Homework #2

1. Use the data in MURDER.RAW to look at the deterrent effect of past executions on murder rate. An unobserved effect model explaining current murder rates in terms of the number of executions in the last three years is

$$mrdrte_t = \alpha_t + \beta_1 exec_{it} + \beta_2 unemp_{it} + a_i + u_{it}$$

(i) Explain why unobserved effects model is appropriate to look at the effect of interest. Discuss expected signs of coefficients.

(ii) Using only 1990 and 1993 years estimate this equation by pooled OLS and FE. Compute heteroskedasticity-robust standard errors. Report, interpret and compare your results.

(iii) Redo (i) excluding the state with largest number of execution variable in 1993 from your analysis. Does this change any key results? Can you explain why?

(iv) Now, estimate the model by FD using all three years of data. Compute the heteroskedasticity robust standard errors. Compare results with (ii).

(v) Estimate the model by FE again using three years of data. Do you find any important differences from the FD estimate?

(vi) Under what circumstances would $exec_{it}$ not be strictly exogenous conditional on a_i?

2. Use the data in WAGEPAN.RAW to estimate the following wage equation for men

$$\ln(wage_{it}) = \alpha_t + \beta_1 educ_i + \beta_2 black_i + \beta_3 hispan_i + \beta_4 \exp er_{it} + \beta_5 \exp er^2{}_{it} + \\ + \beta_6 married_{it} + \beta_7 union_{it} + a_i + u_{it}$$

(i) Estimate this equation by pooled OLS, RE and FE. Report and comment on differences in results. Are usual OLS standard errors reliable even if a_i is uncorrelated with all explanatory variables? Why is $exper_{it}$ redundant in the fixed effect model even though it changes over time?

(ii) Include eight of the occupation dummies into the wage equation and estimate it using FE. Does this change the estimated union wage premium? Why?

(iii) Now add the interaction terms *d81educ, d82educ,…,d87educ* and estimate the equation by FE. What do you conclude about the change in return to education over time?

3. Use the data in CARD.RAW to estimate the return to education.

(i) Estimate a *log(wage)* equation by OLS with *educ, exper,* $exper^2$, *black, south, smsa, smsa66,* and *reg661* through *reg668* as explanatory variables. Next, estimate a reduced form equation for *educ* containing all explanatory variables from the OLS equation and *nearc4*. Do *educ* and *nearc4* have statistically significant partial correlation?

(ii) Estimate the *log(wage)* equation by IV using *nearc4* as IV for *educ.* Compare the results with OLS regression from the part (i).

(ii) Now, IQ score is available for a subset of the men in the sample. Regress *IQ* on *nearc4*. Is IQ score correlated with *nearc4*?

(iii) Regress *IQ* on *nearc4* along with *smsa66,* and *reg661* through *reg668*. Do you find statistically significant partial correlation between IQ score and living near a four-year college in this case?

(iv) Conclude about the importance of controlling for the 1996 location and regional dummies in the wage equation when using *nearc4* as IV for *educ*.

4. Consider a standard unobserved effect panel data model

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}$$

(i) List the methods for estimating this model. Write down the main assumptions and briefly describe estimation procedure for each method.

(ii) Assuming that T=2, show that the FE and the FD estimates are numerically identical.

(iii) Wooldridge 14.3.

(iv) Describe the Hausman test procedure for the FE versus the RE. Explicitly write down the test hypothesis.

5. A simple regression model is

$$y = \beta_0 + \beta_1 x + u$$
$$\mathrm{cov}(x, u) \neq 0$$

(i) State assumptions that instrumental variable for x must satisfy to obtain consistent estimates of β_0 and β_1. Can you test these assumptions?

(ii) Assuming that the IV assumptions are satisfied for some variable z, derive the formula for instrumental variable estimator of β_1.

(iii) Now, let z be a zero-one dummy instrumental variable for x. Show that the instrumental variable estimator of β_1 is $\hat{\beta}_1 = (\bar{y}_1 - \bar{y}_0)/(\bar{x}_1 - \bar{x}_0)$, where $\bar{y}_1$ and $\bar{x}_1$ are sample averages over a sub-sample with z=1; $\bar{y}_0$ and $\bar{x}_0$ are sample averages for the part of sample with z=0.