

## Problem Set 4

### WELCOME

Chapters 8 and 9 problems! Enjoy!

See the exercises below, or you can [download them as a pdf](#).

### WHAT DO I SUBMIT?

- Your written up answers to exercise questions. If you work on a piece of paper, please scan using some sort of phone software (like Microsoft Lens or Adobe Scan) rather than just taking a picture.
- A do-file that runs your Stata analysis (for question 8).
- A log file that includes the output from running your do-file (for question 8).

### EXERCISES

1. The following equation describes the median housing price in a community in terms of amount of pollution ( $nox$  for nitrous oxide) and the average number of rooms in houses in the community ( $rooms$ ):

$$\log(\text{price}) = \beta_0 + \beta_1 \log(\text{nox}) + \beta_2 \text{rooms} + u$$

- (a) What are the probable signs of  $\beta_1$  and  $\beta_2$ ? What is the interpretation of  $\beta_1$ ? Explain.
- (b) Why might  $nox$  [or more precisely,  $\log(\text{nox})$ ] and  $rooms$  be negatively correlated? If this is the case, does the simple regression of  $\log(\text{price})$  on  $\log(\text{nox})$  produce an upward or a downward biased estimator of  $\beta_1$ ?
- (c) Using data, the following equations were estimated:

$$\widehat{\log(\text{price})} = 11.71 - 1.043 \log(\text{nox}), n = 506, R^2 = 0.264$$

$$\widehat{\log(\text{price})} = 9.23 - 0.718 \log(\text{nox}) + 0.306 \text{rooms}, n = 506, R^2 = 0.514$$

Is the relationship between the simple and multiple regression estimates of the elasticity of  $\text{price}$  with respect to  $nox$  what you would have predicted, given your answer in part (b)? Does this mean that 0.718 is definitely closer to the true elasticity than 1.043?

2. Read the box “*The Return to Education and the Gender Gap*” in Section 8.3 of your textbook (Stock & Watson).
  - (a) Consider a man with 16 years of education and 2 years of experience. Use the results from column (4) of Table 8.1 and the method in Key Concept 8.1 to estimate the expected change in the logarithm of average hourly earnings (AHE) associated with an additional year of experience.

- (b) Explain why your answer to (a) does not depend on the region he is from.
- (c) Repeat (a), assuming 10 years of experience.
3. To answer this question, refer to *Table 8.3: Nonlinear Regression Model of Test Scores* in your textbook:
- (a) A researcher suspects that the effect of % Eligible for subsidized lunch has a nonlinear effect on test scores. In particular, he conjectures that increases in this variable from 10% to 20% have little effect on test scores but that changes from 50% to 60% have a much larger effect. i. Describe a nonlinear specification that can be used to model this form of nonlinearity. ii. How would you test whether the researcher's conjecture was better than the linear specification in column (7) of Table 8.3?
- (b) A researcher suspects that the effect of income on test scores is different in districts with small classes than in districts with large classes. i. Describe a nonlinear specification that can be used to model this form of nonlinearity.
4. Labor economists studying the determinants of women's earnings discovered a puzzling empirical result. Using randomly selected employed women, they regressed earnings on the women's number of children and a set of control variables (age, education, occupation, and so forth). They found that women with more children had higher wages, controlling for these other factors. Explain how sample selection might be the cause of this result. (Hint: Notice that women who do not work outside the home are missing from the sample.) [This empirical puzzle motivated James Heckman's research on sample selection that led to his 2000 Nobel Prize in Economics. See Heckman (1974)]
5. This question uses directed acyclic graphs (DAGs), which we will cover in class. You may also find it helpful to read Huntington-Klein, *The Effect*, [Chapter 8: Causal Paths and Closing Back Doors](#), especially Sections 8.3–8.5.
- Consider the relationship between a woman's number of children and her earnings from question 4.
- (a) Draw a DAG that includes the following variables: Earnings, Number of Children, Decision to Work Outside the Home, and Ability/Motivation. Add arrows representing plausible causal relationships. For each arrow, write one sentence explaining why you included it.
- (b) Is "Decision to Work Outside the Home" a confounder, a collider, or a mediator on the path between Number of Children and Earnings? Explain.
- (c) When researchers study only employed women, they are conditioning on "Decision to Work." Using your DAG, explain why this could produce a spurious positive relationship between number of children and earnings — even if children have no direct causal effect on earnings.
6. The demand for a commodity is given by  $Q = \beta_0 + \beta_1 P + u$ , where  $Q$  denotes quantity,  $P$  denotes price, and  $u$  denotes factors other than price that determine demand. Supply for the commodity is given by  $Q = \gamma_0 + \gamma_1 P + v$ , where  $v$  denotes factors other than price that determine supply. Suppose  $u$  and  $v$  both have a mean of 0, have variances  $\sigma_u^2$  and  $\sigma_v^2$ , and are mutually uncorrelated.

- (a) Solve the two simultaneous equations to show how  $Q$  and  $P$  depend on  $u$  and  $v$ . (*Hint: In equilibrium, quantity supplied equals quantity demanded. Set the two equations equal and solve for  $P$  in terms of  $u$  and  $v$ . Then substitute back to find  $Q$ .*)
- (b) Derive the means of  $P$  and  $Q$ . (*Hint: Use your answers from part (a) and the fact that  $E(u) = E(v) = 0$ .*)
- (c) **(Optional)** Derive the variance of  $P$ , the variance of  $Q$ , and the covariance between  $Q$  and  $P$ .
7. Revisit the box “*The Return to Education and the Gender Gap*” in Section 8.3 of your textbook (Stock & Watson). Discuss the internal and external validity of the estimated effect of education on earnings.
8. Use the dataset `CollegeDistance.dta` (described in Empirical Exercise AEE 4.3) to answer the following questions.
- (a) Run a regression of  $ED$  on  $Dist$ ,  $Female$ ,  $Bytest$ ,  $Tuition$ ,  $Black$ ,  $Hispanic$ ,  $Incomehi$ ,  $Ownhome$ ,  $DadColl$ ,  $MomColl$ ,  $Cue80$ , and  $Stwmfg80$ . If  $Dist$  increases from 2 to 3 (that is, from 20 to 30 miles), how are years of education expected to change? If  $Dist$  increases from 6 to 7 (that is, from 60 to 70 miles), how are years of education expected to change?
- (b) Run a regression of  $\ln(ED)$  on  $Dist$ ,  $Female$ ,  $Bytest$ ,  $Tuition$ ,  $Black$ ,  $Hispanic$ ,  $Incomehi$ ,  $Ownhome$ ,  $DadColl$ ,  $MomColl$ ,  $Cue80$ , and  $Stwmfg80$ . If  $Dist$  increases from 2 to 3 (from 20 to 30 miles), how are years of education expected to change? If  $Dist$  increases from 6 to 7 (from 60 to 70 miles), how are years of education expected to change?
- (c) Run a regression of  $ED$  on  $Dist$ ,  $Dist^2$ ,  $Female$ ,  $Bytest$ ,  $Tuition$ ,  $Black$ ,  $Hispanic$ ,  $Incomehi$ ,  $Ownhome$ ,  $DadColl$ ,  $MomColl$ ,  $Cue80$ , and  $Stwmfg80$ . If  $Dist$  increases from 2 to 3 (from 20 to 30 miles), how are years of education expected to change? If  $Dist$  increases from 6 to 7 (from 60 to 70 miles), how are years of education expected to change?
- (d) Do you prefer the regression in (c) to the regression in (a)? Explain.
- (e) Add the interaction term  $DadColl \times MomColl$  to the regression in (c). What does the coefficient on the interaction term measure?
- (f) Mary, Jane, Alexis, and Bonnie have the same values of  $Dist$ ,  $Bytest$ ,  $Tuition$ ,  $Female$ ,  $Black$ ,  $Hispanic$ ,  $Incomehi$ ,  $Ownhome$ ,  $Cue80$ , and  $Stwmfg80$ . Neither of Mary’s parents attended college. Jane’s father attended college, but her mother did not. Alexis’s mother attended college, but her father did not. Both of Bonnie’s parents attended college. Using the regressions from (e):
- What does the regression predict for the difference between Jane’s and Mary’s years of education?
  - What does the regression predict for the difference between Alexis’s and Mary’s years of education?
  - What does the regression predict for the difference between Bonnie’s and Mary’s years of education?
- (g) Is there any evidence that the effect of  $Dist$  on  $ED$  depends on the family’s income?
- (h) After running all these regressions (and any others that you want to run), summarize the effect of  $Dist$  on years of education.