Econometrics I

Lecture 1: Introduction

Paul T. Scott
NYU Stern

Fall 2018

Today's Roadmap

1. Some logistics
2. Overview of econometrics and causality
3. Review of some math and statistics

Course Description

This is an econometrics course for first-year PhD students and advanced undergraduates who are interested in doing quantitative research in the social sciences. The aim of the course is to teach you to **use** popular applied econometric methods while developing your theoretical understanding of those methods. Topics include least squares, asymptotic theory, hypothesis testing, instrumental variables, difference-in-differences, regression discontinuity, treatment effects, panel data, maximum likelihood, discrete choice models, machine learning, and model selection.

Prerequesites for Studying Econometrics

Ideally, students should have experience with

- Calculus
- Basic Probability and Statistics (19/20)
- Linear Algebra (15/20)
- Data analysis software such as R, Python/Numpy, Matlab, Julia, Stata, or similar (Excel doesn't count) (16/20)

Assignments and Grading

- 4 Problem sets: 15% each
  - Turn in write ups and code to me by email.
  - You may use any software you like. When giving examples, I will use R.
  - To compose, consider using software like LaTeX and R Markdown.
- Group project: 40%
  - A research project on any topic (subject to my approval) using econometric analysis.
  - I suggest choosing a published paper to replicate and finding a way to extend, test, or improve on it. Ideally, choose something on a topic that interests you. I will also share a list of suggestions.
  - 2-3 students per group
  - Proposal due in middle of semester. Please discuss ideas with me before submitting your proposal.
  - Presentations in final class session
  - Paper due at end of semester

Getting Help

- Office hours: Monday 4:30-6:30pm, or by appointment (email)
- I will try to respond to emails within 48 hours.
  - I'm not available to help you debug your code.
- Three consultation workshops with Skand Goel (5th year economics PhD student):
  - One before each of the first two problem sets are due (time and location TBA)
  - The class session before the group presentations.

Course outline (subject to change)

1	9/6	Introduction, Math and Statistics Review
2	9/13	Linear Regression, Ordinary Least Squares
3	9/20	Asymptotic Theory, Testing and Inference
4	9/27	Robust Estimation, Delta Method, Bootstrap
5	10/4	(Quasi-)Experiments, Endogeneity, Instrumental Variables
6	10/11	Simultaneity, Generalized Method of Moments, Treatment Effects
7	10/18	Differences in Differences, Regression Discontinuity
8	10/25	Panel Data, Fixed and Random Effects
9	11/1	Nonlinear Estimation, Maximum Likelihood, Sample Selection
10	11/8	Nonparametric Estimation, Machine Learning, Model Selection
11	11/15	Binary Choice, Discrete Choice (guest lecturer: Chris Conlon)
	11/22	No Class - Thanksgiving
12	**11/29**	Project Consultation Workshop with Skand Goel
13	12/6	Group Project Presetations
	12/13	No Class, Group Research Papers due (by email)

What is Econometrics?

- **Econometric Questions:**
  - Often, if not mostly, about *causality*
    - About individuals: *What is the effect of education on wages?*
    - About markets (micro): *How does the price of the iPad affect the number of units that will be sold?*
    - About markets (macro): *How does raising the minimum wage affect employment?*
  - Some studies are primarily descriptive
    - *What is the relationship between growth and inequality?*
  - Sometimes questions are model-driven
    - *How much do people discount the future?*
  - Occasionally pure forecasting is important
    - *What is probability of bankruptcy for firm/individual with a given set of characteristics?*

What is Econometrics?

- **Obstacles:**
  - Selection: Economic agents (people, firms, etc.) are purposeful and know more than we do about their personal situations!
- **Techniques:**
  - We look for data that can establish causality:
    - **Gold Standard:** A randomized controlled trial
    - "Natural experiments" or "Instrumental Variables": something random that happens (not controlled by researchers) that affects behavior
    - Controls: try to measure everything else that might affect the outcome
  - Big emphasis on *inference* – what matters, how much do those things matter, and how confident are we in such claims?

Causality vs. Prediction

What *is* causality?

- Deep philosophical question that we don't have time to chew on...
- Rough concept for this class:
  - **Prediction:** If we *observe* a piece of data x, what is the likely value of y?
  - **Causality:** If we *change* a value of x, what will be the new value of y, holding everything else *constant*?
- Examples:
  - **Prediction:** What is 4-year-old's future SAT score likely to be, given her behavior in the "marshmallow test"?
  - **Causality:** Knowing that school funding and achievement are correlated, determine the effect of increasing funding on achievement.
- This distinction is crucial across the sciences

The calcium-rich food is thought to reduce the risk by up to 14 per cent

ıld or grey at a
greater risk of
e

Because the researchers didn't actually test diet changes of their participants the findings could be a result of healthier people being likely to eat more cheese.

This could be because they're richer and can afford to eat more cheese, or because of their diet, one UK study included in the analysis followed vegetarians who would likely have a diet including lots of plants as well as cheese.

A Causal Question and an Ideal Experiment

Simple question: What is the effect of class size on students' test scores?

- Some potential research designs to answer the question:

A Causal Question and an Ideal Experiment

Simple question: What is the effect of class size on students' test scores?

- Some potential research designs to answer the question:
  - Compare test scores across classrooms (cross-sectional)

A Causal Question and an Ideal Experiment

Simple question: What is the effect of class size on students' test scores?

- Some potential research designs to answer the question:
  - Compare test scores across classrooms (cross-sectional)
  - Compare a student's score over time as class size changes (time-series)

A Causal Question and an Ideal Experiment

Simple question: What is the effect of class size on students' test scores?

- Some potential research designs to answer the question:
  - Compare test scores across classrooms (cross-sectional)
  - Compare a student's score over time as class size changes (time-series)
  - Randomly toss kids into either small or large classes (RCT)

A Causal Question and an Ideal Experiment

Simple question: What is the effect of class size on students' test scores?

- Some potential research designs to answer the question:
  - Compare test scores across classrooms (cross-sectional)
  - Compare a student's score over time as class size changes (time-series)
  - Randomly toss kids into either small or large classes (RCT)
- Often the first two are the best we can do, but then there are big selection concerns!
  - Students who know they benefit from one set up or another will sort themselves
  - Class size might be a function of the school's or student's resources, which will also affect scores
  - Even over time, the same student's resources or desire/fit for a certain class might change
- *Citation:* Example adapted from Krueger (1999) and Angrist & Pischke's *Mostly Harmless Econometrics*

Formalizing the Question at Hand

- Index students by i
- **Treatment:**, D_i, A 0/1 variable for whether i is treated (e.g., whether a student is in a small class)
- **Observed Outcome:** Y_i is the measured outcome for i
- **Potential Outcome:** Y_{di} is the outcome depending on whether D is 0 or 1
- **Treatment Effect:** The change in outcome from being treated, which is the causal effect of treatment:

$$TE_i = Y_{1i} - Y_{0i}$$

- Simplification for example: $TE_i = \beta$ for everyone, but Y_{i0} differs

Group Means, Treatment Effects and Selection

- Parameter of interest is β, the average effect of a small class setting:

$$\beta = E(Y_{i1}) - E(Y_{i0})$$

- Simplest idea: compare mean outcomes across groups:

$$\begin{aligned} E(Y|D=1) - E(Y|D=0) &= E(Y_{1i}|D=1) - E(Y_{0i}|D=0) \\ &= E(Y_{0i} + \beta|D=1) - E(Y_{0i}|D=0) \\ &= \beta + \underbrace{(E(Y_{0i}|D=1) - E(Y_{0i}|D=0))}_{\text{Selection Bias}} \end{aligned}$$

- **Section Bias:** Bias arising from the fact that $E(Y_{0i}|D) \neq E(Y_{0i})$
- Why might there be selection bias in our running example?
  - Wealthy parents may put children in private school (small class size) but those children might have higher baseline test scores for other reasons (books in the home, tutors...)

The Data We See and the Data We Don't

An example table with only two students:

	Alice	Bob	Avg.
Test Score in Small Class (Y_{i1})	6	4	5
Test Score in Large Class (Y_{i0})	5	3	4
Treatment Effect (β)	1	1	1

- Truth: Small class size increases test scores by 1
- In reality we never see this table because Alice and Bob cannot both be in a small class and a big class
- Suppose that Alice enters a small class setting but Bob does not... what would this mean for our estimates?

The Data We See and the Data We Don't

An example table with only two students:

	Alice	Bob	Avg.
Test Score in Small Class (Y_{i1})	**6**		6
Test Score in Large Class (Y_{i0})		**3**	3
Treatment Effect (β)			**3**

- We end up comparing two people with different baselines
- The difference in means here is larger than the actual causal effect
- What about if we *knew* that Bob and Alice looked the same, but Bob still went to a big class?

The Data We See and the Data We Don't

An example table with only two students:

	Alice	Bob	Avg.
Test Score in Small Class (Y_{i1})	**4**	4	4
Test Score in Large Class (Y_{i0})	3	**3**	3
Treatment Effect (β)	1	1	1

- Now the numbers are correct!
- Clearly the key is making sure that Alice and Bob "look the same"
- Much (but not all) of econometrics in a nutshell: when do Bob and Alice look the same?

Recapping and Taking Stock

- When we talk about causality we are talking about **treatment effects**:

$$Y_{i1} - Y_{i0}$$

- The **Average Treatment Effect** is the average effect across all individuals:

$$ATE = E(Y_{i1}) - E(Y_{i0})$$

- Observational data is often plagued by **selection bias**
  - Major Reason: economic entities are purposeful and respond to incentives
  - The solution to this problem is to ensure that treated and untreated entities are drawn from the same population (i.e., *look the same*)
- Next, examine how **Randomized Controlled Trials** solve the selection issue

Randomization and Selection

So how to deal with selection?

- Crux of the problem: people's *choice* of treatment depends on Y_{di}.
- Solution: break this by force
- **Randomized Control Trial**: An experiment where a researcher randomly assigns subjects either to treatment or control group.
  - Idea (and language) comes from medical drug research
  - Experimental Ideal: Perfect compliance with group status
- Random assignment $\Rightarrow$ Y_{di} and D independent $\Rightarrow$ $E(Y_{i0}|D) = E(Y_{i0})$

$$\begin{aligned} E(Y|D=1) - E(Y|D=0) &= E(Y_{1i}|D=1) - E(Y_{0i}|D=0) \\ &= E(Y_{0i}+\beta|D=1) - E(Y_{0i}|D=0) \\ &= \beta + (E(Y_{0i}|D=1) - E(Y_{0i}|D=0)) \\ &= \beta \end{aligned}$$

The Tennessee STAR Experiment

Continuing with example from Krueger (1999)/Angrist & Pischke (2009)...

- From 1985-1986, $\sim 11.6k$ students from 80 schools randomly assigned to small classes or big classes
  - Teachers *also* randomly assigned
  - Assignment was done within schools
- Four cohorts were analyzed: Kindergarten - 3rd grade
- Every year, students were given the Stanford Achievement Test as a measure of outcomes
- No study is perfect:
  - Some students switched classes anyway
  - Some students dropped out
  - Students would switch treatment status over time
- Nevertheless, this is very close to the experimental ideal!

Checking Randomization

A. Students who entered STAR in kindergarten[b]

Variable	Small	Regular	Regular/Aide	Joint *P*-Value[a]
1. Free lunch[c]	.47	.48	.50	.09
2. White/Asian	.68	.67	.66	.26
3. Age in 1985	5.44	5.43	5.42	.32
4. Attrition rate[d]	.49	.52	.53	.02
5. Class size in kindergarten	15.1	22.4	22.8	.00
6. Percentile score in kindergarten	54.7	49.9	50.0	.00

- Blue box demonstrates randomization: students look similar across groups
- Green box is the treatment: average small class size has 7 fewer students
- Red box is the treatment effect: small class size students do about 5 percentile points better
- Source: Krueger, Alan. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics.* 1999.

Results and Findings in Pictures

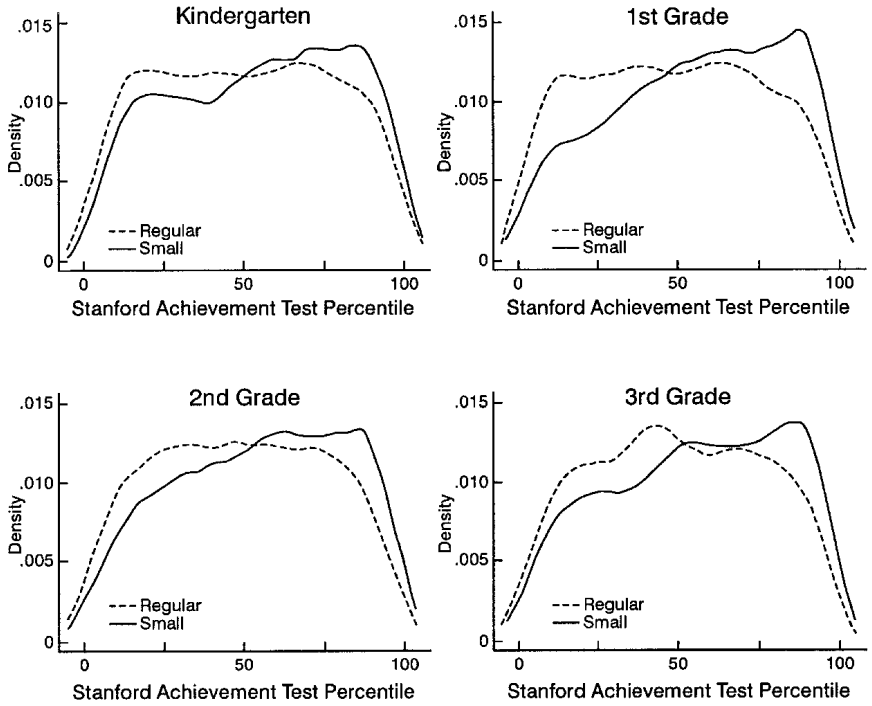

FIGURE I
Distribution of Test Percentile Scores by Class Size and Grade

- Lots of heterogeneity, but it looks like the *average* test scores are larger in small classes (later, we will formalize such comparisons)
- Source: Krueger, Alan. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics.* 1999.

Causality versus Prediction

- Estimate of β is attempting to answer a causal question: What is the effect of moving a student into a small class
- An equally valid question: Can we predict test scores from a student's class size?
  - In this case "selection bias" is a non-issue because we do not care *why* people in different size classes have different scores
  - Indeed, sometimes "causality" barely makes sense in a forecasting setting (e.g., stock prices)
- Prediction and forecasting can be very important and we will learn more about this towards the end of the semester.
- Examples of forecasting questions:
  - Given the price of an asset today, what will be its price tomorrow?
  - What is your best guess of a 4-year old's future SAT score, given their behavior in the Marshmallow Test?

Mapping to a Regression Equation

- An alternative representation of the same setup:

$$\begin{aligned} Y_i &= Y_{i0} + (Y_{i1} - Y_{i0}) \times D_i \\ &= \mu_0 + (Y_{i0} - \mu_0) + (Y_{i1} - Y_{i0}) \times D_i \\ &= \beta_0 + \beta_1 D_i + \varepsilon_i \end{aligned}$$

  where β_0 is the control group average, β_1 is the treatment effect, D_i is a $0/1$ variable for being in control/treatment and $\varepsilon_i = Y_{0i} - \mu_0$ is called the **residual** or **error** term
- The last line is an example of a **regression** equation
- Here the residual's purpose is clear: it is the part of Y_i that is not explained by the treatment

The Regression Framework

- D_i does not to be 0/1, so just let it be X_i for some variable:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \tag{1}$$

  which is what we call a linear regression equation
- Why this representation?
  - Sometimes we care about relationships that aren't 0/1. E.g., the relationship between income and consumption
  - This is very flexible for adding *more* variables than just one
  - Compactly reduces all data for i to one equation
- Subsequent lectures will discuss:
  1. Estimating β when X is continuous or includes several variables
  2. Doing inference on β (how precise is our estimate?)
  3. Causal inference when variation in X is not experimental, i.e., when X and ε might be correlated

External Validity and Structural Econometrics

- Suppose the experimental study on class size involved kindergarten students, but suppose we are interested in the impact of lowering class sized throughout a child's education. How can we predict what the impact is?
- One answer: do more experiments, but sometimes this is infeasible.
  1. Policy changes with long-term and/or macroeconomic effects
  2. Merger analysis
  3. Delivering personally-tailored nutrition advice?
- Another answer: build and validate a theoretical model.
- External validity motivates most **Structural Econometrics**, where the econometric analysis is based on a theoretical model.