EL COLEGIO DE MÉXICO

CEE 60 AÑOS 1964-2024 CENTRO DE ESTUDIOS ECONÓMICOS

CENTRO DE ESTUDIOS ECONÓMICOS

Maestría en Economía 2024–2026

Microeconometrics for Evaluation

5 Regression with Controls II

Disclaimer: I AM NOT the original intellectual author of the material presented in these notes. The content is STRONGLY based on a combination of lecture notes (from Aurora Ramirez), textbook references, and personal annotations for learning purposes. Any errors or omissions are entirely my own responsibility.

Índice

Controls good and bad

Including additional control variables in a regression can often strengthen the *causal interpretation* of the coefficients. By adjusting for relevant background characteristics, we can better isolate the effect of the main explanatory variable.

However, **not all controls are helpful**. Some variables, if added, can introduce bias rather than reduce it.

Specifically, **bad controls** are those that may themselves be influenced by the treatment. If you control for a variable that is affected by the treatment, you might accidentally adjust away some of the effect you are trying to measure.

By contrast, **good controls** are factors that are determined *prior* to the treatment. These are often fixed or pre-existing characteristics that help explain variation in outcomes without being affected by the treatment itself.

Intuition: Suppose you want to estimate the effect of attending college on income. A *good control* might be high school GPA—something determined before college. A *bad control* would be current job satisfaction—something that could already reflect the income effect you're trying to study. Including variables that occur after the treatment can distort your estimates by removing part of the effect you're aiming to capture.

An example

- Imagine we randomly assign individuals a college degree, where $C_i \in \{0, 1\}$.
- Our goal is to estimate the *causal effect* of a college education on income Y_i—that is, the causal conditional expectation function (CEF).
- Suppose individuals can work in two types of jobs: white-collar ($W_i = 1$) or blue-collar ($W_i = 0$).
- So, which regression specification should we use?

- A **natural approach**: find β that satisfies the moment condition $E[C_i(Y_i - \alpha - \beta C_i)] = 0$
- An **alternative approach**: find β that satisfies $E[C_i(Y_i - \alpha - \beta C_i - \gamma W_i)] = 0$.
  Is this better?

Intuition: If college degrees are randomly assigned, then estimating the effect of college on earnings should be straightforward. The first approach uses only college assignment to estimate the effect. The second approach adds occupation as a control—*but this may be problematic if occupation is influenced by education.* Including such a variable could distort the causal estimate, as you're controlling for something on the causal path from education to earnings.

Example: Estimating the Causal Effect of College

$$Y_i = C_i Y_{1i} + (1 - C_i) Y_{0i} \qquad \textbf{Definition of observed outcome} \tag{1}$$

$$\begin{aligned} \mathbb{E}[Y_i \mid C_i = 1] &= \mathbb{E}[C_i Y_{1i} + (1 - C_i) Y_{0i} \mid C_i = 1] && (2)\\ &= \mathbb{E}[1 \cdot Y_{1i} + 0 \cdot Y_{0i} \mid C_i = 1] && (3)\\ &= \mathbb{E}[Y_{1i} \mid C_i = 1] && (4) \end{aligned}$$

$$\begin{aligned} \mathbb{E}[Y_i \mid C_i = 0] &= \mathbb{E}[C_i Y_{1i} + (1 - C_i) Y_{0i} \mid C_i = 0] && (5)\\ &= \mathbb{E}[0 \cdot Y_{1i} + 1 \cdot Y_{0i} \mid C_i = 0] && (6)\\ &= \mathbb{E}[Y_{0i} \mid C_i = 0] && (7) \end{aligned}$$

$$\mathbb{E}[Y_i \mid C_i = 1] - \mathbb{E}[Y_i \mid C_i = 0] = \mathbb{E}[Y_{1i} \mid C_i = 1] - \mathbb{E}[Y_{0i} \mid C_i = 0] \quad \textbf{Difference in means} \tag{8}$$

$$\begin{aligned} C_i \perp (Y_{1i}, Y_{0i}) &\Rightarrow \mathbb{E}[Y_{0i} \mid C_i = 1] = \mathbb{E}[Y_{0i} \mid C_i = 0] = \mathbb{E}[Y_{0i}] && (9)\\ &\Rightarrow \mathbb{E}[Y_{1i} \mid C_i = 1] - \mathbb{E}[Y_{0i} \mid C_i = 0] = \mathbb{E}[Y_{1i} - Y_{0i}] && (10) \end{aligned}$$

$$\beta = \mathbb{E}[Y_{1i} - Y_{0i}] \quad \blacksquare \qquad \textbf{Average Treatment Effect (ATE)} \tag{11}$$

The difference in expected outcomes between those who went to college and those who did not, $\mathbb{E}[Y_i \mid C_i = 1] - \mathbb{E}[Y_i \mid C_i = 0]$, identifies the causal effect when C_i is randomly assigned. Under mean independence, this difference equals the *average treatment effect* $\mathbb{E}[Y_{1i} - Y_{0i}]$.

Intuition: Since college assignment C_i is **random**, it's as if nature ran an experiment: some people got a college degree, others didn't, but not based on anything else about them. That means we can **directly compare their average earnings**, and the difference will reflect the true effect of college. The observed difference in outcomes isn't confounded by other variables—it's just the average gain from going to college. That's why the regression captures the causal effect.

Example: The Bad Control Problem

- Let (W_{1i}, W_{0i}) denote potential occupations.
- We observe:

$$W_i = C_i W_{1i} + (1 - C_i) W_{0i}$$

- **Bad control:** Conditioning on W_i does *not* lead to a valid causal interpretation of the effect of C_i on Y_i.
- To understand why, consider the conditional difference in average earnings between college and non-college individuals among those in white-collar jobs:

$$\mathbb{E}[Y_i \mid W_i = 1, C_i = 1] - \mathbb{E}[Y_i \mid W_i = 1, C_i = 0]$$

- Substitute observed outcomes:

$$= \mathbb{E}[Y_{1i} \mid W_{1i} = 1, C_i = 1] - \mathbb{E}[Y_{0i} \mid W_{0i} = 1, C_i = 0]$$

- Assume joint independence between $(Y_{1i}, W_{1i}, Y_{0i}, W_{0i})$ and C_i. Then:

$$= \mathbb{E}[Y_{1i} \mid W_{1i} = 1] - \mathbb{E}[Y_{0i} \mid W_{0i} = 1]$$

- This expression *does not* identify the causal effect of C_i on Y_i.

Intuition: The problem with conditioning on W_i is that it's a *post-treatment variable*—college can affect occupation. By comparing earnings only within white-collar jobs, we are selecting on a variable influenced by college, which introduces selection bias. It's like only looking at the "successful" people in both groups and comparing them—this distorts the true effect of college. That's why W_i is a **bad control**.

Example: The Bad Control Problem II

- Consider:

$$\mathbb{E}[Y_{1i} \mid W_{1i} = 1] - \mathbb{E}[Y_{0i} \mid W_{0i} = 1]$$

- Add and subtract $\mathbb{E}[Y_{0i} \mid W_{1i} = 1]$:

$$= \mathbb{E}[Y_{1i} \mid W_{1i} = 1] - \mathbb{E}[Y_{0i} \mid W_{1i} = 1] \tag{12}$$

$$+ \mathbb{E}[Y_{0i} \mid W_{1i} = 1] - \mathbb{E}[Y_{0i} \mid W_{0i} = 1] \tag{13}$$

- Group terms:

$$= \mathbb{E}[Y_{1i} - Y_{0i} \mid W_{1i} = 1] + (\mathbb{E}[Y_{0i} \mid W_{1i} = 1] - \mathbb{E}[Y_{0i} \mid W_{0i} = 1]) \tag{14}$$

- First term: **causal effect on treated** Second term: **selection bias**

Important sign intuition:

- If

$$\mathbb{E}[Y_{0i} \mid W_{1i} = 1] < \mathbb{E}[Y_{0i} \mid W_{0i} = 1]$$

then selection bias is negative.

- *Why?* Any college graduate can likely access a white-collar job, so those with low Y_{0i} are still present: $\mathbb{E}[Y_{0i} \mid W_{1i} = 1] \approx \mathbb{E}[Y_{0i}]$.

- But only higher-Y_{0i} individuals manage to get white-collar jobs without a college degree, so:

$$\mathbb{E}[Y_{0i} \mid W_{0i} = 1] > \mathbb{E}[Y_{0i}]$$

- Therefore:

$$\mathbb{E}[Y_{0i} \mid W_{1i} = 1] - \mathbb{E}[Y_{0i} \mid W_{0i} = 1] < 0$$

Intuition: This is an *apples-to-oranges* comparison. Among white-collar workers, those with degrees include everyone—strong and weak performers. But those without degrees who make it into white-collar jobs are unusually high-ability. Comparing these two groups mixes in this ability difference, introducing **selection bias**. That's why the second term pulls down the total difference, making it look like college has less of an effect—or even a negative one.

Example: Bad Control, i.e., Proxy Control

Proxy control. A variable that helps control for omitted variables, but is itself influenced by the treatment. This means it cannot be used to estimate a clean causal effect.

Suppose we estimate the following regression:

$$Y_i = \alpha + \rho s_i + \gamma a_i + \varepsilon_i \tag{15}$$

- a_i is an IQ score observed in eighth grade, which reflects innate ability before any major schooling decisions.
- Since a_i is measured prior to s_i (e.g., years of schooling), it is **not affected** by the treatment.
- By assumption:

$$\mathbb{E}[s_i\varepsilon_i] = \mathbb{E}[a_i\varepsilon_i] = 0$$

  This implies no correlation between either regressor and the error term.
- Therefore, a_i qualifies as a **good control**.

Intuition: A good control is something known *before* the treatment happens. Since IQ is measured before schooling, it can explain variation in outcomes without being influenced by schooling itself. This makes it a useful way to adjust for differences in ability, without introducing bias. But if you used a variable *affected* by education—like current job performance—that would be a bad (or proxy) control, because it already reflects some of education's effect.

Example: Bad Control, i.e., Proxy Control II

- Equation (1) is the causal model of interest, but unfortunately, data on a_i (early ability) are **unavailable**.
- Suppose we have an alternative ability measure taken *after* schooling is completed. Denote this variable by a_{li}:

$$a_{li} = \pi_0 + \pi_1 s_i + \pi_2 a_i \tag{16}$$

- Here, a_{li} represents **late-measured ability**, which is influenced by both schooling (s_i) and innate ability (a_i).
- Since a_i is unobserved, we regress y_i on s_i and a_{li} in an attempt to control for ability.

Intuition: Because a_{li} is measured *after* schooling, it reflects not just innate ability, but also the effect of schooling. By including a_{li} in the regression, you are controlling for a variable that already contains part of the effect of s_i—**you are "soaking up"some of the treatment effect**. This leads to biased estimates of ρ in the regression. In this case, a_{li} acts as a *proxy control*—intended to adjust for unobserved ability, but contaminated by the treatment itself.

Example: Bad Control, i.e., Proxy Control III

- Start from the original regression:

$$Y_i = \alpha + \rho s_i + \gamma a_i + \varepsilon_i \tag{1}$$

- And the proxy relationship:

$$a_{li} = \pi_0 + \pi_1 s_i + \pi_2 a_i \tag{2}$$

- Solve equation (2) for a_i:

$$a_i = \frac{a_{li} - \pi_0 - \pi_1 s_i}{\pi_2}$$

- Substitute into (1):

$$Y_i = \alpha + \rho s_i + \gamma \left(\frac{a_{li} - \pi_0 - \pi_1 s_i}{\pi_2} \right) + \varepsilon_i \tag{17}$$

$$= \alpha + \rho s_i + \frac{\gamma}{\pi_2} a_{li} - \frac{\gamma \pi_0}{\pi_2} - \frac{\gamma \pi_1}{\pi_2} s_i + \varepsilon_i \tag{18}$$

- Group terms:

$$Y_i = \left(\alpha - \frac{\gamma \pi_0}{\pi_2} \right) + \left(\rho - \frac{\gamma \pi_1}{\pi_2} \right) s_i + \frac{\gamma}{\pi_2} a_{li} + \varepsilon_i \tag{3}$$

- Note that if $\gamma > 0$, $\pi_1 > 0$, and $\pi_2 > 0$, then:

$$\left(\rho - \frac{\gamma\pi_1}{\pi_2}\right) < \rho$$

So the estimated coefficient on s_i is biased downward.

- In the extreme case where $\pi_1 = 0$, then:

$$\left(\rho - \frac{\gamma\pi_1}{\pi_2}\right) = \rho$$

and there is no bias.

- **Important:** The OVB formula tells us that if we omit a_i, the regression:

$$Y_i = \alpha + \tilde{\rho} s_i + \tilde{\varepsilon}_i$$

will yield:

$$\tilde{\rho} = \rho + \gamma\delta_{as}$$

where δ_{as} is the coefficient from the regression:

$$a_i = \delta_{as} s_i + \nu_i$$

- If $\delta_{as} > 0$, then this omits a positively correlated variable and overstates the coefficient.
- In contrast, regression (3) biases it downward.
- So together:

$$\rho + \gamma\delta_{as} \quad \text{(overstate)} \quad > \quad \text{true } \rho \quad > \quad \left(\rho - \frac{\gamma\pi_1}{\pi_2}\right) \quad \text{(understate)}$$

This makes regression (3) useful for **bounding** the true effect.

Intuition: We want to control for ability (a_i), but we don't observe it. Instead, we use a later measure (a_{li}), which is partly caused by education. Substituting a_i with a_{li} introduces bias because a_{li} already .ªbsorbs"some of the effect of schooling. As a result, the coefficient on s_i shrinks. On the other hand, if we omit ability entirely, the estimate of ρ grows due to positive correlation between s_i and a_i. So we can think of the true effect as being somewhere in between. That's the power of bounding with a bad control.

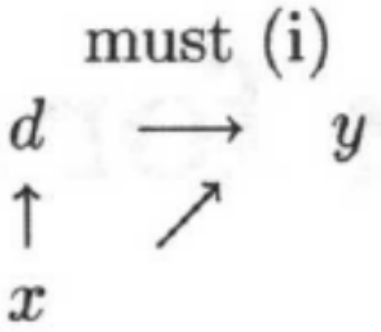
Variables to Control For

- When estimating the effect of a treatment d on an outcome variable y using observational data, the treated group ($d = 1$) and control group ($d = 0$) may differ in observed characteristics x.
- To adjust for these differences, we **control for** x: that is, we compare treated and untreated individuals who share the same value of x.
- But *which* variables should we control for?
  - **General guidance:** We should control for *pre-treatment covariates*—variables that influence the outcome y_j but are not themselves affected by the treatment d.

Intuition: To estimate a causal effect, we want to isolate variation in the outcome that is due only to the treatment—not other differences between groups. If treated and control units differ in important characteristics, comparing them directly will conflate these differences with the effect of d. Controlling for variables x that were determined *before* the treatment helps ensure we're comparing like with like. But we must avoid controlling for anything that's affected by the treatment, as that would distort the very effect we're trying to measure.

Variables to Control For: Must Cases

Consider a causal chain where each arrow represents a direct causal or influencing relationship:



must (i): x influences both d and y

- Here, x is a **pre-treatment variable** that influences both the treatment d and the outcome y.
- In this case, x **must** be controlled for. Why? Because it may differ across the treated ($d = 1$) and control ($d = 0$) groups.
- When this happens, x is called a *confounder*.

Intuition: If x affects both who gets treated and what outcomes they have, failing to control for x would misattribute its effect to d. For example, if smarter students (x) are more likely to go to college (d) and also earn more (y), then college appears more effective than it really is. Controlling for x levels the playing field, so we can isolate the impact of d alone.

Variables to Control For: Must Cases II

Case (i):

- A specific model for this case is:

$$d_i = 1[\alpha_1 + \alpha_x X_i + \epsilon_i > 0], \qquad y_i = \beta_1 + \beta_d d_i + \beta_x X_i + u_i, \qquad u_i \perp\!\!\!\perp \epsilon_i \mid X_i \tag{19}$$

- Suppose $\beta_d = 0$ (i.e., **no effect** of d on y).
- Then:

$$\mathbb{E}(y \mid x, d = 1) = \mathbb{E}(y \mid x, d = 0) \tag{20}$$

$$= \beta_1 + \beta_x x \tag{21}$$

- But as long as $\alpha_x \neq 0$, the distribution of X_i depends on d_i.
- So, taking expectation over x in both cases:

$$\mathbb{E}(y \mid d = 1) = \mathbb{E}[\beta_1 + \beta_x X_i \mid d = 1] \tag{22}$$

$$= \beta_1 + \beta_x \mathbb{E}(X_i \mid d = 1) \tag{23}$$

$$\mathbb{E}(y \mid d = 0) = \beta_1 + \beta_x \mathbb{E}(X_i \mid d = 0) \tag{24}$$

Therefore:

$$\mathbb{E}(y \mid d = 1) \neq \mathbb{E}(y \mid d = 0) \quad \text{unless} \quad \mathbb{E}(X_i \mid d = 1) = \mathbb{E}(X_i \mid d = 0) \tag{25}$$

- **Interpretation:** Even when $\beta_d = 0$, imbalance in X_i across treatment groups creates differences in $\mathbb{E}(y \mid d)$.
- If the arrow $x \to y$ is removed (i.e., $\beta_x = 0$), then:

$$y_i = \beta_1 + \beta_d d_i + u_i \tag{26}$$

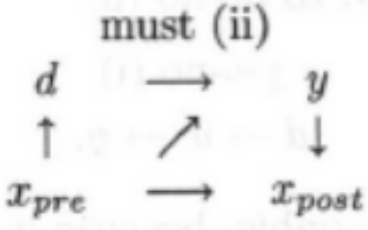
$$\Rightarrow \mathbb{E}(y \mid x, d) = \beta_1 + \beta_d d \tag{27}$$

- In that case, we have the clean path $x \to d \to y$, and imbalance in x no longer confounds the estimate of β_d.

Intuition: This case illustrates why failing to control for x creates bias even when the treatment has *no true effect*. If x influences selection into d and also affects y, then observed differences in y between groups may come entirely from x — not d. That's why we must adjust for x when it influences both treatment and outcome.

Variables to Control For: Must Cases III

Case (ii):



must (ii): x_{pre} affects d and y, x_{post} affected by d

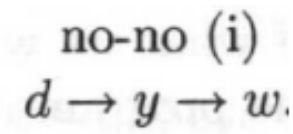
- x_{pre} should be controlled for because it is a **pre-treatment confounder**.
- However, x_{post} should **not** be controlled for, because it is affected by the treatment d.

Intuition: Think of x_{pre} as something like IQ measured *before* schooling — it helps determine both whether someone goes to college (d) and how much they earn (y). It is fair game to control for it. On the other hand, x_{post} might be something like confidence or

skills acquired *after* schooling — and it is partly caused by d. If you control for x_{post}, you're "soaking up"part of the treatment's effect, which biases the estimate of d. So we must avoid controlling for post-treatment variables.

Variables to Control For: No-no Cases

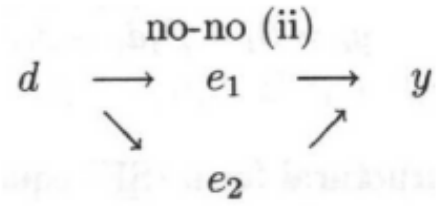
Case (i):



no-no (i): $d \rightarrow y \rightarrow w$

- w is a **post-response variable**. Controlling for w can block part (or even all) of the effect that d has on y.

Case (ii):



no-no (ii): $d \rightarrow e_1 \rightarrow y$, and $e_2 \rightarrow e_1$

- d affects y through multiple channels. If we control for e_i (a post-treatment mediator), we weaken or distort the estimated impact of d.

Intuition: Imagine d is a training program, y is productivity, and w is a performance bonus. If you control for w — which is based on productivity — you're essentially conditioning on the outcome or its consequence. This removes real variation caused by d. Similarly, if d improves productivity partly through boosting motivation e_1, and we control for motivation,

we understate the full effect of the training. In short: *never control for things that happen after the treatment and are part of its effect.*

Variables to Control For: Yes–No Cases

Case (i):

- Consider a causal chain related to the previous "no-no (ii)" situation:

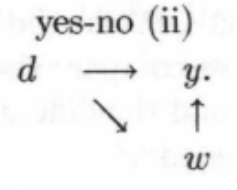
yes-no (i)

$$d \rightarrow w \rightarrow y$$

yes–no (i): $d \rightarrow w \rightarrow y$

- If we want the effect of d on y *excluding* what operates through w, then controlling for w is appropriate.
- But if we're interested in the *total effect* of d on y, then w should not be controlled for.

Case (ii):



yes–no (ii): $d \rightarrow y$, $d \rightarrow w \rightarrow y$

- Controlling for w here will only capture the **direct effect** of d on y.

Intuition: Sometimes we want to ask: "What would the effect of d be on y if we could hold some mediator w fixed?" That's a direct effect, and controlling for w answers it. But if we want to measure the *total impact* of d — including effects that flow through w — then controlling for w blocks part of that path. It's not about right or wrong — it's about whether you want the total or direct effect.

Variables to Control For: Yes–No Cases II

Example. Program affects both graduation and earnings. Specific model for Case (ii):

$$w_i = \alpha_1 + \alpha_d d_i + \epsilon_i \tag{28}$$

$$y_i = \underbrace{\beta_1 + \beta_d d_i}_{\text{Effect of } d} + \underbrace{\beta_w w_i}_{\text{Effect of } w} + u_i \tag{29}$$

$$= \text{Structural form (SF) for } y \tag{30}$$

- Assume $(d, w) \perp u_i$, and define $\epsilon_i \equiv w_i - \mathbb{E}[w_i \mid d_i]$

Substitute the w equation into the SF for y_i:

$$y_i = \beta_1 + \beta_d d_i + \beta_w(\alpha_1 + \alpha_d d_i + \epsilon_i) + u_i \tag{31}$$

$$= \beta_1 + \beta_d d_i + \beta_w \alpha_1 + \beta_w \alpha_d d_i + \beta_w \epsilon_i + u_i \tag{32}$$

$$= (\beta_1 + \beta_w \alpha_1) + (\beta_d + \beta_w \alpha_d) d_i + \beta_w \epsilon_i + u_i \tag{33}$$

$$= \underbrace{(\beta_1 + \beta_w \alpha_1)}_{\text{Intercept}} + \underbrace{(\beta_d + \beta_w \alpha_d)}_{\text{Total effect}} d_i + \underbrace{\beta_w \epsilon_i + u_i}_{\text{Error term}} \tag{34}$$

$$= \text{Reduced form (RF) for } y \tag{35}$$

Therefore, total effect of d on y:

$$\mathbb{E}[y \mid d = 1] - \mathbb{E}[y \mid d = 0] = \beta_d + \beta_w \alpha_d \tag{36}$$

Interpretation: The total effect of d on y consists of:

- A direct effect: β_d

- An indirect effect that passes through w: $\beta_w \alpha_d$

Controlling for w in the regression isolates only the direct effect β_d. If we want the full impact of d on y — including the mechanism through w — we should not control for w.

Variables to Control For: Yes–No Cases III

- **Direct effect:**

$$\mathbb{E}(y \mid d = 1, w) - \mathbb{E}(y \mid d = 0, w) = \beta_d \qquad \text{(from the SF for } y\text{)} \tag{37}$$

- **Indirect effect:**

$$\{\mathbb{E}(w \mid d = 1) - \mathbb{E}(w \mid d = 0)\} \times \{\mathbb{E}(y \mid d, w = 1) - \mathbb{E}(y \mid d, w = 0)\} \tag{38}$$

$$= \alpha_d \cdot \beta_w \tag{39}$$

Interpretation: The direct effect β_d isolates the impact of d on y holding w fixed. The indirect effect captures the change in y that operates through how d shifts w (via α_d), and how w affects y (via β_w).

$$\text{Total effect} = \beta_d + \underbrace{\alpha_d \cdot \beta_w}_{\text{indirect}} \tag{40}$$

What is identified when x is controlled for?

- To show that:

(i) $y_0 \perp d \mid x \quad \Rightarrow \quad \mathbb{E}(y_1 - y_0 \mid d = 1, x)$ *(effect on the treated)*

(ii) $y_1 \perp d \mid x \quad \Rightarrow \quad \mathbb{E}(y_1 - y_0 \mid d = 0, x)$ *(effect on the untreated)*

(iii) $y_0 \perp d \mid x$ and $y_1 \perp d \mid x \quad \Rightarrow \quad \mathbb{E}(y_1 - y_0 \mid x)$ *(effect on the population)*

For (i), observe:

$$\begin{aligned}\mathbb{E}(y \mid d=1,x) - \mathbb{E}(y \mid d=0,x) &= \mathbb{E}(y_1 \mid d=1,x) - \mathbb{E}(y_0 \mid d=0,x) \\ &= \mathbb{E}(y_1 \mid d=1,x) - \mathbb{E}(y_0 \mid d=1,x) \\ &\quad + [\mathbb{E}(y_0 \mid d=1,x) - \mathbb{E}(y_0 \mid d=0,x)] \\ &= \mathbb{E}(y_1 - y_0 \mid d=1,x) + \underbrace{[\mathbb{E}(y_0 \mid d=1,x) - \mathbb{E}(y_0 \mid d=0,x)]}_{y_0 \text{ comparison group bias}}\end{aligned}$$

where

$$y_0 \text{ comparison group bias} \equiv \mathbb{E}(y_0 \mid d=1,x) - \mathbb{E}(y_0 \mid d=0,x)$$

Interpretation: If $y_0 \perp d \mid x$, then the comparison group bias is zero, and the difference in outcomes equals the average treatment effect on the treated.

What is identified when x is controlled for? II

We now analyze the conditional mean difference:

$$\mathbb{E}(y \mid d=1,x) - \mathbb{E}(y \mid d=0,x)$$

Recall the fundamental identity for observed outcome y:

$$y = dy_1 + (1-d)y_0$$

Thus, we can write:

$$\mathbb{E}(y \mid d=1,x) = \mathbb{E}(dy_1 + (1-d)y_0 \mid d=1,x) = \mathbb{E}(y_1 \mid d=1,x)$$

$$\mathbb{E}(y \mid d=0,x) = \mathbb{E}(dy_1 + (1-d)y_0 \mid d=0,x) = \mathbb{E}(y_0 \mid d=0,x)$$

Now consider the decomposition of the original expression:

$$\mathbb{E}(y \mid d=1,x) - \mathbb{E}(y \mid d=0,x) = \mathbb{E}(y_1 \mid d=1,x) - \mathbb{E}(y_0 \mid d=0,x)$$

Add and subtract $\mathbb{E}(y_1 \mid d=0,x)$ inside the expression:

$$= \underbrace{[\mathbb{E}(y_1 \mid d=1,x) - \mathbb{E}(y_1 \mid d=0,x)]}_{\textit{Comparison group bias on } y_1} + \underbrace{[\mathbb{E}(y_1 \mid d=0,x) - \mathbb{E}(y_0 \mid d=0,x)]}_{\textit{Average treatment effect on untreated}}$$

Therefore:

$$\mathbb{E}(y \mid d=1,x) - \mathbb{E}(y \mid d=0,x) = \text{bias} + \text{effect on untreated}$$

Interpretation:

- The first term represents a selection bias: the treated individuals differ from untreated even in their potential outcomes y_1.
- The second term is the causal effect of treatment on the untreated group.
- If $y_1 \perp d \mid x$, then the bias term disappears, and the conditional difference identifies the treatment effect on the untreated.