

ECO 510: Data Analytics III: Causal Inference Spring 2021

Meeting Time: Wednesdays 5:45pm-7:15pm on Zoom

Instructor: Professor Anthony T. Lo Sasso

E-mail: alosasso@depaul.edu

Overview

Some of the most interesting and controversial research in the social and health sciences involves causal relationships. Here are a few recent examples:

- Moving out of a high poverty neighborhood reduces extreme obesity and diabetes (Ludwig et al, 2011).
- Neonatal intensive care units reduce mortality rates among very low birth weight babies by about 1 percentage point, and the cost per life saved in the NICU is about \$550,000 (Almond et al, 2010).
- Recent increases in disability rates among Vietnam veterans are caused by features of the VA disability system and are not evidence of a delayed causal effect of military service (Angrist et al 2010).

The authors of these studies do not want us to think of their findings as mere correlations. They are making causal claims. What assumptions are the authors making? And what do they even mean by causality? What about those warnings that *correlation does not imply causation*? Should we believe their results?

This course is about the research designs and methods that researchers use to support causal inferences in the social sciences. Some specific topics include randomized experiments, instrumental variables, regression discontinuity designs, difference-in-differences models, regression analysis, and propensity score matching.

There are four broad goals of the course. The first is to learn some of the notation and language that social scientists use to describe causal effects. The second is to understand the logic and assumptions that support a set of research designs that are commonly used in quantitative social science research. The third is to gain some experience in implementing the methods using a statistical software package such as Stata. The fourth objective is to develop skill at reading, understanding, and critiquing “technical” scientific articles that make casual claims.

Learning Objectives

By the end of the course, you will be able to:

- Express causal research questions using the formal notation and vocabulary used by economists, statisticians, and other researchers to describe causal relationships.
- Identify and judge the plausibility of various threats to the validity of causal inferences in applied settings.
- Identify and explain the structure, logic, and assumptions associated with several quasi-experimental research designs.

- Analyze data from various quasi-experimental research designs using statistical software and methods that are currently considered the best practice.
- Read and critique quasi-experimental research studies in writing and in conversations.

Prerequisites

Econometrics II

Required Textbook

- *Mastering 'Metrics*. Joshua Angrist and Jorn-Steffen Pischke, Princeton 2015. (MM)

Supplementary (Optional) Textbook

- *Mostly Harmless Econometrics: An Empiricist's Companion*, Joshua Angrist and Jorn-Steffen Pischke, Princeton 2009

Software

- Stata

Grading

- Assignments, generally weekly, some pretty quick, some more involved (50%)
- Paper/empirical project (50%)

A = 93-100, A- = 90-92, B+ = 87-89, B = 83-86, B- = 80-82, C+ = 77-79, C = 73-76,
C- = 70-72, D+ = 67-69, D = 60-66, F = <60

Paper

You have options for your paper.

1. Do a replication of a published paper highlighting one (or more) of the approaches we discuss in this course! Many journals require authors to add data and code to an online repository. Replication is (supposed to be) an integral part of science, but it is not often done. Your work should include basic checks on the robustness of the published results (assuming you've managed to replicate the published results).
2. Do an empirical analysis of an intervention/treatment/policy of interest. My main interest is in your description of the empirical methodology, which should feature at least two of the approaches discussed in this course, with particular emphasis on your assumptions and potential threats to validity. Your empirical work should be presented in clear tables and carefully interpreted. Interpretation and explanation for differences of the results between different approaches is critical. Please note that I'm not expecting a whole paper. Instead of a lengthy introduction, a one sentence description of the question will do. You should look at the literature, but an extended lit review is not necessary. Economic theory is also not required unless it in some way contributes to your empirical model and assumptions.

3. Provide a proposal to study a policy or program of interest. My main interest is in your description of the empirical methodology, which should feature at least four of the approaches discussed in this course, with particular emphasis on your assumptions and potential threats to validity for each.

Other Information: Our class D2L site will include course material and supplementary readings throughout the term. Lectures will be recorded via Panopto and uploaded to D2L. Students can then find the lecture in D2L and view it. We will also use the Discussion feature in D2L to have active discussions regarding the papers and lectures assigned each week. You will be responsible for viewing each lecture and learning the material. This will require you to actively engage with the material. That might mean taking notes while the lecturing is happening—just like you would in a real, meat-space classroom.

As a hybrid class, we will meet weekly on Zoom (synchronously!) to review homework and discuss the week's lecture material and readings.

The Discussion board on D2L will be where we “converse” about the material and you can ask questions. If you send me email questions about course material, I will respond by telling you post it. You may of course email me regarding any personal matters as they relate to the course.

A few caveats: It will be more difficult for me to “read the classroom” to get a sense of whether what I’m saying is clear—so you need to help me! I’m asking—but not requiring—you to leave your cameras on during our weekly Zoom session. It really does make the experience better for everyone. I would also encourage you to do your best to help your fellow students. Answer questions in the Discussion board when you can. Finally, given the uncertainties in the coming term, technical or otherwise, I reserve the right to modify the syllabus in the event of unforeseen circumstances.

Schedule of Topics (Refer to D2L for papers)

1. Introduction and Background Issues: Workflow, Analysis Plans, Identification

Freedman, David A. (1991) “Statistical Models and Shoe Leather.” *Sociological Methodology* 21, 291-313.

Ioannidis JPA (2005) Why Most Published Research Findings Are False. *PLoS Med* 2(8): e124.

Gentzkow, M and Shapiro, JM (2013) *Code and Data for the Social Sciences: A Practitioner’s Guide*

Further Reading:

Rosenbaum, Paul (1999) “Choice as an Alternative to Control in Observational Studies” *Statistical Science* 14(2). Also read the rejoinders by Manski, Robins, Cook and Shadish.

2. Randomized Experiments

MM, Chapter 1

Heckman, JJ. and Smith, J. (1995) Assessing the Case for Social Experiments. *Journal of Economic Perspectives* (9) 2.

Katherine Baicker, Sarah Taubman, Heidi Allen, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Eric Schneider, Bill Wright, Alan Zaslavsky, Amy Finkelstein, and the Oregon Health Study Group, "[The Oregon Experiment – Effects of Medicaid on Clinical Outcomes](#)" *New England Journal of Medicine*, Volume 368, Issue 18. May 2013.

Further Reading:

Ludwig, J et al (2011) Neighborhoods, Obesity, and Diabetes: A Randomized Social Experiment. *The New England Journal of Medicine* 365 (October)

Sarah Taubman, Heidi Allen, Bill Wright, Katherine Baicker, Amy Finkelstein, and the Oregon Health Study Group (2014), "[Medicaid Increases Emergency Department Use: Evidence from Oregon's Health Insurance Experiment](#)", *Science*, 2 January 2014.

Finkelstein, A et al (2012) The Oregon Health Insurance Experiment: Evidence From The First Year. *Quarterly Journal of Economics*, Aug; 127(3): 1057-1106.

3. Instrumental Variables

MM, Chapter 3

Angrist, J, Lavy, V, and Schlosser, A. (2010) Multiple Experiments for the Causal Link between the Quantity and Quality of Children, *The Journal of Labor Economics* (October).

Angrist, J, Chen, S and Frandsen, B. (2010) Did Vietnam Veterans Get Sicker in the 1990s? The Complicated Effects of Military Service on Self-Reported Health, *The Journal of Public Economics*, December 2010.

Kaestner, Robert, and Anthony T. Lo Sasso, "Does Seeing the Doctor More Keep You Out of the Hospital?" *Journal of Health Economics* 2015;39: 259-272.

McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA*. 1994 Sep 21;272(11):859-66.

Further Reading

Heckman, J. (1997) Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. *Journal of Human Resources* 32(3)

4. Regression Discontinuity Design

MM, Chapter 4

Almond, D, Doyle, J, Kowalski, A, and Williams, H. (2010) Estimating the Marginal Returns to Medical Care: Evidence from At-risk Newborns. *Quarterly Journal of Economics* 125(2).

Carpenter, C., & Dobkin, C. (2009). The effect of alcohol consumption on mortality: regression discontinuity evidence from the minimum drinking age. *American Economic Journal: Applied Economics*, 1(1), 164-82.

Elder, Todd E. "The importance of relative standards in ADHD diagnoses: evidence based on exact birth dates." *Journal of health economics* 29.5 (2010): 641-656.

Further Reading

Card, D, Dobkin, C, Maestas, N. (2009) Does Medicare Save Lives? *Quarterly Journal of Economics* 124.

Imbens, G and Lemieux, T (2008) Regression Discontinuity Designs: A Guide To Practice. *Journal of Econometrics* 142(2).

Cook, TD (2008) Waiting for Life to Arrive: A History of the regression discontinuity design in psychology, statistics, and economics. *Journal of Econometrics* 142.

Lee, D. and Lemieux, T (2010) Regression Discontinuity Designs in Economics. *Journal of Economic Literature* 48.

5. Panel Data and Difference in Differences

MM, Chapter 5

Lu Jinks, Lu, Thomas J. Kniesner, John Leeth, and Anthony T. Lo Sasso, "Opting Out of Workers' Compensation: Non-Subscription in Texas and Its Effects," *Journal of Risk and Uncertainty* 2020; 60(1): 53-76.

Lo Sasso, Anthony T., and Ithai Z. Lurie, "Community Rating and the Market for Private Non-Group Health Insurance," *Journal of Public Economics* 2009;93(1-2): 264-279.

Further Reading

Bollinger, B, Leslie, P, and Sorensen, A. (2011) Calorie Posting in Chain Restaurants. *American Economic Journal: Economic Policy*, 3:91-128.

Dickert-Conlin, S., Elder, T., & Moore, B. (2011). Donorcycles: Motorcycle helmet laws and the supply of organ donors. *The Journal of Law and Economics*, 54(4), 907-935.

6. Controlling For Observable Confounders: Regression and Matching

MM, Chapter 2

Further Reading

Lyle, D. (2006) Using Military Deployments and Job Assignments to Estimate the Effect of Parental Absences and Household Relocations on Children's Academic Achievement. *Journal of Labor Economics* 24(2).

Eliason, M and Storrie, D. (2009) Job Loss is Bad For Your Health – Swedish Evidence on Cause-Specific Hospitalization Following Involuntary Job Loss. *Social Science and Medicine* 68(8)