Econometrics I

Lecture 1: Introduction

Paul T. Scott
NYU Stern

Fall 2021

Course Description

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, and discrete choice models.

Prerequisites for Studying Econometrics

Ideally, you should have experience with

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

Assignments and Grading

- 4 Problem sets: 15% each
  - 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.
  - 1-3 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
- Office hours (Zoom): Monday 5-6pm or by appointment (email).

Tentative course outline

Session	Date	Chapters	Topics and deliverables
1	9/9	1-4	Introduction, Least Squares Estimation
	9/16		No Class
2	9/23	1-4	Least Squares Estimation
3	9/30	5, 9	Asymptotics, Inference, Standard Errors
			Assignment 1 due
4	10/7	6	(Quasi-)Experiments, Endogeneity, Treatment Effects
5	10/14	8.1-8.5	Instrumental Variables, Simultaneity
6	10/21	10, 13	Instrumental Variables, Simultaneity
			Assignment 2 due
7	10/28	11	Panel Data, Fixed and Random Effects
8	11/4	14, 17, 18	Maximum Likelihood
			Final project proposals due
9	11/11		Model Selection, Machine Learning
10	11/18	14, 17, 18	Binary and Discrete Choice
	11/25		No Class - Thanksgiving
11	12/2		Structural Estimation
			Assignment 4 due
12	12/9		Structural Estimation
13	12/15		**Final project presentations (Wednesday!)**
	12/20		**Final projects due (Monday)**

What is Econometrics?

- **Experiments and Research Design:**
  - In natural sciences, randomized controlled trials are considered the gold standard.
  - In the social sciences, it's often hard to run experiments: macroeconomic policy, mergers and antitrust policy, major programs like universal basic income.
  - This is perhaps the main reason we have econometrics: in the absence of controlled experiments, we need to figure out how to learn what we want to learn from naturally occurring data.

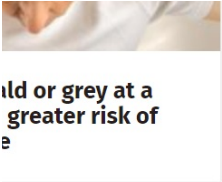
What is Econometrics?

- **Econometric Questions:**
  - Often 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
    - How many home runs will Shohei Ohtani hit in 2022?
    - *What is the relationship between growth and inequality?*

What is Econometrics?

- **Obstacles:**
  - Endogeneity: general term for an observed variable's being correlated with things we can't observe. Related: omitted variables.
  - Selection: Economic agents (people, firms, etc.) are purposeful and know more than we do about their personal situations!
  - Simultaneity: is the relationship between price and quantity increasing (supply) or decreasing (demand)? What is driving the changes? The world is complicated, and it's often too simplistic to say *''this is the effect of X on Y''*.
  - External validity:
    - What will happen if we raise the minimum wage to levels not before seen?
    - Is the price variation in the data short-run or long-run? Do consumers/firms respond differently to the two types of variation?
    - Related: structural econometrics, theory building.

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



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:**, T_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_{Ti} is the outcome depending on whether T 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_{0i} differs

Group Means, Treatment Effects and Selection

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

$$\beta = E(Y_{1i}) - E(Y_{0i})$$

- Simplest idea: compare mean outcomes across groups:

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

- **Section Bias:** Bias arising from the fact that $E(Y_{0i}|T) \neq E(Y_{0i})$
- 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_{1i})	6	4	5
Test Score in Large Class (Y_{0i})	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_{1i})	**6**		6
Test Score in Large Class (Y_{0i})		**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_{1i})	**4**	4	4
Test Score in Large Class (Y_{0i})	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**:

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

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

$$ATE = E(Y_{1i}) - E(Y_{0i})$$

- 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.
  - Experimental Ideal: Perfect compliance with group status
- Random assignment $\Rightarrow$ Y_{Ti} and T independent $\Rightarrow$ $E(Y_{0i}|T) = E(Y_{0i})$

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

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

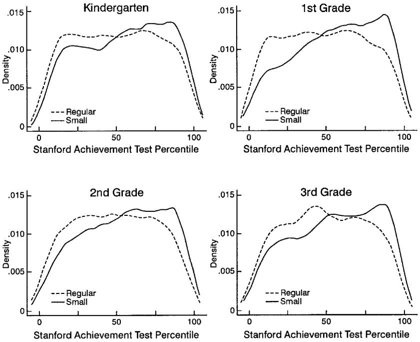


FIGURE I
Distribution of Test Percentile Scores by Class Size and Grade

- 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
- 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:

$$\begin{aligned} Y_i =& Y_{0i} + (Y_{1i} - Y_{0i}) \times T_i \\ =& \mu_0 + (Y_{0i} - \mu_0) + (Y_{1i} - Y_{0i}) \times T_i \\ =& \beta_0 + \beta_1 T_i + \varepsilon_i \end{aligned}$$

  where β_0 is the control group average, β_1 is the treatment effect, T_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. In this context, it comes from the difference between mean ability and the student's own ability.

The Regression Framework

- T_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