Problem Set 2

EC 204: Empirical Economics 2

Due on: Thursday July 23, 2020

Be sure to submit your responses directly to Blackboard. Include your .do file for the computational questions. Lastly, if you are working with a partner or as a group of 3, be sure to include everyone's name so that we can assign the grades accordingly.

True/False/Uncertain (20 points)

For the each of the following statements, indicate whether it is true, false, or if the assertion is indeterminate. In all cases, defend your answer with a concise explanation. **To earn full credit, you must include an explanation with your answer.**

1. If homoskedasticity does not hold, then OLS is biased.
2. Suppose you wish to run a regression of $\log(profits)$ on *capital* and *labor*, and. Unfortunately, *capital* is not in your dataset (so you do not observe this variable). You decide to then regress $\log(profits)$ on *labor* only. Since *capital* is positively related to *labor*, running this regression results in an OLS coefficient on *labor* to be biased upward.
3. If you run a regression of *output* on *labor* and *capital*, and then run a regression of *output* on *labor*, *capital*, and *land*, the R^2 in the second regression will likely be lower than the R^2 in the first regression.
4. Suppose the true population model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$. Less variation in x_3 in your sample will lead to a higher variance of the OLS estimator, $\hat{\beta}_3$.

Longer Questions (20 Points)

1. **(Omitted variable bias)** For all the handwringing we do about the omitted variable bias, there are some (special) scenarios under which it exists but may not be a huge problem. Consider the following set up

$$\begin{aligned} \text{True Model}: &\quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \\ \text{Biased Model}: &\quad y = \widetilde{\beta}_0 + \widetilde{\beta}_1 x_1 + v \end{aligned}$$

   (a) State the mathematical definition of the omitted variable bias and describe (in words) what it means.

   (b) Construct an example in which $\widetilde{\beta}_1$ is biased downward. State the assumptions needed in order to reach this conclusion

(c) How could you change your assumptions in (b) in order to obtain the same result (i.e. that $\widetilde{\beta}_1$ is biased downward)?

(d) Now suppose that x_1 and x_2 are uncorrelated. What happens to the bias?

(e) What if β_2 is as you had it in (b), but the correlation between x_1 and x_2 is nearly zero?

2. (**Sales Regression**) Consider the following regression model

$$\log(sales) = \beta_0 + \beta_1 emp + \beta_2 tenure + u$$

where *sales* denotes firm sales in millions of US dollars, *emp* denotes the number of employees working the firm, and *tenure* denotes the number of years that CEO has been at the firm.

(a) What are the dependent variable and independent variables in this model?

(b) What are the parameters of the model?

(c) How do you interpret β_1?

(d) Give a specific example of a variable that can be included in the error term in this model.

(e) State the zero conditional mean assumption in this case and discuss whether it is likely to hold.

Computational Exercises (20 Points)

1. **(Wooldridge Ch 3, C2)** Use the data in HPRICE.DTA to estimate the model

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u$$

where *price* is the house price measured in thousands of dollars

(a) Write out the results in equation form.

(b) What is the estimated increase in price for a house with one more bedroom, holding square footage constant?

(c) What is the estimated increase in price for a house with an additional bedroom that is 140 square feet in size? Compare this to your answer in part (b)

(d) What percentage of the variation in price is explained by square footage and number of bedrooms?

(e) The first house in the sample has $sqrft = 2,438$ and $bdrms = 4$. Find the predicted selling price of for this house from the OLS regression line.

(f) The actual selling price of the first house in the sample was $\$300,000$ (so $price = 300$). Find the residual for this house. Does it suggest that the buyer underpaid or overpaid for the house?

2. **(Wooldridge Ch 3, C6)** Use the data in WAGE2.DTA to answer this question.

(a) Run a simple regression of `IQ` on `educ` to obtain the slope coefficient, say $\widetilde{\delta}_1$.

(b) Run the simple regression of log(`wage`) on `educ` and obtain the slope coefficient, $\widetilde{\beta}_1$.

(c) Run the multiple regression of log(`wage`) on `educ` and `IQ`, and obtain the slope coefficients $\widehat{\beta}_1$ and $\widehat{\beta}_2$, respectively.

(d) Verify that $\widetilde{\beta}_1 = \widehat{\beta}_1 + \widehat{\beta}_2 \delta_1$.