Problem Set 1

EE426 Econometrics 2

Due Febuary 11, 2014

Please report the regression results in each problem and print STATA .do file attached at the end of your answer.

1. Use the data KIELMC.dta [Note: every time you start using the data, please be sure to see what the data look like, i.e., describe, browse, tabulate in STATA.]

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

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

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 $\beta_1 > 0$?

(1.2) Estimate the model from (1.1) and report the results. Interpret the coefficient on $y81 \cdot log(dist)$. What do you conclude?

(1.3) Add age,age^2,$rooms$, $baths$, $log(intst)$, $log(land)$, and $log(area)$ to the equation. Report the result. Now, what do you conclude about about the effect of the incinerator on housing values?

2. Use the data WAGEPAN.dta

(2.1) Use pooled OLS to estimate a *log(wage)* equation using explanatory variales *educ, black, hisp, exper, married, union*, and a full set of year dummies (using 1980 as the base year). Interpret and disscuss the coefficients on the *married* and *union* variable.

(2.2) Now do the first difference. Write the equation for the first difference. Which variables are dropped out of the estimation and why? Be sure to exclude diffrences for the first year, 1980, as there is no earlier year.

[STATA Hint for data: sort nr year, by nr: gen l_lwage=lwage[_n-1], gen d_lwage = lwage-l_lwage]

(2.3) Run the regression $\triangle lwage_{it}$ on $\triangle married_{it}$,$\triangle union_{it}$, $d82_t$, ..., $d87_t$, $t = 2, ..., T$; $i = 1, ..., N$, being sure to include a constant.

(2.4) Compare the estimated marriage and union premiums from the levels and first difference estimations (both 3.2 and 3.3), and comment.

.

3. Again, use the data WAGEPAN.dta.

(3.1) Estimate the model

$$lwage_{it} = \beta_0 + \beta_1 educ_i + \beta_2 black_i + \beta_3 hisp_i + v_{it}$$

by OLS, and report the estimates and standard errors in the usual form.

(3.2) Estimate the model in (3.1) by random effects (thinking that $v_{it} = a_i + u_{it}$). How do the RE and pooled OLS estimates of the β_j compare?

(3.3) Are the RE and pooled OLS standard errors the same? Which ones are more reliable, and why?

(3.4) Add a full set of year dummies to the estimations in (3.1) and (3.2). Do any of your conclusions from (3.1) and (3.2) change?

(3.5) Now estimate the model from (3.4) by FE, recognizing that all explanatory variales but the year dummies drop out. How do the FE coefficients on the year dummies compare with the RE estimates?

(3.6) Can you draw some general conclusions from the particular example?

4. Suppose that the idiosyncratic errors in $y_{it} = \beta_1 x_{1it} + ... + \beta_k x_{kit} + a_i + u_{it}$, $\{u_{it} : t = 1, 2, ..., T\}$ are serially uncorrelated with constant variance, σ_u^2. Show that the correlation between adjacent differences, $\triangle u_{it}$ and $\triangle u_{i,t+1}$, is -0.5. Therefore, under the ideal FE assumptions ($u_{it} \sim iid(0, \sigma_u^2)$), first differencing induces negative serial correlation of a known value, and that's FE is more efficient.

5. In a random effects model, define the composite error $v_{it} = a_i + u_{it}$, where a_i is uncorrelated with u_{it} and the u_{it} have constant variance σ_u^2 and are serially uncorrelated. Define $e_{it} = v_{it} - \lambda \overline{v}_i$, where $\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)]^{1/2}$.

(5.1) Show that $E(e_{it}) = 0$.

(5.2) Show that $Var(e_{it}) = \sigma_u^2$, t = 1, ..., T.

(5.3) Show that for $t \neq s$, $Cov(e_{it}, e_{is})$= 0.