Problem Set: Panel Data

Problem 1. *Suppose that the idiosyncratic errors in*

$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \cdots + \beta_k x_{itk} + f_i + u_{it}, \quad t = 1, 2, \ldots, T$$

are serially uncorrelated with constant variance, σ_u^2*. Show that the correlation between adjacent differences,* Δu_{it} *and* $\Delta u_{i,t+1}$*, is* -0.5*. Therefore, under the ideal FE assumptions, first differencing induces negative serial correlation of a known value.*

Problem 2. *Download RENTAL (description). The data on rental prices and other variables for college towns are for the years 1980 and 1990. The idea is to see whether a stronger presence of students affects rental rates. The unobserved effects model is*

$$\log(rent_{it}) = \beta_0 + \delta_0 y90 + \beta_1 \log(pop_{it}) + \beta_2 \log(avginc_{it}) + \beta_3 pctstu_{it} + f_i + u_{it},$$

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

1. *Estimate the equation by pooled OLS and report the results in standard form. What do you make of the estimate on the 1990 dummy variable? What do you get for* $\hat{\beta}_{pctstu}$*?*
2. *Are the standard errors you report in part (i) valid? Explain.*
3. *Now, difference the equation and estimate by OLS. Compare your estimate of* β_{pctstu} *with that from part (i). Does the relative size of the student population appear to affect rental prices?*
4. *Estimate the model by fixed effects to verify that you get identical estimates and standard errors to those in part (iii).*

Problem 3. *Papke (1994) studied the effect of the Indiana enterprise zone (EZ) program on unemployment claims. She analyzed 22 cities in Indiana over the period from 1980 to 1988. Six enterprise zones were designated in 1984, and four more were assigned in 1985. Twelve of the cities in the sample did not receive an enterprise zone over this period; they served as the control group. Download EZUNEM (description).*

A simple policy evaluation model is

$$\log(uclms_{it}) = \mu_t + \beta_1 ez_{it} + f_i + u_{it},$$

where $\texttt{uclms}_{it}$ *is the number of unemployment claims filed during year* t *in city* i*. The parameter* μ_t *just denotes a different intercept for each time period. Generally, unemployment claims were falling statewide over this period, and this should be reflected in the different year intercepts. The binary variable* $\texttt{ez}_{it}$ *is equal to one if city* i *at time* t *was an enterprise zone; we are interested in* β_1*. The unobserved effect* f_i *represents fixed factors that affect the economic climate in city* i*. Because enterprise zone designation was not determined randomly—enterprise zones are usually economically depressed areas—it is likely that* $\texttt{ez}_{it}$ *and* f_i *are positively correlated (high* f_i *means higher unemployment claims, which lead to a higher chance of being given an EZ).*

1. *Estimate* β_1 *using first-difference method and interpret the coefficient estimate.*

2. *Let's allow each city to have its own time trend:*

$$\log(uclms_{it}) = f_i + c_i t + \beta_1 ez_{it} + u_{it}, \tag{1}$$

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

3. *Estimate the differenced equation by fixed effects. What is the estimate of* β_1*? Is it very different from the estimate obtained in part 1? Is the effect of enterprise zones still statistically significant?*

4. *Add a full set of year dummies to the estimation in part 3. What happens to the estimate of* β_1*?*

   If we add year dummies, γ_t*, to the differenced equation, it becomes:*

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