

Problem Set 4
Multiple Regression

INTRODUCTORY ECONOMETRICS, UCM3

1. (Exercise 3.1, Wooldridge (2006)). Using the data in GPA2.RAW on 4,137 college students, the following equation was estimated by OLS:

$$\widehat{colgpa} = 1.392 - 0.0135hsperc + 0.00148sat$$

$$n = 4,137, R^2 = 0.273$$

where *colgpa* is measured on a four-point scale, *hsperc* is the percentile in the high school graduating class (defined so that, for example, *hsperc* = 5 means the *top* five percent of the class), and *sat* is the combined math and verbal scores on the student achievement test.

- a) Why does it make sense for the coefficient on *hsperc* to be negative?
- b) What is the predicted college GPA when *hsperc* = 20 and *sat* = 1050?
- c) Suppose that two high school graduates, A and B, graduated in the same percentile from high school, but Student A's SAT score was 140 points higher (about one standard deviation in the sample). What is the predicted difference in college GPA for these two students? Is the difference large?
- d) Holding *hsperc* fixed, what difference in SAT scores leads to a predicted *colgpa* difference of .50, or one-half of a grade point? Comment on your answer.

2. (Exercise 3.3, Wooldridge (2006)). The following model is a simplified version of the multiple regression model used by Biddle and Hamermesh (1990) to study the tradeoff between time spent sleeping and working and to look at other factors affecting sleep:

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + u$$

where *sleep* and *totwrk* (total work) are measured in minutes per week and *educ* and *age* are measured in years. (See also Problem 2.12.)

- a) If adults trade off sleep for work, what is the sign of β_1 ?
- b) What signs do you think β_2 and β_3 will have?
- c) Using the data in SLEEP75.RAW, the estimated equation is

$$\widehat{sleep} = 3638.25 - 0.148totwrk - 11.13educ + 2.20age$$

$$n = 706, R^2 = 0.113$$

If someone works five more hours per week, by how many minutes is *sleep* predicted to fall? Is this a large tradeoff?

- d) Discuss the sign and magnitude of the estimated coefficient on *educ*.
- e) Would you say *totwrk*, *educ*, and *age* explain much of the variation in *sleep*? What other factors might affect the time spent sleeping? Are these likely to be correlated with *totwrk*?

3. (Exercise 3.4, Wooldridge (2006)). The median starting salary for new law school graduates is determined by

$$\log(\widehat{salary}) = \beta_0 + \beta_1 LSAT + \beta_2 GPA + \beta_3 \log(libvol) + \beta_4 \log(cost) + \beta_5 rank + u$$

where $LSAT$ is median LSAT score for the graduating class, GPA is the median college GPA for the class, $libvol$ is the number of volumes in the law school library, $cost$ is the annual cost of attending law school, and $rank$ is a law school ranking (with $rank = 1$ being the best).

- Explain why we expect $\beta_5 \leq 0$.
- What signs do you expect for the other slope parameters? Justify your answers.
- Using the data in LAWSCH85.RAW, the estimated equation is

$$\log(\widehat{salary}) = 8.34 + 0.0047 LSAT + 0.248 GPA + 0.095 \log(libvol) + 0.038 \log(cost) - 0.0033 rank$$

$$n = 136, R^2 = 0.842$$

What is the predicted ceteris paribus difference in salary for schools with a median GPA different by one point? (Report your answer as a percent.)

- Interpret the coefficient on the variable $\log(libvol)$.
- Would you say it is better to attend a higher ranked law school? How much is a difference in ranking of 20 worth in terms of predicted starting salary?

4. (Exercise 3.5, Wooldridge (2006)). 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 they 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.

- In the model

$$GPA = \beta_0 + \beta_1 study + \beta_2 sleep + \beta_3 work + \beta_4 leisure + u$$

does it make sense to hold *sleep*, *work*, and *leisure* fixed, while changing *study*?

- Explain why this model violates Assumption MLR.4.
- How could you reformulate the model so that its parameters have a useful interpretation and it satisfies Assumption MLR.4?