Assignment V

1. Use the data in KIELMC for this exercise.

   (i) The variable *dist* is the distance from each home to the incinerator site, in feet. Consider the model

   $$\log(price) = \beta_0 + \delta_0 y81 + \beta_1 \log(dist) + \delta_1 y81 \cdot \log(dist) + u.$$

   If building the incinerator reduces the value of homes closer to the site, what is the sign of δ_1? What does it mean if $\beta_1 > 0$?

   (ii) Estimate the model from (i) and report the results. Interpret the coefficient on $y81 \cdot \log(dist)$. What do you conclude?

   (iii) Add *age, age2, rooms, baths*, $\log(intst)$, $\log(land)$, and $\log(area)$ to the equation. Estimate the model and report the results. Now, what do you conclude about the effect of the incinerator on housing values? Now, what do you conclude about the effect of the incinerator on housing values?

   (iv) Why is the coefficient on $\log(dist)$ positive and statistically significant in part (ii) but not in part (iii)? What does this say about the controls used in part (iii)?

2. We use JTRAIN for this exercise. Unlike what we talked about in class, we are going to add the data in year 1989 to study the effect of the job training grant on hours of job training per employee. The basic model for the three years is

   $$hrsemp_{it} = \beta_0 + \delta_1 d88_t + \delta_2 d89_t + \beta_1 grant_{it} + \beta_2 grant_{i,t-1} + \beta_3 \log(employ_{it}) + a_i + u_{it}.$$

   (i) Estimate the equation using first differencing (FD). Write down the model you use for estimation and report the results.

   (ii) Interpret the coefficient on *grant* and comment on its significance.

   (iii) Is it surprising that $grant_{-1}$ is insignificant? Explain.

   (iv) Do larger firms train their employees more or less, on average? How big are the differences in training?

   (v) Reestimate the model using fixed-effects (FE). Compare the estimation results with FD. Are there any major differences for (ii)-(iv) based on FE estimation?

3. Use the data in RENTAL for this exercise. 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

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

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

(i) 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 $\hat{\beta}_3$?

(ii) Are the standard errors you report in part (i) valid? Explain.

(iii) Now, difference the equation and estimate by OLS. Compare your estimate of $\hat{\beta}_3$ with that from part (i). Does the relative size of the student population appear to affect rental prices?

(iv) Obtain the heteroskedasticity-robust standard errors for the first-differenced equation in part (iii). Does this change your conclusions?

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

4. (*optional*) Consider the dynamic panel data model

$$y_{i,t} = c_i + \rho y_{i,t-1} + u_{i,t}, \quad i = 1, 2, \cdots, N, \;\; t = 1, 2, \cdots, T, \tag{1}$$

where $|\rho| < 1$, c_i is the fixed effect parameter capturing individual heterogeneity and $u_{i,t}$ is the idiosyncratic error term satisfying weakly exogeneity assumption and is *i.i.d.* across i.

(i) Consider the fixed effects (FE) estimator:

$$\hat{\rho}_{NT} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t} - \overline{y}_i\right)\left(y_{i,t-1} - \overline{y}_i\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t-1} - \overline{y}_i\right)^2},$$

where $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{i,t}$. Is it a consistent estimator of ρ? Please explain.

(ii) Consider the model (1) in first difference:

$$\Delta y_{i,t} = \rho \Delta y_{i,t-1} + \Delta u_{i,t}, \quad i = 1, 2, \cdots, N. \;\; t = 1, 2, \cdots, T. \tag{2}$$

Can we use the pooled OLS on (2) to obtain a consistent estimator of ρ? Please explain.