

Instructor	Prof. Vivek Bhattacharya vivek.bhattacharya@northwestern.edu Office Hours: Wednesdays 3:45–5:30 pm, KGH 3379 (Although I should be around, please shoot me a quick email if you plan to stop by office hours.)
TA	Ryan Lee RyanLee2014@u.northwestern.edu Office Hours: Thursdays 5:00-7:00 pm, KGH 3198

Course Description: The primary purpose of this course is to teach you some simple tools to take a dataset and learn something about the patterns hidden in it. In the first part of the course, we will go over tools to estimate statistical correlations. These correlations are informative, and it's important to know how to get them right. But, such statistical correlations are often not the object of economic interest. Academic economists, businesspeople, and policy makers alike often care about the *causal* effect a policy instrument and an outcome, which need not coincide with the correlation between the policy and the outcome. We'll discuss why we care about causal effects, what sort of questions can be answered with a precise causal estimate rather than simply a correlation, and why it can be misleading to confuse correlation with causation.

This takes us to the second part of the course—and indeed the meat of it. We will develop formal models that can be used to estimate causal effects. While we will discuss the statistical properties of these models and develop estimators, the goal is not so much to learn how to plug-and-chug but rather to learn how to critically assess (i) what conceptual quantity a model is designed to estimate, and (ii) whether a model is applicable for a particular question. As such, the theory will be interspersed with applications. We'll devote a few classes at the end of the course to critiquing empirical analysis in depth.

Through the problem sets, the recitations, and the final project, you will develop some familiarity with the “practical” parts of econometrics as well. You'll learn how to implement these estimators in statistical software and have a chance to critique them in the context of a recent study of an important policy issue.

Prerequisites: While 381-1 is the formal prerequisite, you should be fine if you've taken introductory statistics, probability, linear regression, and linear algebra.

Textbooks and Software: I will distribute lecture notes. If you'd like supplemental material, I'd recommend the following textbooks. We won't use these textbooks formally in class (i.e., I won't assign reading from them). So if you are thinking about purchasing them, feel free to buy whichever edition is cheapest. I would strongly recommend acquiring a copy of *Mostly Harmless Econometrics* if it fits with your budget; the others are certainly optional.

- Angrist and Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. A lot of the material in my lecture notes will come from this book, although the book is geared towards incoming graduate students rather than advanced undergraduates. If you are interested in purchasing one book for this course, I'd recommend this one. It's cheap for a

textbook, fun to read, and it will also provide you with more material to digest if you decide to keep going with economics, econometrics, or any sort of policy analysis.

- Stock and Watson (2011). *Introduction to Econometrics*. This is a fairly comprehensive overview of the material we cover in class. It is maybe a more “basic” undergraduate textbook, though, and the focus is slightly different. Most of you should have this textbook since it was required in 381-1, but you’ll be fine in this class even if you don’t have it.
- Angrist and Pischke (2014). *Mastering ‘Metrics: The Path from Cause to Effect*. This book is geared towards undergraduates and covers much of the material from *Mostly Harmless Econometrics* that we will cover in class. It’s pitched in a somewhat basic (and unfortunately silly) way, but it could be good supplemental material.
- Cunningham (2018). *Causal Inference: The Mixtape*. This book also covers much of the material we’ll cover, and you might want to look at it if you want a different treatment from the lecture notes. It’s also free online at http://scunningham.com/cunningham_mixtape.pdf.

Problem sets will require access to Stata. Stata is unfortunately commercial software, but you can find it at many computing labs around the university, as well as through NUWorkspace at <http://nuworkspace.northwestern.edu/>. If you do not have an account, I recommend you request one immediately. We will usually provide datasets in Stata format, and Ryan and I will primarily provide help for Stata.

If you are interested in learning a broader-use, open-source language, you are welcome to complete the problem sets in R or Python. And you will get more limited help from Ryan and me with technical issues.

Grading: Problem Sets: 15%, Midterm 1: 35%, Midterm 2: 35%, Final Project: 15%.

- Problem Sets: There will be 6 problem sets. I’ll drop your lowest score, so you can feel free to miss one assignment without penalty if things get busy during the semester.

Problem sets will be submitted on Crowdmark. I will not give extensions on problem sets, nor will I accept late problem sets, other than in truly exceptional situations (with a Dean’s note).

You can work in groups of up to four (although I’d recommend no more than three) on each problem set. Groups can turn in a single copy of their solutions through Crowdmark. Make sure you have an active role in solving the problems. It is not in your best interest to free-ride off others: you’ll have to do the exams yourself!

A randomly selected subset (and possibly all) of the problems will be graded carefully for accuracy. The remainder of the problem set will be graded on completion. Show all your work. For computational problems, submit the (well-documented) code and the output, but also incorporate your results into your writeup in a clean, readable fashion.

Be neat. If Ryan and I can’t read your work, we won’t grade it. I’d recommend typing your problem sets, especially if you have messy handwriting.

- Exams: The midterms will be on **April 29** and **May 27**, during regular class hours. I will not give any make-up midterm exams (again, except in exceptional circumstances). If you anticipate having legitimate conflicts with the midterm date, please let me know by **April 6**. You will be allowed one sheet (8.5×11 inches) of notes for each exam. It can be hand-written or typed, but you must make it yourself. (I’ll ask you to hand it in with your exam, and I’ll return it with the exam.) Nothing else: no books, no computers, no phones, no calculators.

- *Final Project:* A goal of the class is to prepare you to think critically about analyses done in various settings—academic papers, policy briefs, popular press, etc. The final project will give you an opportunity to conduct such an analysis. Approximately three weeks before the end of the class, I will distribute a piece of empirical analysis about a current policy issue. You will write a 3–5 page (single-spaced) critique of this paper. What question does it answer? Do you find the answer convincing? Why or why not? What other analyses could the authors conduct to convince you of the results? What do you think their empirical strategies are missing? How would you do the analysis differently?

You will have to work by yourself on this critique. You cannot talk to anyone else (other than Ryan or me) about the paper, but you can research anything online (as long as it is cited). The project should be emailed to me by **Tuesday, June 2, at 11:59 pm**. More details will be distributed about the final project closer to the end of the quarter.

Schedule

March 31	Introduction and Discussion of Empirical Work
April 1	Causality and Potential Outcomes
April 6	Ordinary Least Squares
April 8	Ordinary Least Squares
April 13	Limited Dependent Variables
April 15	Limited Dependent Variables and MLE
April 20	MLE + Nonparametric Regression
April 22	Nonparametric Regression
April 27	Review for Midterm 1
April 29	Midterm 1
May 4	Panel Data and Fixed Effects
May 6	Instrumental Variables
May 11	Instrumental Variables and Heterogeneity
May 13	No Class
May 18	Differences-in-Differences
May 20	DID and Regression Discontinuity
May 22 (FRIDAY)	Regression Discontinuity
May 22	Review for Exam 2 + Discussion of Empirical Topic I
May 25	Memorial Day, No Class
May 27	Midterm 2
June 1	Discussion of Empirical Topic
June 2 (TUESDAY)	Final Project Due at 11:59 pm

Problem sets will be due on April 10, April 24, April 27, May 8, May 15, and May 25. I will try to stick to the proposed schedule of topics as closely as possible, but we may substitute some topics for detailed discussions of some research papers. Recitations will cover additional practice problems as well as (potentially) discussions of empirical papers.

AccessibleNU Notice: Any student requesting accommodations related to a disability or other condition is required to register with AccessibleNU (accessiblenu@northwestern.edu; 847-467-5530) and provide professors with an accommodation notification from AccessibleNU, preferably within the first two weeks of class. All information will remain confidential.