Computer Toolkit for Econometricians

© A. Colin Cameron
Department of Economics, U.C.-Davis

November 1, 2000

1. Overview

The econometrician's toolkit goes beyond statistical programs. Here I give a guide to various non-statistical programs that are useful to econometrics researchers. I assume a Wintel platform with internet connection to a mainframe/workstation running Unix for some specialized software. The general issues raised below, however, pertain to all platforms and some of the software can be obtained for other platforms.

The prices in the following software list are academic prices, followed by student prices where relevant.

Software Type	**Program**	**Price**	**Web Address**
2. File Manager	Windows Commander	$30	www.ghisler.com
3. Text Editor	UltraEdit	$30	www.ultraedit.com
4. Word Processors	Scientific Word	$450/$160	www.mackichan.com
(LATEX-based)	WinEdt	$40/$30	www.winedt.com
5. Graphics Package	?? See below ??	??	??
6. Internet Tools	See below	Mostly free	various
7. Programming Language	C++, Fortran	–	various
8. Matrix Language	Gauss, MATLAB, SAS/IML	$$$	various
9. File transfer	See below	$40	various
10. Statistical Software	See below	$$$	various

2. File Manager

Organization of files on a computer is as important as organization of an office. One ideally has separate directories and sub-directories for each project. For

example, one may have a teaching directory with sub-directories for each class taught and a research directory with sub-directories for each research project. There will then usually be reason to create separate sub-sub-directories within each sub-directory. Even with such organization it can be difficult to locate a file.

The Windows file manager is Windows Explorer. Better file managers permit

- simultaneous display of two directories (target and source side-by-side) rather than just one directory;
- easy viewing of the content of the files (not just easy execution of the files);
- easy searches for particular strings of text within files, directories or collections of directories (so a file can be obtained by remembering a reasonably unique text string in the file);
- easy compression of selected collections of files or directories;
- calculation of the size of selected collections of files or directories;
- easy re-ordering of directories by file-name, file-extension, file-date or file size.

I use **Windows Commander**. This is shareware that costs around US$30 and can be obtained at www.ghisler.com.

3. Text Editor

Econometricians often write programs, view program output and view data as plain text. Great time savings are possible if one uses a specialized editor rather than simpler editors such as Notepad.

[Some Background: All files are ultimately a collection of zeroes and ones. Text files are a subset of all files where the entries are all text, meaning the 26 alphabetical letters in lower and upper case, the ten digits and standard punctuation marks and some symbols such as * and %. Special keys such as F5 are not included. Because a Microsoft Word document includes special keys it is not a text file and can only be read using proprietary software (MS Word) rather than generic text software. Text is usually stored using ASCII code, or sometimes ANSII, which is a special code for translating characters such as the letter "a" into a particular combination of 8 zeroes and ones, or a unique byte where a byte is 8 bits. A text editor is capable of working with such text. Data is most often stored as text data, the advantage being that then it can be read by a wide variety

of programs and platforms. By contrast an Excel workbook saved on a Wintel PC, for example, cannot be interpreted by all software on all platforms. Many programs are also written as text. Most statistical programs that an econometrician might write, such as for SAS, Stata and Gauss are written in text and the non-graphical output is in text. And web-page construction software such as HTML is written in text.]

The Windows text editor is Notepad. This will only read in small files, does not give line numbers, and can in some cases fail to recognize "end-of-line" markers in text files, making for very lengthy lines that are difficult to read. A good text editor:

- can handle very large files (at least 10 megabytes);
- gives the current line and column number (useful for debugging programs that give error messages such as "error at line 51 column 23");
- permits printing two pages to a page (saves paper and fewer pages of paper makes it easier to read hardcopy version of a program);
- permits multiple windows;
- permits replacing text in several files simultaneously.

There are several good text editors out there. An excellent one is **UltraEdit** available from www.ultraedit.com. I use **Programmer's File Editor** (PFE) which is not as good but is one that I am used to using. It is freeware at www.winsite.com/info/pc/win95/misc/pfe101i.zip/

4. Word Processor

The most commonly-used word processor is Microsoft (MS) Word. This does a pretty good job but does not handle mathematics well. [Word documents also take up very large files, which is less of a problem now unless one wishes to transfer files using 1.4 mb disks]. I use MS Word for correspondence and little else.

Econometricians inevitably use mathematical text. The standard programs for mathematical text are TEX, due to Donald Knuth, and its successor LATEX. These are remarkable programs. They are written in text and one can create LATEX documents simply using a text editor (see above), but this requires remembering a lot of commands. For example, \$\QTR{bf}{x}_{i}^{\prime }\QTR{bf}{\beta }\$ gives bold-face x subscript i transpose beta, or $\mathbf{x}_i'\boldsymbol{\beta}$.

A windows interface to LATEX can allow this to be created in a way more similar to MS Word. For example, in Scientific Word the sequence is CTRL-M, F5, x, CTRL_L, i, →, F5, GTRL-G, b where CTRL-M is for mathematical text, F5 is boldface, CTRL-L is for subscript and CTRL-G is for greek. As an alternative to these various shortcut keys one can instead click on icons.

The best windows interface to LATEX is **Scientific Word**, available from www.mackichan.com at a price of $450 (academic) or $160 (student). The student price is excellent and was not initially available. An enhanced version of Scientific Word called Scientific Workplace includes Maple and costs $650 or $230. A restricted version of Scientific Word called Scientific Notebook is also available but is too restrictive for graduate and post-graduate use.

A cheaper windows interface, which is not as close to WYSIWYG (what you see is what you get) is **WinEDT** from www.winedt.com which is available for $40 (academic) or $30 (student). The main attraction of this over Scientific Word is lower price.

While the windows face is much more convenient, one may still occasionally need to refer to a LATEX manual. I use Leslie Lamport (1994), *LATEX: A Document Preparation System User's Guide and Reference Manual*, Addison-Wesley, and Michel Goossens, Frank Mittelbach and Alexander Samarin (1994), *The LATEX Companion*, Addison-Wesley.

5. Graphics Package

There are two uses for graphics.

One use is to create graphs from data. A starting point is Microsoft Excel. Better is the graphics provided within a statistical package. The packages vary in terms of flexibility and ease of use. I use the graphics within Stata (www.stata.com) and find it quite good for a statistical package. The best Windows interface specialized graphics software is **SigmaPlot** which can be obtained for $499 from www.spssscience.com.

A second use is to draws diagrams with labels for teaching purposes, such as supply and demand curves and the t-distribution for hypothesis testing. At this stage I do not have adequate software for this. I use MS Draw which is provided free within MS Word. I find it very slow to use and I am need of something better.

Once a graphic is created in Windows the easiest way is to click on the graphic and copy it into a word processor, e.g. Scientific Word or MS Word, using Paste Special. This usually saves the picture as a Windows Meta File (.wmf) of BitMap (.bmp). These files can be large, especially bmp files. More space-efficient is to

save the graphic as a .gif file or .jpg (or .jpeg) file, which are formats recognized across many programs, and to import as a picture.

6. Internet Tools

The presentation here is very brief.

6.1. Web browser

Use Microsoft Internet Explorer or Netscape Communicator.

6.2. Email

Use Microsoft Outlook or some equivalent such as Eudora.

6.3. Web Search Engine

There are many out there and this will change very rapidly. The key is an engine that is fast and which usually finds what you want on the first few entries rather than on the second and later pages. At the moment Google at www.google.com is the best.

6.4. Web Publishing

Simple web pages are written in a program called HTML. These can created using windows software that has an interface similar to MS Word, and is as easy to use as MS Word. In theory one can save MS Word documents as HTML files, but I find it much better to use **Netscape Composer** within Netscape Communicator which is free. For some money one instead use **Microsoft FrontPage** within MS Office. For even more money one can by programs specialized web design programs such as Dreamweaver at www.macromedia.com. More complex web pages use Java but I have found no need to get into this.

The way to create an initial web page is to copy someone else's, read it into a program such as Netscape Composer or MS FrontPage, and then edit using cut and paste, etc. A key distinction is made between relative references, for example, howto.html which means file howto.html in whatever the current directory is, and absolute references, for example, http://www.econ.ucdavis.edu/faculty/cameron/howto.html which means file howto.html in directory www.econ.ucdavis.edu/faculty/cameron . As much as possible avoid hard references in web-page design. The easiest way to start is to put all your web files into your root web directory and use relative

references. A brief summary of using Netscape Composer to create a homepage is at http://www.econ.ucdavis.edu/graduate/kaoru/howto/howtoComposer.html

6.5. Adobe Acrobat

HTML is good for basic text and for incorporating pictures. For example, I write a class syllabus directly in HTML using Netscape Composer, rather than using a word processor and then converting a word processor document to HTML. But HTML is a very primitive language which handles mathematics very poorly.

For mathematical documents the norm is to save the document as a pdf file (portable document file). These can be read by virtually anyone nowadays, using Adobe Reader which can be downloaded free from www.adobe.com. But to write a pdf file requires purchase of software. The software to get is **Adobe Acrobat**, available from www.adobe.com at a price of $429. It can be purchased much more cheaply at academic institutions. For example at the UCD Bookstore it costs less than $100. The software is very easy to use once installed. Instead of printing to a desktop printer one instead prints to a printer called Acrobat PDF Writer and saves as a file with extension .pdf.

7. Programming Languages

Statistical packages cannot handle all estimation problems, so one may need to use of a programming language. The 1960's norm was the programming language **Fortran**. Over time this has been superceded by C and its current successor **C++**. The code underlying most statistical packages is actually written in C++.

For econometrics the norm is to instead use a matrix programming language, though for many problems code written in C++ may execute faster.

8. Matrix Programming Languages

Econometrics makes great use of vectors and matrices. It is therefore useful to use programs where vectors, rather than scalars are the primitive. One advantage is ease of reading the program, as it more closely follows the mathematics. For example if $\mathbf{x}$ is a $n \times 1$ vector one can compute $\mathbf{x}'\mathbf{x}$ using the command x'x. By contrast, if the scalar is a primitive then using $\mathbf{x}'\mathbf{x} = \sum_{i=1}^{n} x_i^2$ we would have code such as: xtranx = 0; do i = 1 to n; xtx = xtx + x(i)*x(i); end. Econometricians therefore use matrix programming languages, even though other programming languages may perform faster due, for example, to faster looping.

The most commonly-used matrix programming languages are:

Program	Price	Web Address
Gauss	$$$	www.aptech.com
Matlab	$$$	www.mathworks.com
SPlus	$$$	www.splus.mathsoft.com
SAS/IML	$$$	www.sascom

GAUSS was written for econometricians and was the initial entrant. **Matlab** is written for engineers and mathematicians and is becoming increasingly popular in economics. **SPlus** is written for statisticians and is the norm for JASA articles, for example. **SAS/IML** is a matrix programming add-on for SAS.

These packages are expensive - of the order of $500 without site-licenses and site-licenses are not always available. For econometrics one usually wants to additionally by an optimization routine for nonlinear estimation, such as Maxlik in GAUSS, which pushes the cost of these programs up even further.

Which should one choose? GAUSS was the initial entrant in econometrics, though MATLAB is making inroads. A reasonable decision rule is to use what is most often used by immediate colleagues as colleagues can be of great assistance in providing initial programs and programming tips. Programming in GAUSS or MATLAB is much more difficult than any other aspect of computer use an econometrician may encounter.

9. File Transfer / Networking

Ideally all software and data are on one's own PC. In practice there can be a need to use another computer, due to considerations such as expensive software only available on the other computer, problem to large for a desktop PC, or desire to free up one's own computer to work on other problems.

9.1. File Transfer

The easiest way to communicate between two computers is to be directly networked, in which case a directory on the remote compute becomes just another drive on the PC, such as drive H: and one can transfer files in the same way that they would be transferred to a floppy disk.

Without a network the way to transfer files between computers is via ftp (file transfer protocol). This is done using a windows ftp interface program that on one side has the directory on the remote computer and on the other side the directory on the local compute. To transfer files just highlight the files and hit the copy

button. There are many good programs. I use **WS-FTP** from www.ipswitch.com which costs around $40. The one thing to be aware of in such file transfer is that there is a distinction between binary files and text (or ascii) files. Binary files, which are non-text files have to be sent as binary files. Text files should be sent as text files. Text files can instead be sent as binary files but this will add a blank line after every line, which can cause problems if read in as data where the first line is real data and the second line is mistakenly read in as real data with all variables taking the value zero.

9.2. Running Programs on a Remote Computer

Running programs interactively on a remote computer can be difficult unless one has the correct software to be able to provide the correct interface on the local computer, especially if a graphic user interface is used. And running programs interactively has the general drawback of being inefficient if it does not store the commands in a file for future re-use if modifications are to be done.

Generally I run programs as batch jobs (even if on my own PC). There is then a cycle of

1. Edit input file of batch commands.
2. Run the program of batch commands which produces an output file.
3. Edit the output file.

Here step 2 is necessarily done on the remote computer. Steps 1 and 3 can either be done on the remote computer or on the local PC. They are easier done on the local PC, but then one has to transfer files back and forth which is inconvenient if this requires use of ftp. They can be done on the remote computer, but this requires learning how to use an editor on the remote computer plus some file management commands. On Unix machines, for example, the basic editors are inconvenient to use. The editor Pico is often available and is relatively easy to use, but it does not provide line numbers.

10. Statistical Software

There is a lot out there and it is generally not possible to find that one package that will do everything. The major programs used by econometricians include

Program	Specialty	Price	Web Address
SAS	General Statistics	$$$	www.sas.com
SPSS	General Statistics	$$$	www.spss.com
Stata	Cross-section econometrics, biostatistics	$$$	www.stata.com
Limdep	Cross-section econometrics	$$$	www.limdep.com
Eviews	Time series econometrics	$$$	www.eviews.com
TSP	General econometrics	$$$	www.tspintl.com

SAS is a universal program that had first-mover advantage and the ability of handling large data sets as it makes greater use of writing to and from files rather than putting all data into the memory at once. With increased memory available this advantage of SAS is less important. **SPSS** is another universal package, widely used in social sciences but less used in economics than SAS.

Within econometrics, commonly-used packages include **STATA** (also used in biostatistics) and **LIMDEP** for cross-section data; **EVIEWS** for time series data; and **TSP** for both cross-section and time series data.