13.6 Problem Set 5

Question 1, page 452, 7th Edition

In Example 13.1, assume that the averages of all factors other than educ have remained constant over time and that the average level of education is 12.2 for the 1972 sample and 13.3 in the 1984 sample. Using the estimates in Table 13.1, find the estimated change in average fertility between 1972 and 1984. (Be sure to account for the intercept change and the change in average education.)

Question 2, page 452, 7th Edition

2 Using the data in KIELMC, the following equations were estimated using the years 1978 and 1981:

$$\widehat{\log(price)} = \underset{(.26)}{11.49} - \underset{(.058)}{.547}\, nearinc + \underset{(.080)}{.394}\, y81 \cdot nearinc$$

$$n = 321,\ R^2 = .220$$

and

$$\widehat{\log(price)} = \underset{(.27)}{11.18} + \underset{(.044)}{.563}\, y81 - \underset{(.067)}{.403}\, y81 \cdot nearinc$$

$$n = 321,\ R^2 = .337.$$

Compare the estimates on the interaction term *y81·nearinc* with those from equation (13.9). Why are the estimates so different?

Question 3, page 452, 7th Edition

Why can we not use first differences when we have independent cross sections in two years (as opposed to panel data)?

Question 4, page 453, 7th Edition

If we think that β_1 is positive in (13.14) and that Δu_i and $\Delta unem_i$ are negatively correlated, what is the bias in the OLS estimator of β_1 in the first-differenced equation?

Question 6, page 453, 7th Edition

In 1985, neither Florida nor Georgia had laws banning open alcohol containers in vehicle passenger compartments. By 1990, Florida had passed such a law, but Georgia had not.

(i) Suppose you can collect random samples of the driving-age population in both states, for 1985 and 1990. Let *arrest* be a binary variable equal to unity if a person was arrested for drunk driving during the year. Without controlling for any other factors, write down a linear probability model that allows you to test whether the open container law reduced the probability of being arrested for drunk driving. Which coefficient in your model measures the effect of the law?

(ii) Why might you want to control for other factors in the model? What might some of these factors be?

(iii) Now, suppose that you can only collect data for 1985 and 1990 at the county level for the two states. The dependent variable would be the fraction of licensed drivers arrested for drunk driving during the year. How does this data structure differ from the individual-level data described in part (i)? What econometric method would you use?

Question C2, page 454, 7th Edition

C2 Use the data in CPS78_85 for this exercise.

(i) How do you interpret the coefficient on *y85* in equation (13.2)? Does it have an interesting interpretation? (Be careful here; you must account for the interaction terms *y85·educ* and *y85·female*.)

(ii) Holding other factors fixed, what is the estimated percent increase in nominal wage for a male with 12 years of education? Propose a regression to obtain a confidence interval for this estimate. [*Hint:* To get the confidence interval, replace *y85·educ* with *y85*·(*educ* − 12); refer to Example 6.3.]

(iii) Reestimate equation (13.2) but let all wages be measured in 1978 dollars. In particular, define the real wage as *rwage* – *wage* for 1978 and as *rwage* = *wage*/1.65 for 1985. Now, use log(*rwage*) in place of log(*wage*) in estimating (13.2). Which coefficients differ from those in equation (13.2)?

(iv) Explain why the *R*-squared from your regression in part (iii) is not the same as in equation (13.2). (*Hint:* The residuals, and therefore the sum of squared residuals, from the two regressions *are* identical.)

(v) Describe how union participation changed from 1978 to 1985.

(vi) Starting with equation (13.2), test whether the union wage differential changed over time. (This should be a simple *t* test.)

(vii) Do your findings in parts (v) and (vi) conflict? Explain.

Question C7, page 455, 7th Edition

C7 Use GPA3 for this exercise. The data set is for 366 student-athletes from a large university for fall and spring semesters. [A similar analysis is in Maloney and McCormick (1993), but here we use a true panel data set.] Because you have two terms of data for each student, an unobserved effects model is appropriate. The primary question of interest is this: Do athletes perform more poorly in school during the semester their sport is in season?

(i) Use pooled OLS to estimate a model with term GPA (*trmgpa*) as the dependent variable. The explanatory variables are *spring*, *sat*, *hsperc*, *female*, *black*, *white*, *frstsem*, *tothrs*, *crsgpa*, and *season*. Interpret the coefficient on *season*. Is it statistically significant?

(ii) Most of the athletes who play their sport only in the fall are football players. Suppose the ability levels of football players differ systematically from those of other athletes. If ability is not adequately captured by SAT score and high school percentile, explain why the pooled OLS estimators will be biased.

(iii) Now, use the data differenced across the two terms. Which variables drop out? Now, test for an in-season effect.

(iv) Can you think of one or more potentially important, time-varying variables that have been omitted from the analysis?

Question C16, page 458, 7th Edition

C16 Use the data in COUNTYMURDERS for this exercise. The data set covers murders and executions (capital punishment) for 2,197 counties in the United States.

(i) Find the average value of *murdrate* across all counties and years. What is the standard deviation? For what percentage of the sample is *murdrate* equal to zero?

(ii) How many observations have *execs* equal to zero? What is the maximum value of *execs*? Why is the average of *execs* so small?

(iii) Consider the model

$$murdrate_{it} = \theta_t + \beta_1 execs_{it} + \beta_2 execs_{i,t-1} + \beta_3 percblack_{it} + \beta_4 percmale_i + \beta_5 perc1019 + \beta_6 perc2029 + a_i + u_{it},$$

where θ_t represents a different intercept for each time period, a_i is the county fixed effect, and u_{it} is the idiosyncratic error. What do we need to assume about a_i and the execution variables in order for pooled OLS to consistently estimate the parameters, in particular, β_1 and β_2?

(iv) Apply OLS to the equation from part (ii) and report the estimates of β_1 and β_2, along with the usual pooled OLS standard errors. Do you estimate that executions have a deterrent effect on murders? What do you think is happening?

(v) Even if the pooled OLS estimators are consistent, do you trust the standard errors obtained from part (iv)? Explain.

(vi) Now estimate the equation in part (iii) using first differencing to remove a_i. What are the new estimates of β_1 and β_2? Are they very different from the estimates from part (iv)?

(vii) Using the estimates from part (vi), can you say there is evidence of a statistically significant deterrent effect of capital punishment on the murder rate? If possible, in addition to the usual OLS standard errors, use those that are robust to any kind of serial correlation or heteroskedasticity in the FD errors.