ECON 4151
Lab Session: OVB

Wonjin Lee

Washington University in St. Louis

October 9, 2020

Omitted variable bias

- Structural equation:

$$y_i = \beta_0 + \beta_1 yeared_i + \beta_2 AFQT_i + \nu_i$$

- β_1 has a causal interpretation.
- Instead, we run

$$y_i = b_0 + b_1 yeared_i + \epsilon_i$$

- On average, what will this regression identify? What will $\hat{b}_1$ be an estimate of?

Omitted variable bias

- The bivariate regression formula:

$$
\begin{aligned}
b_1 &= \frac{\text{Cov}(yeared, y)}{\text{Var}(yeared)} \\
&= \frac{\text{Cov}\left(yeared, \beta_0 + \beta_1 yeared_i + \beta_2 AFQT + \nu_i\right)}{\text{Var}(yeared)} \\
&= \frac{\beta_1 \text{Var}(yeared) + \beta_2 \text{Cov}(yeared, AFQT)}{\text{Var}(yeared)} \\
&= \beta_1 + \beta_2 \frac{\text{Cov}(yeared, AFQT)}{\text{Var}(yeared)} \\
&= \beta_1 + \beta_2 \delta_1
\end{aligned}
$$

where δ_1 is coming from

$$AFQT_i = \delta_0 + \delta_1 yeared + \text{ residual}$$

Omitted variable bias

- $b_1 = \beta_1 + \beta_2\delta_1$.
- You should know the following table by heart and get good at signing the bias.

	$\text{Corr}(x_1, x_2) > 0$	$\text{Corr}(x_1, x_2) < 0$
$\beta_2 > 0$	positive bias	negative bias
$\beta_2 < 0$	negative bias	positive bias

- In our previous example, we would have a positive omitted variable bias. Why?
    1. β_2 was likely positive - i.e. intelligence has a positive effect on earnings, conditional on years of schooling, and
    2. δ_1 was likely positive, i.e. people with higher intelligence have, on average, more schooling.