POLS 6481, Spring 2020

Professor Scott Basinger

Reading Assignment Week 3

Distributed Monday, February 1 (*sorry!*)

Due Thursday, February 4

Required reading: Wooldridge 1.4 + 3.1 + 3.2 + 3.3 + Appx 3.A(.1–.4) + 5.1

1. Write the equations for the *sample regression function* and the *population regression function* for a linear model with two explanatory variables (i.e., *x*1 and *x*2). What are two differences between them?

2. In Wooldridge’s Example 3.1 (5th edition), he shows the following two equations:

|  |  |
| --- | --- |
|  |  |

A. Why are the slope coefficients for *ACT* different from each other?

B. What does the difference in slope coefficients tell us about the associations between…

(i) *ACT* scores and high school GPA

(ii) high school GPA and college GPA?

3. Theorem 3.1 states that OLS estimates of the coefficients are unbiased when four assumptions are met; omitted variable bias violates which one of the four assumptions? Explain why in your own words or with an example.

4. Suppose you have two variables, *x*1 and *x*2, that are positively correlated and that each has a positive correlation to *y* – yielding positive coefficients in simple regressions. Suppose further that *x*1 has a stronger association with *y* than *x*2 has. With these assumptions in place, do the following:

A. Show that if both *x*1 and *x*2 are included in a multiple regression, then the coefficient of *x*1 will be smaller than when *x*1 is included in alone a simple regression.

B. Show that in some circumstances, if both *x*1 and *x*2 are included in a multiple regression, the coefficient of *x*2 might switch to negative, in which case the coefficient on *x*1 actually gets larger!

Hint for 4A and 4B: numerical examples might be easier, and an arithmetic demonstration can – in this case – take the place of an algebraic demonstration.

5. Wooldridge includes the following table in chapter 3:



Choose one of the four cells in the table, and then choose (or invent) an example that illustrates the problem. Name the dependent variable (*y*), the included independent variable (*x*1), and the potentially omitted independent variable (*x*2). Explain how the correlation between *x*1 and *x*2 is either positive or negative, and stipulating a correlation between *y* and *x*2. Finally, explain the effect on estimates of *β*1 in a simple (bivariate) and multiple (multivariate) regression model.

6. According to section 5.1, is the problem of omitted variable bias made better or made worse by adding more observations? Why?