Take a long and short regression: OVB notes

$$y_i = \alpha^l + \beta^l T_i + \gamma X_i + \varepsilon_i \qquad \text{(long)}$$

$$y_i = \alpha^s + \beta^s T_i + u_i \qquad \text{(short)}$$

$$X_i = \alpha^A + \delta T_i + v_i \qquad \text{(auxillary)}$$

The OVB formula is a connection between the coeffients:

$$\beta^s = \beta^l + \gamma \delta$$

The use of it:

- when we don't have data on X_i, but would ideally control for it (ie, it's a confounder), we can reason that if (for example) we think $\gamma > 0$ and $\delta > 0$ then $\beta^l < \beta^S$
- To argue the signs of γ and δ we need to argue the correlation is positive (between X_i and Y_i, fixing T_i, for γ; and between T_i and X_i for δ)
- But we only use the OVB formula in

this way if we care about β^l instead of β^s.
This happens when X_i is a confounder – and to argue X_i is a confounder we need a <u>causal</u> story

(ie.

$$\begin{tikzcd}
T \arrow[rr] & & Y \\
& X \arrow[ul] \arrow[ur] &
\end{tikzcd}
\quad \underline{\text{not}} \quad
\begin{tikzcd}
T \arrow[rr] \arrow[dr] & & Y \\
& X \arrow[ur] &
\end{tikzcd}$$

)

remember: correlation between X and T cannot distinguish these!

<u>why does the OVB formula hold?</u>

The <u>formal proof</u> is in the lecture slides. The intuition I presented in class was that the OVB formula breaks the way that T_i correlates with Y_i in the short regression into two parts:

- A direct effect, holding X_i constant (β^l)
- An indirect effect "via" the way it correlates with X_i ($\gamma\delta$)

<u>NOTE:</u> NOTHING CAUSAL;
ALL CORRELATION

The reason I talked about the chain rule is because this has the same intuition (breaking into a direct and indirect part). note that you cannot derive the OVB formula from the chain rule: they're just analogous to each other.

The analogy is that:

$\beta^s = \dfrac{dY_i}{dT_i}$ → The total way in which different T_i is associated with different Y_i

$\beta^l = \dfrac{\partial Y_i}{\partial T_i}$ → The partial effect: keeping X_i constant

$\gamma : \dfrac{\partial Y_i}{\partial X_i}$ → same as above.

$\delta : \dfrac{\partial X_i}{\partial T_i}$ → same as above

Then just like the chain rule:

$$\underbrace{\frac{dY_i}{dT_i}}_{\text{total}} = \underbrace{\frac{\partial Y_i}{\partial T_i}}_{\text{direct}} + \underbrace{\frac{\partial Y_i}{\partial X_i} \cdot \frac{\partial X_i}{\partial T_i}}_{\text{indirect}} \iff \beta^s = \beta^l + \gamma\delta$$