University of Essex
2015-16

Department of Government
TA: Stavros Poupakis

GV903: Advanced Research Methods

Class 11 Panel Data

Load the data `psid_baltagi` and use `des` and `list` to see what kind of data you have. Can you recognize whether this is panel data?

1. Define the panel structure of the dataset using the command `xtset id t`. Then use `xtdes` and `xtsum` to have a first look at the data.

2. Create a scatter plot of the logarithm of wage and years of full-time work experience. Also create a scatter plot of the save variables, but **only** for the first 4 individuals in the dataset. Do a third scatter plot, again for the first 4, but using different color for each individual. Is the legend helpful? Take it off.

3. Using the command `xtline` plot the logarithm of wage for the first 4 individuals over time. See what will happen if you specify the option `overlay`.

4. Run an OLS regression of the logarithm of wage on years of full-time work experience, gender, whether wage was set through a union contract and years of education. Add also a **quadratic term** for full-time work experience. You will need to generate first a new variable.

5. Fit the same model, but use clustered standard error with the option `vce()`. You should find out what to put in the brackets.

6. Fit a Fixed Effects and a Random Effects model. This can be done using the `xtreg` command with the options `fe` and `re`, respectively.

7. Store the estimates of each of the two regressions to use them in the Hausman test. To do this type something like `estimates store choose_name` after each of **both** regressions. Then perform the Hausman test by

   ```
   hausman one_name_here other_name_here , sigmamore
   ```

   Note that in the syntax first needs to be the 'always consistent' estimator and then the efficient which is not consistent under the alternative. **What the test concludes? Which is assumption is required by both that if is not true both estimators are invalid?**

8. Going a step further you can model panel data with more flexible models using the `xtmixed` command. Another rich source of estimation solutions for panel data, and in general multilevel model is the package `gllamm`. We will discuss in class about them.