PANEL DATA

File inputting

It might sometimes occur that STATA will not open a dataset despite its appropriate extension and construction. Dealing with bigger datasets, we need to change STATA default settings. First we need to approach `matsize` by using the command `set matsize X` (where `X<800`). We then move to `memory size` by the command `set memory X` (where `X` can be defined either in bites directly or in megabites if you use `Xm` instead. You are now ready to open the file using STATA, which should be done using Open function from the File menu.

These settings will be lost if you close STATA - reopening it requires typing the commands once again.

Defining panels

Even if you have panel data, you do not need to perform panel analysis. Using the standard command `regress` or `reg` you will obtain the pooled regression results.

If you decide to do the panel analysis (sometimes a smart choice), STATA needs to be informed about the structure of the panel. We should thus define the time dimension and the grouping variable. Command `tis variable_name` defines the time dimension. Obviously, sometimes we do not have time in our dataset and then you are free to choose any variable you like to be your 'time' dimension. Command `iis variable_name` serves the purpose of defining a grouping variable. In most of our research, it is either a country or a pair of countries or a household.

It may sometimes occur that due to inappropriate construction of some records, STATA will not accept our preferred variable to be the grouping one. Unfortunately, in most cases, we are forced to edit the dataset and eliminate the inconsistencies. Gladly, sometimes actual data may serve the purpose of grouping variable. This was the case for the output equation dataset, where instead of incorrectly specified country and country code area variable could be used (a variable pretty characteristic for a country and rarely changing with time).

Specifying your `tis` and `iis` variables one should also have the question in mind. Sometimes we want to compare cross-section of countries in every moment in time. An example here would be for example the gravity equations, where we want to find a 'true' mechanism of determining international trade. In this case, standard approach of `tis` for time and `iis` for an item definitely makes sense. Sometimes, however, we might be interested in the differences in the processes across countries. For instance, if we wanted to trace the growth patterns and we use many countries to augment the statistical quality of the results, one might try the opposite.

Regression

After letting STATA know the panel structure, we can let it do its job. This is done by typing the command:

```
xtreg dep_var expl_var_1 expl_var_2 ....
```

By default STATA estimates random effect panel regression. If you want to obtain random effects, you need to type `,fe` in the end of the regression command.

To verify whether one should use FE or RE on a particular dataset, Hausman test should be used (`xthausman`). Alternatively, one might also use the Breusch-Pagan LM test (`xttest0`). Their interpretation is similar.

For random effect models, you might want to observe the effects. STATA is well programmed to generate and report both year and grouping dummies, which is done automatically using the `xi` command. Instead of `xtreg` use the following:

```
xi:xtreg dep_var expl_var_1 expl_var_2 ... i.chosen_dummy, re.
```

You are free to choose any of the previously defined `tis` or `iis` variables in this command. STATA will automatically generate the dummies, eliminate the excessive one and report their significance in the regression output. Note that in the output you will usually get both between and within R^2 statistic and pay attention to their interpretation.

Diagnostics

Panel data or pooled regression, dataset might be troubled by autocorrelation or heteroscedasticity. While testing for the latter is rater easy (type `xttest` and that's it), autocorrelation is problematic in panel datasets. For that reason, you are advised to use GLS regression right away, to eliminate the problem of $p-values$ inapplicability.

This is obtained by the command `xtgls` instead of `xtreg`. All other commands apply in the same way.

You can also use `PCSE` (panel-corrected standard errors), which are computed basing on Prais-Winsten transformation. This is obtained by typing `xtpcse` in place of `xtgls`.