TSP Exercise Session - Problem Set 5

Multiple Regression II

Preparations

Please create a new folder for this exercise session with your name in directory T:. Then go to L:\Intermediate Econometrics\PC2 and copy the files into your folder.

1) Effects of smoking during pregnancy on infant health

(Based on Wooldridge, Computer Exercise C 3.1, p.110, omitted variable bias)

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:

\[ bwght = \beta_0 + \beta_1 \cdot cigs + \beta_2 \cdot faminc + u \]

with \( bwght \) birth weight, ounces; \( cigs \) cigs smoked per day while pregnant; \( faminc \) 1988 family income, $ 1000s

(a) What are the most likely signs for \( \beta_1 \) and \( \beta_2 \)?

(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 “pc2_01.xls”. Report the results in equation form, including the sample size and \( R^2 \). Discuss your results, focusing on whether adding \( faminc \) substantially changes the estimated effect of \( cigs \) on \( bwght \).
2) Effects of firm performance on CEO salary

(Based on Wooldridge, Computer Exercise C 3.3, p.111, Multicollinearity)

The data set “pc2_02.xls” contains data on 177 chief executive officers and can be used to examine the effects of firm performance on CEO salary. The data set contains following variables of interest:

- salary: 1990 compensation, $1000s
- sales: 1990 firm sales, millions
- mktval: market value, end 1990, mills.
- ceoten: years as ceo with company

(a) Estimate a model relating annual salary to firm sales and market value. 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 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?

3) Discrimination against certain customer groups by fast food restaurants

(Based on Wooldridge, Computer Exercise C 3.8, p. 112)

Use the data in “pc2_03.xls” 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 larger concentration of blacks. The data set contains the following variables:

- psoda: price of medium soda
- prpbck: proportion black, zipcode
- prppov: proportion in poverty, zipcode
- income: median family income, zipcode
(a) Find the average values of prplck and income in the sample, along with their standard deviations. What are the units of measurement of prplck 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:

$$psoda = \beta_0 + \beta_1 \cdot prplck + \beta_2 \cdot income + u$$

Estimate this model by OLS and report the results in equation form, including the sample size and $R^2$. Interpret the coefficient on prplck. Do you think it is economically large?

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

(d) A model with a constant price elasticity with respect to income may be more appropriate. Report the estimates of the model:

$$\log(psoda) = \beta_0 + \beta_1 \cdot prplck + \beta_2 \cdot \log(income) + u$$

If prplck increases by 0.20 (20 percentage points), what is the estimated change in psoda?

(e) Now add the variable prppov to the regression in part (d). What happens to $\hat{\beta}_{prplck}$?

(f) Find the correlation between log(income) and prppov. Is it roughly what you expected?

(g) Evaluate the following statement: “Because log(income) and prppov are so highly correlated, they have no business being in the same regression.”