FINM3123 Introduction to Econometrics

Chapter 9 Exercises

1. Consider a simple model relating the annual number of crimes on college campuses ($crime$) to student enrollment ($enroll$):

$$\log(crime) = \beta_0 + \beta_1 \log(enroll) + u.$$

   The sample we used was not a random sample of colleges in the United States, because many schools in 1992 did not report campus crimes. Do you think that college failure to report crimes can be viewed as exogenous sample selection? Explain.

2. The point of this exercise is to show that tests for functional form cannot be relied on as a general test for omitted variables. Suppose that, conditional on the explanatory variables x_1 and x_2, a linear model relating y to x_1 and x_2 satisfies the Gauss-Markov assumptions:

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \\ E(u|x_1, x_2) &= 0 \\ Var(u|x_1, x_2) &= \sigma^2. \end{aligned}$$

   To make the question interesting, assume $\beta_2 \neq 0$.
   Suppose further that x_2 has a simple linear relationship with x_1 :

$$\begin{aligned} x_2 &= \delta_0 + \delta_1 x_1 + r \\ E(r|x_1) &= 0 \\ Var(r|x_1) &= \tau^2. \end{aligned}$$

   (a) Show that

$$E(y|x_1) = (\beta_0 + \beta_2\delta_0) + (\beta_1 + \beta_2\delta_1)x_1.$$

   Under ramdom sampling, what is the probability limit of the OLS estimator from the simple regression of y on x_1? Is the simple regression estimator generally consistent for β_1? [$Hint$: If $plim(T_n) = \alpha$ and $plim(U_n) = \beta$, then $plim(T_n + U_n) = \alpha + \beta$ and $plim(T_n U_n) = \alpha\beta$.]

   (b) If you run the regression of y on x_1, x_1^2, what will be the probability limit of the OLS estimator of the coefficient on x_1^2? Explain.

   (c) Using substitution, show that we can write

$$y = (\beta_0 + \beta_2\delta_0) + (\beta_1 + \beta_2\delta_1)x_1 + u + \beta_2 r.$$

   Also show that, if we define $\nu = u + \beta_2 r$ then $E(\nu|x_1) = 0$, $Var(\nu|x_1) = \sigma^2 + \beta_2^2\tau^2$. What consequences does this have for the t statistic on x_1^2 from the regression in part (b)?

(d) What do you conclude about adding a nonlinear function of x_1—in particular, x_1^2—in an attempt to detect omission of x_2?

3. Use the data in TWOYEAR.RData for this exercise.

   i) The variable *stotal* is a standardized test variable, which can act as a proxy variable for unobserved ability. Find the sample mean and standard deviation of *stotal*.

   ii) Run simple regressions of *jc* and *univ* on *stotal*. Are both college education variables statistically related to *stotal*? Explain.

   iii) Add *stotal* to the following equation

   $$\log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u,$$

   and test the hypothesis that the returns to two- and four-year colleges are the same against the alternative that the return to four-year colleges is greater.

   iv) Add $stotal^2$ to the equation estimated in part (iii). Does a quadratic in the test score variable seem necessary?

   v) Add the interaction terms $stotal \cdot jc$ and $stotal \cdot univ$ to the equation from part (iii). Are these terms jointly significant?

   vi) What would be your final model that controls for ability through the use of *stotal*? Justify your answer.

4. Use the data in MURDER.RData only for the year 1993 for this question, although you will need to first obtain the lagged murder rate, say $mrdrte_{-1}$.

   i) Run the regression of *mrdrte* on $exec, unem$. What are the coefficient and t statistic on *exec*? Does this regression provide any evidence for a deterrent effect of capital punishment?

   ii) How many executions are reported for Texas during 1993? (Actually, this is the sum of executions for the current and past two years.) How does this compare with the other states? Add a dummy variable for Texas to the regression in part (i). Is its t statistic unusually large? From this, does it appear Texas is an "outlier"?

   iii) To the regression in part (i) add the lagged murder rate. What happens to $\hat{\beta}_{exec}$ and its statistical significance?

   iv) For the regression in part (iii), does it appear Texas is an outlier? What is the effect on $\hat{\beta}_{exec}$ from dropping Texas from the regression?