Syllabus for POLI 891-002: Advanced Topics in Causal Inference

Ye Wang
University of North Carolina at Chapel Hill

Fall 2022

Course description

This graduate-level seminar aims to equip participants with a modern perspective on the progress of causal inference and the skill to apply frontier methods to social science studies. It follows the design-based perspective— we analyze datasets as if they were generated by hypothetical experiments. We focus on assumptions imposed on the treatment assignment process rather than those on the outcome model.

The course consists of two separate yet interdependent parts. The first part includes four lectures on the foundation of causal inference and causal analysis of cross-sectional data. They cover basic concepts such as the Neyman-Rubin model, counterfactuals, and the "fundamental problem of causal inference." They also introduce various experimental designs, how to conduct estimation and statistical inference in each of them, as well as the connection between experiments and observational studies. We conclude this part with an overview of covariate adjustment techniques in observational studies, including matching, weighting, regression, balancing, and doubly robust estimation.

In the second part, students will give presentations on five selected topics in causal inference: the application of machine learning in causal inference (2 weeks), time-series cross-sectional (TSCS) data analysis (3 weeks), interference and contagion (2 weeks), and shift-share designs (1 week). There will be two to three presenters each week, depending on the difficulty of the materials.

Time and location

Classes: 3:35pm-6:15pm on Mondays, Room 0112 in Murphey Hall.
There will no class on Sept. 5 (Labor Day), Sept. 26 (Well-being Day), or Nov. 21 (Thanksgiving).
Office hours: 3pm-6pm on Tuesdays, Room 322 in Hamilton Hall.

Requirements

Students who take the course should have working knowledge of probability theory, matrix algebra, and calculus, as well as some background in writing scripts in R.

Grading of the course is based on class participation (20%), presentation (40%), and the final proposal (40%). In the second part of the course, you are expected to read the assigned papers carefully prior to class and participate actively in the discussion of these papers. Each student needs to give two to three presentations during the seminar. The presentation should cover the basic ideas of the method, assumptions and contexts for it to work, how to implement it in R, and its potential applications in political science.

You are required to submit a two-page description of a research proposal in the middle of the semester (before October 10) and a complete draft of the proposal before December 1. There is no restriction on the research question in the proposal. But it must include a section in which you justify your research design from the design-based perspective, such as what assumptions you have imposed for causal identification, why they are convincing in your context and what methods are proper in this case. In the last week of the course, each student will give a 10-minute presentation on the proposal that is going to be submitted.

Course outline

Section I: Lectures on the foundation of causal inference

Lecture 1: The fundamental problem of causal inference

Week 1, August 15

The Neyman-Rubin model
The fundamental problem of causal inference: two solutions
Estimand vs. estimator
Sampling uncertainty vs. design uncertainty
Design-based perspective vs. model-based perspective

References: Holland (1986), Samii (2016), Ch1 and Ch2 of Imbens and Rubin (2015), Abadie et al. (2020)

Lecture 2: Experimental design

Week 2, August 22

Complete experiment vs. Bernoulli trial
Horvitz-Thompson estimator, Hajek estimator, and regression estimator
Unbiasedness and consistency
Asymptotic inference
Fisher's randomization test and bootstrap

References: Ch3-Ch7 of Imbens and Rubin (2015), Lin (2013), Samii and Aronow (2012)

Lecture 3: From experiments to observational studies

Week 3, August 29

Block randomization and clustering experiments
Power analysis
Pre-analysis plan and pilot study
Justifications for covariate adjustment
Propensity score
Matching

References: Ch9, Ch10 and Ch12 of Imbens and Rubin (2015), Abadie et al. (2017), Blair et al. (2019), Abadie and Imbens (2006), Abadie and Spiess (2021), Imai and Ratkovic (2014)

Lecture 4: Covariate adjustment

Week 4, September 12

Weighting
Regression
Balancing
Doubly robust estimation
Sensitivity analysis

References: Hahn (1998), Aronow and Samii (2016), Hainmueller (2012), Blackwell (2014), Cinelli and Hazlett (2020)

Section II: Selective topics

Topic 1: Applying machine learning to causal inference

Week 5, September 19: Basic algorithms in machine learning

References: Grimmer, Roberts, and Stewart (2021), Athey and Imbens (2019)

Week 6, October 3: Estimating heterogeneous treatment effects

References: Grimmer, Messing, and Westwood (2017), Wager and Athey (2018)

Topic 2: Time-series cross-sectional data analysis

Week 7, October 10: Problems of fixed effects models

References: Strezhnev (2017), Imai and Kim (2019)

Week 8, October 17: Flexible models

References: Liu, Wang, and Xu (2020), Arkhangelsky et al. (2019)

Week 9, October 24: Weighting methods

References: Blackwell and Glynn (2018), Kim, Imai, and Wang (2019)

Topic 3: Interference and contagion

Week 10, October 31: Exposure mapping

References: Hudgens and Halloran (2008), Todd et al. (2021), Aronow and Samii (2017)

Week 11, November 7: Unknown exposure mapping and contagion

References: Aronow, Samii, and Wang (2020), Egami (2018)

Topic 4: Shift-share designs

Week 12, November 14

References: Adao, Kolesár, and Morales (2019), Autor, Dorn, and Hanson (2013)

Student presentation

Week 13, November 28

References:

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. 2017. "When Should You Adjust Standard Errors for Clustering?" National Bureau of Economic Research.

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2020. "Sampling-Based Versus Design-Based Uncertainty in Regression Analysis." *Econometrica* 88 (1): 265–96.

Abadie, Alberto, and Guido W Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica* 74 (1): 235–67.

Abadie, Alberto, and Jann Spiess. 2021. "Robust Post-Matching Inference." *Journal of the American Statistical Association*, 1–13.

Adao, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. "Shift-Share Designs: Theory and Inference." *The Quarterly Journal of Economics* 134 (4): 1949–2010.

Arkhangelsky, Dmitry, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager. 2019. "Synthetic Difference in Differences." National Bureau of Economic Research.

Aronow, Peter M, and Cyrus Samii. 2016. "Does Regression Produce Representative Estimates of Causal Effects?" *American Journal of Political Science* 60 (1): 250–67.

———. 2017. "Estimating Average Causal Effects Under General Interference, with Application to a Social Network Experiment." *The Annals of Applied Statistics* 11 (4): 1912–47.

Aronow, Peter M, Cyrus Samii, and Ye Wang. 2020. "Design-Based Inference for Spatial Experiments with Interference." *arXiv Preprint arXiv:2010.13599*.

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

Autor, David H, David Dorn, and Gordon H Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *The American Economic Review*, 2121–68.

Blackwell, Matthew. 2014. "A Selection Bias Approach to Sensitivity Analysis for Causal Effects." *Political Analysis* 22 (2): 169–82.

Blackwell, Matthew, and Adam N Glynn. 2018. "How to Make Causal Inferences with Time-Series Cross-Sectional Data Under Selection on Observables." *American Political Science Review* 112 (4): 1067–82.

Blair, Graeme, Jasper Cooper, Alexander Coppock, and Macartan Humphreys. 2019. "Declaring and Diagnosing Research Designs." *American Political Science Review* 113 (3): 838–59.

Cinelli, Carlos, and Chad Hazlett. 2020. "Making Sense of Sensitivity: Extending Omitted Variable Bias." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82 (1): 39–67.

Egami, Naoki. 2018. "Identification of Causal Diffusion Effects Using Stationary Causal Directed Acyclic Graphs." *arXiv Preprint arXiv:1810.07858*.

Grimmer, Justin, Solomon Messing, and Sean J Westwood. 2017. "Estimating Heterogeneous Treatment Effects and the Effects of Heterogeneous Treatments with Ensemble Methods." *Political Analysis* 25 (4): 413–34.

Grimmer, Justin, Margaret E Roberts, and Brandon M Stewart. 2021. "Machine Learning for Social Science: An Agnostic Approach." *Annual Review of Political Science* 24: 395–419.

Hahn, Jinyong. 1998. "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects." *Econometrica*, 315–31.

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis*, 25–46.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–60.

Hudgens, Michael G, and M Elizabeth Halloran. 2008. "Toward Causal Inference with Interference." *Journal of the American Statistical Association* 103 (482): 832–42.

Imai, Kosuke, and In Song Kim. 2019. "When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?" *American Journal of Political Science* 63 (2): 467–90.

Imai, Kosuke, and Marc Ratkovic. 2014. "Covariate Balancing Propensity Score." *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 243–63.

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

Kim, In Song, Kosuke Imai, and Erik Wang. 2019. "Matching Methods for Causal Inference with Time-Series

Cross-Sectional Data."
Lin, Winston. 2013. "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique." *The Annals of Applied Statistics* 7 (1): 295–318.
Liu, Licheng, Ye Wang, and Yiqing Xu. 2020. "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data." *Available at SSRN 3555463*.
Samii, Cyrus. 2016. "Causal Empiricism in Quantitative Research." *The Journal of Politics* 78 (3): 941–55.
Samii, Cyrus, and Peter M Aronow. 2012. "On Equivalencies Between Design-Based and Regression-Based Variance Estimators for Randomized Experiments." *Statistics & Probability Letters* 82 (2): 365–70.
Strezhnev, Anton. 2017. "Generalized Difference-in-Differences Estimands and Synthetic Controls." *Unpublished Manuscript*.
Todd, Jason Douglas, Edmund J Malesky, Anh Tran, and Quoc Anh Le. 2021. "Testing Legislator Responsiveness to Citizens and Firms in Single-Party Regimes: A Field Experiment in the Vietnamese National Assembly." *The Journal of Politics* 83 (4): 1573–88.
Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113 (523): 1228–42.