Econ 140 - Spring 2016
Section 6

GSI: Fenella Carpena

March 3, 2016

1 Omitted Variable Bias

Exercise 1.1. (Stock & Watson, Review the Concepts, Exercise 6.1) A researcher is interested in the effect on test scores of computer usage. Using school district data like that used in this chapter, she regresses district average test scores on the number of computers per student. Will $\widehat{\beta}_1$ be an unbiased estimator of the effect on test scores of increasing the number of computers per student? Why or why not? If you think $\widehat{\beta}_1$ is biased, is it biased up or down? Why?

2 Multiple Regression Model

Exercise 2.1. (Stock & Watson, Review the Concepts, Exercise 6.2) A multiple regression includes two regressors: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$.

(a) What is the expected change in Y if X_1 increases by 3 units and X_2 is unchanged?

(b) What is the expected change in Y if X_2 decreases by 5 units and X_1 is unchanged?

(c) What is the expected change in Y if X_1 increases by 3 units and X_2 decreases by 5 units?

Exercise 2.2. (Stock & Watson, Exercise 6.5) Data were collected from a random sample of 220 home sales from a community in 2013. Let $Price$ denote the selling price (in \$1000), BDR denote the number of bedrooms, $Bath$ denote the number of bathrooms, $Hsize$ denote the size of the house (in square feet), $Lsize$ denote the lot size (in square feet), Age denote the age of the house (in years), and $Poor$ denote a binary variable that is equal to 1 if the condition of the house is reported as "poor." An estimated regression yields

$$\widehat{Price} = 119.2 + 0.485BDR + 23.4Bath + 0.156Hsize + 0.002Lsize + 0.090Age - 48.8Poor$$

(a) Suppose that a homeowner converts part of an existing family room in her house into a new bathroom. What is the expected increase in the value of the house?

(b) Suppose that a homeowner adds a new bathroom to her house, which increases the size of the house by 100 square feet. What is the expected increase in the value of the house?

(c) What is the loss in value if a homeowner lets his house run down so that its condition becomes "poor"?

3 $\overline{R}^2$ (Adjusted R^2)

Exercise 3.1. (Stock & Watson, Review the Concepts, Exercise 6.3) How does $\overline{R}^2$ differ from R^2? Why is $\overline{R}^2$ useful in a regression model with multiple regressors?

4 Perfect and Imperfect Multicollinearity

Exercise 4.1. (Stock & Watson, Review the Concepts, Exercise 6.4) Explain why two perfectly multicollinear regressors cannot be included in a linear multiple regression. Give two examples of a pair of perfectly multicollinear regressors.

Exercise 4.2 (Stock & Watson, Review the Concepts, Exericse 6.5) Explain why it is difficult to estimate precisely the partial effect of X_1 holding X_2 constant, if X_1 and X_2 are highly correlated.

5 Additional Exercises

Question 1. Consider the following regression model to explain city crime rates ($crimerate$) in terms of the probability of conviction ($prbconv$) and average sentence length ($avgsen$).

$$crimerate_i = \beta_0 + \beta_1 * prbconv_i + \beta_2 * avgsen_i + u_i.$$

What are some factors contained in u_i? Do you think that the first least squares assumption in this model is likely to hold?

Question 2. Consider the regression $Y_i = \beta_0 + \beta_1 X_i + u_i$. Does a higher R^2 mean that it is more likely that X_i causes Y_i? What about a higher $\overline{R}^2$?

Question 3. Suppose we want to estimate the effect campaign spending on campaign outcomes. Let $voteA_i$ denote the percent of the vote for candidate A in county i, let $expendA_i$ denote the campaign expenditures for candidate A, let $expendB_i$ denote the campaign expenditures for candidate B. Let $totexp_i$ be the total expenditures across both candidates, and let $shareA_i$ denote the percentage of total campaign expenditures made by candidate A (in other words $shareA_i = 100 \cdot (expendA_i/totexp_i)$. Consider the regression model

$$voteA_i = \beta_0 + \beta_1 expendA_i + \beta_2 expendB_i + \beta_3 shareA_i + u_i.$$

Does this regression violate the fourth least squares assumption? Explain why or why not.

Question 4. Which of the following can cause OLS estimators to be biased? Select all that apply.
(a) Heteroskedasticity
(b) Omitting an important variable
(c) X_1 and X_2 are both included in the regression, but their sample correlation is very close to 1

Question 5. In a study relating college grade point average to time spent in various activities, you distribute a survey to several students. The students are asked how many hours the spend each week in four activities: studying, sleeping, working, and leisure. Any activity is put into one of the four categories, so that for each student, the sum of hours in the four activities must be 168.

(a) In the model $GPA_i = \beta_0 + \beta_1 study_i + \beta_2 sleep_i + \beta_3 work_i + \beta_4 leisure_i + u_i$, does it make sense to hold *sleep*, *work* and *leisure* fixed, while changing *study*?

(b) Explain why this model violates the fourth least squares assumption.

(c) How could you reformulate the model so that its parameters have a useful interpretation and it satisfies the fourth least squares assumption? Provide an interpretation of one of the parameters in your proposed model.