Problem Set 2
Econometrics 718
Chris Taber
Due: Thurs. Oct . 20

Problem 1. Consider a model in which

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \mu_i + \xi_{it}$$

Assume further that you have a variable X_{it} which varies within i (across t).

You can pick any data set and regression model you like, but I would like you to run the fixed effect model a number of different ways

a) Use xtreg, fe command in stata (using straight standard errors, and also clustering by person)

b) Do the fixed effects regression, i.e. regress $(Y_{it}-\bar{Y}_i)$ on $(X_{it}-\bar{X}_i)$. You can construct the $\bar{Y}_i$ variable by using the egen command with by. Get the standard errors two ways-the standard way and clustering by person.

c) Also run the regression with individual dummy variables for each person (construct the standard errors the standard way and also clustering by person). The xi command may be useful here.

How do all of these results compare? What happens if you only use two periods?

Problem 2 Now take the data set jtrain1 from http://www.stata.com/texts/eacsap/. This has data on firms and the amount of job training they get.

a) Only use the data from 1987 and 1988. Construct the difference in differences estimator in two different ways:

i) Construct the 4 means (control,treatment×before,after)

ii) Run the regression

$$hrsemp_{it} = \beta_0 + \beta_1 grant_{it} + \beta_2 1(year = 1988) + \beta_3 E_i + u_{it}$$

where E_i is a dummy variable for being a treatment (i.e. someone who would receive the grant in 1988).

iii) Run the fixed effect regression:

$$hrsemp_{it} = \theta_i + \beta_1 grant_{it} + \beta_2 1(year = 1988) + u_{it}$$

Do you get exactly the same answer, why or why not?

b) Now (using all the data rather than 2 years) include a firm specific time trend in the model in two different ways:

i) use the xi command (something like xi: reg y x i.fcode*year)

ii) For each firm, run a regression of x and y on an intercept and a time trend, take the residuals and run them on eachother (not sure the cleanest way to do this, but you could again use egen with by)

Problem 3. Now use the data set regm.raw.gz that you can get from the computer software part of my website.

You can read it into stata using the command: infile coll merit male black asian year state chst using regm.raw

Now run the difference in difference model 4 different ways

a) Standard regression using all data (construct standard errors 3 ways, robust, cluster by state year, cluster by state)

b) Standard regression using all data but weighted so that all states get the same weight

c) Now take the mean of all variable by $state \times year$ and run the diff in diff regression (robust se, and clustering by states)

d) Do the same as in c, but weight by state so it looks like the population

How does this all compare?

e) Now use this data to do an event study looking at the effects of the merit aid program both before and after it was implemented. You will need to figure out when the merit aid program was passed for each state that had one. There are multiple ways you can implement this-you can choose yourself.