

Syllabus for Causal Inference and Research Design
Scott Cunningham, PhD (Baylor University, Waco Texas USA)

Course Description

Confusing correlation and causality is a dangerous error. On the one hand, believing that naive correlations reflect a causal relationship can lead people to make incorrect decisions. Policymakers may eliminate programs, for instance, because they observe poor outcomes associated with the program, when in fact it is those with the poorest outcomes who are choosing the program in the first place!

But the opposite error is just as bad. An overly skeptical view of causality such that no correlation can ever be causal is also naïve and can also lead to incorrect decisions if only because the policymakers discounts the value of quantitative information altogether.

The purpose of this class is to correct both types of errors by introducing students to the modern field of “causal inference”. If I am successful, then you will have a better understanding of when and why it is justified to believe that discovered correlations are in fact a reflection of true causal effects, and when they are not.

The class is a very hands-on course. Students will learn to write programs in the econometrics software packages, Stata and R, in addition to learn the numerous research designs economists and statisticians have developed to estimate causal inference when randomized experiments are impossible, unethical or financially unreasonable. Students will also develop a basic understanding of the logic contained in directed acyclical graphical models, a powerful tool for thinking through causal relationships and planning your research projects.

We will cover some programming etiquette, regression, basic causality tools like potential outcomes and directed acyclical graphs, randomization inference, regression discontinuity, instrumental variables, panel methods, differences-in-differences, synthetic control, and if time permits, matching/weighting/subclassification. The majority of the class will focus on selection bias and treatment assignment, and ways to address problems associated with both.

Course Objectives

The primary objective of this course is for students to understand a variety of econometric estimators and research strategies for inferring causal effects in observational data. But a secondary course objective is help the student develop some basic competency in programming in Stata and/or R.

Course Outcomes

Course objectives are measures via the course assignments which assess acquired substantive knowledge and analytical ability via written work. See below under “Coursework, Grades, and Grading Policies”.

Credit

Students will be evaluated based on one test and three replications

Textbooks and Readings

There are two required textbooks for the class and one recommended. The main textbook we will use is my free online book Causal Inference: The Mixtape (contracted with Yale University Press). That can be downloaded from my website <https://www.scunning.com>. I also will supply all of my slides which covers material not in the current free version.

The second book is entitled Mostly Harmless Econometrics Princeton University Press, 1st edition by Angrist and Pischke. It is not expensive.

A third book I like but which is not required is Counterfactuals and Causal Inference: Methods and Principles for Social Research. Cambridge University Press (2nd edition) by Morgan and Winship. It's particularly good at explaining the potential outcomes model, the directed acyclical graphical models, IV, matching, and partial identification, as well as something called the front door criterion. What I like about this book is the two authors are sociologists who are nonetheless conversant in the causal inference methodological toolkit, and probably because of their background as much as their talent, they're excellent communicators. They make a lot of this seem easy (which it is!).

We will also be reading a lot of articles this semester. I will post links to these on Canvas as well as upload them. Some of the readings are technical pieces from economics journals. The degree to which a student needs to be familiar with the details of a paper will be clear from the emphasis given to the paper in lecture.

Coursework, Grades, and Grading Policies

Graduate credit is weighted equally across midterm 1, midterm 2, 4 replication projects and an original research project:

Final exam (40%)

Replications (50%)

Slack registration (10%)

Exams

The exam covers any material from the assigned readings in the text, as well as any additional material that I cover in lecture, including any articles I cover in class. The exam is scheduled for the last day of class. As we are doing this remotely, we will set

up a Zoom session. Everyone will have their video on and I will keep an eye so that I can see enough. That's the best we can do. You will submit these to me by emailing a Word document after the time is up.

Replications

You will be required to conduct three empirically intensive “replication projects” over the course of the class. You will be given minimal to moderate guidance. In some cases, I have not personally replicated the paper and in other cases I have several times. This is a very hands-on course, and I will be using Stata for all of these replications. You are welcome to use R, though I cannot be as much help for questions about R as I can about Stata. For projects we do together, I have R and Stata code though.

I will be holding both a language agnostic programming lecture as well as an introduction to Stata lecture for those of you for whom programming is not your first language. I unfortunately cannot hold an R workshop, but I am going to see if I can't find some assistance from a former student who I think would be good at this.

Slack channel

Once everyone is registered, I would like everyone to immediately register for Slack on your desktop and smart phone, and then join a Slack channel to be created. This is your main class participation credit. It's worth 10% of your grade. I want you to have a way to start threads on different topics, including replications, because replications are hard and we need to be able to help one another as much as we can. You can also more easily reach me through direct message in Slack. I prefer in-time interactive communication over emails these days.