CAUSAL INFERENCE AND RESEARCH DESIGN: SYLLABUS

Course code: Economics 395M
Term: Spring 2020
Location: BRB2.136
Time: 2:00 - 3:15pm

Instructor: Prof. Scott Cunningham
Department of Economics
UT Austin visiting professor
Phone: (254) 537-2239
Email: scunning@gmail.com
Office: Room 2.138
Office hours: Tuesday/Thursday
3:30pm–5:00pm, or by appointment

COURSE DESCRIPTION

Confusing correlation and causality is a dangerous error. On the one hand, believing that a naive correlation is a causal relationship can lead people to make incorrect decisions. But the opposite error is also there, which is an overly skeptical view of causality such that no correlation can ever be causal can also lead to incorrect decisions. The purpose of this class is to correct both types of errors by introducing students to the modern theory of "causal inference" which I hope can help you better understand why you are justified to draw particular beliefs about causal relationships between sets of events.

The class is a very hands-on course. Students will learn to write programs in the econometrics software package, STATA, in addition to learn the numerous research designs economists and statisticians have developed to estimate causal inference when experiments are not feasible. Students will also develop a basic understanding of the logic contained in directed acyclical graphical models. We will cover matching estimation, linear and nonlinear regression models, panel methods, differences-in-differences, synthetic control, instrumental variables and regression discontinuity. The majority of the class will focus on selection bias and treatment assignment.

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. Complementary course objectives consist of developing programming skills in STATA.

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 midterm, four replications, and a research project.

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 `http:www.scunning.com`. I also will supply all of my slides on Canvas so that students can see the new material we will be covering not in the book. 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 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.

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:

- Midterm (30%) Thursday March 12, 2020
- Replications (40%) TBD
- Research Project (30%) Last Day of Semester

EXAMS (30%)

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. Students will be excused from the midterm exams only for valid medical or family emergencies. These excuses must be identified before the midterm and students must produce signed evidence verifying the reason why they cannot attend.

The exam is scheduled for **Thursday March 12, 2020**. It will be blue books, essay questions, covering both concepts, analytical problems, and papers we discussed and read in class together. I will be grading these assignments anonymously randomized to me and my TAs to ensure fairness and consistency.

REPLICATIONS (40%)

You will be required to conduct four empirically intensive "replication projects" over the course of the semester, each of which is worth 10% of your grade. 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. A student version is available through STATA Gradplan for $69, or students can use the version installed in the computer labs. Please contact Stata directly at the following url if you are interested in purchasing your own copy.

- `http://www.stata.com/order/new/edu/gradplans/campus-gradplan/`

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.

Many students these days prefer R to Stata. I really don't care. So if you want to use R, that's fine. Just know I don't personally know the language, and therefore really cannot be of much help trouble shooting or problem solving. You will be on your own.

One last thing - replications should be considered group activities. You turn in your own work, but you work with whoever you need to work with to get the project completed. There's benefits from working together, but there's also free riding issues inherent in any work project. Still, most graduate assignments benefit from getting together with someone else and plowing through it.

RESEARCH PROJECTS (TOTAL: 30%)

In addition to the midterms and replications, students are required to write an original empirical research project.

APPROVED TOPIC - SUBTOTAL: 5%

Your research project will contain a tractable research question, data at some low order of aggregation (individual, state panel data, etc.) and a directed acyclical graphical model.

DAG - SUBTOTAL: 5%

The directed acyclical graphical model which has been approved should be rich enough that there are testable predictions which can be verified in your data. If we are unable to convincingly show that the DAG holds for your data, we will need to come up with a solution, which may involve using an alternative estimation strategy, alternative data, or a different question.

ROUGH DRAFT - SUBTOTAL: 10%

Your first rough draft will be due at TBD (roughly early April).

FINAL DRAFT, INCLUDING ABSTRACT - SUBTOTAL: 10%

Your final draft will be due on last day of class. There will be no exam.

CLASS PARTICIPATION

With a class of approximately 40 students, class participation is more challenging. The one thing that I will be trying to do is call on you and ask for your feedback and critical opinions about topics we are discussing, so please come prepared for class. My TA will be creating name tags for everyone, which will be distributed and passed out each class. Please make them visible. Whatever you've heard about the absent minded professor being bad with names, I'm in the right tail of the distribution of even those people! So please be patient with me as I struggle to remember even my kids' names sometimes.

TENTATIVE SCHEDULE AND TOPICS

The following is a rough outline of the topics we will cover in class:

- Statistics review, linear regression
- Potential outcomes model, directed acyclic graphical modeling
- Programming review and the hidden curriculum
- Selection on observables: back-door criterion
- Selection on observables: OLS, stratification, propensity score matching, covariate matching
- Selection on observables: regression discontinuity design
- Selection on unobservables: panel methods, differences-in-differences, triple differences,

synthetic control
- Selection on unobservables: instrumental variables, heterogenous treatment effects, LATE
- Other topics (remainder)