FINM3123 Introduction to Econometrics

Chapter 3 Exercises

1. Suppose that the population model determining y is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u,$$

and this model satisfies Assumptions MLR.1 through MLR.4. However, we estimate the model that omits x_3. Let $\tilde{\beta}_0$, $\tilde{\beta}_1$, and $\tilde{\beta}_2$ be the OLS estimators from the regression of y on x_1 and x_2. Show that the expected value of $\tilde{\beta}_1$ (given the values of the independent variables in the sample) is

$$E(\tilde{\beta}_1) = \beta_1 + \beta_3 \frac{\sum_{i=1}^{n} \hat{r}_{i1} x_{i3}}{\sum_{i=1}^{n} \hat{r}_{i1}^2},$$

where the $\hat{r}_{i1}$ are the OLS residuals from the regression of x_1 on x_2. [*Hint:* The formula for $\tilde{\beta}_1$ comes from equation

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2}.$$

Plug $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$ into this equation. After some algebra, take the expectation treating x_{i3} and $\hat{r}_{i1}$ as nonrandom.]

2. The file CEOSAL2.RData (or CEOSAL2.xls) contains data on 177 chief executive officers and can be used to examine the effects of firm performance on CEO salary.

   i) 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.

   ii) Add *profits* to the model from part (i). 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?

   iii) Add the variable *ceoten* to the model in part (ii). What is the estimated percentage return for another year of CEO tenure, holding other factors fixed?

   iv) 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. Use the data in DISCRIM.RData (or DISCRIM.xls) to answerthis question. These are ZIP code-level data on prices for various items at fast-food restaurant, 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 a larger concentration of blacks.

i) Find the average values of *prpblck* and *income* in the sample, along with their standard deviations. What are the units of measurement of *prpblck* and *income*?

ii) Consider a model to explain the price of soda, psoda, in terms of the proportion of the population that is black and median income:

$$psoda = \beta_0 + \beta_1 prpblck + \beta_2 income + u.$$

Estimate this model by OLS and report the results in equation form, including the sample size and R-squared. (Do not use scientific notation when reporting the estimates.) Interpret the coefficient on *prpblck*. Do you think it is economically large?

iii) Compare the estimate from part (ii) with the simple regression estimate from *psoda* on *prpblck*. Is the discrimination effect larger or smaller when you control for income?

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

$$\log(psoda) = \beta_0 + \beta_1 prpblck + \beta_2 \log(income) + u.$$

If *prpblck* increases by .20 (20 percentage points), what is the estimated percentage change in *psoda*?

v) Now add the variable *prppov* to the regression in part (iv). What happens to $\hat{\beta}_{prpblck}$?

vi) Find the correlation between $\log(income)$ and *prppov*. Is it roughly what you expected?

vii) Evaluate the following statement: "Because $\log(income)$ and *prppov* are so highly correlated, they have no business being in the same regression."