Advanced Econometrics: Introduction

Huan Deng

April 29, 2025

Outline

Basics

Learning Goals

In this course, we have four main goals:

1. Develop language and concepts to discuss econometrics and empirical research
2. Grasp the potential outcome framework
3. Understand what is identification
4. Learn popular research designs (For example, RCT, IV, DID)

Hopefully, after this semester, you can freely read empirical papers in leading journals

Estimand, Estimate and Estimator

- Estimand: the quantity to be estimated, typically population parameters that characterize the population
  - e.g., $E(Y) \equiv \mu$, or β from $E(Y|X) = X\beta$
- Estimate: the approximation of the estimand using a finite data sample: $Y_1, \ldots, Y_N$
- Estimator: the mapping from the data to the estimate

$$\hat{\mu} = f(Y_1, \ldots, Y_N) = \frac{1}{N} \sum_{i=1}^{N} Y_i$$

Terminology and Basic Statistics

Distributions of estimates:

1. **Finite Sample Distribution:** Exact distribution for a given sample size N.
2. **Asymptotic Distribution:** Approximation as $N \to \infty$.
3. **Bootstrap:** Provides a practical approximation to finite sample distribution.

Examples:

- Finite sample: For i.i.d. samples, $E(\hat{\mu}) = \mu$, $V(\hat{\mu}) = \frac{\sigma^2}{N}$.
- Asymptotic: $\hat{\mu} \xrightarrow{p} \mu$, $\sqrt{N}(\hat{\mu} - \mu) \xrightarrow{d} N(0, \sigma^2)$.

Terminology and Basic Statistics

The estimand is something we are interested in:

- For the sample mean $\hat{\mu}$, the estimand is the population mean μ.
- Often, estimands are defined as limits of estimates: $\hat{\mu} \xrightarrow{p} \mu$.

Example: OLS Estimator

$$\hat{\beta}_{\text{OLS}} \xrightarrow{p} \frac{\text{Cov}(Y, X)}{\text{Var}(X)} = \beta + \frac{\text{Cov}(\epsilon, X)}{\text{Var}(X)}$$

- Bias occurs if $\text{Cov}(\epsilon, X) \neq 0$.
- Estimand is the ideal, while all we have is the estimate

A Simple Example

Does education increase one's earning?

- Economic model: $Y_i = \beta X_i + \alpha A_i$, where Y_i is earning, X_i is years of schooling, and A_i represents ability
- Observables: Y, X; Unobservables: A.
- The econometric model: $Y_i = \beta X_i + \epsilon_i$, where $\epsilon_i = \alpha A_i$.

OLS estimate converges to:

$$\hat{\beta}_{\text{OLS}} \xrightarrow{p} \beta + \frac{\text{Cov}(\epsilon, X)}{\text{Var}(X)} = \beta + \alpha \frac{\text{Cov}(A, X)}{\text{Var}(X)}$$

- Bias arises if $\alpha \neq 0$ (ability matters) and $\text{Cov}(A, X) \neq 0$ (ability is correlated with schooling).

Omitted Variable Bias

- We suffer from omitted variable bias (OVB) if our econometric model fails to include a variable that correlates with both the dependent and independent variables
- Example: Ability A is omitted but correlates with schooling X and earnings Y.
- If we can find and include the omitted variable, then we are done
- Tons of papers claim "we include as many controls as possible"
- Good Controls v.s. Bad Controls

Selection Bias

- Human beings make decisions based on their self-interests and private information
- This leads to the self-selection bias: people with higher unobserved ability might be more likely to have more years of schooling and earn more
- Even schooling has no effect on earning, we still observe positive correlation between schooling and earning
- But correlation is not causality!

Strategies to Establish Causality

- Assume away unobserved heterogeneity in ability: $A_i = A$ for all i
  - Very strong assumption, not desirable!
- Selection on observables: we know and have access to all variables that affect schooling choices
  - It requires that the econometrician knows as much as the decision maker and has really good data.
  - In some settings, selection on observables might be plasuable, more on this later.

Strategies to Establish Causality

- Suppose we can randomly assign schooling levels X
- IV or natural experiment, for example, distance to schools
- Directly model people's self-selection, but we need exclusion restriction (some variable only affects schooling decision but not earnings), more on this later

Identification

Identification

Basic Definitions

Identification

Loosely speaking, identification analysis maps data to parameters:

- Define the parameters of interest
- Use consistent estimators to estimate these parameters

There are various identification arguments, here we use Lewbel (2019) to discuss relevant concepts

Historical Context

Key milestones in econometric identification:

- **Ceteris Paribus:** Alfred Marshall (1890) and William Petty (1662).
- **Simultaneous equations:** Philip Wright (1915) and Sewall Wright (1928).
- **Instrumental variables:** Early developments by the Cowles Foundation (1940s-50s).
- **Causal inference:** Haavelmo (1943) and Rubin (1974).

Basic Definitions

Let θ denote the parameter(s) to be identified, and let ϕ represent observable information derived from data. Identification requires that θ be uniquely determined by ϕ under the given model M.

Definition

A parameter θ is point identified if no two distinct values of θ are observationally equivalent, i.e., ϕ uniquely determines θ.

Point Identification

Example: Linear Regression

$$Y = X\theta + e,$$
$$\text{where } \mathbb{E}[eX] = 0 \text{ and } \mathbb{E}[X^2] \neq 0.$$

The parameter θ is identified as $\theta = \mathbb{E}[XY]/\mathbb{E}[X^2]$.

Set Identification

Set Identification: A parameter is set identified if its possible values form a set rather than a single point.

- For example, we might not know the exact causal effect of a medicine, but if we know the sign or upper/lower bound of the effect, we can still inform the policymakers
- Set identification might require fewer assumptions, thus with high credibility

Identification

Some Examples of Identification

Wright–Cowles Identification

Identification using reduced-form regression coefficients, common in structural econometrics. For instance, in a supply-demand system:

$$Y = bX + cZ + U,$$
$$Y = aX + \varepsilon.$$

Instrument Z aids in identifying a via exclusion restrictions.

Distribution-Based Identification

- ϕ is the distribution function of the data. Identification relies on the uniqueness of parameters θ under the assumed distributional form
- For this, we assume we know the entire distribution of the data

Extremum-Based Identification

Parameters θ are identified as solutions to extremum problems:

$$\theta = \arg\max_{\zeta} G(\zeta),$$

where $G(\zeta)$ is an objective function such as a likelihood function or moment condition.

Identification Argument

- For some simple problems, we can prove identification in a rigorous way
- Mathematically express the estimand as a function of the observables
- Argument looks like: some carefully extracted variation provides identifying power
- For some complicated settings, especially structural papers, we provide heuristic argument such as " xx parameter is mainly identified by xx variation," despite we can't explicitly write out the formula

Potential Outcome Framework

Potential Outcome

- We all know that the rooster's crow does not cause the sun to rise. **Correlation does not imply causation**.
- We adopt the **potential outcome framework** proposed by Rubin (1974) to define causal effects.
- The concept of potential outcome originated in randomized experiments but is now widely used in observational studies
- Causal effects of a policy is the comparison of potential outcomes.

Potential Outcome

Key Points:

- Focus on a binary treatment D:
  - $D_i = 1$: Agent receives treatment.
  - $D_i = 0$: Agent does not receive treatment.
- Outcomes:
  - $Y_i(1)$: Outcome under treatment.
  - $Y_i(0)$: Outcome without treatment.

Examples:

- Surgery/medication as treatment
- Welfare program as treatment
- College attendance as treatment

Potential Outcome

- The causal effect of the policy is $Y_i(1) - Y_i(0)$. A few clarifications:
  - The definition of causal effect depends on the potential outcomes, but it **does not** depend on which potential outcome is realized.
  - the causal effect is the comparison of potential outcomes, for the **same unit**, at the **same moment** in time.
  - **Stable Unit Treatment Value Assumption (SUTVA)**: no interference, and no hidden variations of treatments

Average Treatment Effects

Key Parameters:

- Average Treatment Effect (ATE):

$$\text{ATE} = E[Y_i(1) - Y_i(0)]$$

- Average Treatment Effect for the Treated (ATT):

$$\text{ATT} = E[Y_i(1) - Y_i(0) \mid D_i = 1].$$

- Average Treatment Effect for the Untreated (ATU):

$$\text{ATU} = E[Y_i(1) - Y_i(0) \mid D_i = 0].$$

Conditional ATE:

$$\text{ATE}(x) = E[Y_i(1) - Y_i(0) \mid X_i = x].$$

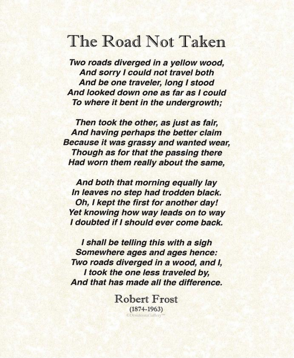
The Fundamental Problem in Causal Inference

- Only one of the potential outcomes can be realized and thus observed. This is the fundamental problem in causal inference (Holland, 1986):

$$Y_i^{obs} = Y_i(1)D_i + Y_i(0)(1 - D_i) \tag{1}$$

- "Causal inference is a missing data problem." — Donald Rubin

The Fundamental problem in Causal Inference



The Road Not Taken

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim
Because it was grassy and wanted wear,
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I,
I took the one less traveled by,
And that has made all the difference.

Robert Frost
(1874-1963)

Figure: We won't know "The Road Not Taken"

Figure: Unless an angel shows us the counterfactual

The Fundamental problem in Causal Inference

ATT can be decomposed into two parts:

$$
\begin{aligned}
ATT &= E[Y_i(1) - Y_i(0)|D_i = 1] \\
&= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] \\
&= \underbrace{\{E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0]\}}_{\text{observed difference}} \\
&\quad + \underbrace{\{E[Y_i(0)|D_i = 0] - E[Y_i(0)|D_i = 1]\}}_{\text{selection bias}}
\end{aligned} \tag{2}
$$

Random Assignment

- So what can we do? If we know the assignment of treatment is random, then the selection bias will disappear.
- Random assignment leads to:

$$(Y_i(0), Y_i(1)) \perp D_i, \tag{3}$$

  then we can estimate causal effects from the data.
- We will discuss more details later

Some Suggestions

Some Personal Take

- No Free Lunch: you can't find an empirical method that outperforms all other methods for all contexts
- Yet, not every identification strategy is born equal
  - "I control for as many confounders as possible" is really weak
  - Identification by functional form is rarely accepted
  - Nonparametric identification allowing for unobserved heterogeneity is well accepted

Some Personal Take

- For each research question, focus on one identification strategy and leave other estimators or methods to the “robustness checks”
- Identification is fundamental, an empirical paper can’t get published in a good journal if the identification argument is weak
- Aim for nonparametric identification + simple yet robust estimators (OLS, for example)

Some Personal Take

- Be an expert on the research topic you work on, at least, know the institutional background
  - For example, if you use DID, then you should know the relevant policy so that you can better defend the "parallel trend" and "no anticipation" assumptions
- Good identification arguments lie in a deep understanding of policies and economic theories, but not just statistics

Some Personal Take

- Reduced form or structural modeling? Why not both? More on this later
- In addition to Stata, learning a programming language will help you go a long way
- Python is overall the best, R is good for data wrangling and visualization, Julia is fast for computation but lacks community support
- Don't reinvent the wheel.