Assignment #4

ECON 2023 Introductory Econometric

March 25, 2023

1 Multiple Regression

1. Which of the following can cause OLS estimators to be biased? Explain why and why not.
   (a) Heteroskedasticity.
   (b) Omitting an important variable.
   (c) A sample correlation coefficient of .95 between two independent variables both included ni the model.

2. (a) Consider the simple regression model $y = \beta_0 + \beta_1 x + u$under the first four Gauss-Markov assumptions. For some function $g(x)$, for example $g(x) = x^2$ or $g(x) = \log(1 + x^2)$, define $x_i = g(x_i)$. Define a slope estimator as
   $$\tilde{\beta_1} = \left(\sum_{i=1}^{n}(z_i - \bar{z})y_i\right) / \left(\sum_{i=1}^{n}(z_i - \bar{z})x_i\right)$$
   Show that β_1 is unbiased.
   (b) Add the homoskedasticity assumption, MLR.5. Show that
   $$Var(\tilde{\beta_1}) = \sigma^2 \left(\sum_{i=1}^{n}(z_i - \bar{z})^2\right) / \left(\sum_{i=1}^{n}(z_i - \bar{z})x_i\right)^2$$
   (c) Show directly that under the Gauss-Markov assumptions, $Var(\hat{\beta_1}) \leq Var(\tilde{\beta_1})$, where is the OLS estimator.[Hint: The Cauchy-Schwartz inequality implies that
   $$\left(\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})\right)^2 \leq \left(\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right);$$
   notice that we can drop $\bar{x}$ from the sample covariance.]

3. Consider an equation to explain salaries of CEOs in terms of annual firm sales, return on equity (roe, ii percentage form), and return on the firm's stock (ros, in percentage form):
   $$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roe + \beta_3 ros + u.$$
   (a) In terms of the model parameters, state the null hypothesis that, after controlling for $sales$ and roe, ros has no effect on CEO salary. State the alternative that better stock market performance increases a CEO's salary.
   (b) Using the data ii CEOSALI, the following equation was obtained by OLS:
   $$\widehat{\log(salary)} = \underset{(0.32)}{4.32} + \underset{(0.035)}{0.28}\log(sales) + \underset{(0.0041)}{0.0174} roe + \underset{(0.00054)}{0.00024} ros + u.$$
   By what percentage is salary predicted to increase if ros increases by 50 points? Does ros have a practically large effect on salary?
   (c) Test the null hypothesis that ros has no effect on salary against the alternative that ros has a positive effect. Carry out the test at the 10% significance level.
   (d) Would you include ros in a final model explaining CEO compensation in terms of firm performance? Explain.

Software Problem Set

1. Use the data in HTV to answer this question. The data set includes information on wages, education, parents' education, and several other variables for 1,230 working men in 1991.

   (a) What is the range of the *educ* variable in the sample? What percentage of men completed twelfth grade but no higher grade? Do the men or their parents have, on average, higher levels of education?

   (b) Estimate the regression model

   $$educ = \beta_0 + \beta_1 motheduc + \beta_2 fatheduc + u$$

   by OLS and report the results in the usual form. How much sample variation in *educ* is explained by parents' education? Interpret the coefficient on *motheduc*.

   (c) Add the variable *abil* (a measure of cognitive ability) to the regression from part (a), and report the results in equation form. Does **ability** help to explain variations in education, even after controlling for parents' education? Explain.

2. Use the data ni DISCRIM to answer this question. These are ZIP code-level data on prices for various items at fast-food restaurants, along with characteristics of the zip code population, in New Jersey and Pennsylvania. The idea is to see whether fast-food restaurants charge higher prices in areas with a larger concentration of blacks.

   (a) Find the average values of *prpblck* and *income* in the sample, along with their standard deviations. What are the units of measurement of *prpblck* and *income*?

   (b) Consider a model to explain the price of soda, *psoda*, in terms of the proportion of the population that is black and median income:

   $$psoda = \beta_0 + \beta_1 prpblck + \beta_2 ncome + u.$$

   Estimate this model by OLS and report the results in equation form, including the sample size and R-squared. (Do not use scientific notation when reporting the estimates.) Interpret the coefficient on *prpblck*. Do you think it is economically large?

   (c) Compare the estimate from part (b) with the simple regression estimate from *psoda* on *prpbick*. Is the discrimination effect larger or smaller when you control for income?