Lecture 02: Paramaters of interest and regression

PPHA 34600

Prof. Fiona Burlig

Harris School of Public Policy
University of Chicago

From last time: that pesky fundamental problem

Let's go back to our model:

- We have $i \in \{1, ..., N\}$ units
- $D_i \in \{0, 1\}$ is the treatment indicator for unit i
  - → Treated units: $D_i = 1$
  - → Untreated units: $D_i = 0$
- $Y_i(D_i)$ is the outcome for unit i with treatment status D_i
- The **treatment effect** for unit i is just:

$$\tau_i = Y_i(1) - Y_i(0)$$

From last time: that pesky fundamental problem

Let's go back to our model:

- We have $i \in \{1, ..., N\}$ units
- $D_i \in \{0, 1\}$ is the treatment indicator for unit i
  - $\rightarrow$ Treated units: $D_i = 1$
  - $\rightarrow$ Untreated units: $D_i = 0$
- $Y_i(D_i)$ is the outcome for unit i with treatment status D_i
- The **treatment effect** for unit i is just:

$$\tau_i = Y_i(1) - Y_i(0)$$

... but we can never observe both $Y_i(1)$ and $Y_i(0)$ simultaneously!

Okay, we can't see τ_i. What *can* we see?

Social science generally agrees that estimating τ_i is impossible.

$\rightarrow$ Should we give up and go home now?

Obviously not!

(Sorry?)

Okay, we can't see τ_i. What *can* we see?

Social science generally agrees that estimating τ_i is impossible.

$\rightarrow$ Should we give up and go home now?

Obviously not!
(Sorry?)

We can still make progress on **functions** of τ_i

- Which ones depend on what questions we want to answer!

We can imagine several different "treatment parameters"

Our bread and butter is the **Average Treatment Effect (ATE)**:

$$\tau^{ATE} = E[\tau_i] = E[Y_i(1) - Y_i(0)] = E[Y_i(1)] - E[Y_i(0)]$$

The ATE tells us the *average* impact of treatment across a [sample]...

... but this might not be the only object of interest

Interlude: The naive estimator vs the ATE

Note that the **Average Treatment Effect (ATE)...**

$$\tau^{ATE} = E[\tau_i] = E[Y_i(1) - Y_i(0)] = E[Y_i(1)] - E[Y_i(0)]$$

And the **naive estimator...**

$$\tau^{N} = \overline{Y(1)} - \overline{Y(0)}$$

are **not** the same!

ATE: **potential** outcomes
Naive estimator: **observed** outcomes

We can imagine several different "treatment parameters"

We can also think of treatment effects for particular groups.

Heterogeneous treatment effects:

$$\tau^X = E[\tau_i | X_i = x]$$

As an example, consider:

$$\begin{aligned}\tau^{Female} &= E[\tau_i | \text{Gender}_i = \text{female}] = E[Y_i(1) - Y_i(0) | \text{Gender}_i = \text{female}] \\ &= E[Y_i(1) | \text{Gender}_i = \text{female}] - E[Y_i(0) | \text{Gender}_i = \text{female}]\end{aligned}$$

Nothing stops this from being extremely general...

... but we still face that pesky fundamental problem

A natural extension of heterogeneous treatment effects

We can also consider heterogeneity *by treatment status*.

Average treatment effect on the treated (ATT)

- The impact of treatment *on treated units*:

$$\begin{aligned} \tau^{ATT} &= E[\tau_i | D_i = 1] = E[Y_i(1) - Y_i(0) | D_i = 1] \\ &= E[Y_i(1) | D_i = 1] - E[Y_i(0) | D_i = 1] \end{aligned}$$

- AKA the ATET, TOT, or TT

Why might this differ from the ATE?

A natural extension of heterogeneous treatment effects

We can also consider heterogeneity *by treatment status*.

Average treatment effect on the treated (ATT)

- The impact of treatment *on treated units*:

$$\begin{aligned}\tau^{ATT} &= E[\tau_i | D_i = 1] = E[Y_i(1) - Y_i(0) | D_i = 1] \\ &= E[Y_i(1) | D_i = 1] - E[Y_i(0) | D_i = 1]\end{aligned}$$

- AKA the ATET, TOT, or TT

Why might this differ from the ATE?

→ That pesky selection thing returns

Note that we still don't observe $E[Y_i(0) | D_i = 1]$!

A natural extension of heterogeneous treatment effects

We can also consider heterogeneity *by treatment status*.

Average treatment effect on the *un*treated (ATN)

- The impact of treatment *on untreated units*:

$$\begin{aligned}\tau^{ATN} &= E[\tau_i | D_i = 0] = E[Y_i(1) - Y_i(0) | D_i = 0] \\ &= E[Y_i(1) | D_i = 0] - E[Y_i(0) | D_i = 0]\end{aligned}$$

- AKA the ATEN, TONT, or TNT

Why might this differ from the ATE?

A natural extension of heterogeneous treatment effects

We can also consider heterogeneity *by treatment status*.

Average treatment effect on the *un*treated (ATN)

- The impact of treatment *on untreated units*:

$$\begin{aligned}\tau^{ATN} &= E[\tau_i | D_i = 0] = E[Y_i(1) - Y_i(0) | D_i = 0] \\ &= E[Y_i(1) | D_i = 0] - E[Y_i(0) | D_i = 0]\end{aligned}$$

- AKA the ATEN, TONT, or TNT

Why might this differ from the ATE?

→ That pesky selection thing returns

Note that we still don't observe $E[Y_i(1) | D_i = 0]$!

What can we use these parameters for?

These objects might all be useful - but for different things!

Consider a voluntary EPA emissions monitoring program:

- Suppose the relevant population is *all* firms, but...

  $\rightarrow$ Not every firm will participate

- Let $D_i = 1$ be firms who participate

- Let $Y_i(1)$ and $Y_i(0)$ be measures of emissions

  - When might we want τ^{ATT}? τ^{ATN}? τ^{ATE}?

  - Do we expect these three treatment parameters to be the same?

Not everyone need have the same treatment effect

We can define...

Homogenous treatment effects:

- Treatment effects that are the same for **everyone**
- Note that this includes untreated units!

Heterogenous treatment effects:

- Treatment effects that are **not** the same for everyone
- Note that this includes untreated units!

With homogenous treatment effects:

$$\tau^{ATE} = \tau^{ATT} = \tau^{ATN}$$

What happens when treatment effects differ?

Typically, because of **selection**:

$$\tau^{ATE} \neq \tau^{ATT} \neq \tau^{ATN}$$

- Units that *choose* to take up treatment are different
- This extends to treatment effects as well
- If reducing your emissions is cheap for you, are you likely to enroll?
- If reducing your emissions is expensive for you, are you likely to enroll?

In this case, the ATE is a **function** of the ATT and ATN:

$$\tau^{ATE} = Pr(D_i = 1)\tau^{ATT} + (1 - Pr(D_i = 1))\tau^{ATN}$$

Zooming in on the ATT

We defined this a few minutes ago as:

$$\begin{aligned}\tau^{ATT} &= E[Y_i(1) - Y_i(0)|D_i = 1] \\ &= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]\end{aligned}$$

Good news:

- $E[Y_i(1)|D_i = 1]$ is easily observable from data, because $E[Y_i(1)|D_i = 1] \approx \overline{Y(1)}$

Bad news:

- $E[Y_i(0)|D_i = 1]$ is unobservable

→ We never see untreated outcomes for treated units!

The selection problem in ATT-land

We can do some math to think about selection here:

$$\tau^{ATT} = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]$$

The selection problem in ATT-land

We can do some math to think about selection here:

$$\tau^{ATT} = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]$$

$$= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] + \underbrace{E[Y_i(0)|D_i = 0] - E[Y_i(0)|D_i = 0]}_{\text{just add and subtract this}}$$

The selection problem in ATT-land

We can do some math to think about selection here:

$$\tau^{ATT} = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]$$

$$= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] + \underbrace{E[Y_i(0)|D_i = 0] - E[Y_i(0)|D_i = 0]}_{\text{just add and subtract this}}$$

$$\approx \underbrace{\overline{Y(1)}}_{\text{sample mean}} - E[Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 0] - \underbrace{\overline{Y(0)}}_{\text{sample mean}}$$

The selection problem in ATT-land

We can do some math to think about selection here:

$$\tau^{ATT} = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]$$

$$= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] + \underbrace{E[Y_i(0)|D_i = 0] - E[Y_i(0)|D_i = 0]}_{\text{just add and subtract this}}$$

$$\approx \underbrace{\overline{Y(1)}}_{\text{sample mean}} - E[Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 0] - \underbrace{\overline{Y(0)}}_{\text{sample mean}}$$

$$= \underbrace{\overline{Y(1)} - \overline{Y(0)} - E[Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 0]}_{\text{rearranged}}$$

The selection problem in ATT-land

We can do some math to think about selection here:

$$\tau^{ATT} = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]$$

$$= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] + \underbrace{E[Y_i(0)|D_i = 0] - E[Y_i(0)|D_i = 0]}_{\text{just add and subtract this}}$$

$$\approx \underbrace{\overline{Y(1)}}_{\text{sample mean}} - E[Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 0] - \underbrace{\overline{Y(0)}}_{\text{sample mean}}$$

$$= \underbrace{\overline{Y(1)} - \overline{Y(0)} - E[Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 0]}_{\text{rearranged}}$$

$\Rightarrow$ the ATT is a combination of what we see and selection:

$$\tau^{ATT} \approx \underbrace{\overline{Y(1)} - \overline{Y(0)}}_{\text{data}} - \underbrace{E[Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 0]}_{\text{selection}}$$

Roy's parable illustrates selection ruining everything

Let's think about education:

- Suppose only 2 types of people: college-educated and non-college-educated
- Non-attendees: $Y_i(0) \sim N(65,000; 5,000^2)$
- Attendees: $Y_i(1) \sim N(60,000; 10,000^2)$
- Assume the correlation in incomes is high: 0.84

Roy's parable illustrates selection ruining everything

Let's think about education:

- Suppose only 2 types of people: college-educated and non-college-educated
- Non-attendees: $Y_i(0) \sim N(65,000; 5,000^2)$
- Attendees: $Y_i(1) \sim N(60,000; 10,000^2)$
- Assume the correlation in incomes is high: 0.84

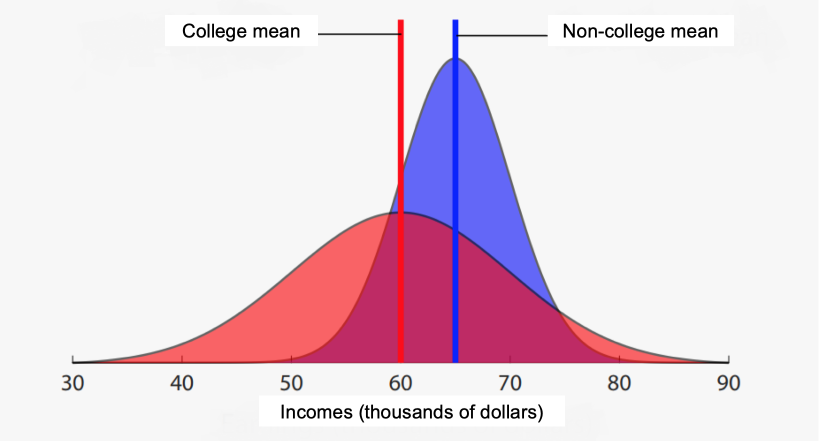
Economists like models. Here's a simple one:

- Each person picks her maximum income:

$$y_i = max(y_i(0), y_i(1))$$

  where these are lowercase because they're now *realizations* of Y_i
- If person i's income is higher with college, she'll go to school

Visualizing data is often useful



We can use simulated data to think about this

	Non-attendees	Attendees	Mean
Non-college income	**63,985**	*68,690*	*65,001*
College income	*56,599*	**72,317**	*59,992*
# of obs	78,414	21,586	100,000

What do our estimators tell us is going on?

We want to know the **treatment effect** of college.

Let's start with the naive estimator:

$$\tau^N = \overline{y}(1 \mid d_i = 1) - \overline{y}(0 \mid d_i = 0) = 72,317 - 63,985 = \mathbf{8,332}$$

What do our estimators tell us is going on?

	Non-attendees	Attendees	Mean
Non-college income	**63,985**	*68,690*	*65,001*
College income	*56,599*	**72,317**	*59,992*
# of obs	78,414	21,586	100,000

What do our estimators tell us is going on?

We want to know the **treatment effect** of college.

Let's start with the naive estimator:

$$\tau^N = \overline{y}(1 \mid d_i = 1) - \overline{y}(0 \mid d_i = 0) = 72,317 - 63,985 = \mathbf{8,332}$$

This suggests college causes incomes to rise...but it assumes
$E[Y_i(1)] = E[Y_i(1)|D_i = 1]$ and $E[Y_i(0)] = E[Y_i(0)|D_i = 0]$

What do our estimators tell us is going on?

We want to know the **treatment effect** of college.

What about the <u>ATE</u>?

$$\tau^{ATE} = \overline{y}(1) - \overline{y}(0) = 59,992 - 65,001 = \mathbf{-5,009}$$

What do our estimators tell us is going on?

	Non-attendees	Attendees	Mean
Non-college income	**63,985**	*68,690*	*65,001*
College income	*56,599*	**72,317**	*59,992*
# of obs	78,414	21,586	100,000

What do our estimators tell us is going on?

We want to know the **treatment effect** of college.

What about the <u>ATE</u>?

$$\tau^{ATE} = \overline{y}(1) - \overline{y}(0) = 59,992 - 65,001 = -\mathbf{5,009}$$

This suggests college causes incomes to drop!

This is the impact from **forcing everyone** to go to college.

What do our estimators tell us is going on?

We want to know the **treatment effect** of college.

What about the ATT: the effect of college on people who go?

$$\tau^{ATT} = \overline{y}(1 \mid d_i = 1) - \overline{y}(0 \mid d_i = 1) = 72,317 - 68,690 = \mathbf{3,627}$$

What do our estimators tell us is going on?

	Non-attendees	Attendees	Mean
Non-college income	**63,985**	*68,690*	*65,001*
College income	*56,599*	**72,317**	*59,992*
# of obs	78,414	21,586	100,000

What do our estimators tell us is going on?

We want to know the **treatment effect** of college.

What about the ATT: the effect of college on attendees?

$$\tau^{ATT} = \overline{y}(1 \mid d_i = 1) - \overline{y}(0 \mid d_i = 1) = 72,317 - 68,690 = \mathbf{3,627}$$

College caused incomes to rise **for those that went**!

What do our estimators tell us is going on?

We want to know the **treatment effect** of college.

What about the ATT: the effect of attendance on non-college-educated?

$$\tau^{ATN} = \overline{y}(1 \mid d_i = 0) - \overline{y}(0 \mid d_i = 0) = 56,599 - 63,985 = -\mathbf{7,386}$$

What do our estimators tell us is going on?

	Non-attendees	Attendees	Mean
Non-college income	**63,985**	*68,690*	*65,001*
College income	*56,599*	**72,317**	*59,992*
# of obs	78,414	21,586	100,000

What do our estimators tell us is going on?

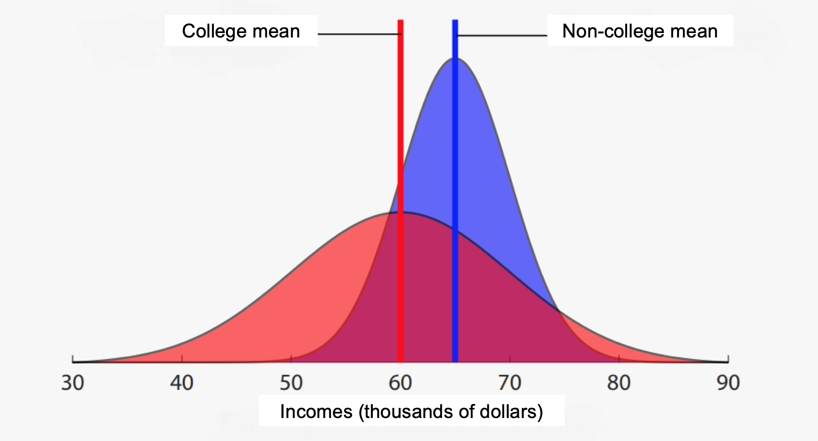
We want to know the **treatment effect** of college.

What about the ATT: the effect of attendance on non-college-goers?

$$\tau^{ATN} = \overline{y}(1 \mid d_i = 0) - \overline{y}(0 \mid d_i = 0) = 56,599 - 63,985 = -\mathbf{7,386}$$

College would have caused incomes to drop **for those that chose not to go**!

Looking at the distributions is illuminating



People chose what was best for them – and messed up our naive estimator.

These treatment parameters teach us something interesting

Naive estimator: "effect size" of 8,332

- This tells us nothing!

Average treatment effect: effect size of -5,009

- Forcing college on everyone would be a bad idea

Average treatment on the treated: effect size of 3,627

- Attendees benefitted from their schooling

Average treatment on the untreated: effect size of -7,386

- Non-attendees were right not to go

Using OLS regression to estimate treatment parameters

We begin with an extremely general model:

$$Y_i(1) = g_1(X_i, \varepsilon_i)$$
$$Y_i(0) = g_0(X_i, \varepsilon_i)$$

where X_i are observed characteristics and ε_i, an error term, contains unobserved characteristics.

For tractability, assume the errors are **additively separable**:

$$Y_i(1) = g_1(X_i, \varepsilon_i) = g_1(X_i) + \varepsilon_{1i}$$
$$Y_i(0) = g_0(X_i, \varepsilon_i) = g_0(X_i) + \varepsilon_{0i}$$

Using OLS regression to estimate treatment parameters

To make our lives easier, let's assume **linearity**:

$$Y_i(1) = \beta_1 X_i + \varepsilon_{1i}$$
$$Y_i(0) = \beta_0 X_i + \varepsilon_{0i}$$

Linearity isn't as restrictive as it seems:

- If the underlying conditional expectation function is linear, this regression estimates it
- If the underlying conditional expectation function is non-linear, regression is its best linear approximation
- We can include non-linear terms [e.g. $\beta_1^A X_i + \beta_1^B X_i^2$]

Using OLS regression to estimate treatment parameters

We'll also assume that $Y_i(1)$ and $Y_i(0)$ only differ by a **constant treatment effect**:

$$Y_i(1) = Y_i(0) + \tau$$

This yields:

$$Y_i = \beta X_i + \tau D_i + \varepsilon_i$$

This is starting to look like a nice OLS regression!

Using OLS regression to estimate treatment parameters

We'll also assume that $Y_i(1)$ and $Y_i(0)$ only differ by a **constant treatment effect**:

$$Y_i(1) = Y_i(0) + \tau$$

This yields:

$$Y_i = \beta X_i + \tau D_i + \varepsilon_i$$

This is starting to look like a nice OLS regression!

What assumptions on ε_i do we need to interpret τ as the causal effect of treatment (D_i) on our outcome (Y_i)?

Modeling selection in regression land

Start by writing observed outcomes as a function of potential outcomes. (We'll now omit X_i for simplicity and only think about D_i)

Since

$$Y_i = \begin{cases} Y_i(1) & \text{if } D_i = 1 \\ Y_i(0) & \text{if } D_i = 0 \end{cases}$$

we can write:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

Now assume constant treatment effects: $\tau_i = Y_i(1) - Y_i(0) = \tau$

Modeling selection in regression land

We can now write:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

Modeling selection in regression land

We can now write:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

$$Y_i = D_i Y_i(1) + \underbrace{Y_i(0) - D_i Y_i(0)}_{\text{expand}}$$

Modeling selection in regression land

We can now write:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

$$Y_i = D_i Y_i(1) + \underbrace{Y_i(0) - D_i Y_i(0)}_{\text{expand}}$$

$$Y_i = \underbrace{Y_i(0) + (Y_i(1) - Y_i(0)) D_i}_{\text{rearrange}}$$

Modeling selection in regression land

We can now write:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

$$Y_i = D_i Y_i(1) + \underbrace{Y_i(0) - D_i Y_i(0)}_{\text{expand}}$$

$$Y_i = \underbrace{Y_i(0) + (Y_i(1) - Y_i(0)) D_i}_{\text{rearrange}}$$

$$Y_i = Y_i(0) + (Y_i(1) - Y_i(0)) D_i + \underbrace{E[Y_i(0)] - E[Y_i(0)]}_{\text{add \& subtract}}$$

Modeling selection in regression land

We can now write:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

$$Y_i = D_i Y_i(1) + \underbrace{Y_i(0) - D_i Y_i(0)}_{\text{expand}}$$

$$Y_i = \underbrace{Y_i(0) + (Y_i(1) - Y_i(0)) D_i}_{\text{rearrange}}$$

$$Y_i = Y_i(0) + (Y_i(1) - Y_i(0)) D_i + \underbrace{E[Y_i(0)] - E[Y_i(0)]}_{\text{add \& subtract}}$$

$$Y_i = \underbrace{E[Y_i(0)] + (Y_i(1) - Y_i(0)) D_i + Y_i(0) - E[Y_i(0)]}_{\text{rearrange}}$$

Modeling selection in regression land

...

$$Y_i = \underbrace{E[Y_i(0)]}_{\beta} + \underbrace{(Y_i(1) - Y_i(0))}_{\tau} D_i + \underbrace{Y_i(0) - E[Y_i(0)]}_{\varepsilon_i}$$

$$Y_i = \underbrace{\beta + \tau D_i + \varepsilon_i}_{\text{redefine}}$$

where

- β: mean (expectation) of $Y_i(0)$
- τ: constant treatment effect, $\tau_i = \tau = Y_i(1) - Y_i(0)$
- ε_i: random component of $Y_i(0)$: $Y_i(0) - E[Y_i(0)]$

This looks familiar!

Modeling selection in regression land

$$Y_i = \beta + \tau D_i + \varepsilon_i$$

Modeling selection in regression land

$$Y_i = \beta + \tau D_i + \varepsilon_i$$

Taking conditional expectations of Y_i on $D_i = 1$ and $D_i = 0$:

$$E[Y_i|D_i = 1] = \beta + \tau + E[\varepsilon_i|D_i = 1]$$

$$E[Y_i|D_i = 0] = \beta + E[\varepsilon_i|D_i = 0]$$

Modeling selection in regression land

$$Y_i = \beta + \tau D_i + \varepsilon_i$$

Taking conditional expectations of Y_i on $D_i = 1$ and $D_i = 0$:

$$E[Y_i|D_i = 1] = \beta + \tau + E[\varepsilon_i|D_i = 1]$$

$$E[Y_i|D_i = 0] = \beta + E[\varepsilon_i|D_i = 0]$$

Now, a familiar enemy:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = \tau + \underbrace{E[\varepsilon_i|D_i = 1] - E[\varepsilon_i|D_i = 0]}_{\text{☠☠☠}}$$

Modeling selection in regression land

$$Y_i = \beta + \tau D_i + \varepsilon_i$$

Taking conditional expectations of Y_i on $D_i = 1$ and $D_i = 0$:

$$E[Y_i | D_i = 1] = \beta + \tau + E[\varepsilon_i | D_i = 1]$$

$$E[Y_i | D_i = 0] = \beta + E[\varepsilon_i | D_i = 0]$$

Now, a familiar enemy:

$$E[Y_i | D_i = 1] - E[Y_i | D_i = 0] = \tau + \underbrace{E[\varepsilon_i | D_i = 1] - E[\varepsilon_i | D_i = 0]}_{\text{☠☠☠}}$$

This is just the selection term, written as a function of regression errors ε_i!

For regression to estimate τ, we need...

$$E[\varepsilon_i|D_i = 1] - E[\varepsilon_i|D_i = 0] = 0$$

In general, we require

$$E[\varepsilon_i|D_i] = 0$$

(if you include a constant, you can always get $E[\varepsilon_i|D_i] = \mu$ to be okay)

Note that by the Law of Total Expectations,

$$E_b[E[a|b]] = E[a]$$

therefore:

$$E[\varepsilon_i|D_i] = E[\varepsilon_i] = 0$$

What is this assumption, anyway?

$$E[\varepsilon_i | D_i] = 0$$

What is this assumption, anyway?

$$E[\varepsilon_i | D_i] = 0$$

In words:
The expectation of the error term, *conditional on treatment*, is zero.

In other words:
Once we condition on treatment, there is *no additional information* in ε_i.

In different other words:
The errors are uncorrelated with the treatment variable

In more different other words:
There is no selection bias

In even more different other words:
We observe everything that is correlated with treatment and affects Y_i

In even fewer different other words:
D_i is *exogenous*

Selection as omitted variable bias

We can think about selection as a form of *omitted variable bias*.

Consider the *true* model:

$$Y_i = \alpha + \tau D_i + \beta X_i + \varepsilon$$

where D is treatment (college), and X is learning ability (unobservable!)
We can't see X, so we instead run:

$$Y_i = \alpha + \tau D_i + \nu$$

where $\nu = \beta X_i + \varepsilon_i$.

What happens now? Recall that:

$$\hat{\tau} = \frac{Cov(Y_i, D_i)}{Var(D_i)}$$

Selection as omitted variable bias

$$\hat{\tau} = \frac{Cov(Y_i, D_i)}{Var(D_i)}$$
$$= \underbrace{\frac{Cov(\alpha + \tau D_i + \beta X_i + \varepsilon_i, D_i)}{Var(D_i)}}_{\text{plug in for } Y_i}$$

Selection as omitted variable bias

$$\hat{\tau} = \frac{Cov(Y_i, D_i)}{Var(D_i)}$$

$$= \underbrace{\frac{Cov(\alpha + \tau D_i + \beta X_i + \varepsilon_i, D_i)}{Var(D_i)}}_{\text{plug in for } Y_i}$$

$$= \underbrace{\frac{Cov(D_i, \alpha) + \tau Cov(D_i, D_i) + \beta Cov(D_i, X_i) + Cov(D_i, \varepsilon)}{Var(D_i)}}_{\text{laws of } E[]}$$

Selection as omitted variable bias

$$\hat{\tau} = \frac{Cov(Y_i, D_i)}{Var(D_i)}$$

$$= \underbrace{\frac{Cov(\alpha + \tau D_i + \beta X_i + \varepsilon_i, D_i)}{Var(D_i)}}_{\text{plug in for } Y_i}$$

$$= \underbrace{\frac{Cov(D_i, \alpha) + \tau Cov(D_i, D_i) + \beta Cov(D_i, X_i) + Cov(D_i, \varepsilon)}{Var(D_i)}}_{\text{laws of } E[]}$$

$$= \underbrace{\frac{0 + \tau Var(D_i) + \beta Cov(D_i, X_i) + 0}{Var(D_i)}}_{\text{definitions}}$$

Selection as omitted variable bias

$$\hat{\tau} = \frac{Cov(Y_i, D_i)}{Var(D_i)}$$

$$= \underbrace{\frac{Cov(\alpha + \tau D_i + \beta X_i + \varepsilon_i, D_i)}{Var(D_i)}}_{\text{plug in for } Y_i}$$

$$= \underbrace{\frac{Cov(D_i, \alpha) + \tau Cov(D_i, D_i) + \beta Cov(D_i, X_i) + Cov(D_i, \varepsilon)}{Var(D_i)}}_{\text{laws of } E[]}$$

$$= \underbrace{\frac{0 + \tau Var(D_i) + \beta Cov(D_i, X_i) + 0}{Var(D_i)}}_{\text{definitions}}$$

$$= \underbrace{\tau + \beta \frac{Cov(D_i, X_i)}{Var(D_i)}}_{\text{simplify}}$$

Selection as omitted variable bias

In other words:

$$\hat{\tau} = \tau + \underbrace{\beta \frac{Cov(D_i, X_i)}{Var(D_i)}}_{\text{☠☠☠}}$$

Selection as omitted variable bias

In other words:

$$\hat{\tau} = \tau + \underbrace{\beta \frac{Cov(D_i, X_i)}{Var(D_i)}}_{\text{☠☠☠}}$$

We wanted to have

$$\hat{\tau} = \tau$$

But because we didn't observe X_i, we are left with ugliness!

To put this back in selection terms, recall that X_i is learning ability: the thing that determines whether college will be good for you.

Once again: selection messes everything up!

Recap

TL;DR:

1. There are many parameters we might want to estimate
2. Selection bias is a big problem for estimation
3. We can use regression to estimate these parameters