ECON 6511: Advanced Applied Econometrics
Homework 4
Due in class February 7, 2017

1. (Wooldridge, Chapter 14, Problem 5) Use the data in RENTAL.dta for this exercise. The data for the years 1980 and 1990 include rental prices and other variables for college towns. The idea is to see whether a stronger presence of students affects rental rates. The unobserved effects model is:

$$\log(rent_{it}) = \beta_0 + \delta_0 y90_t + \beta_1 \log(pop_{it}) + \beta_2 \log(avginc_{it}) + \beta_3 pctstu_{it} + a_i + u_{it}$$

where *pop* is city population, *avginc* is average income, and *pctstu* is student population as a percentage of city population (during the school year).

(a) Estimate the equation by pooled OLS and report the results in equation form. What do you make of the estimate on the 1990 dummy variable? What do you get for $\hat{\beta}_{pctstu}$?

(b) Are the standard errors you report in part (a) valid? Explain.

(c) Now, difference the equation and estimate by OLS. Compare your estimate of β_{pctstu} with that from part (b). Does the relative size of the student population appear to affect rental prices?

(d) Estimate the model by fixed effects to verify that you get identical estimates and standard errors to those in part (c).

2. (Wooldridge, Chapter 14, Problem 4) Papke (1994) studied the effect of the Indiana enterprise zone (EZ) program on unemployment claims. The author uses a model that allows each city to have its own time trend:

$$\log(uclms_{it}) = a_i + c_i t + \beta_1 ez_{it} + u_{it},$$

where a_i and c_i are both unobserved effects. This allows for more heterogeneity across cities.

(a) Show that, when the previous equation is first differenced, we obtain

$$\Delta \log(uclms_{it}) = c_i + \beta_1 \Delta ez_{it} + \Delta u_{it}, \quad t = 2, \ldots, T.$$

Notice that the differenced equation contains a fixed effect, c_i.

(b) Estimate the differenced equation by fixed effects using the data in EZUNEM.DTA. What is the estimate of β_1? Is it very different from the estimate in textbook Example 13.8 of -0.182? Is the effect of enterprise zones still statistically significant?

(c) Add a full set of year dummies to the estimation in part (b). What happens to the estimate of β_1?

3. (Wooldridge, Chapter 14, Problem 7) Use the state-level data on murder rates and executions in MURDER.dta for the following exercise

   (a) Consider the unobserved model

   $$mrdrte_{it} = \eta_t + \beta_1 exec_{it} + \beta_2 unem_{it} + a_i + u_{it},$$

   where η_t simply denotes different year intercepts and a_i is the unobserved state effects. If past executions of convicted murderers have a deterrent effect, what should be the sign of β_1? What sign do you think β_2 should have? Explain.

   (b) Using just the years 1990 and 1993, estimate the equation from part (a) by pooled OLS. Ignore the serial correlation problem in the composite errors. Do you find any evidence for a deterrent effect?

   (c) Now, using 1990 and 1993, estimate the equation by fixed effects. You may use first differencing since you are only using two years of data. Is there evidence of a deterrent effect? How strong?

   (d) Compute the heteroskedasticity-robust standard error for the estimation in part (c).

   (e) Find the state that has the largest number for the execution variable in 1993. (The variable *exec* is total executions in 1991, 1992, and 1993.) How much bigger is this value than the next highest value?

   (f) Estimate the equation using first differencing, dropping Texas from the analysis. Compute the usual and heteroskedasticity-robust standard errors. Now, what do you find? What is going on?

   (g) Use all three years of data and estimate the model by fixed effects. Include Texas in the analysis. Discuss the size and statistical significance of the deterrent effect compared with only using 1990 and 1993.

4. (Wooldridge, Chapter 14, Problem 14) Use the data set in AIRFARE.dta to answer this question. The estimates can be compared with those at the end of the last lecture.

   (a) Compute the average of the variable *concen* for each route; call these *concenbar*. How many different time averages can there be? Report the smallest and the largest. Note: You can generate a variable containing the average value of *concen* for each route using the Stata command: "egen concenbar = mean(concen), by(id)"

(b) Estimate the equation:

$$\begin{aligned} lfare_{it} = &\beta_0 + \delta_1 y98_t + \delta_2 y99_t + \delta_3 y00_t + \beta_1 concen_{it} + \beta_2 ldist_t + \beta_3 ldistsq_i \\ &+ \gamma_1 concenbar_i + a_i + u_{it} \end{aligned}$$

by random effects. Verify that $\hat{\beta}_1$ is identical to the FE estimate computed in class.

(c) Using the equation from part (b) and the usual RE standard error, test $H_0 : \gamma_1 = 0$ against the two-sided alternative. Report the p-value. What do you conclude about RE versus FE for estimating β_1 in this application?

(d) Obtain a t-statistic (and, therefore, p-value) that is robust to arbitrary serial correlation and heteroskedasticity using ", cluster(id)". Does this change the conclusion reached in part (c)?