Economics 113

Lecture Module 4

- Omitted Variables Bias

Multivariate Regression

Omitted variable bias

- What happens when we omit an important variable?
- Need to conjecture regarding the relationship between the omitted variable and included x and y variables
- Upward bias:
  - Estimate is on average *higher* than the true parameter: $E(\widehat{\beta}|x) > \beta$
- Downward bias:
  - Estimate is on average *lower* than the true parameter: $E(\widehat{\beta}|x) < \beta$
- Let us derive rigorously when these cases occur.

Multivariate Regression

Omitted variable bias

- Suppose that a *population* has the following relationship:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

- However, we forget about x_2 and estimate:

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1$$

- Our equation is *misspecified*.
- What is the relationship between the true β_1 and $\widehat{\beta}_1$?
- How do we quantify the bias?

Multivariate Regression

Omitted variable bias

- Population relationship between x_1 and x_2:

$$x_2 = \delta_0 + \delta_1 x_1 + \epsilon$$

- Plug x_2 into the population equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 \left(\delta_0 + \delta_1 x_1 + \epsilon\right) + u$$

- Expand:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 \delta_0 + \beta_2 \delta_1 x_1 + \beta_2 \epsilon + u$$

- Group terms:

$$y = \underbrace{\beta_0 + \beta_2 \delta_0}_{\widehat{\beta}_0} + \underbrace{\left(\beta_1 + \beta_2 \delta_1\right)}_{\widehat{\beta}_1} x_1 + \underbrace{\widetilde{u}}_{\beta_2 \epsilon + u}$$

- $\widehat{\beta}_1$ is an estimate of $\beta_1 + \beta_2 \delta_1$, not β_1.

- Simple formula for the bias:

$$Bias \quad = \quad \widehat{\beta}_1 - \beta_1 = \beta_1 + \beta_2 \delta_1 - \beta_1 = \boxed{\beta_2 \delta_1}$$

Multivariate Regression

Omitted variable bias

- To quantify the bias, we need the sign of β_2 and δ_1.
- If x_2 and x_1 are positively correlated, $\delta_1 > 0$. If not, then $\delta_1 < 0$.
- If x_2 has a positive effect on y, $\beta_2 > 0$. If not, then $\beta_2 < 0$.
- Put them together:

	Corr(x_1,x_2)>0 ($\delta_1 > 0$)	Corr(x_1,x_2)<0 ($\delta_1 < 0$)
x_2 has a positive effect on y ($\beta_2 > 0$)	positive bias ($\beta_2\delta_1 > 0$)	negative bias ($\beta_2\delta_1 < 0$)
x_2 has a negative effect on y ($\beta_2 < 0$)	negative bias ($\beta_2\delta_1 < 0$)	positive bias ($\beta_2\delta_1 > 0$)

Multivariate Regression

Omitted variable bias - Examples

- Example: Effect of class attendance on grades
- Population follows:

$$final = \beta_0 + \beta_1 attend + \beta_2 study + u$$

- We instead forget about *study* and estimate:

$$\widehat{final} = \widehat{\beta}_0 + \widehat{\beta}_1 attend$$

- Suppose we estimate $\widehat{\beta}_1 > 0$, and conclude that attendance increases your grade ($\beta_1 > 0$). Is this right?
- Positive correlation between *study* and *final*
- Positive correlation between *study* and *attend*
- $\widehat{\beta}_1$ suffers from an upward bias: $\beta_1 < \widehat{\beta}_1$

Multivariate Regression

Omitted variable bias - Examples

- Intuition
  - $\widehat{\beta}_1 > 0$ suggests that higher attendance improves your grade
  - However, students who attend class often tend to study more
  - Thus, *attend* may actually be accounting for the effects of studying, and not attendance.
- Overall, given $\beta_1 < \widehat{\beta}_1$, the result $\widehat{\beta}_1 > 0$ is *insufficient* to guarantee that $\beta_1 > 0$.

Multivariate Regression

Omitted variable bias - Examples

- Example: Effect of drugs on crime
- Population follows:

$$crime = \beta_0 + \beta_1 educ + \beta_2 drugs + u$$

- We instead forget about *drugs* and estimate:

$$\widehat{crime} = \widehat{\beta}_0 + \widehat{\beta}_1 educ$$

- Suppose we estimate $\widehat{\beta}_1 < 0$, and conclude education reduces your likelihood of committing a crime ($\beta_1 < 0$)
- Positive correlation between *drugs* and *crime*
- Negative correlation between *drugs* and *educ*
- $\widehat{\beta}_1$ suffers from an downward bias: $\widehat{\beta}_1 < \beta_1$

Multivariate Regression

Omitted variable bias - Examples

- Intuition
  - $\widehat{\beta}_1 < 0$ suggests that education reduces your likelihood of committing a crime
  - However, people who go to school are less likely to abuse drugs
  - Thus, *educ* may actually be accounting for the propensity of drug use, not the effects of education
- Overall, given $\widehat{\beta}_1 < \beta_1$, the result $\widehat{\beta}_1 < 0$ is *insufficient* to guarantee that $\beta_1 < 0$.

Multivariate Regression

Omitted variable bias - Examples

- Example: Effect of graduate education on wages
- Population follows:

$$\log(wage) = \beta_0 + \beta_1 geduc + \beta_2 Exper + u$$

- We instead forget about *Exper* and estimate:

$$\widehat{\log(wage)} = \widehat{\beta}_0 + \widehat{\beta}_1 \log(geduc)$$

- Suppose we estimate $\widehat{\beta}_1 > 0$, and conclude that graduate education increases your wage ($\beta_1 > 0$)
- Positive correlation between *Exper* and $\log(wage)$
- Negative correlation between *Exper* and *geduc* (by construction)
- $\widehat{\beta}_1$ suffers from an downward bias: $\widehat{\beta}_1 < \beta_1$

Multivariate Regression

Omitted variable bias - Examples

- Intuition
  - $\widehat{\beta}_1 > 0$ suggests that graduate education of some sort increases your wage
  - However, people who pursue graduate education have lower levels of experience
  - Thus, people with no graduate education may earn relatively high wages since they have lots of experience.
- Overall, given $\widehat{\beta}_1 < \beta_1$, the result $\widehat{\beta}_1 > 0$ is *sufficient* to guarantee that $\beta_1 > 0$.