Computers and Econometrics - Preliminaries

John C. Frain.

September 1, 2009

1   Introduction

This note is an introduction to the computer aspects of Econometrics I and II.

It is assumed that all MSc students have a working knowledge of Microsoft WIN-DOWS, WORD and EXCEL. They should be able to use WORD to draft a report, EXCEL to manage data and draw graphs and be able to copy, move, delete and otherwise manage files with WINDOWS EXPLORER. You should also know how to use the internet browser and your college email. Most students will have used these tools during their undergraduate courses and be familiar with them. If anyone has any concerns about their knowledge of this software let me know urgently. You will have plenty of opportunities to use and develop these skills throughout the year, not only in your econometrics, but in your macroeconomics and microeconomics. WORD, EXCEL Internet access, email and many other programs are available to you on College computers. Details of these and other computer facilities available on College computers are set out at `http://isservices.tcd.ie/`.

All students and other College users have an allocation of 100MB of personal file storage in the form of a folder on a central server. This file storage can be used to hold email attachments, word documents, picture files or any other data and programs that you wish to store. After you have logged on to the system a computer room PC will have a shortcut to your file storage on the desktop. Double click on this to view the contents of your file storage. Alternatively it can be used in other programs much the same as a second hard disk, the `S:` drive.

Each user is obliged to monitor and maintain the amount of space they have used in their personal file storage to ensure that they do not exceed the 100MB quota. They should also make back-up copies of any critical files that they have stored there. IS Services is not responsible for backing up the data in users' file storage folders. If you accidentally delete your data, analysis, reports or other valuable material or it is otherwise lost (e.g. a network crash) you will almost certainly be unable to get it back. Experience is that such accidents occur just before a report is due. **It is essential that you make copies of all the important files on your `S:` drive**. IS recommend the use of memory sticks or zip drives. The same consideration holds if you are writing reports on your home computer. In such a case you will find that a memory stick very useful for transferring data between your home computer and

a college computer. Standard back-up procedures would use two memory sticks which one would use on alternative days. In this way one would not lose mere than a days work even if one was unfortunate to lose one of the sticks and have a data catastrophe on the same day.

I would strongly recommend that you set up a separate directory on your `S:` drive for each practical exercise or project. Here you can store and easily find any data, program and result files that you may need for a project. This practice will make econometric packages easier to manage and facilitate backup. We will return to this point later. Files do get lost if they are not organised properly.

2 Reporting the results of a project

Each project may be accompanied by specific instructions as to the material to be presented at the end of the project. Three items will in general be required.

1. A report containing the results of the analysis

2. A listing of the program or set of instructions used in the production of the report and

3. the computer output generated by the program or set of instructions.

2.1 The report

Good advice on writing and presentations for economists is given in "Writing Tips for Ph. D Students" by John D. Cochrane, `http://faculty.chicagogsb.edu/john.cochrane/research/Papers/phd_paper_writing.pdf`. This paper is a description of what is required in academia while the description here is more typical of a research paper in an organisation such as a Central Bank.

A practical requirement here is that your report should be submitted in MS WORD[1] or Adobe pdf (portable document format).

In real life the content and format of a report will vary with the audience for whom it is being written. In an outside agency you will probably be required to write an executive summary of your report. This executive summary may be your only contact with senior management. Concentrate on your main results and policy implications. In my experience senior management will stop reading when they encounter an equation. If you make your executive summary interesting senior management may ask you to explain it further and may even read some of your paper.

---

[1]reports prepared in LaTeX or a word processor other than MS WORD should be presented in Adobe pdf (portable document format). If you use another word processor at home and translate to WORD format please ensure that the document is fully and properly translated. If you run your programs on a MAC computer make sure that the report, listings and computer output can be read on a Windows PC. I do not have access to a MAC computer and may not be able to read or process your submissions. If it is not fully translated you can not get marks for what I can not read

While all reports do not necessarily have the same structure most empirical studies will contain the following, (not necessarily in this order) -

Executive Summary (optional) As stated above an executive summary is aimed at non-technical management. In this course you will probably not require an executive summary in your econometric exercises/projects although it may be good practice to do so.

Abstract The abstract is **not** an executive summary. It is a very short technical summary and is aimed at other experts. Most journals require that this be very short. Just state your main conclusion

Introduction This section should set out the motivation and basic aims of the study. You might be able to illustrate these with some graphs or summary statistics. For example if you were studying the relationship between tobacco consumption and income you might you might showing a graph relating tobacco consumption to income. You might also consider one main aspect of the previous literature that is of particular interest. Explain the function of each of the following components of the paper. Finally give a brief summary of your main results. If you are writing a paper pay particular attention to this section. A busy reader may glance at your introductory section and put the paper aside if he does not see anything of interest. If you attract his attention he may skip to material in a later section and eventually read the whole paper. Avoid, if possible, technical details in your introduction. Keep it short and concise.

Literature Review This section should present a review of any other previous work that is relevant to your analysis. In your literature review tell a story – don't just list papers.

Theoretical Framework and Econometric Model This section should present the economic theory that you are using. It should show how the econometric model estimated in the following section is derived. If there is an established mathematical notation use it. In any case describe your notation. If the theory is likely to interrupt the flow of your reasoning postpone it to an appendix.

Estimation of Econometric Model This section gives the results of the econometric model and gives them economic meaning.

Summary and Conclusions, Policy Recommendations This section gives you another chance to set out the importance of your study. This is probably the second most important section in your paper. If your introduction attracts management they will then go to your conclusions.

References

Data Appendix With the details in your data appendix a reader should be able to reconstruct your data set. Many journals make it a condition of publication

that data sets (and programs) be supplied. Some journals often hold these data sets and programs on their web sites where they are available to other researchers.

Theory Appendix (optional) If the economics or econometrics are very complicated or include a considerable amount of hard maths including them within the main body of the report may take from the general arguments put forward. In such a case it is better to but such material in an appendix.

The default font in MS word is Times Roman. In general text looks better and is easier to read in Times Roman or another proportional font. Computer listings and tabular material do not look well in such a font and should be given in Courier or another fixed size font.

When you use courier to show computer output or tables some of the lines may be to long and may fold to the next line. Use a smaller size font (10pt) or edit the table to fit on a line. Do not include spurious decimals in your output. A t-statistic of `3.15` is statistically the same as one of `3.146723` and is easier to read.

Tables transferred from computer output and set in Times Roman are almost impossible to read. I know of cases in an American university where such scripts are returned unmarked. In a table numbers should ideally be aligned on decimal places. Cutting and pasting from EXCEL to WORD may produce hard to read tables. Please reformat these so that they can be easily read.

Use the standard spell and syntax check in your word processor before submission. Finally print and read your report before submission and amend as necessary.

2.2 Computer Listings and Output

My working methodology and the one I recommend to you is as follows. In the project working directory I create a file (the program) or set of files containing my initial trial instructions for the project. I then run this set of instructions through the econometric package and review the output. I then make necessary amendments and additions to the program file to amend or extend the analysis and re-run the program. This process of extension and amendment may be re-run many times until I am satisfied with the output.

As I go through this process I add comments to the program to explain what I am doing and why I am doing it. Including a list of variables and their definitions, as comments, is also good practice

Redundant instructions should be deleted or "commented out" in this final program. Comments should also be revised as necessary. The end result of the process is the required program listing. Each of the econometric packages that you will use has a way of helping you to compose your program. This will be demonstrated to you as each package is introduced.

The output produced by this program is the required program output. You are expected to read and examine this before submission and ensure that the program

does not produce large amounts of redundant output. The computer listings and output should be plain text files. If they are processed and saved as .doc files in MS WORD lots of formatting material is added to the file. As part of the marking exercise I may try to run your program. Our econometric packages can not read WORD files and again no marks can be given for programs that I can not run.

3 Econometric Packages

Three econometric packages will be introduced in class

STATA STATA is a general purpose statistical package. It can manage large data sets. It offers good facilities for the analysis of cross-section and panel data. In economics a cross-section data set contains data on collection of economic agents at a given point in time. The Irish Household Budget Survey is an example of a cross-section data set The last household budget survey conducted by the CSO collected consumption data on 8000 households and details of their Principal Economic Status, Occupation, Industry, Household Composition, Social Economic Groups, Social Class, Farm Acreage, Town Size and Regional Authority, Accommodation type, Household Tenure. A panel data set follows the economic agents over a period of time and repeats the measurement process for each agent in each period.

MATLAB. MATLAB is an interactive system for numerical computation. From our point of view the basic variable in MATLAB is a MATRIX and you transcribe the matrix algebra from your text into MATLAB and it does the calculations

EVIEWS EVIEWS is an econometric package which has good facilities for time series analysis.

Various other econometric/statistical packages are also available. I have used most of the packages in the following lists at some stage or another.

SAS or SPSS. These are large general statistical packages. SPSS is available on the TCD network

MICROFIT. MICROFIT is a very easy to use econometric package which is available on the TCD network. It is good for time series analysis. Some of you may have used it in undergraduate econometrics.

RATS. RATS like EVIEWS is particularly good for time-series analysis.

Maple or Mathematica. These are packages for symbolic mathematical manipulation which also can be used for numerical analysis. They have widespread use in Finance and Economics. Mathematica is available on Public Access computers in TCD and you may find it useful for checking some symbolic mathematical calculations.

There are also various other packages that are of interest to economists. The following programs are distributed under various free licences and are available for download.

LaTeX LaTeX is a typesetting program for scientific papers and books. Most journals and many publishers can accept and some insist on submission in LaTeX. If you are likely to produce papers and books with a large mathematical content some variant of TeX or LaTeX is the best choice. This document, my transparencies, and some other material for class was or will be produced in LaTeX.

OpenOffice. OpenOffice is a free office suite that is almost 100% compatible with Microsoft Office. Calc the spreadsheet in the OpenOffice suite is recommended for transferring data between programs and I use it for that purpose in preference to Excel. The new version 3 of OpenOffice can read MS Office 2007 files.

Gnumeric. Gnumeric is a free Excel type program. The numeric methods in Gnumeric are more accurate than those in other spreadsheets.

Abiword. Abiword is a free word processor. Abiword and Gnumeric integrate better with LaTeX than Microsoft office or OpenOffice

Scilab or Octave These are similar in function to Matlab. The recently launched version 3 of Octave runs much easier in Windows than the earlier versions. Base Octave has more data analysis and statistical functions than base Matlab. Scilab has the same functionality as Matlab but a slightly different syntax. It does contain an automatic translator for Matlab programs. It has a very good econometric toolbox (GROCER) which can complete most econometric estimation procedures.

R R is a general statistical package which is the result of worldwide collaboration between a large number of statisticians. R and its associated packages is, to my knowledge, the most powerful and up to date of all data analysis/statistical packages. The editor Tinn-R provides a graphical interface to R (see also SciViews, RCommander and JGR for alternative GUIs). Several PhD students in the Department are using R for econometric analysis.

Gretl Gretl is a free econometric package that has been considerably developed over the last years. In certain respects it is like STATA but is much simpler as it concentrates on econometrics. It has interfaces to R and Octave for problems not covered by Gretl.

jMulti jMulti is a relatively new package with a strong emphasis on time-series analysis. It works through a modern graphical user interface and is very easy to use.

Maxima Maxima is a package for symbolic manipulation. Unless your work is very esoteric or you have a specific need for some of the special packages facilities in Maple or Mathematica, Maxima is an excellent alternative and will meet the needs of many users for free.

Many of these free packages provide facilities that are on a par with or even better than their commercial equivalents. They are also available on many varieties of Linux and on MAC PCs. They can be downloaded from the internet and put into operation in a matter of minutes. Acquiring a copy of Stata may or similar may take days or longer - particularly if a request has to submitted through a central purchasing department.