

EC501 Econometric Methods and Applications

Lab Lecture

In this session, we will have a gentle introduction to Stata, using the data from Problem set 3.

1. Download the file `Problem_Set_03_Data.xls`, open it in Excel, and save it in text (.txt) format.
2. Locate the link to Stata and open it. If you cannot find the link. Right click on the Desktop, choose New and then Shortcut, write `\\sernt27\stata11$\Stata.exe` in the box, and then choose Next and Finish.
3. Take a moment to look at the way Stata organizes the screen. You should see 4 windows:
 - (a) The top-left window keeps an history of the commands you (will) type;
 - (b) The bottom-left window will show the variables you are using;
 - (c) The long window in the bottom to the right is where you can type commands;
 - (d) The larger window (top-right) is where you will see the commands that are given, and the corresponding outputs.
 - (e) Also, in the top-right you will see the drop-down menus and some useful icons.
4. In Stata, go to File, Import, and choose the first option. Then select the text file you have just created and click OK. You have now imported the data into Stata! You can see a description of the variables in the Variables window.
5. Go again to File, Save as, and save the file in Stata format (.dta) with an appropriate name, for example `Problem_Set_03_Data.dat`.
6. You can see the data by going to Data, Data Editor, and then choosing Edit (if you want to make changes to the data) or Browse (if you just want to see it).
7. From now on, the data is ready to be used in Stata and the next time you want to use the data just double click on this file.
8. Notice that when you imported and saved the file using the drop-down menus, Stata printed the corresponding commands in the main window.
9. You can now use Stata in essentially three different ways:
 - (a) You can continue to use drop-down menus (this is useful if you do not know the commands, if you want to use the help facility, or if you want to see the options available);

- (b) You can type the commands in the command window (note that Stata is case sensitive!);
 - (c) You can write a file (.do) with the commands and execute them (this is the best option when you are familiar with the commands).
10. We will use a mixture of the three modes today so that you can taste all of Stata's flavours.
 11. One thing that is useful to do is to keep a record (a log) of what is done. For that, go to File, Log, Begin, and choose an appropriate name. Note that in many ways it is better to use the .log format rather than the default format.
 12. Before going any further, it is useful to compute some summary statistics. You can do that using the drop-down menus (Statistics; Summary Tables and tests; Summary and descriptive statistics; Summary statistics), or you can just give the command `summarize y x3 x2`, which can be abbreviated as `su y x3 x2`.
 13. If you want more information, you can use the option `detail`, abbreviated as `d`. Options are typed after a comma at the end of the command. For example, `su y x3 x2, d`.
 14. Looking at the data is also a very good practice. To do a scatter plot, say of y versus x_3 , you can use the command `twoway (scatter y x3)`. Of course, you can also do this with the drop-down menu.
 15. To obtain the corresponding picture for y and x_2 , just click on the last command in the window with the command history and Stata will copy it to the command window. You can just edit the command (by replacing `x3` with `x2`) and press the Enter key. So, you have now learned a quick way of recycling your old commands. This is very useful if you make a mistake or if you need to issue many similar commands.
 16. In the problem set, you are asked to use OLS to estimate the model

$$\log y_i = \beta_1 + \beta_2 \log x_{i2} + \beta_3 \log x_{i3} + \epsilon_i,$$

using the sample period 1960–1994. If you inspect the data, you will see that we have data from 1959. So, we need to restrict our sample. We can do this in several ways:

- (a) we can just restrict the sample when we actually estimate the model by adding `if year>1959` after the command;
 - (b) or we can delete the observation for the year 1959 from the data. You can do that with the command `drop if year==1959`. Notice the double equal sign; this is not a typo! Alternatively, you could have used the command `drop if year<1960`;
 - (c) rather than using the `drop` command, you can use its inverse: `keep`. In this case, you could use, for example, `keep if year>=1960`.
17. You can also use `drop` and `keep` to delete variables from your dataset.

18. To be able to estimate the model we need to create the logs of the variables. Notice that in econometrics we only use natural logs so `log` and `ln` are generally synonyms!
19. To transform, or create, variables, the main command is `generate`, abbreviated as `gen`. Alternatively, you can use the drop-down menu: Data; Create or change variable; Create new variable.
20. If the variable already exists and you want to redefine its content, you can do it with the command `replace`.
21. One way or another, create the new variables `ly`, `lx2` and `lx3` as the logs of `y`, `x2` and `x3`. This can be done, for example, with commands of the form `gen ly=ln(y)`. Notice the single equal sign. The double equal sign is only used in logical conditions of the type do something if something else is equal (double equal) to something.
22. At this point, it is important to know that Stata stores variables in different formats. For continuous variables, you should use double precision as much as possible. To do it, you need to specify after the `gen` command that you want to create a variable in double precision. You can use, for example, `gen double ly=ln(y)`.
23. You are now just one step away from estimating your (first?) regression with stata. To do it, just give the command `reg ly lx2 lx3` and savour the moment!
24. Obviously, you can do the same with the drop-down menus. Go to Statistics, Linear models and related; Linear models, and fill in the right boxes.
25. Notice that by default Stata includes a constant in the model. If you do not want the constant you have to say it, either by ticking the right box if you are using the menus, or by using the option `noconstant`.
26. Also note that for Stata, the constant is always the last regressor. So, the order of the estimates is b_2, b_3, b_1 .
27. After estimating the model, there are lots of things you can do:
 - (a) For example, you can do all the tests you were asked to do in the problem set. Just use the following commands (these are always computed as F-statistics):
 - i. `test (lx2=1);`
 - ii. `test (lx3=0);`
 - iii. `test (lx2+lx3=1);`
 - iv. `test (lx2=0) (lx3=0);`
 - (b) You can also save the residuals and the fitted values. You can do this from the drop-down menus (Statistics; Postestimation; Predictions, residuals, etc.), or just use the commands: `predict y_hat, xb` (for the fitted values), or `predict e, res` (for the residuals).
 - (c) When you do a regression, Stata saves lots of results. You can see them with the command `ereturn list`. For example, if you want just to see the R^2 with more decimal places, you can type `display e(r2)`, or just `di e(r2)`.

(d) To get the covariance matrix we need to use matrix commands. Do it like this:

```
matrix v = e(V)
matrix list v
```

28. We can now see how to create a do file. To do that, go to Window; Do-file Editor; New Do-file Editor. Once the new window opens, you can write Stata commands in it. These commands will not be immediately executed. You can also create a do file with any text editor, like Notepad, and then open it with Stata.

(a) For example, write the commands

```
reg ly lx2
reg ly lx3
```

(b) Once you have your do file (do not forget to save it with an informative name!), you can then execute the commands in it with Ctrl+D.

(c) Of course, you can always make changes to the commands in the do file and run it again.

29. When you are done, you can close the log file (either using the drop-down menu or with the command `log close`), and exit Stata!

30. Finally, note that you can get lots of information on Stata in the web. Either just Google the thing you want to learn about, e.g., `reg stata`. Otherwise, a good starting point is here: <http://www.essex.ac.uk/economics/resources/statainf.aspx>.