EC501- ECONOMETRIC METHODS AND APPLICATIONS

(INTRODUCTION TO STATA)

OVERVIEW

- PURPOSE OF STATA LECTURE AND CLASSES
- 2. WHAT IS STATA
- 3. STARTING AND EXITING STATA
- 4. THE STATA USER INTERFACE
- 5. SETTING THE SIZE OF MEMORY
- 6. USING THE VIEWER
- 7. THE HELP SYSTEM
- 8. -MORE- CONDITIONS, ERROR MESSAGES AND RETURN CODES, THE BREAK KEY
- 9. LOADING AND SAVING DATA
- 10. USING THE DATA EDITOR
- 11. LABELING DATA
- 12. CREATING NEW VARIABLES
- 13. DELETING VARIABLES AND OBSERVATIONS
- 14. USING THE DO-FILE EDITOR
- 15. GRAPHS
- 16. LOGS: PRINTING AND SAVING OUTPUT
- 17. MORE ON STATA

1. PURPOSE OF STATA LECTURE AND CLASSES

- The main objectives of Stata Lecture and Classes are:
 - The introduction of Stata as a statistical package that permits the implementation of the econometric methods studied in EC501;
 - The Lecture introduce some of the main features of Stata;
 - The Classes are designed to implement the econometric methodologies studied in the EC501 Course Lectures and Classes.

2. WHAT IS STATA

- Stata is a statistical package for managing, analysing, and graphing data;
- Stata is command-driven. It is opperated in a user-interface way. Type a little get a little, so that the user is always in control;
- Stata's model for a dataset is that of a table- the rows are the observations and the columns are the variables (observations are numbered; variables are named);
- Stata is very fast. Partly because of clever programming, and partly because Stata keeps the data in memory;
- You do not work on the data on disk, but just a copy of it on memory, which makes Stata safe for interactive use;
- The only way to harm the permanent copy of your data on disk is if you explicitly save over it;
- To check Stata power: http://www.stata.com/capabilities;
- Which version: **about**

Latest version: Stata 9.2 (currently installed in the University of Essex labs)

| | Stata/SE | Intercooled | Small |
|---------|----------|-------------|-------|
| Maxvar | 32,767 | 2,047 | 99 |
| Matsize | 11,000 | 800 | 40 |

3.STARTING AND EXITING STATA

- To start Stata:
 - Click on the Start button, then Programs and then Stata;
 - or double-click on a Stata data file, which have the extension .dta (in this case, Stata is launched and the data file is loaded into Stata).
- To exit Stata:
 - Enter **exit** from the command line;
 - or select File > Exit from Stata's menus.
 - If you have changed the dataset in memory and have not saved it, you can do one of two things:
 - save the dataset and then exit Statasave *filename*exit
 - or you can force Stata exit without saving the dataset by typing exit, clear.

4. THE STATA USER INTERFACE

- The windows
 - The main windows contained within the main Stata window are:
 - The Results window [Ctrl+1] or click the results icon]
 - The Command window [*Ctrl*+4]:
 - Type commands here (use PgUp and PgDn buttons for past commands).
 - Press *Enter* to implement the command.
 - The Review window [*Ctrl*+5]:
 - Past commands appear here (click on command, and it will appear in the Command window, double-click on it to execute).
 - The Variables window [*Ctrl*+6]:
 - Variables appear here (click on the variable, and it will appear in the Command window).
 - The menus and toolbar permit to access the following Stata's additional windows:
 - The Graph window [*Ctrl*+2]
 - The Viewer window [*Ctrl*+3 or click the viewer icon]:
 - Permits to observe the results of: help, search, net search, view
 - The Data editor [*Ctrl*+7 or click the data editor icon or type **edit** in the Command window]
 - The Data browser [click the data browser icon or type **browse** in the Command window]
 - The Do-file editor [*Ctrl*+8 or click the do-file editor icon]

4. THE STATA USER INTERFACE

- The toolbar (icons)
 - Open: open a Stata dataset.
 - Save: save a Stata dataset.
 - Print: print Results, Viewer, or Graph..
 - Log: to start or stop, append to an existing log, close, suspend or resume a log file.
 - Viewer: open Viewer window, or bring window to the front.
 - Results: open Results window, or bring to the front.
 - Graph: open Graph window, or bring to the front.
 - Do-file editor: open do-file editor, or bring to the front.
 - Data editor: open data editor, or bring to the front.
 - Data browser: open data browser, or bring window to the front.
 - More: command to continue when paused in the middle of long output.
 - Break: stop the current task. It returns the system to as it was before you issued the command.
- Closing and opening windows:

All windows but the Results and the Command windows can be closed. To open a closed window, click the Window menu, and select the desired window.

5. SETTING THE SIZE OF MEMORY

Memory size considerations:

By default, Intercooled Stata allocates 1 megabyte to Stata's data areas and you can change it with the command **set memory**;

You can even change the allocation to be larger than the physical amount of memory in your computer using virtual memory- set virtual on (but you are reducing the amount of memory for computations).

- There are only two restrictions on the set memory command:
 - The amount of memory in your computer;
 - Whenever you change the total, there cannot be any data already in memory. If you have a dataset in memory, save it, clear memory, reset the total, and then use it again.

save mydata, replace drop _all set memory 30m use mydata

6. USING THE VIEWER

- Functions of the Viewer:
 - view the contents of the help system;
 - view the help file for any Stata command;
 - search Stata's documentation and FAQs by keywords;
 - search net resources by keywords;
 - obtain advice on how to use search;
 - find and install Stata Journal, STB, and user-written programs;
 - review, manage, and uninstall user-written programs;
 - check for and optionally install official Stata updates;
 - view logs and other SMCL and ASCII files;
 - launch a browser;
 - see the latest news from http://www.stata.com/

6. USING THE VIEWER

Viewer buttons

- Back: return to the previous contents viewed.
- Refresh: reload the current contents.
- Search: perform a keyword search for help files, FAQs, Stata Journal articles and programs, STB articles and programs, and user-written programs.
- Help!: gives links, advice, and detailed instructions for using the Viewer.
- Contents: contains a table of contents for all the Stata help files.
- What's New: describe the new features of Stata in the current release.
- News: list recent news and information of interest to Stata users.

7. THE HELP SYSTEM

- Stata has help on-line and a lot of it. You can access this by pulling down Help or typing the help and search commands;
- Both methods access the same underlying information, but the first method is better because it invokes the Viewer, displaying the help file in a separate window;
- Blue text indicates a hypertext link, so you can click on it to go directly to the help;
- Rule 1: If you know the name of the command and want to learn more about it, type help followed by the command name;
 The help command provides access to Stata's interactive help files. These files are a shortened version of what is in the printed manuals;
- Rule 2: If you do not know the name of the command, type **search** anything you want or you may pull down Help, select Search..., check Search documentation and FAQs, and then type anything you want in the Keywords input field.

7. THE HELP SYSTEM

- **search** responds by providing a list of references in the Results window-references to the on-line help, references to printed documentation, references to FAQs at the www.stata.com web site, and references to articles that have appeared in the Stata Technical Bulletin (STB) now the Stata Journal (SJ);
- The **help** command has the advantage that permits you to continue to work while keeping the Viewer openned;
- Stata is Internet aware. When you type **net search** *keywords*, Stata searches all user-written programs on the Internet, including the entire collection of STB and SJ programs and FAQs posted on the Stata web site;
- This permits you to **upgrade** Stata with new commands from the Internet;
- Stata has on-line tutorials which will help you learn about Stata's features. For example, you can type

tutorial graphics

• Stata has the *Reference Manuals* and the *User's Guide*. The former explain each Stata command. The latter is organized by subjects and can be viewed as a summary from the Reference Manuals.

8. -MORE- CONDITIONS, ERROR MESSAGES AND RETURN CODES, THE BREAK KEY

- _more_ is Stata's way of telling you it has something more to show you, which will cause the information on the screen to scroll off;
- When you see _more_ at the bottom of the screen:

Press *letter l* or *Enter* in order to go to the next line;

Press *letter q* equal to press the Break button;

Press space bar or any other key in order to display the next screen;

Press the clear _more_ condition button, the button labeled Go with a circle around it.

- If you type set more off, _more_ conditions will never arise and Stata's output will scroll by at full speed;
- If you type set more on, _more_ conditions will be restored.

8. -MORE- CONDITIONS, ERROR MESSAGES AND RETURN CODES, THE BREAK KEY

• When an error occurs, Stata produces an error message and a return code. For instance,

```
. list waggeno variables definedr(111);
```

- The "no variables defined" is called the error message.
- The r(111) is called the return code. You can click on blue return codes to get a detailed explanation of the error.
- Mistakes are forgiven
- Having said "no variables defined" and r(111), all is forgiven; it is as if the error never occurred.
- Whenever an error occurs in a do-file (user-written program), the program immediately stops execution and the error message and return code are displayed (despite there is a command- capture- which permits to run the do-file with errors).

8. -MORE- CONDITIONS, ERROR MESSAGES AND RETURN CODES, THE BREAK KEY

- When you want to make Stata stop its current task and return to the Stata dot prompt, you press Break- click the Break button (it is the button with the big red X);
- Stata cancels what it is doing and returns control to you as soon as possible.
- In general, the side effects of pressing Break are none, so the state of the system is as if you had never issued the command.

9. LOADING AND SAVING DATA

To load a Stata data file:

```
Select File > Open...
Or type use Wages2006
Or click the Open button
```

When loading a dataset, the current dataset in memory is discarded. Either save the current dataset first or allow it to be discarded: Choosing a file after selecting File > Open... will prompt you before discarding it

Or type use Wages2005, clear
Or type clear and then use Wages1

To save a new dataset (or an old dataset under a new name):

Select File > Save As...

Or type save Newwages in the Command window

9. LOADING AND SAVING DATA

To save a dataset for use with Stata 7 or Stata 8:

```
Select File > Save As..., and select
Stata 7 Data (*.dta) (or Stata 8 Data (*.dta)) from the Save as type list
```

To resave a dataset that has been changed (overwriting the original data file):

Select File > Save

Or click the Save button

Or type save Wages 2006, replace

Or simply type save, replace

- Stata data files are named *filename*.dta. When saving or loading data files, the extension .dta is automatically assumed.
- Important note: changes are not permanent until you save them. You work with a copy of the dataset in memory, not with the data file itself.
- There is no way to recover your original data file once you have done a File > Save or save, replace.
- Additional commands: insheet, infile, infix, and others

- To enter the Data Editor:
 - Click the Data Editor button
 - Or type edit in the Command window and press Enter.
- The editor is like a spreadsheet.
 Columns correspond to variables and rows to observations. You navigate by clicking on a cell or by using the arrow keys.
- To copy and paste data between Stata's editor and other spreadsheets:
 Highlight the data to copy in either Stata or the other spreadsheet, right-click on the window, and then select Copy.
 - After copying the data, paste it into Stata's editor or into another spreadsheet by selecting the top left cell of the area to which you wish to paste, right-clicking on the window, and select Paste.
- To modify or enter data:
 Choose the cell, type the value, and press *Enter* or *Tab Enter* highlights the next cell in the column

- Numeric data and string data (i.e. data consisting of characters) are entered the same way (it is not necessary to type double quotes around strings).
- Missing values for numeric variables are recorded as "." (i.e., a period). To enter a missing numeric value:

Press *Enter* or *Tab*Or type "." and press *Enter* or *Tab*.

- The editor initially names variables var1, var2, ...

 To rename a variable: double-click anywhere in the variable's column. This opens the Variable Properties dialog. Then type the new name of the variable, and click OK. (Or in the Command window type rename var1 Year).
- A variable name must be 1 to 32 characters long. The characters may be any letters (uppercase or lowercase), digits, underscores, but no spaces or other characters. The first character must be a letter or an underscore (but an underscore should be avoided).
- Stata is case sensitive: Education, education and EDUCATION are different names.

To select the variables that appear in the editor:

```
Type in the Command window: edit age (selects the variable age) or edit wage age (selects the variables wage age)
Any number of variables can be selected.
```

To restrict the number of observations that appear in the editor:
 To restrict to only one observation, type in the Command window:

```
edit in 1 (uses only the first observation)edit in 2 (uses only the second observation)
```

edit in -2 (uses only the second from the last observation)

edit in -1 (uses only the last observation)

To restrict to a series of observations:

```
edit in 1/10 (uses observations 1 through 10)edit in 2/-2 (uses observations 2 through second from the last)
```

To restrict to those observations satisfying a mathematical expression:

```
edit if exp (edit observations for which exp is true)
ex: edit if wage^2>100
edit if age = =10 (edit observations for which age equals 10)
```

edit if wage<20 (edit observations for which var1<20)

You can combine in and if (the order does not matter)

```
edit in 1/10 if wage>=20 edit if wage<100 in 6/-2
```

You can select variables and restrict observations at the same time:

```
edit var1 in 2/-2
edit var1 var2 in 1/10 if var2>=20
```

To delete variables and observations

Click the Delete... button (and then you can delete the current variable, or the current observation, or all the observations that have the same value as the current observation). Later we will study the **drop** and **keep** commands.

The Data Editor can be used to view data. To enter the editor in browse mode: Click the Data Browser button, or type browse in the Command window. The browse mode is useful when you just want to look the data. It does not permit you to change it.

Buttons on the Data Editor

- Preserve: save changes made to the data in the Data Editor, in the backup copy.
- Restore: when you enter the editor, Stata automatically makes a backup copy of your data. If you want to cancel the changes that you made and restore the backup copy, click Restore.
- Sort: sort the observations of the current variable in ascending order.
- <<: shift the current variable to be the first variable.</p>
- >>: shift the current variable to be the last variable.
- Hide: hide the current variable.
- Delete...: Opens a dialog that allows to delete the current variable, or the current observation, or all the observations with the same value as the current observation.

The logical operators are

By default, & is evaluated before |. Therefore a|b&c means a|(b&c). To change the order of evaluation type (a|b)&c.

| < | less than |
|----|---|
| <= | less than or equal |
| == | equal |
| > | greater than |
| >= | greater than or equal |
| != | not equal (~= can also be used) |
| & | and |
| | or |
| ! | not (logical negation; ~ can also be used |
| () | parentheses specify order of evaluation |

11. LABELING DATA

- Describe, summarize and label
 - To obtain a description of the data currently in memory: describe
 - To summarize the data (e.g. get descriptive statistics of the variables): summarize
 - To set or reset the data label: label data "text"

ex: label data "Hourly wages referring to year 2006"

- To set or reset a variable's label: label var varname "text"

ex: label var age "Age of the worker in 2006"

- To label a variable's values:

Create a value label: **label define** *lblname* # "text" # "text" ...
Attach value label to variable: **label values** varname lblname

ex: label define codegender 0 "male" 1 "female" label values gender codegender

11. LABELING DATA

- To unlabel the data: label data

a variable: label var varname

a variable's values: label values varname

- To delete a value label: label drop lblname

ex: label drop codegender

- To change a value label:

Delete it: label drop lblname

Redefine it: label define lblname # "text" # "text" ...

- To make the changes permanent: resave the dataset: select File > Save or type: save *filename*, replace

12. CREATING NEW VARIABLES

Stata operators are:

| Arithmetic | Logical | Relational (numeric and strings) |
|------------------------|---------|--|
| + addition | ! not | > greater than |
| - subtraction | or | < less than |
| * multiplication | & and | >= greater or equal |
| / division | | <= less or equal |
| ^ power | | == equal |
| | | != not equal |
| + string concatenation | | |

12. CREATING NEW VARIABLES

- Generate and replace
 - To create a new variable that is an algebraic expression of other variables:

generate newvar = exp

(generate may me abbreviated)

(egen is an extension of gen)

examples:

(to generate newvar = exp

egen newvar = exp

generate newvar = exp

egen newvar = exp

gen education1=ln(education)

egen averagewage=mean(wage)

g silly = ((age + 150)/ln(wage - 40))^2

- To change the contents of an existing variable:

replace oldvar = exp

(replace may not be abbreviated)

(*exp* is an algebraic expression that is a combination of existing variables, operators, and functions).

replace silly=10 if silly<=20

Stata has many mathematical, statistical, string, date, time-series, and programming functions. Type search functions to get a better picture.

12. CREATING NEW VARIABLES

Create a dummy variable:

```
gen dummyname = 0
replace dummyname = 1 if exp
```

Example:

Suppose you want to create a new dummy variable, e.g. a dummy variable taking the value 1 if average hourly earnings are above £15:

```
gen blablabla = 0
replace blablabla = 1 if wage>15
```

13. DELETING VARIABLES AND OBSERVATIONS

- Clear, drop and keep
 - To clear all the data in memory: **drop** _all
 - To drop all the data in memory and clear other memory areas used by graphs, dialogs, matrices, etc.: **clear**
 - If you want to drop a single variable: **drop** varname
 - To drop any number of variables: **drop** *varname1 varname2 varnameN*

ex: drop wage education

- To drop *varname*_i through *varname*_i: **drop** *varname*_i- *varname*_i

ex: drop var1-var10

- To drop all variables starting with pop: **drop pop***
- You may combine all the above: **drop** varname_i-varname_i varname2 pop*

13. DELETING VARIABLES AND OBSERVATIONS

- The syntax previously presented to the edit command is also valid to the drop command.

drop in -3 drop in 2/-2 drop pop* area in 2/-2 drop if *exp* drop if wage>20 in 10/-10

keep works like drop, except you specify the variables or observations to keep.

keep if wage>10 in 3/-3

- To make changes permanent, resave the dataset: Select File > Save or type in the Command window save filename, replace.
- The commands drop and keep can not be abbreviated. The reason is that they are destructive commands of the dataset. This is a general rule in Stata.

- The Do-file Editor lets you submit several commands to Stata at once. If there is an error in some command, Stata stops the implementation of the do-file and reports the error.
 - Type the commands into the editor;
 - Then click the Do button.
- The Do-file Editor permits you to edit do-files and other text files with it. Do-files have the extension .do associated with them.
- To enter the Do-file Editor: click the Do-file Editor button, or type **doedit** and press *Enter* in the Command window.
- The Do-file Editor has standard features found in other text editors. Cut, Copy, Paste, Undo, Open, Save, and Print are some examples.
- There are more advanced features in the Do-file Editor that we will learn in this course. For instance, you may call other Do-files from a Do-file.
- You may have several Do-file Editors open.

The Do-file Editor toolbar

The Do-file Editor has twelve buttons. If you ever forget what a button does, hold the mouse pointer over the button for a moment, and a box will appear with a description of that button.

- New: start a new do-file.
- Open: open a do-file from disk.
- Save: save to disk the current do-file.
- Print: print the current do-file.
- Find: search for a string in the current do-file.
- Cut: cut the selected text from the current do-file and copy it to the Clipboard.
- Copy: copy the selected text to the clipboard.
- Paste: paste text from the Clipboard into the current do-file.
- Undo: undo the last change.
- Do: execute the commands in the current do-file.
- Run: execute the commands in the current do-file without showing any output.
- Preview: preview the current file in the Viewer. The current file must be saved first.

- Suppose you want to issue many commands to Stata during your session and that you want to be able to reproduce your work later without having to type each of the commands again.
- You may place commands in a text file (in the Do-file Editor directly, or in other text file); Stata can read the file and execute each command in sequence. This file is know as a Stata "do-file".
- Example of a do-file:

use Wage2006.dta
list education wage gender in -3/3 if wage>15
generate lneducation=ln(education)
label var lneducation "Logarithm of Years of Education"
sort lneducation
save, replace

To implement this do-file, first you would click in the Do-file editor button to open the Do-file Editor. Then, you would type the previous commands. Finally, you would click the Do button to implement it. Or type **do** *mydofile*.

- Comments can be added to the do-file as you work. Just start the respective line with a "*".
- If there was some error in any command in the do-file, Stata would stop its implementation, and would report the error. You could then go back to the do-file, correct the error and implement it again. But you could run it with **capture**.
- You might want to select File > Save As... from the Do-file editor to save this do-file. Later, you could select File > Open... to open it and implement it again in other dataset, or add more commands as you move forward with your analysis.
- The File menu
 - The File menu of the Do-file Editor includes standard features found in most text editors. You may choose to start a New file, Open... an existing file, Save the current file, or save the current file under a new name with Save As.... You may also Print... the current file. You may select Insert File... to insert the contents of another do-file at the current cursor position in the Do-file Editor.
- Saving interactive commands from Stata as a do-file While working interactively with Stata, you may want to rerun the last several commands that you typed interactively. You can save the contents of the Review window as a do-file and open that file in the Do-file Editor.

15. GRAPHS

- The Graph button brings the Graph window to the front of all other windows. You can have multiple graph windows open.
- Like almost other Stata cababilities, you can construct graphs in Stata in the Command window or from the Stata menu. The syntax of graphs can become relatively more complicated than other syntaxes in Stata.
- To print a graph: The Graph window must be open, right-click on the window, and select Print..., or select File > Print > "graph title".
- To save a graph to disk: The Graph window must be open, right-click on the window, and select Save Graph..., enter the *filename*, click OK.
 If you do not specify an extension, .gph will be added, and the file will be named *filename*.gph.
- To load a graph from disk: Select File > Open Graph..., select the graph filename, click OK. Or double-click on the .gph file from the Desktop, or type graph use *filename*.
- There is much more to learn about graphs in Stata. Stata permits to construct a very large number of graph types. Type **search graphs**.

Using logs in Stata

A log is a recording of the Stata session.

Logs are stored as Stata formatted (SMCL) files or text (ASCII) files. They can be viewed in the Viewer, they can be printed, they can be translated from formatted (SMCL) logs into text (ASCII) logs.

ASCII logs can be loaded into a text editor or word processor.

To start a log:

Click the Begin Log button, and then enter a *filename*, or select File > Log > Begin..., then select the log type and enter a *filename*, or type

log using filename

The file will be named filename.smcl for SMCL logs filename.log for ASCII logs

If you specify a file that already exists, you can choose one of three options

- view the existing file
- append the new log file to the file
- overwrite the file with the new log

To close the log file:

- the log file is closed automatically when you exit Stata
- to close the log file sooner: click Close/Suspend Log, and choose Close log file, or type: log close.

To temporarily suspend output from being written to the log:

- click Close/Suspend Log, and choose Suspend log file, or type: log off.

To resume the suspended log:

- click Close/Resume Log and choose Resume, or type: log on.

To view a snapshot of the current log in the Viewer:

- Select File > Log > View...
- The filename of the current log is filled in automatically.

To view an updated snapshot of the current log in the Viewer:

- Click the Refresh button at the top of the Viewer

To print the current log:

- First, view a snapshot of the current log in the Viewer. Then, right-click on the Viewer window and select Print...

To add comments to the log file, type a "*" at the beginning of the line:

* this is a comment

Alternatively, you can use the log command to create a log file. To start a new log file:

- Type: log using filename

- Type: log using filename, smcl

- Type: log using filename, text

To append to *filename*.smcl or *filename*.log:

- Type: log using *filename*, append

To overwrite *filename*.smcl or *filename*.log:

- Type: log using *filename*, replace

To translate a SMCL log to a text log:

- Select File > Log > Translate..., enter the SMCL log file in the Input File field, enter the new file in the Output File field, click Translate

To create a log that contains only the command lines that you type:

- Type: cmdlog using filename
- This type of file is saved as *filename*.txt
- Later, you can run the do-file to reproduce the output by typing
 do filename.txt

17. MORE ON STATA

• Stata is Internet aware. You can download files from the Internet from inside Stata. Not just data files, but also the official updates to Stata provided by StataCorporation, from the STB (now Stata Journal), User-written Programs (which are files with the extension .ado).

use http://www.stata-press.com/data/r9/auto

- This introduction lectures provide you the essential tools to begin using Stata. You should start playing with it.
- If you make a mistake and cannot figure out what is wrong, look at the command's help file: Select Help > Stata Command..., and enter *commandname*. Compare the syntax of the command with what you typed.
- If you want to find the Stata command that produces a particular result, try: Select Help > Search..., choose Search documentation and FAQs, and enter *topic*.

17. MORE ON STATA

- You can get help from
 - The User's Guide, which is designed to be read cover to cover.
 - The Reference manuals are not designed to be read cover to cover, but be sampled when needed. They present an alphabetical explanation of most Stata commands, including examples of their applications.
 - The datasets used in the examples in the reference manuals are posted at: http://www.stata-press.com/data/
- Look at the Stata web site, which has much useful information, including answers to frequently asked questions (FAQs):

http://www.stata.com/support/faqs/

Many useful links to Stata resources are found at:

http://www.stata.com/links/resources.html