Potential Outcomes Framework

Prof. Tzu-Ting Yang
楊子霆

Institute of Economics, Academia Sinica
中央研究院經濟研究所

March 26, 2026

Causal Effect and Potential Outcomes Framework

- Estimating causal effect of treatment is a challenging task
  - Because **we can NOT observe counterfactual outcomes** if one had chosen different treatments
- In order to obtain causal effect, we need to compare **observed outcomes** with **counterfactual outcomes**
- The **potential outcomes framework** provides a way to think about causal effects in a structured way

Potential Outcomes Framework

Potential Outcomes Framework

Treatment Status

- The potential outcomes framework is developed by statistician Donald Rubin
  - Rubin, Donald. 1974. **Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies**. *Journal of Educational Psychology*, 66(5).

- Treatment: the intervention/variable whose effect we are interested in

- D_i: a dummy that indicate whether individual i receive treatment or not
$$D_i = \begin{cases} 1 & \text{if individual } i \text{ received the treatment} \\ 0 & \text{otherwise.} \end{cases}$$

- Examples:
  - Attend graduate school or not
  - Have health insurance or not
  - Win a lottery or not
  - Increase corporate tax rate or not
  - Democracy v.s. Dictatorship

Potential Outcomes Framework

Treatment Status

- D_i can be a multiple valued (continuous) variable

$$D_i = s$$

- Examples:
  - Schooling years
  - Number of children
  - Number of polices
  - Number of advertisements
  - Money supply
  - Income tax rate
- **In the following slides, we focus on the case when a treatment D_i is a binary variable**

Potential Outcomes Framework

Potential Outcome

- A potential outcome is the outcome that would be realized according to which treatment an individual received
- Suppose there are **two treatments** for each individual:
  - $D_i = 1$
  - $D_i = 0$
- Thus, each individual *i* has **two potential outcomes** and one for each value of the treatment Y_i^D
  - Y_i^1: Potential outcome for an individual *i* if getting treatment
  - Y_i^0: Potential outcome for an individual *i* if not getting treatment

- **Example:**
  - Annual earnings if attending graduate school
  - Annual earnings if not attending graduate school
- Again, potential outcome can be Y_i^s:
  - s can be continuous
  - More than two potential outcomes
- **How many treatments we have, how many potential outcomes will be**

Potential Outcomes Framework

Observed Outcome

- Observed outcome: for each particular individual, we only can observe one potential outcome
- Observed outcome Y_i is realized as

$$\mathrm{Y}_i = \mathrm{Y}_i^1 D_i + \mathrm{Y}_i^0 (1 - D_i)$$

or

$$\mathrm{Y}_i = \begin{cases} \mathrm{Y}_i^1 & \text{if } D_i = 1 \\ \mathrm{Y}_i^0 & \text{if } D_i = 0 \end{cases}$$

- **Only one potential outcome can be realized**
- The unobserved outcome is called the "**counterfactual**" outcome

Casual Effects

Casual Effect

- **Causal effect:** the comparisons between the potential outcomes under each treatment
  - The differences between **observed (potential) outcome** and **counter-factual (potential) outcome**

Casual Effect for an Individual

Casual Effect for an Individual

- Individual Treatment Effect (ITE):

$$\tau_i = \mathrm{Y}_i^1 - \mathrm{Y}_i^0$$

  - **Interpretation:** The difference between an individual i's potential outcome under treatment v.s. without treatment
  - **Example:**
    - ⋆ The difference in individual i's earnings if he/she attends graduate school v.s. not attending graduate school
  - We usually cannot identify the ITE

Individual Treatment Effect (ITE)

An Example

- Imagine a population with 4 people

i	D_i	Y_i^1	Y_i^0	Y_i	$\mathrm{Y}_i^1 - \mathrm{Y}_i^0$
David	1	3	?	3	?
Tina	1	2	?	2	?
Mary	0	?	1	1	?
Bill	0	?	1	1	?

- We want to evaluate the effect of attending graduate school on the annual earnings
  - D_i: Attending graduate school $D_i = 1$, otherwise $D_i = 0$
  - Y_i^1: (Potential) annual earnings if individual i attend graduate school
  - Y_i^0: (Potential) annual earnings if individual i do not attend graduate school
  - Y_i: Observed annual earnings for individual i

Individual Treatment Effect (ITE)

An Example

- What is Individual causal effect (ITE) of attending graduate school for David?
  - We only observe the annual earnings for David who attended graduate school
  - Only observe Y^1
- What is Individual causal effect (ITE) of attending graduate school for Bill?
  - We only observe the annual earnings for Bill who did not attend graduate school
  - Only observe Y^0

Individual Treatment Effect (ITE)

An Example

- Suppose **we can observe counterfactual outcomes**

i	D_i	Y_i^1	Y_i^0	Y_i	$Y_i^1 - Y_i^0$
David	1	3	2	3	1
Tina	1	2	1	2	1
Mary	0	1	1	1	0
Bill	0	1	1	1	0

- The ITE for David: $\alpha_{David} = 1$
- The ITE for Bill: $\alpha_{Bill} = 0$

Causal Effect for General Population

Causal Effect for General Population

- People might be more interested in the **causal effect for general population**
- Understand the treatment effect (causal effect) for general population:
  - Estimate the **population average of the individual treatment effects**

Review: Expectation

- We usually use $\mathrm{E}[Y_i]$ (the expectation of a variable Y_i) to denote **population average** of Y_i
- Suppose we have a population with N individuals

$$\mathrm{E}[\mathrm{Y}_i] = \frac{1}{N}\sum_{i=1}^{N} \mathrm{Y}_i$$

Causal Effect for General Population

- Average Treatment Effect (ATE):

$$\alpha_{\text{ATE}} = \mathrm{E}[\tau_i] = \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0] = \frac{1}{N}\sum_{i=1}^{N}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0]$$

  - **Interpretation:**
    - Average difference in potential outcomes for the whole population
  - **Example:**
    - Average effect of attending graduate school on annual earnings for whole population
    - Average difference between the earnings of the same individuals if they attend graduate schools v.s. if not attending graduate schools
  - We'll spend a lot time trying to identify/estimate ATE

Average Treatment Effect (ATE)

An Example

- Missing data problem also arises when we estimate ATE

i	D_i	Y_i^1	Y_i^0	Y_i	$Y_i^1 - Y_i^0$
David	1	3	?	3	?
Tina	1	2	?	2	?
Mary	0	?	1	1	?
Bill	0	?	1	1	?
$E[Y_i^1]$		?			
$E[Y_i^0]$			?		
$E[Y_i^1 - Y_i^0]$					?

- What is the effect of attending graduate school on average annual earnings of whole population (ATE)?
- $\alpha_{\text{ATE}} = E[Y_i^1 - Y_i^0] = ?$

Average Treatment Effect (ATE)

An Example

- Suppose **we can observe counterfactual outcomes**

i	D_i	Y_i^1	Y_i^0	Y_i	$Y_i^1 - Y_i^0$
David	1	3	2	3	1
Tina	1	2	1	2	1
Mary	0	1	1	1	0
Bill	0	1	1	1	0
$E[Y_i^1]$		1.75			
$E[Y_i^0]$			1.25		
$E[Y_i^1 - Y_i^0]$					0.5

- What is the effect of attending graduate school on average annual earnings of whole population (ATE)?

- $\alpha_{\text{ATE}} = \dfrac{1+1+0+0}{4} = 0.5$

Causal Effect for a Specific Sub-population

Review: Conditional Expectation

- We usually use $\mathrm{E}[Y_i|X_i = 1]$ to denote the average of Y_i in the population that has $X_i = 1$
- Suppose the population has N_1 individuals with $X = 1$

$$\mathrm{E}[Y_i|X_i = 1] = \frac{1}{N_1} \sum_{i:X=1} \mathrm{Y}_i$$

Causal Effect for a Specific Sub-population

- Conditional average treatment effect (CATE) for a subpopulation:

$$\alpha_{\text{CATE}} = \mathrm{E}[\tau_i | X_i = f] = \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0 | X_i = f] = \frac{1}{N_f} \sum_{i:X_i=f} [\mathrm{Y}_i^1 - \mathrm{Y}_i^0]$$

  - N_f is the number of units in the subpopulation
  - **Interpretation:**
    - Average difference in potential outcomes for the specific subgroup
  - **Example:**
    - Average effect of attending graduate school on annual earnings for **female** ($X_i = f$)
    - Average difference between the earnings of **female** if they attend graduate schools v.s. if not attending graduate schools

Causal Effect for Treatment Group

- Average treatment effect on the treated (ATT):

$$\alpha_{\text{ATT}} = \mathrm{E}[\tau_i \mid D_i = 1] = \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0 \mid D_i = 1] = \frac{1}{N_1} \sum_{i:D_i=1} [\mathrm{Y}_i^1 - \mathrm{Y}_i^0]$$

- Note that ATT is a special case of CATE
- **Interpretation:**
  - Average difference in potential outcomes for those who were treated
- **Example:**
  - Average effect of attending graduate school on annual earnings for those attending graduate school ($D_i = 1$)
  - Average difference between the earnings of those attending graduate schools vs. earnings if they had not attended graduate schools

Average Treatment Effect on Treated (ATT)

- Missing data problem also arises when we estimate ATT

i	D_i	Y_i^1	Y_i^0	Y_i	$Y_i^1 - Y_i^0$
David	1	3	?	3	?
Tina	1	2	?	2	?
Mary	0	?	1	1	?
Bill	0	?	1	1	?
$E[Y_i^1 \mid D_i = 1]$		2.5			
$E[Y_i^0 \mid D_i = 1]$			?		
$E[Y_i^1 - Y_i^0 \mid D_i = 1]$					?

- What is the effect of attending graduate school on average annual earnings for those who choose to attend graduate school (ATT)?
- $\alpha_{\text{ATT}} = E[Y_i^1 - Y_i^0 \mid D_i = 1] = ?$

Average Treatment Effect on Treated (ATT)

- Suppose **we can observe counterfactual outcomes**

i	D_i	Y_i^1	Y_i^0	Y_i	$Y_i^1 - Y_i^0$
David	1	3	2	3	1
Tina	1	2	1	2	1
Mary	0	1	1	1	0
Bill	0	1	1	1	0
$E[Y_i^1 \vert D_i = 1]$		2.5			
$E[Y_i^0 \vert D_i = 1]$			1.5		
$E[Y_i^1 - Y_i^0 \vert D_i = 1]$					1

- What is the effect of attending graduate school on average annual earnings of those who choose to attend graduate school (ATT)?

- $\alpha_{\text{ATT}} = \dfrac{1+1}{2} = 1$

Causal Effect for a Control Group

- Average treatment effect on the untreated (ATU):

$$\alpha_{\text{ATU}} = \mathrm{E}[\tau_i | D_i = 0] = \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0 | D_i = 0] = \frac{1}{N_0} \sum_{i:D_i=0} [\mathrm{Y}_i^1 - \mathrm{Y}_i^0]$$

- Note that ATU is a special case of CATE
- **Interpretation:**
  - Average difference in potential outcomes for those who were untreated
- **Example:**
  - Average effect of attending graduate school on annual earnings for those NOT attending graduate school ($D_i = 0$)
  - Average difference between the earnings of those NOT attending graduate school vs. earnings if they had attended graduate school

Average Treatment Effect on Untreated (ATU)

- Missing data problem also arises when we estimate ATU

i	D_i	Y_i^1	Y_i^0	Y_i	$Y_i^1 - Y_i^0$
David	1	3	?	3	?
Tina	1	2	?	2	?
Mary	0	?	1	1	?
Bill	0	?	1	1	?
$E[Y_i^1 \mid D_i = 0]$		?			
$E[Y_i^0 \mid D_i = 0]$			1		
$E[Y_i^1 - Y_i^0 \mid D_i = 0]$					?

- What is the effect of attending graduate school on average annual earnings for those who choose NOT to attend graduate school (ATU)?

- $\alpha_{ATU} = E[Y_i^1 - Y_i^0 \mid D_i = 0] = ?$

Average Treatment Effect on Untreated (ATU)

- Suppose **we can observe counterfactual outcomes**

i	D_i	Y_i^1	Y_i^0	Y_i	$Y_i^1 - Y_i^0$
David	1	3	2	3	1
Tina	1	2	1	2	1
Mary	0	1	1	1	0
Bill	0	1	1	1	0
$E[Y_i^1 \| D_i = 0]$		1			
$E[Y_i^0 \| D_i = 0]$			1		
$E[Y_i^1 - Y_i^0 \| D_i = 0]$					0

- What is the effect of attending graduate school on average annual earnings of those who choose NOT to attend graduate school (ATU)?

- $\alpha_{\text{ATU}} = \dfrac{0+0}{2} = 0$

Selection into Treatment

- In this numerical example, we have $\alpha_{\text{ATT}} > \alpha_{\text{ATE}} > \alpha_{\text{ATU}}$
- This may indicate selection into treatment:
  - Those who benefit most from the treatment (attending graduate school) are most likely to take it

Summary

- Individual Treatment Effect (ITE):

$$\alpha_{\mathsf{ITE}} = \mathrm{Y}_i^1 - \mathrm{Y}_i^0$$

- Average treatment effect (ATE):

$$\alpha_{\mathsf{ATE}} = \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0] = \frac{1}{N} \sum_i [\mathrm{Y}_i^1 - \mathrm{Y}_i^0]$$

- Conditional average treatment effect (CATE):

$$\alpha_{\mathsf{CATE}} = \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0 | X_i = f] = \frac{1}{N_f} \sum_{i:X_i=f} [\mathrm{Y}_i^1 - \mathrm{Y}_i^0]$$

Summary

- Average treatment effect on the treated (ATT):

$$\alpha_{ATT} = \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0 | D_i = 1] = \frac{1}{N_1} \sum_{i:D_i=1} [\mathrm{Y}_i^1 - \mathrm{Y}_i^0]$$

- Average treatment effect on the untreated (ATU):

$$\alpha_{ATU} = \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0 | D_i = 0] = \frac{1}{N_0} \sum_{i:D_i=0} [\mathrm{Y}_i^1 - \mathrm{Y}_i^0]$$

Which Causal Effect is most Relevant?

- There is no correct answer to this question
- ITE is the most specific effect
  - It is hard to identified
- ATE is the most general parameter
  - What if we give a treatment to the average person/firm/unit
- ATT is often interesting for policy evaluation
  - Policy makers might want to know the effect on those who took up the policy
- ATU is sometimes interesting for policy evaluation
  - We may be concerned about the people who did not take up the policy
  - How would they be affected if they took up the policy

Fundamental Problem of Causal Inference

Fundamental Problem of Causal Inference

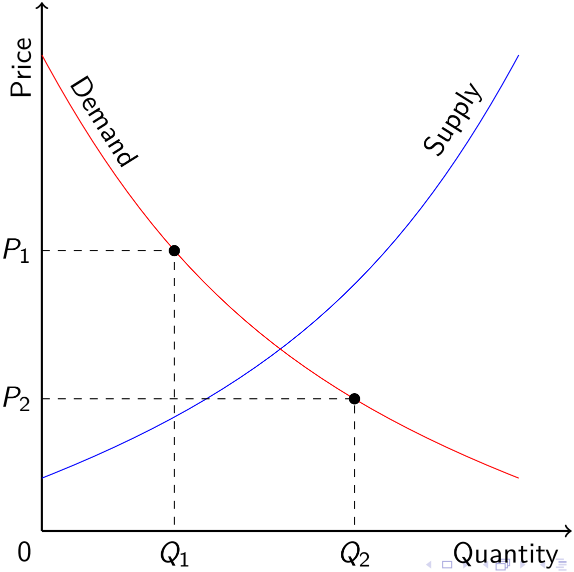
- We can never **directly observe** causal effects
  - ITE, ATE, CATE, ATT or ATU
- Because we can never observe both potential outcomes (Y_i^1, Y_i^0) for any individual
- For someone receiving the treatment $(D_i = 1)$
  - Y_i^1 is observed
  - But Y_i^0 is the **unobserved** counterfactual outcome
    - ★ It represents what would have happened to an individual i if assigned to control
- We need to compare **potential outcomes**, but we only have **observed outcomes**
- Causal inference is a set of statistical tools that deal with a **missing data problem**

Potential Outcome Framework in Economics

- The concepts of potential and observed outcomes are deeply ingrained in economics
  - A demand function represents the potential quantity demanded as a function of price
  - Only the quantity under equilibrium price is realized
  - Other quantities along demand curve are counterfactual

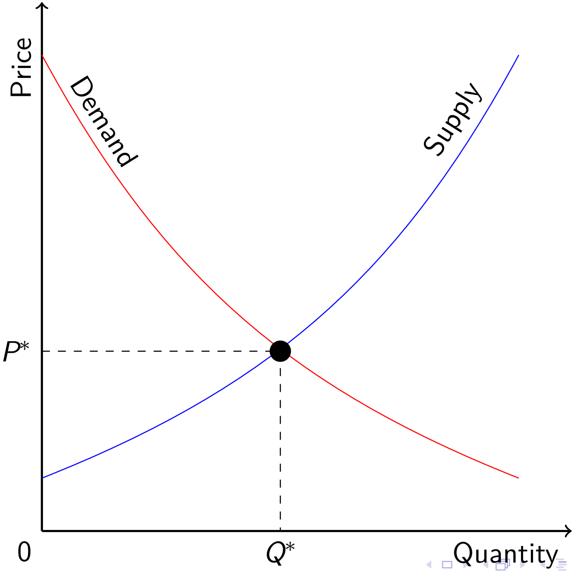
Potential Outcome Framework in Economics

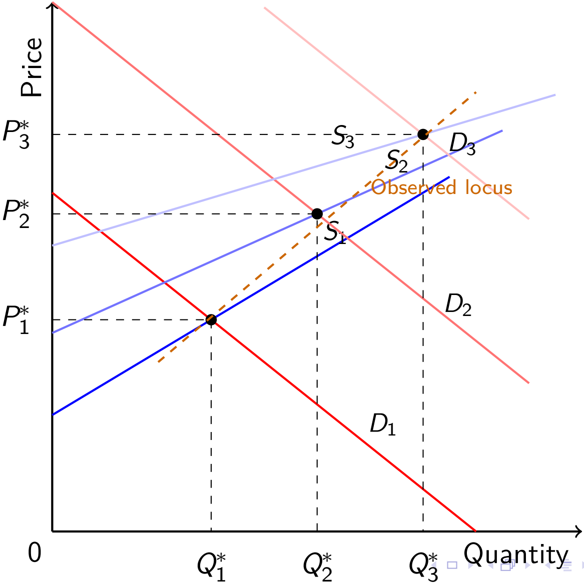
Demand and Supply



Potential Outcome Framework in Economics

Demand and Supply





Price
P_3^*
P_2^*
P_1^*
0
S_3
S_2
S_1
D_3
D_2
D_1
Observed locus
Q_1^*
Q_2^*
Q_3^*
Quantity

The Identification Problem

Why Observed Data Are Not Enough

- Each observed (P^*, Q^*) is an **equilibrium** — jointly determined by supply *and* demand
- When both curves shift simultaneously, the observed price–quantity locus can slope **upward** — the opposite of the Law of Demand
- Observed data alone **cannot verify** the downward-sloping demand curve
- Causal inference constructs the **counterfactual**: what quantity would be demanded at a different price, *holding the demand curve fixed*?
- A classic solution: use an **instrumental variable** that shifts supply only $\Rightarrow$ traces out the true demand curve

Stable Unit Treatment Value Assumption

Stable Unit Treatment Value Assumption (SUTVA)

Assumption

Observed outcomes are realized as

$$\mathrm{Y}_i = \mathrm{Y}_i^1 D_i + \mathrm{Y}_i^0 (1 - D_i)$$

- Implies that observed outcomes for an individual i are **unaffected** by the treatment status of other individual j
- Individual i's observed outcomes are only affected by his/her own treatment
- Rules out possible treatment effect from other individuals (spillover effect/externality)

Stable Unit Treatment Value Assumption (SUTVA)

- Could write out potential outcomes in a more extensive fashion, taking into account both one's own treatment status and the treatment status of others

$$\begin{cases} Y_i^{11} & \text{if } D_i = 1 \text{ and } D_j = 1 \\ Y_i^{10} & \text{if } D_i = 1 \text{ and } D_j = 0 \\ Y_i^{01} & \text{if } D_i = 0 \text{ and } D_j = 1 \\ Y_i^{00} & \text{if } D_i = 0 \text{ and } D_j = 0 \end{cases}$$

- **Example:**
  - Your health status depends on whether you smoke and your father/mother smoke

- **Contagion**:
  - The effect of being vaccinated on one's probability of contracting a disease depends on whether others have been vaccinated
- **Displacement**:
  - Police interventions designed to suppress crime in one location may displace criminal activity to nearby locations.
- **Communication**:
  - Interventions that convey information about commercial products, entertainment, or political causes may spread from individuals who receive the treatment to others who are nominally untreated

Stable Unit Treatment Value Assumption (SUTVA)

- SUTVA may be problematic, so we should choose the units of analysis to minimize interference across units.
- Recent literatures on causal inference are trying to deal with this assumption

Suggested Readings

- Chapter 1, Mastering Metrics: The Path from Cause to Effect
- Chapter 2, Mostly Harmless Econometrics
- Chapter 4, Causal Inference: The Mixtape