

How to read these notes

First of all, it should be noted that these notes are not indented as a substitute to the software's manual. In fact, a great deal of information was adapted from the manual for compiling these notes. Stata-press (www.stata-press.com) publishes an excellent set of manuals with extensive postings that cover all automatically-downloaded commands, programming language, expressions, as well as statistical and econometrical methods. They provide detailed explanation of syntax, a number of illustrative examples, analytical theory, technical notes for programmers and references.

To illustrate the structure of what you are about to read, consider the following extract:

```
help codebook
Stata has a useful command for describing the contents of the data and alert the user on potential red flags.
codebook thoroughly examines variable names, labels and data points:
codebook          Examines _all variables in the dataset.
```

A section enclosed by two horizontal lines indicates the introduction of a new command or expression. In the example above, this section introduces the command `codebook`. The first line of each section is set in blue font and refers to the appropriate help file or search command, *i.e.* by typing in Stata `help codebook` you invoke the associated help file for `codebook`. Note that sometimes there are references to more than one help sources. All words that are set in *Lucinda Console* font indicate Stata commands, Stata expressions or user-defined variables that are understood by Stata within the context of these notes. This font is set in three different colours in order to distinguish the following situations:

- *Lucinda Console* regular black indicates official Stata syntax or syntax that is understood by Stata within the context of these notes. For example, in the above example, `_all` signifies valid Stata syntax expression which in fact can be used to indicate the inclusion of all variables in the dataset, *i.e.* typing `codebook` is equivalent to typing `codebook _all`.
- *Lucinda Console* bold red prompts the reader to type in the command in Stata as shown precisely and execute accordingly.
- *Lucinda Console* regular green surrounded by dashed borders indicates code that must be copied and pasted in the Do-File Editor.
- A Stata command with some of its first letters underlined, such as `help` and `browse`, indicates that the commands can be conveniently abbreviated into `h` and `br` respectively. However, not all commands have an abbreviated form, such as `codebook`. Regularly, the manual uses abbreviations in order to save space and truncate lengthy command lines.

You are expected to read these notes in the order they are provided. Do not skip any pages and do not read in random order. To a great extent, the notes are structured in a way to rely on the execution of previously discussed commands. To maintain consistency and avoid undesirable outcomes, all you need to do is type in and execute from Stata all commands that are set in **red** and **green** fonts.

Lastly, throughout the notes you will encounter a number of graph commands typed with *Lucinda Console italics navy fonts* and highlighted within light shade grey cells, as in the following example:

```
twoway (hist r1, fc(gs14) lc(gs8)) (hist r2, fc(none) lc(orange_red) lw(medium))
(function y=normalden(x), ra(-4 4)), leg(pos(11) ring(0) col(1)
lab(1 "n=100") lab(2 "n=10,000") lab(3 "Normal"))
```



Graph code is often long and spans in more than one visual line, but you must make sure to enter and execute as a single command-line. Graphing is not covered in these notes, yet for the sake of illustration you will be asked to execute a number of graph commands. All you have to do is copy the command as it appears in the cell and paste it in the Command window of Stata in a single command-line.

The notes are specifically designed to accompany the following financial datasets each of which is stored under a different format (notice the file-extensions):

Datasets	<i>EU_ni.dta</i> <i>US_ni.txt</i>	<i>EU_p.csv</i> <i>US_p.raw</i>
Unique Variables	company firm full name country 14 pre-enlargement EU states plus the US sic1 1-digit SIC codes(10 industry classes) currency values DKK, EUR, GBP, SEK, USD ta Total Assets te Total Equity ni Net income	ns Number of Shares outstanding P Closing Price at fiscal year-end
Common Variables	sic4 4-digit SIC codes(802 industry classes) year takes the values 1999-2004 (5 years) code database code identifier per firm	
More datasets	Datasets suffix ' <i>ni</i> ' indicate accounting data and datasets with suffix ' <i>p</i> ' indicate stock market data. Dataset with additional suffixes such as ' <i>F2</i> ' or ' <i>HI</i> ' indicate the manual's section at which the dataset is called for executing certain commands. The dataset <i>rates_E4.dta</i> is created in section E4, whose content is then used in subsequent section under different suffix.	
<i>iso3166.dta</i>	This dataset is downloaded from the International Organization for Standardization (ISO, www.iso.org) and contains the following variables: full names of countries <i>country_name</i> ; 2-digit codes <i>iso3166_2</i> ; 3-digit codes <i>iso3166_3</i> ; number codes <i>iso3166_n</i> .	

During the installation of Stata, the software creates the directory *C:/data* which is henceforth recognised as default working path directory for all Stata sessions. Following installation, make sure to transfer all five datasets above within *C:/data*.

Acknowledgements: I am grateful for many constructive comments from Karl Keesman (Survey Design & Analysis P/L, Australia), Stuart McLeay (University of Wales Bangor, the UK), MEAFA members and various users of this manual from more than 30 countries.

Suggested citation:

Christodoulou D. (2008), *Introduction to Stata 10 and the Management of Financial Data*, MEAFA research network (<http://meafa.econ.usyd.edu.au/>), The University of Sydney.

Contact information: for any questions, suggestions, further clarification or for reporting problems:

Demetris Christodoulou, MEAFA Director
 E-mail: meafa@econ.usyd.edu.au

This version was released in July 2008

