Introductory Econometrics I

A Short Introduction to

Causal Inference and Program Evaluation

Yingjie Feng

School of Economics and Management

Tsinghua University

May 14, 2024

Outline

1. Causal Inference
2. Potential Outcomes
3. An Example: Vaccine Effectiveness
4. Causal Inference and Linear Regression

Causal Inference

- So far we have focused on OLS regression from a technical perspective:
  - Algebraic properties
  - Unbiasedness, Consistency
  - Hypothesis testing, confidence intervals
  - Functional form, dummy variables
  - etc.
- But do not forget OLS is just a "tool".
- In today's class we "deviate" from OLS technicalities and try to formally talk about *causality*, which is key to economics and many other disciplines.
  - But you will see how our causal analysis is related to OLS.

Causal Inference

- The goal of program evaluation is to assess the causal effect of program or policy interventions. Examples:
  - Job training programs on earnings and employment
  - Minimum wage on employment
  - Military service on earnings and employment
- In addition, we may be interested in the effect of variables that do not represent policy interventions. Examples:
  - Interest rate on credit card usage
  - Terrorist risk on economic behavior
  - German reunification on economic growth

Causes and Effects

- Important distinction between *cause* and *effect*:
  - *Cause*: an event that generates some phenomenon.
  - *Effect*: the consequence (or one of the consequences) of the cause.
- Asymmetry in the difficulty of learning about the cause of an effect versus learning about the effect of a cause.
  - Learning causes of an effect is unclear and hard to be formalized as a research question (in economics).
    - There might be 1 000 causes of an effect...
    - We can keep searching for a more "fundamental" cause
  - But learning the effect of a cause is a well-defined question, as we'll see.

Causal Effects in a Potential Outcome Framework

- We will employ a **causal inference framework** in today's class.
- Two key ingredients:
  - **Potential Outcomes**: each individual has a different outcome corresponding to each level that the treatment takes.
    - Example: your potential income in the labor market if you had studied at Tsinghua, and the potential income if you had *not* studied at Tsinghua
    - Of course, in the real world we can only observe *one* of the potential outcomes
  - **Assignment Mechanism**: each individual is assigned treatment based on some mechanism, which guides how estimation and inference will be conducted.
    - This is analogous to the question in the regression model, "how is the key explanatory variable of interest determined?"

Outline

Potential Outcomes: Causation as Manipulation

- Causal analysis: must have ability to expose or not expose each unit to action of cause.
- Essential "each unit be potentially exposable to any one of the causes" [Holland, 1986].
  - If units could have been exposed to cause but they were not in practice: no problem.
  - If units could not have been exposed to cause in any state of world: our cause might not really be a cause.
  - Example: worker's education level versus worker's gender.
- **Important**: manipulability may require some imagination and has to be understood in context.

Potential Outcomes: Causation as Manipulation

- Each unit has as many potential outcomes as different possible treatments there are.
  - Called "potential" outcomes because **only one** of them is observed.
  - Observed outcome: the one corresponding to level of the treatment actually selected by (or assigned to) the unit.
- This introduces the idea of **counterfactual**: what would the outcome of this unit look like if the unit had been exposed to a different treatment?
  - "... effect of a cause is always **relative** to another cause." (Holland, 1986)
  - Good research needs to clearly specify what the counterfactual is.

Potential Outcomes: Causation as Manipulation

- What is the counterfactual in your analysis?
- Examples:
  - "Reduced infection probability due to vaccination"?
    - Counterfactual seems to be very clear: do not receive a vaccine (probably get a placebo in a randomized experiment)
  - "High income due to studying at Tsinghua?"
    - Less clear: study at PKU? study at any other university in China? do not go to any college? go to army?
  - "Lower wage due to being a woman"?
    - Very hard to understand, unless a special design is possible.
    - An excellent example is "Are Emily and Greg More Employable Than Lakisha and Jamal?" (Bertrand and Mullainathan, 2004, AER).

Basic Binary Treatment Setup

- Each unit i is exposed to a binary treatment.
  - $d_i = 1$ if unit i received treatment cause
  - $d_i = 0$ if unit i received the control cause
- Each unit i has two potential outcomes:
  - $y_i(1)$: outcome that would be observed if i were exposed to treatment cause
  - $y_i(0)$: outcome that would be observed if i were exposed to control cause
- Observed data: $(y_i, d_i)'$ where

$$y_i = d_i \cdot y_i(1) + (1 - d_i) \cdot y_i(0)$$

- This setup can be extended to multi-valued or even continuous treatment.

Treatment Effects

- For each unit i, the causal effect of the treatment is $y_i(1) - y_i(0)$
- In this framework we can define many parameters of interest for the population
  - Average treatment effect (ATE): $\mathbb{E}[y_i(1) - y_i(0)]$
  - Average treatment effect on the treated (ATT): $\mathbb{E}[y_i(1) - y_i(0)|d_i = 1]$
  - Average treatment effect on the untreated (ATU): $\mathbb{E}[y_i(1) - y_i(0)|d_i = 0]$
  - Quantile treatment effect (QTE): $Q_\tau[y_i(1)] - Q_\tau[y_i(0)]$
  - etc.

Outline

Vaccines Effectiveness

- Vaccine development procedure often consists of an experiment to check its effectiveness in the real world.
- Does a vaccine decrease the probability of infection with a virus?
  - n participants: $1 \leq i \leq n$
  - Binary explanatory variable ("treatment")
    - $d_i = 1$ if vaccinated; $d_i = 0$ if not
    - $n_1 = \sum_{i=1}^{n} d_i$, $n_0 = \sum_{i=1}^{n}(1 - d_i)$
  - Binary dependent variable ("outcomes")
    - **Each person** has **two potential outcomes**: $y_i(1)$ if vaccinated; $y_i(0)$ if not vaccinated
    - For each $d = 0, 1$, $y_i(d) = 1$ if i is infected with Covid; $y_i(d) = 0$ if not
  - Observed outcome: $y_i = d_i y_i(1) + (1 - d_i) y_i(0)$
    - $y_i = y_i(1)$ if $d_i = 1$; $y_i = y_i(0)$ if $d_i = 0$

Vaccines Effectiveness

- Vaccine development procedure often consists of an experiment to check its effectiveness in the real world.
- Does a vaccine decrease the probability of infection with a virus?
  - n participants: $1 \leq i \leq n$
  - For each person i, we observe either $y_i(1)$ or $y_i(0)$

	Vaccinated	Not vaccinated
infected	$y_i(1) = 1$	$y_i(0) = 1$
not infected	$y_i(1) = 0$	$y_i(0) = 0$

  - Given a sample,
    - ⋆ $\sum_{i:d_i=1} y_i$: # of infected people in the treatment group
    - ⋆ $\sum_{i:d_i=0} y_i$: # of infected people in the control group
  - Difference in "infection rate" between two groups:

$$\frac{1}{n_1} \sum_{i:d_i=1} y_i - \frac{1}{n_0} \sum_{i:d_i=0} y_i$$

What does OLS estimate in this example?

- A "binary-binary" regression:

$$y_i = \beta_0 + \delta_0 d_i + u_i$$

  - In PS1 you show that the OLS estimator $\hat{\delta}_0$ is equivalent to a simple difference-in-means estimator:

$$\hat{\delta}_0 = \frac{1}{n_1} \sum_{i:d_i=1} y_i - \frac{1}{n_0} \sum_{i:d_i=0} y_i$$

  - Statistically, $\hat{\delta}_0$ is estimating the difference in population means

$$\tilde{\delta}_0 = \mathbb{E}[y_i(1)|d_i = 1] - \mathbb{E}[y_i(0)|d_i = 0]$$

  - This amounts to assuming $\mathbb{E}[u_i|d_i] = 0$ in the regression framework, and thus

$$\beta_0 = \mathbb{E}[y_i|d_i = 0], \quad \beta_0 + \delta_0 = \mathbb{E}[y_i|d_i = 1], \quad \delta_0 = \tilde{\delta}_0$$

Selection Bias

$$y_i = \beta_0 + \delta_0 d_i + u_i$$

- But does it make sense to assume $\mathbb{E}[u_i|d_i] = 0$? Or equivalently, are we interested in $\tilde{\delta}_0 = \mathbb{E}[y_i(1)|d_i = 1] - \mathbb{E}[y_i(0)|d_i = 0]$?

$$\mathbb{E}[y_i(1)|d_i = 1] - \mathbb{E}[y_i(0)|d_i = 0]$$

$$= \underbrace{\mathbb{E}[y_i(1) - y_i(0)|d_i = 1]}_{\text{average treatment effect on the treated}} + \underbrace{\mathbb{E}[y_i(0)|d_i = 1] - \mathbb{E}[y_i(0)|d_i = 0]}_{\text{Selection bias}}$$

- Maybe interested in average treatment effect on the treated (ATT)
  - If selection bias=0, then the answer is yes.
  - Otherwise, the estimator is biased (for ATT)!
  - Probably, healthier people self-selected to get a vaccine; risk-averse people self-selected not to get a vaccine

Randomized Experiments v.s. Observational Data

- In the vaccine example, researchers may remove bias by randomized experiment which makes the following hold:

$$(y_i(0), y_i(1)) \perp\!\!\!\perp d_i \quad \Rightarrow \quad \mathbb{E}[y_i(0)|d_i = 1] = \mathbb{E}[y_i(0)|d_i = 0]$$

- However, in many cases we cannot randomize and have to rely on observational data.
- One possible solution is "selection on observables".

$$(y_i(0), y_i(1)) \perp\!\!\!\perp d_i \mid \mathbf{x}_i$$

  - "Controlling for" some covariates $\mathbf{x}_i$, treatment becomes (as if) randomized
  - This is analogous to the reason why you want to run multiple regression rather than a simple regression

Outline

Relationship with Linear Regression

- How is the potential outcome framework related to the linear regression

$$y_i = \beta_0 + \delta \cdot d_i + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

- OLS theory itself does not tell us if δ can be interpreted as an average treatment effect (or other causal parameters).
- Some general conclusions are $d_i \perp\!\!\!\perp (Y_i(0), Y_i(1))$
  - In a randomized experiment, OLS estimator $\hat{\delta}$ can identify the average treatment effect. We have actually shown that.
  - Under $\underbrace{\text{selection on observables}}$, OLS estimator does *not* identify a causal parameter unless very restrictive assumptions are imposed.

$$d_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid X_i$$

Relationship with Linear Regression

- What is the "correct" regression model we should work with under selection on observables?
- Recall "zero conditional mean" assumption in our OLS theory implies we try to estimate the conditional expectation of outcome given the covariates.
- Write $v_i(0) = y_i(0) - \mathbb{E}[y_i(0)]$ and $v_i(1) = y_i(1) - \mathbb{E}[y_i(1)]$. Then, $\rightarrow g(X_i) = 0.$

$$\mathbb{E}[V_i(1) - V_i(0) \mid X_i] = 0$$

$$\Longleftrightarrow \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i] = \underbrace{\mathbb{E}[Y_i(1) - Y_i(0)]}_{\text{constant}}$$

$$\begin{aligned}
\mathbb{E}[y_i|d_i, \mathbf{x}_i] &= \mathbb{E}[y_i(0) + d_i(y_i(1) - y_i(0))|d_i, \mathbf{x}_i] \\
&= \mathbb{E}[y_i(0)|d_i, \mathbf{x}_i] + d_i\mathbb{E}[y_i(1) - y_i(0)|d_i, \mathbf{x}_i] \\
&= \mathbb{E}[y_i(0)|\mathbf{x}_i] + d_i\mathbb{E}[y_i(1) - y_i(0)|\mathbf{x}_i] \\
&= \underbrace{\mathbb{E}[y_i(0)]}_{\beta_0} + d_i\underbrace{\mathbb{E}[y_i(1) - y_i(0)]}_{d_i \times \delta \to \text{ATE}} + \underbrace{\mathbb{E}[v_i(0)|\mathbf{x}_i]}_{f(X_i)} + \underbrace{d_i\mathbb{E}[v_i(1) - v_i(0)|\mathbf{x}_i]}_{d_i \times g(X_i)}
\end{aligned}$$

$f(X_i) = \beta_1 X_1 + \cdots + \beta_k X_k$?

$g(X_i) = 0$??

Relationship with Linear Regression

- Many other methods are available. Regression as a "technique" may still be useful.
- For example, note the following fact under selection on observables:

$$
\begin{aligned}
\mathbb{E}[y_i(1) - y_i(0)] &= \mathbb{E}\Big[\mathbb{E}[y_i(1) - y_i(0)|\mathbf{x}_i]\Big] \\
&= \mathbb{E}\Big[\mathbb{E}[y_i(1)|\mathbf{x}_i] - \mathbb{E}[y_i(0)|\mathbf{x}_i]\Big] \\
&\overset{?}{=} \mathbb{E}\Big[\mathbb{E}[y_i(1)|\mathbf{x}_i, d_i = 1] - \mathbb{E}[y_i(0)|\mathbf{x}_i, d_i = 0]\Big] \\
&= \mathbb{E}\Big[\mathbb{E}[y_i|\mathbf{x}_i, d_i = 1] - \mathbb{E}[y_i|\mathbf{x}_i, d_i = 0]\Big]
\end{aligned}
$$

$$
y_i = \begin{cases} y_i(1) & \text{if } d_i = 1 \\ y_i(0) & \text{if } d_i = 0 \end{cases}
$$

- The last line suggests another way to estimate ATE based on regression techniques. (How?)

① reg y_i on X_i if $d_i = 1$

$X_i \hat{\beta}$: estimate of $\mathbb{E}[y_i \mid X_i, d_i = 1]$.

② reg y_i on X_i if $d_i = 0$.

$X_i \hat{\gamma}$: estimate of $\mathbb{E}[y_i \mid X_i, d_i = 0]$.

③ $$\widehat{ATE} = \frac{1}{n} \sum_{i=1}^{n} \underbrace{X_i (\hat{\beta} - \hat{\gamma})}.$$

$$\hat{\mathbb{E}}\left[\hat{\mathbb{E}}[y_i \mid X_i, d_i = 1] - \hat{\mathbb{E}}[y_i \mid X_i, d_i = 0] \right].$$