Homework 3

STUDENT NAME

Due 7/11 11:59pm

NOTE: If you would like to answer the written portions using RMarkdown, simply type your answers below the questions. If you would like to print out the pdf to do the written portions, you should add `\vspace{50mm}` to add space in the knitted pdf for written work (adjust the 50mm as necessary). Alternatively, you can do the written portions on a separate paper.

Part 1: Theory

1

(Based on Wooldridge 9.4) Consider regressing the hours of television watched by a child on that child's age, their parent's education, and number of siblings they have, like so:

$$HoursTV = \beta_0 + \beta_1 Age + \beta_2 Age^2 + \beta_3 Parent1Eductation + \beta_4 Parent2Education + \beta_5 Siblings + u$$

You estimate this model using survey data where parents report the hours of TV that their children watch. Parental estimates of their children's TV watching may not be completely accurate, so call this data for the dependent variable $HoursTV^* = HoursTV + e$.'

1.1

What would we need to assume about e in order to conclude that our parameter estimates will be unbiased in the face of this measurement error?

1.2

Do you expect the assumption in 1.1 to be true in this case? Why or why not?

1.3

Assume that the assumption you named in 1.1 as well as the rest of the Gauss-Markov assumptions are true, except that the sample you estimate the model with only includes children over the age of five. Would this bias the parameter estimates, why or why not?

2

You want to know how proximity to highway pollution influences human health. You collect data from many city blocks and run the following regression

$$Cancer = \beta_0 + \beta_1 MilesToHighway + u$$

where *Cancer* is the percent of residents that have had cancer in the past 5 years and *MilesToHighway* measures the distance in miles from that block to a highway.

2.1

Name an omitted variable that could bias your estimates of β_1.

2.2

Sign the bias caused by the omission of the variable you named in 2.1 and explain why it has that sign.

3

Suppose some California cities introduce a new drought charge that increases the price of water during drought conditions. You collect data on household water consumption in various California cities (some of which do an some of which do not have these charges) during a drought. You run the following regression:

$$WaterConsumption = \beta_0 + \beta_1 DroughtCharge + \beta_2 Income + \beta_3 Pool + \beta_4 DroughtCharge \times Pool + u$$

where DroughtCharge is the per-gallon drought charge (0 in places without the charge) and *Pool* is an indicator variable for whether the household has a pool.

3.1

What is the predicted change in water consumption when a city's drought charge goes up by one dollar?

3.2

Carefully interpret $\hat{\beta}_3$.

3.3

Carefully interpret $\hat{\beta}_4$.

3.4

In places where more households have large lawns, cities may impose drought charges more frequently and household water consumption will be higher due to lawn irrigation. This would be an omitted variable. What would be the sign of the bias this would cause for $\hat{\beta}_1$?

3.5

Say you find that the t-statistic for $\hat{\beta}_1$ is -3.2. If $\hat{\beta}_1$ has the bias that you described in 3.4, are you comfortable rejecting the null hypothesis $H_0 : \beta_1 = 0$? Why or why not?

Part 2: Application

Preliminaries

Generate 10,000 observations, including an x variable that is distributed normal with mean 0 and standard deviation 2 and an error term u that is distributed normal with mean 0 and standard deviation 1. Using this, generate the y values as $y = 10 - 5x + u$.

1

Take 500 separate random draws of 100 observations each, running the regression $y = \beta_0 + \beta_1 x + u$ on each sample and saving the coefficient estimates $\hat{\beta}_1$. (You should end up with 500 different parameter estimates for β_1.) Plot a histogram of these parameter estimates (don't worry about making it pretty) and calculate their mean. Is $\hat{\beta}_1$ unbiased?

2

Now define x^* as a version of x with measurement error, specifically let $x^* = x + e$ where e is normally distributed with mean 0 and standard deviation 1. Using x^* instead of x will cause a bias in $\hat{\beta}_1$. The true value of β_1 is -5. Show (using theory) what $\mathbb{E}(\hat{\beta}_1)$ is when estimated using x^* instead of x.

3

Take 500 separate random draws of 100 observations each, running the regression $y = \beta_0 + \beta_1 x^* + u$ on each sample and saving the coefficient estimates $\hat{\beta}_1$. Plot a histogram of these parameter estimates (don't worry about making it pretty). Does this validate what you showed in 1.2?