Department of Fundamentos del Análisis Económico.
University of Alicante
2014/15 Academic Year
STATISTICS AND INTRODUCTION TO ECONOMETRICS
Problem Set 7

1. The following model relates the (average) number of cigarettes smoked per day ($cigs$) with the age in years (age) and the years of schooling ($educ$)

$$cigs = \beta_0 + \beta_1 age + \beta_2 age2 + \beta_3 educ + u$$

where $age2 = age^2$. Based on a sample of 807 individual we have obtained the following results

$$\widehat{cigs} = \underset{}{-0.331} + \underset{(3.51)}{0.824}\ age - \underset{(0.155)}{0.0095} age2 - \underset{(0.0017)}{0.471} educ - \underset{(0.162)}{}$$
$$n = 807 \qquad R^2 = 0.044$$

   (a) Interpret the estimated coefficient of $educ$.

   (b) What is the partial effect of age on the number of cigarettes smoked per day? Find the estimated partial effect for 25 years old individuals. Would the estimated effect be the same for someone who is 30 years old? Comment the results.

   (c) Predict cigarette consumption for a 30 years old individual with 12 years of schooling.

   (d) Ann and Ben have the same age, which difference in years of schooling between Ann and Ben would predict that Ann smokes 2 cigarettes more per day than Ben?

   (e) Suppose now that the dependent variable is measured in packs of cigarettes smoked per day (one pack = 20 cigarettes). Find the new estimated coefficients and standard errors.

2. Consider the following model to study the trade-off between time spent sleeping and working and to look at other factors affecting sleep

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

where $sleep$ and $totwrk$ (total work) are measured in minutes per week and $educ$ and age are measured in years.

   (a) If adults trade off sleep for work, what is the sign of β_1?

(b) Using the data in SLEEP75 from Wooldridge book, the estimated equation is

$$\begin{aligned} \widehat{sleep} &= 3638.25 - 0.148 torwrk - 11.13 educ + 2.2\, age \\ n &= 706, \qquad R^2 = 0.113 \end{aligned}$$

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

(c) Discuss the sign and magnitude of the estimated coefficient on *educ*.

(d) 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*?

(e) If we now measure *sleep* and *totwrk* in hours per week, find the estimated parameters, the standard errors and the R^2.

3. The median initial salary of the newly graduated lawyers in the U.S. is determined by:

$$\log(salary_t) = \beta_0 + \beta_1 LSAT_t + \beta_2 GPA_t + \beta_3 \log(libvol_t) + \beta_4 \log(cost_t) + \beta_5 rank_t + u_t$$

where *salary* is the median initial salary of recently graduated lawyers, $LSAT$ is the median result of the test and GPA is the mean grade in the university for that same group of students, *libvol* is the number of books in the library of the law faculty, *cost* is the yearly tuition cost of the law faculty and *rank* is the position that the law faculty has in the law faculty ranking of the U.S. (being $rank = 1$ the best one).

(a) Explain why we expect $\beta_5 \leq 0$.

(b) What signs to you expect for the other slope parameters? Explain.

(c) Using the data in LAWSCH85 from Wooldridge book, the estimated equation is

$$\begin{aligned} \widehat{\log(salary)} &= 8.34 + 0.0047 LSAT + 0.248 GPA + 0.095 \log(libvol) \\ &\quad + 0.38 \log(cost) - 0.0033 rank \\ n &= 136, \qquad R^2 = 0.842 \end{aligned}$$

What is the predicted ceteris paribus difference in salary for schools with a median GPA different by one point?

(d) Interpret the coefficient on the variable $\log(libvol)$.

(e) 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. 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.

   (a) 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*?

   (b) Explain why this model violates Assumption MLR.4.

   (c) How could you reformulate the model so that its parameters have a useful interpretation and it satisfies Assumption MLR.4?

5. Suppose that firm sales in the perfume industry depend on perfume prices and advertising expenditures.

   (a) Propose a constant elasticity model relating sales to prices and advertising expenditures. Interpret the parameters of such model.

   (b) Suppose that the model in part **a)** satisfies MLR.1 to MLR.4. To estimate the model we have a random sample of firms in the perfume industry and we observe sales and prices but not advertising expenditures. We know that if firms expend more on advertising the price of their perfumes goes up, so that advertising expenditures and prices are positively correlated. Let $\widetilde{\beta}_1$ be the OLS estimator of the log-price parameter on the simple linear regression of log-sales on log-prices, would you expect $\widetilde{\beta}_1$ to overestimate or underestimate the causal effect of prices on sales? Justify your answer

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

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

   (a) What are the probable signs of β_1 and β_2? What is the interpretation of those parameters? Explain.

   (b) Why might nox and rooms be negatively correlated? If this is the case, does the simple regression of $\log(price)$ on $\log(nox)$ produce an upward or downward biased estimator of β_1?

(c) Using a sample of 506 neighborhoods in the Boston area (United States) we have obtained the following results:

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

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

Is the relationship between the simple and multiple regression estimates of the elasticity of *price* with respect to *nox* what you would have predicted, given your answer in part **b**?

7. Consider the following model relating house prices in dollars (*price*) with house age (*age*), number of rooms (*rooms*) and number of bathrooms (*baths*) .

$$\log(price) = \beta_0 + \beta_1 rooms + \beta_2 baths + \beta_3 age + u$$

The estimated model is:

$$\widehat{\log(price)} = \underset{(0.160)}{10.358} + \underset{(0.028)}{0.095}\ rooms + \underset{(0.036)}{0.285}baths - \underset{(0.001)}{0.0025}age$$
$$n = 142 \qquad R^2 = 0.70$$

(a) Interpret the estimated coefficient of *rooms*.

(b) Predict $\log(price)$ for a 10 years old house with three rooms and one bathroom.

(c) What is the estimated percentage variation in house price if baths increases in 2 holding the other factors fixed?

(d) Type A houses have 3 rooms and one bathroom and type B houses have 2 rooms and 2 bathrooms and the same age than type A houses. Find the difference in expected log price between these two types of houses. Compute the estimated difference.

(e) If we add to this model the total number of rooms (that is, $rooms + baths$), what would happen? Why?

8. A problem of interest to health officials (and others) is to determine the effects of smoking during pregnancy on infant health. One measure of infant health is birth weight; a birth rate that is too low can put an infant at risk for contracting various illnesses. Since factors other than cigarette smoking that affect birth weight are likely to be correlated with smoking, we should take those factors into account. For example, higher income generally results in access to better prenatal care, as well as better nutrition for the mother. An equation that recognizes this is

$$bwght = \beta_0 + \beta_1 cigs + \beta_2 faminc + u$$

(a) What is the most likely sign for β_2?

(b) Do you think $cigs$ and $faminc$ are likely to be correlated? Explain why the correlation might be positive or negative.

(c) Now estimate the equation with and without $faminc$, using the data in BWGHT from Wooldridge book. Report the results in equation form, including the sample size and R-squared. Discuss your results, focusing on whether adding $faminc$ substantially changes the estimated effect of $cigs$ on $bwght$ (note that $bwght$ is measured in ounces (1 ounce is approximately 28.35 grams), $cigs$ is measured in cigarettes per day and $faminc$ is annual income in thousands of dollars).

9. Use the data in HPRICE1 from Wooldridge book to estimate the model

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

where $price$ is the house price in thousands of dollars, $sqrft$ the house size in square feet (1 square foot is 0.093 m^2) and $bdrms$ is the number of bedrooms.

(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 fixed?

(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 **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 for this house from the OLS regression line.

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

10. The file CEOSAL2 in Wooldridge book contains data on 177 chief executive officers, which can be used to examine the effects of firm performance on CEO salary.

(a) Estimate a model relating annual salary to firm sales and market value ($mktval$). Make the model of the constant elasticity variety for both independent variables. Write the results out in equation form.

(b) Add profits to the model from part **a)**. Why can this variable not be included in logarithmic form? Would you say that these firm performance variables explain most of the variation in CEO salaries?

(c) Add the variable *ceoten* (prior number of years as company CEO) to the model in part **b)**. What is the estimated percentage return for another year of CEO tenure, holding other factors fixed?

(d) Find the sample correlation coefficient between the variables $\log(mktval)$ and $profits$. Are these variables highly correlated? What does this say about the OLS estimators?