

Handout for Lecture 18

Omitted Variable Bias

ECON 340: Economic Research Methods

Instructor: Div Bhagia

Consider the following linear regression model:

$$Y = \beta_0 + \beta_1 X + u$$

Omitted variable bias occurs when both of the following conditions are true:

- (1) The omitted variable is correlated with X
- (2) The omitted variable is a determinant of Y

Consider the following model:

$$TestScore = \beta_0 + \beta_1 \cdot STR + u$$

Here, $TestScore$ represents the average test score on standardized tests for a school, and STR denotes the student-teacher ratio for that school.

Which of these omitted factors do you think will introduce bias in our OLS estimate for β_1 ? Explain your reasoning.

- (a) percentage of English learners
- (b) time of day when tests were conducted
- (c) parking lot space per pupil (number of teacher parking spots divided by the number of students)
- (d) computers per student

Omitted variable bias implies that the exogeneity assumption is violated and hence $E(u|X) \neq 0$. In which case, we have that:

$$\hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(X, u)}{\text{Var}(X)}$$

So the direction and strength of bias $\hat{\beta}_1 - \beta_1$ depends on the correlation between u and X .

In our example:

$$\text{TestScore} = \beta_0 + \beta_1 \cdot \text{STR} + u$$

What should be the direction of bias due to the following omitted variables?

- (a) percentage of English learners
- (b) computers per student