Problem 1. Lets think about the problem from the Wooldridge data set crime4. Suppose the true causal model is

$$lcrmrte = \beta_0 + \beta_1 polpc + \beta_2 taxpc + u$$

where

- lcrmrte: log of crime rate per capita
- polpc: police per capita
- taxpc: tax revenue per capta

a) Explain how to interpret β_1 as a causal effect.

b) Now think about running the regression

$$lcrmrte = \widehat{\beta_0} + \widehat{\beta_1} polpc + \widehat{\varepsilon}$$

where you have omitted taxpc. Which direction do you think the bias will go in this regression? What does that mean about how the coefficient on polpc will compare if you ran this regression versus the correct specification above?

c) Using the stata data set now run these two regressions. Did you find what you expected?

d) Also try running a regression of taxpc on polpc. Does this make sense?

e) What are other potential problems you can think about with this regression?

Problem 2. Now think about the problem from the Wooldridge data set nbasal which is a data set on wages of NBA players. Suppose the true causal model is

$$wage = \beta_0 + \beta_1 draft + \beta_2 points + u$$

where

- wage: Annual Salary in thousands of dollars per year
- draft: draft number
- points: points per game

In the NBA there is a draft for the best players in college. In this data we want to know the relationship between a players draft status and their annual income. The best players are drafted first so we would probably expect β_1 to be negative. We think about points as a measure of the players quality.

a) Explain how to interpret β_1 as a causal effect.

b) Now think about running the regression

$$wage = \widehat{\beta}_0 + \widehat{\beta}_1 draft + \widehat{\varepsilon}$$

where you have omitted points. Which direction do you think the bias will go in this regression? What does that mean about how the coefficient on draft will compare if you ran this regression versus the correct specification above.

c) Using the stata data set now run these two regressions. Did you find what you expected?

d) What are other potential problems you can think about with this regression?

Problem 3 Wooldridge 3.3 (i)-(iv).

Problem 4 Wooldridge problem C3.4 (i)-(iii). In doing this think of this model as a causal model. Run the regression without including priGPA. What happens to the estimate of $\widehat{\beta}_1$? Is this what you would have expected, why or why not?