods for program evaluation." *Annual Review of Economics* 10:465–503.

Angrist, Joshua and Jörn-Steffen Pischke. 2009. "Mostly Harmless Econometrics.".

Athey, Susan and Guido W. Imbens. 2017. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31(2):3–32.

Athey, Susan and Guido W. Imbens. 2019. "Machine Learning Methods That Economists Should Know About." *Annual Review of Economics* 11(1):685–725.

Calonico, Sebastian, Matias D Cattaneo, Max H Farrell and Rocio Titiunik. 2019. "Regression discontinuity designs using covariates." *Review of Economics and Statistics* 101(3):442–451.

Calonico, Sebastian, Matias D. Cattaneo and Rocio Titiunik. 2014. "Robust Data-Driven Inference in the Regression-Discontinuity Design." *Stata Journal* 14(4):909–946.

Calonico, Sebastian, Matias D. Cattaneo and Rocio Titiunik. 2015*a*. "Optimal Data-Driven Regression Discontinuity Plots." *Journal of the American Statistical Association* 110(512):1753–1769.

Calonico, Sebastian, Matias D. Cattaneo and Rocio Titiunik. 2015*b*. "`rdrobust`: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs." *R Journal* 7(1):38–51.

Card, David, David S. Lee, Zhuan Pei and Andrea Weber. 2015. "Inference on Causal Effects in a Generalized Regression Kink Design." *Econometrica* 83(6):2453–2483.

Cattaneo, Matias D., Luke Keele, Rocio Titiunik and Gonzalo Vazquez-Bare. 2016. "Interpreting Regression Discontinuity Designs with Multiple Cutoffs." *Journal of Politics* 78(4):1229–1248.

Cattaneo, Matias D., Luke Keele, Rocio Titiunik and Gonzalo Vazquez-Bare. 2021. "Extrapolating Treatment Effects in Multi-Cutoff Regression Discontinuity Designs." *Journal of the American Statistical Association* 116(536):1941–1952.

Cattaneo, Matias D., Michael Jansson and Xinwei Ma. 2018. "Manipulation Testing based on Density Discontinuity." *Stata Journal* 18(1):234–261.

Cattaneo, Matias D, Michael Jansson and Xinwei Ma. 2020. "Simple local polynomial density estimators." *Journal of the American Statistical Association* 115(531):1449–1455.

Cattaneo, Matias D., Nicolás Idrobo and Rocío Titiunik. 2020. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge Elements: Quantitative and Computational Methods for Social Science, Cambridge University Press.

Cattaneo, Matias D., Nicolás Idrobo and Rocío Titiunik. 2023. *A Practical Introduction to Regression Discontinuity Designs: Extensions*. In preparation for Cambridge Elements: Quantitative and Computational Methods for Social Science, Cambridge University Press.

Cattaneo, Matias D., Rocio Titiunik and Gonzalo Vazquez-Bare. 2016. "Inference in Regression Discontinuity Designs under Local Randomization." *Stata Journal* 16(2):331–367.

Cattaneo, Matias D., Rocio Titiunik and Gonzalo Vazquez-Bare. 2017. "Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality." *Journal of Policy Analysis and Management* 36(3):643–681.

Cattaneo, Matias D., Rocio Titiunik and Gonzalo Vazquez-Bare. 2020. "Analysis of Regression Discontinuity Designs with Multiple Cutoffs or Scores." *Stata Journal* 20(4):866–891.

de Chaisemartin, Clément and Xavier D'Haultfœuille. 2022. "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey." *The Econometrics Journal* .

de Chaisemartin, Clément and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110(9):2964–96.

Hahn, Jinyong, Petra Todd and Wilbert van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69(1):201–209.

Hastie, Trevor, Robert Tibshirani and Jerome Friedman. 2009. *The elements of statistical learning: data mining, inference and prediction*. 2 ed. Springer.

Hernán MA, Robins JM. 2020. "Causal Inference: What If.".

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396):945–960.

Imbens, Guido W. and Donald B. Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

James, Gareth, Daniela Witten, Trevor Hastie and Robert Tibshirani. 2014. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated.

James, Gareth, Daniela Witten, Trevor Hastie and Robert Tibshirani. 2023. *An Introduction to Statistical Learning: with Applications in R*. Springer.
**URL:** *https://www.statlearning.com/*

Keele, Luke J. and Rocio Titiunik. 2015. "Geographic Boundaries as Regression Discontinuities." *Political Analysis* 23(1):127–155.

Lee, David S. 2008. "Randomized Experiments from Non-random Selection in U.S. House Elections." *Journal of Econometrics* 142(2):675–697.

Lee, David S. and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48(2):281–355.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics* 142(2):698–714.

Mullainathan, Sendhil and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31(2):87–106.

Rosenbaum, P. R. 2010. "Design of Observational Studies.".

Rosenbaum, P.R. 2023. "Causal Inference.".

Roth, Jonathan, Pedro H.C. Sant'Anna, Alyssa Bilinski and John Poe. 2023. "What's trending in difference-in-differences? A synthesis of the recent econometrics literature." *Journal of Econometrics* .

Steigerwald, Douglas G., Gonzalo Vazquez-Bare and Jason Maier. 2021. "Measuring Heterogeneous Effects of Environmental Policies Using Panel Data." *Journal of the Association of Environmental and Resource Economists* 8(2):277–313.

Sun, Liyang and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics* 225(2):175–199. Themed Issue: Treatment Effect 1.

Tazhitdinova, Alisa and Gonzalo Vazquez-Bare. 2023. Difference-in-Differences with Unequal Baseline Treatment Status. Working Paper 31063 National Bureau of Economic Research.

Thistlethwaite, Donald L. and Donald T. Campbell. 1960. "Regression-discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment." *Journal of Educational Psychology* 51(6):309–317.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross-Section and Panel Data*. Cambridge, MA: MIT Press.