Trends in Applied Econometrics Software Development 1985-2008, an analysis of Journal of Applied Econometrics research articles, software reviews, data and code

Marius Ooms*†
VU University Amsterdam, Department of Econometrics
De Boelelaan 1105, NL-1081 HV Amsterdam

November 25, 2008

Abstract

Trends in software development for applied econometrics emerge from an analysis of the research articles and software reviews of the Journal of Applied Econometrics, appearing since 1986. The data and code archive of the journal provides more specific information on software use for applied econometrics since 1995. GAUSS, Stata, MATLAB and Ox have been the most important softwares after 2001. I compare these higher level programming languages and R in somewhat more detail. An increasing number of packages is being used. A surprisingly low number of products has been discontinued since 1987. I put the time series count data on the number of articles using different softwares and on the number of reviews discussing different products in a historical perspective, where I distinguish several software types. Two waves of new products showed up in the period under study, the first associated with the introduction of the personal computer and new graphical interfaces, the second one with the appearance of the Internet. The Journal of Applied Econometrics has reviewed 77 packages. I shortly discuss thirteen other relevant packages. A table with all mentioned packages, their authors and latest versions provides a comprehensive overview of the relevant softwares in June 2008.

*Associate Professor of Econometrics at VU University Amsterdam, Correspondence to: mooms@feweb.vu.nl

†I would like to thank Christopher Baum, Jurgen Doornik, Christian Kleiber, Bill Rising and Ronald Schoenberg for helpful comments. All errors are my own. Softwares: I used MS Excel 2000, Windows XP, 5.1 Service Pack 2, MikTeX 2.4, OxEdit 5.0, Firefox 3, OxMetrics 5.0, GAUSS 7, R 2.7, Google, Google Scholar, Google Books, JSTOR and Wiley Interscience to prepare this chapter, **forthcoming as Ooms, M. (2009), Trends in Applied Econometrics Software Development 1985-2008, In: Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics, Edited by T. C. Mills and K. P. Patterson, pp ??-??, Palgrave MacMillan, Basingstoke, U.K.**

1 Introduction

I provide an overview of academic applied econometrics software development, deriving time series count data from the Journal of Applied Econometrics (JAE) software reviews (1987-2008), JAE research articles and the JAE Data archive (1995-2008). The JAE has promoted documentation and indexing of softwares and codes for applied econometrics by publishing software reviews and replication studies. Most importantly, James MacKinnon has patiently, successfully and consistently added software codes of JAE authors to the JAE Data archive.

I first provide a contingency table of used data type versus year of publication. The types of data used indicate a gradual shift from traditional macroeconometrics and time series analysis to microeconometric applications and panel data research. Second, I present the distribution of reviews per software category per two years. and I check which of softwares still exist in June 2008. Third, I present the yearly distribution of software use over the 25 specifically mentioned softwares.

In the observation period the JAE has reviewed the usefulness of 77 different packages for applied econometrics research and education. Surprisingly, only a handful of these products have been discontinued before June 2008 and a large majority have recently been updated. Trends in general and individual applied econometric software development emerge from the corresponding tables. In recent years the range of effective specific softwares in applied econometric research has increased. GAUSS, Stata and MATLAB dominate. Freely downloadable alternatives like R and Ox have not had a similar impact yet.

Econometric programs like LIMDEP, SHAZAM, TSP, RATS and Ox are also used for scientific research outside applied econometrics, not only in traditionally related areas like econometric theory, applied statistics and applied economics but also in marketing, finance, management science, accounting, regional science and transportation science. For example, Micah Altman and McDonald (2001) survey the use of softwares in Political Science, including many econometrics packages. My analysis is therefore admittedly very focussed. Many interesting applied econometrics articles have been published outside the JAE, but data on software use and development for other journals are not easy to obtain and results are therefore difficult to check.

This chapter implicitly defines applied econometrics as the econometrics that leads to publication in the JAE. Cleaning and preparing complicated empirical data sets, writing code for advanced estimation procedures or new types of inference, and presenting and interpreting results for JAE articles involves expert knowledge that distinguishes applied econometrics both from applied economics and from econometric theory.

The remainder of this chapter is organised as follows. The JAE research articles, data archive, software reviews, and software use are discussed in sections 2, 3, 4 and 5, respectively. The most intensively used high level programming languages are treated in more detail in section 6.

A deeper understanding of the tables is obtained by a selective description of the history and characteristics of the packages, given in section 7. This part draws heavily on Ooms and Doornik (2006) and on the extensive account of Renfro (2004b), who corresponded with many econometric software developers, preparing his article and in editing Renfro (2004a). It also reflects my experience as editor of the Econometric Software Links of the Econometrics Journal at www.econometriclinks.com. Section 8 discusses the combination of softwares and the concluding section 9 looks into future aspects of econometric modelling software.

2 JAE Research Articles

The Journal of Applied Econometrics is an important source of information on trends in software development. The founding editor, Hashem Pesaran, has been based at the Cambridge (UK) Department of Applied Economics (DAE) for most of the time since 1986. Richard Stone, the founder of the DAE wrote the first JAE article, his Nobel prize lecture on national accounts, Stone (1986). Stone's methods still underpin the basic data source for applied macroeconometric research today. Whereas Stone pioneered mainframe econometric software development in Cambridge, Pesaran was one of the first to produce user-friendly software for the PC, Data-Fit and Microfit as reviewed in Ericsson (1988). He initiated the software review section and a replication section for the JAE. He has written influential publications in theoretical and applied time series econometrics, and in theoretical and applied microeconometrics for cross section and panel data.

The JAE publishes applied econometric research in all important areas in the field. Special issues of the journal indicate the wide range of topics and methods: time series and cross section model specification as in McAleer (1989) and Magnus and Morgan (1997), event counts as in Trivedi (1997), nonlinear dynamics as in Pesaran and Potter (1992), simulation based inference (frequentist and Bayesian) as in Brown, Monfort, and van Dijk (1993), macro time series as in Pagan (1994), Diebold and Watson (1996), Hendry and Pesaran (2001) and Franses, van Dijk, and van Dijk (2005), microeconometric structural dynamics as in Kapteyn, Kieffer, and Rust (1995) and Christensen, Gupta, and Rust (2004), semiparametric microeconometrics as in Horowitz, Lee, Melenberg, and van Soest (1998), statistical decision making (Bayesian and frequentist, macro, micro and finance) as in Geweke, Rust, and Van Dijk (2000), financial time series analysis as in (Franses and McAleer (2002)), social and spatial interactions as in Durlauf and Moffitt (2003) and finally empirical industrial organisation as in Bauwens, Escribano, and Lubrano (2007).

TABLE 1 AROUND HERE.

The JAE co-editors have worked at both sides of the Atlantic and Pacific and represent the major fields and schools of applied econometrics. Table 1 also illustrates this point. It shows the frequency distributions of the dataset types, over three main categories, panel data, time series data and cross section data for the years 1995-2008, with only 4 issues of 2008 covered. The basic source of these counts were the JAE authors' readme files on the JAE Data archive. If these were unclear I checked the corresponding articles on the JSTOR archive and on Wiley Interscience. The gradual shift from traditional macroeconometrics and time series analysis to microeconometric applications and panel data research emerges. Time series articles are overrepresented in the years with corresponding special issues: 1996, 2001 and 2005. Four articles are based on simulated (Monte Carlo) data, reflecting the research interest of James MacKinnon. Two articles use data from economic experiments and four from auctions, new fields for serious applied econometrics. One article uses cross section metadata in a traditional way and Baltagi (1999) uses bibliographical panel metadata to construct rankings of authors and departments in applied econometrics. Finally, Meddahi (2002) is the only pure econometric theory JAE article I have come across. In computing the total number of research articles I included articles from the JAE's replication section, edited by Badi Baltagi.

3 JAE Software Reviews

The JAE software reviews have been edited by Pravin Trivedi (88-92) and James MacKinnon. The reviews vary greatly in length. Most reviews concentrate on one package, others compare up to six different packages on many features (data management, model formulation, simulation, availability of procedures, speed, help-functions and documentation) as in Brillet (1989), Cribari-Neto (1997). Other reviews compare specific functions like Survival modelling as in Goldstein, Anderson, Ash, Craig, Harrington, and Pagano (1989), GARCH modelling as in Brooks, Burke, and Persand (2003), or properties like numerical reliability as in McCullough (1999).

TABLES 2, 3, 4 AROUND HERE.

Many packages have been reviewed only once, but dedicated widely used (inside and outside the JAE) econometric packages show up several times in these 20 years. Table 2 details the reviews of dedicated econometric softwares since 1987, split in two-year periods to show the distributions over time for each package. Repeated reviews of the same product occur because the package receives a major update (in the beginning of its life) or because it is interesting, important and accessible enough to include in a comparison. JSTOR provides extensive bibliographical information on archived JAE articles in data base entries like 'Reviewed work(s)', but so far, this information is inaccurate and incomplete for the JAE reviews, so the numbers in Table 2 are based on the full text of the 92 articles.

I also checked the latest update and version number to make the table interesting as a reference for the state of relevant softwares in June 2008. I was first surprised to find recent updates for most of the packages. This may have been caused by the introduction of Windows Vista and Excel 2007, which made updating necessary for users who are not able to choose between operating systems. Table 2 also shows the current software companies and main author names. This entry is not relevant for the modern freely downloadable 'team' softwares and therefore missing. The last column gives the country code of the workplace of the company and main authors. Most companies and software developers work in the US. Some are in the UK and nearly all others are in Canada and mainland Europe. None are in South America and Asia, although econometrics is now a well established field of (social) science in those continents. Irregular updates of Internet links to the packages will be provided on the Econometric Software links of the Econometrics Journal at www.econometriclinks.com.

The popular econometrics package Eviews (formerly: Micro-TSP) has been discussed most often. LIMDEP, SHAZAM and PcGive and Microfit received most attention in the 20th century. Gretl is the latest general econometrics package to appear on the JAE pages and S-PLUS-FinMetrics is the latest time series econometrics package that has been reviewed. Three reviewed econometrics packages have been discontinued, or at least I could no longer trace them on the Internet: ESP, PERM and SIMPC. All other packages have been updated since the first review. I include three unreviewed packages in the list. TSM for GAUSS was mentioned in the code archive. Dynare, by Michel Juillard, is widely used in modern applied macroeconomics. Juillard (1996) is often cited. The 2008 version is available as a standalone program, but also in the form of GAUSS, MATLAB and Scilab packages. JMulti is a teaching package for Multivariate time series analysis, see Lütkepohl and Krätzig (2004). It previously required GAUSS to run. Markus Krätzig developed a graphical user interface (GUI), JStatCom, see also Table 3 and Krätzig (2006). Using this GUI and GRTE (the GAUSS RunTime Engine) JMulti is now also available as a free standalone package.

Table 3 shows the corresponding review counts of programs and two packages for Bayesian econometrics (Micro-EBA and BACC) specific panel data econometrics (Frontier, DPD and ExPEnd), and the econometric programming language Ox, econometric programming, scientific word processing and mathematics and computer science. I added the DPD package for Dynamic Panel Data analysis. This code by Manuel Arellano and Stephen Bond has been instrumental for the breakthrough of dynamic panel data econometrics, catering for large unbalanced panels as encountered in practical applications. The fundamental article, Arellano and Bond (1991), has the exceptional econometrics citation scores of 900+ in the ISI Web of Knowledge and 4000+ in Google Scholar. Their procedures have now been implemented in most econometric packages, both in the original time series oriented packages (PcGive) and in the original cross section packages (LIMDEP). Fortran in Table 3 and BIOGEME, Excel and SPSS in Table 4 have been included to keep consistency with Table 5 below. BIOGEME and SPSS are discussed later in this chapter. Stat/Transfer has not been reviewed, but it is referred to on the JAE Data Archive. It allows for user-friendly transfer of data sets between statistics packages, and LIMDEP, GAUSS, MATLAB and Excel.

Finally, Table 4 considers the statistical software reviews and provides summary statistics. Here I also added BUGS by Lunn, Thomas, Best, and Spiegelhalter (2000), because it is widely used in Bayesian Econometrics teaching, WinBUGS is a popular version for Windows. The preferred version is called OpenBUGS. The summary shows that 77 different packages have been reviewed 128 times in 92 articles. Thirteen packages have not been reviewed. The number of 'reviews' in the table equals or exceeds the number of 'articles' by definition. As explained above, a large difference between these two numbers indicates the discussion of several packages in single articles. This phenomenon occurred in 1989-1990 when many PC packages for econometrics became fit for review and in 1999-2000 when the first "GNUwares" came into use among econometricians.

4 JAE Data and Code Archive and reproducibility

The Data (and code) Archive of JAE, `www.econ.queensu.ca/jae/`, consistently coordinated by James MacKinnon, contains detailed references of all articles published since 1995. Most authors (85%) have complied with the policy to provide the data in a well documented human-readable format, fit for different operating systems and econometric softwares and usable for many years to come. This is a high success rate, compared to journals in economics or statistics who intend to have a similar policy. Authors who do not provide data for no reason whatsoever, receive the remark: "Contrary to the policy of the Journal, the author has failed to submit the data used in this paper.".

In recent years a growing number of microeconometric data sets and even some software codes are confidential for reasons of privacy, so the overall coverage of the data archive will go down in the coming years. On the other hand, the number of articles providing details on used software and codes has been high and increasing. This is the main motivation for choosing JAE articles, data and code as the main sources of information for this chapter.

The existence of a carefully managed and indexed data and code archive is an essential prerequisite for the scientific ideal of effortless reproducibility of key results in applied econometrics. Anderson, Greene, McCullough, and Vinod (2008) set the JAE Data and Code Archive as an example. William Greene is a leading econometric software developer and Bruce McCullough and Hrishikesh Vinod are influential reviewers. They discussed the disappointing compliance rates for leading American economics journals for a recent American Economic Association meeting. The situation is hardly better for leading statistics journals, like the Journal of Business and Economic Statistics, where the latest instructions for the FTP (File Transfer Protocol) data archive are now eight years old.

Of course, thanks to automatic indexing by Google and free specific Internet aggregators of economic and econometric research (papers, articles, books, citations, data, and software) like RePEc, www.repec.org, it is relatively easy to find properly documented econometric source code outside official peer reviewed archives. Unsurprisingly, given the working environment of most econometricians, robust, high quality econometric procedures seldom come for free. Here, the situation in computer science and statistics seems to be much better as the (much larger) programmer communities are funded in a different way.

Buckheit and Donoho (1995) gave a lively discussion of the difficulties in reproducing (even) one's own computer intensive results in computer science. Koenker and Zeileis (2007) elaborate on the difficulties in reproducing exact econometric results using codes from data archives. This is a nontrivial exercise, even using the original econometric software and a similar operating system. They advocate the use of Internet based tools for subversion control (SVN) for programmer communities and recent R applications to consistently develop reproducible econometric results. Roger Koenker is the father of quantile regression in econometrics, see Koenker (2005). Achim Zeileis is a key R developer.

The good news to derive from Tables 2-4 is that it is now unlikely that the current software and code will become completely useless because of the discontinuation of products.

5 Softwares used in JAE Research Articles

Table 5 details the time-varying impact of the main softwares in applied econometrics research since 1995. The softwares are ordered by first mentioned use to get a clear picture of the growing range of products used. Up to three softwares were mentioned per article, for example S-PLUS, Fortran and Stata for a cross section study. The basic sources of the counts were the readme files on the Data archive. If these were unclear I checked the corresponding articles on the JSTOR archive and on Wiley Interscience.

TABLE 5 AROUND HERE.

The 'Range' indicates the number of different products per year, which reached a maximum of 14 in 2006. The row labelled 'Missing' counts the number of articles that don't mention specific software. This number has increased in absolute terms, but it has decreased compared with the number of (Research) 'Articles' mentioned in the bottom row. Twenty five packages have been used. I distinguish seven general econometrics packages (E), four statistical programming languages (SPL), three econometrics time series packages (ETS), two mathematical matrix programming languages (MPL), two third generation numerical programming languages (NPL), Ox as an Econometric matrix programming language (EMPL), BACC as an econometric MCMC (Markov Chain Monte Carlo) package (EMC2) and finally, SPSS and Excel.

GAUSS is number one and consistently mentioned over time. In addition, two specific GAUSS applications figure once. Stata and MATLAB have only become attractive for applied econometrics after 2000. SAS and Ox (in later years) appear regularly. RATS has been the most important econometrics package for time series applications. Fortran has been consistently more important than C. Other packages appear less than five times. R and S-PLUS both appear three times. SHAZAM, Xplore and PcGive do not reappear after 2001. The other packages cannot be written off as tools for applied econometrics software development. They may well have been used in the preparation of the articles, but the authors did not develop new programs or procedures that they would like to publish.

In sum, Table 5 shows that the programming languages GAUSS, MATLAB, Stata, and Ox are the most important tools for applied econometrics software development. GAUSS, MATLAB, Stata are apparently widely available in economics and econometrics departments, all over the world.

6 High level programming languages in Econometrics

Table 6 illustrates some characteristics of the dominating languages GAUSS, MATLAB, Stata, and Ox. The table displays very short programs that load a simple data set from a human-readable ASCII file, estimate regression coefficients using OLS and show these on screen. The examples are adapted from first lessons of course notes available on the net. The table also includes code for R (and S-PLUS) as this is an increasingly important alternative as discussed below.

TABLE 6 AROUND HERE.

The codes for the matrix programming languages GAUSS and MATLAB are very similar. Beginning is easy, because variables don't have to be declared. The 'default type' is a matrix (of double-precision floating-point numbers). Statements end with a semicolon. MATLAB uses square brackets for concatenation, GAUSS has special concatenation operators. GAUSS uses square brackets for indexing, MATLAB indexes with parentheses. Indexing in GAUSS and MATLAB starts at one. Fortunately, arguments in clear function calls are in parentheses. GAUSS provides the least squares solution for the coefficients by the 'divide symbol' `/` which looks a bit weird and mathematically incorrect, but is easy to use, MATLAB uses the more sensible `\` operator instead. Neither GAUSS nor MATLAB use a formal print function to show the regression coefficients.

The Stata code is totally different and is reminiscent of many command line driven packages in the early 1980s. Stata is, as Baum (2003) put it, "on the middle ground" between econometric packages and matrix languages. The default regression method requires variable names (of columns of data set, rather than a matrix) to read the data. OLS is the default estimator of the easy-to-read-and-remember `regress` command which also adds a constant term and computes standard errors and p-values by default. The `matrix` command extracts the regression coefficients in vector format. The mathematical structure is hidden from the programmer. The standard output of `regress` (not shown in the table) is in the ANOVA format, rather than the standard regression output of econometrics programs.

Like Stata, R starts with a data set, rather than a matrix. In the R example we assume that the variable names are on the first line of the data file, so that 'header=T(rue)'. OLS is performed using a challenging call of `lm()` (linear model). This function creates an model object, and the corresponding function `coefficients` extracts the coefficient estimates from the model. The model is specified with the names of the variables and the data set. The operator `~` separates regressand and regressor, the operator `+` separates the regressors. The 'dollar' operator makes sure we use the coefficients from the linear model object.

Ox has a syntax similar to C++ (and Java), so all statements are executed in a `main()` function, square bracket pairs index a matrix, and indexing starts at zero. Doornik (2003) discusses differences and similarities of C++ and Ox. Variables have to be declared, but they automatically get a type (double, matrix) when they are assigned. Ox uses the same matrix concatenation operator as GAUSS. A dedicated least squares function `olsc()` is provided, but the (slower and less robust) explicit matrix formula for OLS could have been used instead. The ampersand (reference) is used to deliver the coefficients directly at the memory address of the vector variable `b`. This reduces memory use. I added a statement for computing fitted values $\widehat{y}$ to clarify the matrix programming nature of Ox. I use a slightly adapted Hungarian notation, where vector names start with `v` and matrix names with `m`, but this is not required. The Object Oriented (OO) nature of Ox does not appear in this example, but a Modelbase class, derived from a Database class is standard. Both classes are extendible. Unlike R and S-Plus, Ox does not force the OO features upon novice users.

Table 6 probably shows that Stata code is the most easy to use for students and researchers with limited programming experience. GAUSS and MATLAB require knowledge of matrix algebra and numerical programming, but this should not be a problem for econometricians. R is harder to get into as it requires a profound knowledge of statistical terminology and object oriented programming. Ox is easy to learn if a basic programming language and matrix algebra are known.

Readability and complexity are not the only selection criteria for high level programming languages. Large models require an extendible modelling language (like Stata and R) and new models require an efficient programming language in which to code new algorithms to estimate and evaluate new model types (like GAUSS and Ox). The programming language should also cater for effective data management, robust optimisation methods, state-of-the-art stochastic simulation, decent easily adaptable graphical and textual output facilities.

For maintenance and reproducibility one requires explicit documentation facilities, that can transform comments in the code into context-sensitive, clearly structured and indexed help functions for existing and new procedures. One should be able to integrate existing numerical procedures from low level languages. For business use, the econometric programming language should be applicable as an engine within other software, so that econometric procedures can be called by and feed results to programs like Excel, Access, or commercial front-office and back-office applications written in lower level languages like C++ or Perl. In computing intensive simulation based methods, one wants to automatically optimise code for parallel computing or for specific hardwares at a low level to increase speed. Multiprocessor computers are now standard. Efficient computing will probably return as a very important issue as electricity prices go through the roof.

The econometric language should have an interface to the Structured Query Language (SQL), a standard language that provides an interface to many relational database systems, and to specific economic, financial, and energy data management software, like FAME, www.fame.com, or HAVER, www.haver.com. For example, none of the above mentioned languages can be used to to manipulate and select data from the vast datasets on the WRDS (Wharton Research Data Service), which is now the leading academic archive for econometric time series data. Only SAS, C and Fortran can be used on the WRDS server.

The next section discusses other important aspects of a large array of packages and adds a historical perspective, concentrating on the period since 1980. This discussion should help the reader in interpreting the preceding tables on the historical impact of the different products.

7 Historical development of econometric softwares

Over the last fifty years, econometric software development has developed from writing complicated sets of computer specific instructions into coding in structured purpose built programming languages and into interactive GUI based model development. Increased backward compatibility, cross-platform and cross-operating system applicability of new software and low cost of maintaining existing software has increased the lifetime of packages and procedures. Less than 10% of the 77 packages reviewed in the JAE has been discontinued.

Econometric software development started around 55 years ago. Renfro (2004b) gives a detailed account of the history of econometric software development in the English speaking world. Early econometric software development was labour intensive and served only a few institutions that could manage and pay the substantial capital input for the required programmable computers. Moreover, software was very computer specific and served only a few institutions that could manage and pay the substantial capital input for the required programmable computers. Today, this situation has completely changed. Modern econometric software is written by a few individuals and thousands of users perform econometric estimations, forecasts and tests on thousands of machines. The joint cost of standard econometric software and hardware is low and dropping. Thanks to a concentration in hardware and software development, a few developers now serve an entire community. However, expert support and tailored innovative development of user-friendly platform-independent applications is still expensive.

Three structural changes affected econometric software development in a major way in the period 1985-2008. The first was the breakthrough in hardware development: the onset and subsequent quick improvement in computer power and graphical displays of personal computers (PC or Micro computer) since the 1980s opened opportunities for new developers. Many textbook authors wrote their own packages. Cheap standard storing devices for the PC (floppy disks) made distribution (and copying) of econometric software easy. This change is reflected in the large number of different softwares reviewed in 1990 as detailed in the summary statistics of Table 4.

The second change was the introduction and standardisation of effective graphical user interfaces (GUI) for data analysis, programming and operating systems. Graphing became easy and it was no longer necessary to memorise a list of basic commands and options.

The third change was the development and widespread use of Internet since the 1990s, more specifically the WWW standard and the later development of powerful search engines like Google. This led to the development of "free" products in mathematics, statistics and computer science. These products have now become powerful, stable and easier-to-use so that they are effectively applied in econometric software development and in innovative research in econometrics, leading to JAE publications.

The interfaces of many computer programs for data input, programming, text processing, formula and graph editing become more and more similar, due to the worldwide concentration in operating systems and standardisation of other scientific applications like LaTeX. Only three operating systems remain important, MS-Windows (Microsoft), Mac OS X (Apple) and Linux (Many distributions), where Ubuntu/Linux is the most popular version of late. Products developed on one platform can be ported to or recompiled on other platforms, although this is far from trivial for most econometricians. Racine (2000) discusses some aspects of Cygwin ports of basic Unix tools to Windows.

Hendry and Doornik (2000) discuss and illustrate the necessary changes of the time-series econometrics program PcGive in 1980s and 1990s: from command interaction to menu interaction and IDE (Integrated Development Environment), from text menus to mouse-pointer driven drop down menus and dialogs of a WIMP (Windows, Icons, Menus, Pointing) graphical user interface (GUI), from black and white text graphs to coloured bitmap to high quality, adjustable publication ready figures, from a static manual to a context-sensitive help system, from static presentation to live presentations of simulation exercises, from basically one program code in Fortran, and later in C++, to a modular object oriented architecture allowing user built extensions with an up-to-date user interface with the same look and feel as the standard applications. PcGive was extended with an independent Windows interface, GiveWin. Doornik (1998) also developed the object oriented econometric matrix programming language Ox, which allowed independent development of new packages and which was later integrated within OxMetrics, Doornik (2007) together with PcGive and the time series programs STAMP and G@RCH. The new interface for OxMetrics was built with the free cross-platform GUI wxWidgets. Other softwares have provided similar updates in order to keep old users and get new customers. For example, Stata introduced objected oriented features and GUI programming in Stata 8 and the matrix language Mata in Stata 9.

In the remaining subsections I make a distinction between five admittedly overlapping categories of software, macro-econometrics software, (pure) time-series-econometrics software, micro-econometrics software, statistical software for econometrics and mathematical software for econometrics. I treat them in turn.

7.1 Macro-econometrics software

Back in the 1960s Robert Hall laid the foundations of TSP (Time Series Processor) software. At the end of the 1970s TSP already had many of the characteristics of a modern econometric software package, it read and wrote a variety of data formats, it included a matrix language, it made use of symbolic differentiation, it contained good nonlinear solvers, a powerful optimiser and simulation procedures. In this sense TSP can be considered as the most original econometric software on the market.

In the PC era of the 1980s, TSP was split into two separate programs, Micro-TSP, headed by David Lilien and PC-TSP, headed by Bronwyn Hall. Micro-TSP later became the Windows-program Eviews, Econometric Views, whereas PC-TSP is now simply called TSP, see Hall and Cummins (2005) and Eviews (2004). TSP retained the numerical and algebraic programming features. Eviews later introduced its own object oriented programming language. One of the main attractions of Micro-TSP and Eviews was the timely interface for the first univariate econometric time series models: ARCH and GARCH. This user-friendly implementation of Generalised Autoregressive Conditional Heteroskedasticity models was developed in close cooperation with Robert Engle, the father of ARCH. A special issue of the JAE, Franses and McAleer (2002), was published to celebrate Engle's seminal ARCH article, Engle (1982).

Ken White started the package SHAZAM at Wisconsin and is now at UBC in Vancouver, where SHAZAM is now updated by a small team. Whistler, White, Wong, and Bates (2004) describe the latest version. Nobel Laureate Lawrence Klein founded the Wharton Econometric Forecasting Association (WEFA) a U. Penn, WEFA is now part of Global Insight and markets the econometric software AREMOS, which was strongly influenced by Klein's modelling methodology. AREMOS is not frequently updated, but it is still being used.

In the UK, at the Department of Applied Economics of the University of Cambridge, Hashem and Bahram Pesaran used their expertise in econometric estimation and testing for the development of Data-FIT, later called Microfit, for the PC. At the department of Statistics at the London School of Economics, econometric software development was inspired by the hands-on tradition of Denis Sargan. David Hendry, a student and later a colleague of Sargan, developed the programs AUTOREG and GIVE. In Oxford Hendry developed PCGIVE (Generalised Instrumental Variable Estimator) and PCFIML (Full Information Maximum Likelihood) on the IBM PC. Jurgen Doornik modernised and extended PcGive as explained in the first part of this section.

More recently Michel Juillard developed a standalone version of Dynare, previously only available for GAUSS and MATLAB. Dynare implements modern small-scale, but very computer intensive DSGE (Dynamic Stochastic General Equilibrium) modelling. These highly nonlinear structural models are difficult to solve and estimate and require Bayesian econometric techniques to do inference. DSGE models are introduced and used at central banks throughout the world.

On the educational side of the spectrum, Gretl, by Allin Cottrell and Ricardo Lucchetti is an international GNU (GNU's Not Unix: a free open source Unix-like operating system) econometrics program, with menus in French, Italian, Spanish, Polish and German as well as English. It is based on code for a textbook by Ramu Ramanathan. As in other packages mentioned in this, the traditional macroeconometric procedures are being supplemented with microeconometric functions, dynamic panel data (DPD) procedures in particular.

7.2 Time-series-econometrics software

One can no longer imagine applied econometrics without implementations of ARMA (Autoregressive Moving Average), VAR (Vector Autoregression) and GARCH (Generalized Autoregressive Conditional Heteroskedasticity) time series models. The Box-Jenkins methodology is a standard procedure in many fields of science. Under the direction George Box, the first special software for

ARMA analysis was written by David Pack. David Reilly turned this into AutoBox. He also coded the Multivariate Time Series (VARMA) program MTS. AutoBox and MTS are now marketed by Reilly's company AFS.

Chris Sims developed SPECTRE at the end of the 1970s. This was one of the first econometrics programs offering spectral analysis. Subsequently, Chris Sims's Vector AutoRegressive modelling (VAR) methodology of Sims (1980) was made available in RATS (Regression Analysis of Time Series) by Thomas Doan, see Doan (2004). CATS in RATS (shortly after PcGive) was one of the first widely available softwares for Søren Johansen's likelihood based analysis of the concept of cointegration, eventually published as Johansen (1995).

The Census Bureau in Washington DC produced the first reliable software for seasonal adjustment of economic time series, Census X-11, implementing a methodology (updated to X-12 ARIMA) that is now an international standard and available in most time series econometrics softwares, see Ladiray and Quenneville (2001).

At the London School of Economics, Andrew Harvey initiated the development of STAMP, for structural time series modelling, implementing an econometric methodology which serves both as an alternative to Box-Jenkins forecasting models and as an alternative to Census X-11 seasonal adjustment. Siem Jan Koopman now develops the (multivariate) STAMP software at the VU University in Amsterdam, see Koopman, Harvey, Doornik, and Shephard (2007).

At the Bank of Spain, Victor Gómez and Augusín Maravall developed the second influential alternative software for seasonal adjustment: TRAMO/SEATS. Their procedures are also available in many time series programs.

Herman Bierens is the independent author of EasyReg International, a free software package (developed in visual Basic), primarily developed for econometrics education but equipped with many advanced procedures in Bierens' area of research (nonparametric methods, first for time series and later for cross sections), and therefore also featuring in a recent JAE research article.

7.3 Micro-econometrics software

This subsection is short as there is only one surviving dedicated econometric software for non-standard econometric models for cross section data, LIMDEP. Micro-econometricians have mainly been using lower level programming languages and statistical packages, discussed below.

William Greene based the first versions of LIMDEP for LIMited DEPendent variable econometrics on code for multinomial logit models by Marc Nerlove and James Press at the University of Wisconsin. Greene (2007) describes the current features of the program. Previous versions of Greene's influential and popular textbook, now in its sixth version, Greene (2008), contained a special student edition, EA/LIMDEP of the software. Over the last 20 years most standard econometric procedures (time series and panel data) have been added. Greene also authored the packages ET and NLOGIT. Greene is now at New York University.

7.4 Statistical software for econometrics

In the last twenty five years several statistical programs have become more geared towards econometrics and subsequently widely used by econometricians. The general statistics package SAS, SAS (2004), has a long tradition (starting in the 1960s) of implementing macroeconometric and microeconometric procedures for large data sets. In academic research and education in econometrics, SAS/ETS has lost ground from its strong position at the end of the 1980s, though its econometrics features are still being developed, recently in state space procedures, in generalised maximum entropy estimation and in automatic model selection for forecasting. Of course, SAS is widely used in official institutions and in business applications, but few modern econometrics textbooks continue to use SAS examples.

SPSS, dating back to the 1970s, is not particularly suited for econometrics, but it is used for handling large and complicated data sets. Interesting third party packages for SPSS exist, like Jeroen Vermunt's LATENT GOLD for Latent Class models and event history modelling in marketing and social sciences. It is also suitable for modern microeconometrics problems (as other packages which were primarily developed for the social sciences).

The beginning of the PC era saw the birth of the 'Data Analysis and Statistical Software' Stata. Stata, by William Gould, was not an instant success among econometricians, whereas it was for statistics in medicine. At first, it did not have extensive programming facilities and specialised in applications for survival data, see Goldstein, Anderson, Ash, Craig, Harrington, and Pagano (1989). It was not suited for dynamic econometric modelling. Peterson (1991) correctly predicted: "this shortcoming could be mitigated substantially in future versions". Later Stata introduced more programming tools and eventually a matrix language and it was completed with more and more econometric models. Stata's data management features made it well suited for the econometric analysis of complicated panel data like event histories. Time series procedures have been added. Stata is now a popular package in applied economics and econometrics and a large number of introductory econometric textbooks present examples using Stata. Kit Baum maintains a large Statistical Software Components (SSC) archive within RePEc, www.repec.org, with over 1000 free open source Stata procedures and programs for statistics, economics and econometrics. Baum (2006) also wrote an applied econometric textbook for Stata.

S-PLUS and corresponding packages cater for financial econometrics and operations research: financial time series analysis, modelling credit risks and optimising asset allocation. S-PLUS, originally a product of StatSci, founded by R. Douglas Martin in Seattle, Washington, is a commercial version of the object oriented statistical programming language S, which Martin learned at Bell Laboratories in Murray Hill, New Jersey, now Lucent technologies. The software was primarily developed for statistical data analysis of many types, see Venables and Ripley (2002), with excellent graphs. Martin added robust estimation procedures, inspired by John Tukey, inventor of the term "bit", FFT (Fast Fourier Transform) and EDA (Exploratory Data Analysis). The current owner of S-PLUS, Insightful, focuses on data mining and risk management. Zivot and Wang (2005), also in Seattle, Washington, develop the S-PLUS FinMetrics software for financial econometric time series analysis. The package also includes financial engineering procedures developed by Carmona (2004) and efficient Kalman filters state space procedures by Siem Jan Koopman, see Koopman, Shephard, and Doornik (1999). The popular financial time-series textbook by Tsay (2005) makes intensive use of S-PLUS FinMetrics.

The sudden popularity of the Internet motivated the start of the statistical software Xplore in the later 1990s. There was great optimism about online cooperative development and use of software for advanced statistical computations. Härdle and Horowitz (2000) envisaged that the establishment of well documented method archives, central common platform independent compilers and new web user interfaces would give easy access to the most advanced nonparametric methods. One of their suggested 'Method and Data technology centres' was created and a (Java based) web interface, Xplore Quantlet Client (XQC), was realised. Online electronic books with econometric and financial time series applications were provided educational purposes. Online web based econometric computing has not caught on yet. Xplore is now freely downloadable from www.xplore-stat.de.

In recent years, Michel Bierlaire has developed BIOGEME, an open source package (in C++ and Python) for modern random coefficient (or mixed) discrete choice modelling. He cooperates with Moshe Ben-Akiva and Nobel Laureate Daniel McFadden. Train (2003) treats this important topic in a textbook.

Young and old econometricians are switching from S-PLUS and other packages to the freely-available statistical system R, an open source statistical system that was initiated by statisticians Ross Ihaka and Robert Gentleman from Auckland, New Zealand. R has the S syntax (and is also known as GNU S). Graphs in R are provided via Gnuplot (which is also used in SHAZAM, discussed above and TSMod, discussed below). R is part of the free GNU operating system (OS) and is part of all standard installations of this OS and therefore of many Linux installations. Officially, Gnuplot does not belong to GNU. Over 1200 packages are available for R at the CRAN (Comprehensive R Archive Network) on www.r-project.org. Cribari-Neto and Zarkos (1999) reviewed an early version of R from an econometric research point of view and Racine and Hyndman (2002) took a teaching perspective. Shumway and Stoffer (2006) provided up-to-date R code for their time series textbook. Rossi, Allenby, and McCulloch (2005) developed an R package (`bayesm`) for their marketing statistics textbook. Li and Racine (2007) wrote the `np` package for a text in nonparametric econometrics. Modern statistical methods are often made available in R. For example, Hastie, Tibshirani, and Friedman (2001) discuss their well known automatic model selection methods for regression and classification implemented in R.

Most R developers seem to work under the Linux OS and choose short Unix-style package names. Many R packages are not difficult to use under Windows and Mac OS. Developing R packages under MS Windows has not been too easy though, as Rossi (2006) reports in his 15 page tutorial on this topic: "There is a sense in which the Windows R environment is a house of cards that must be carefully assembled or it won't work!" A specialised archive of R for econometrics does not exist. A comprehensive package for financial engineering, www.rmetrics.org, which encompasses many econometric time series functions, has been built by Diethelm Würtz at the ETH in Zürich.

7.5 Mathematical software for econometrics

The beginning of the PC-era also witnessed the start of the matrix programming language GAUSS developed by Lee Edlefsen and Sam Jones in Washington State. GAUSS did not offer a new econometric methodology, but it did have a very appealing combination of price and features for econometricians and economists, see GAUSS (2005). It soon became popular and has remained popular ever since. A simple language with short matrix expressions as illustrated in Table 6, decent graphs, fast numerical algorithms, tools to handle large data sets with limited memory and a wide range of free and powerful packages implementing econometric applications for cross section models and time series. Schoenberg (1997), affiliated with Washington University, developed early procedures for constrained Maximum Likelihood for GAUSS, which found widespread application in the estimation of GARCH models. Ron Schoenberg also wrote FANPAC, a financial time series analysis package with early applications of multivariate GARCH models.

The matrix programming language and signal processing tools of MATLAB, MATLAB (2004), of the Mathworks, founded by Clive Moler, are used by many econometricians to implement model solvers and estimation methods. Econometricians use the free an comprehensive archive of econometric tools, spatial-econometrics.com, administered by James P. LeSage at the university of Toledo, Ohio. Although the archive is set up for spatial econometrics procedures, LeSage and Pace (2004), it contains many "estimation functions that provide printed and graphical output similar to that found in RATS, SAS or TSP".

Table 3 lists seven other mathematical programming languages which have not been used for JAE research articles so far, but code for these languages is provided by prominent econometricians. For example, Scilab code can be obtained for Dynare. Christopher Sims serves recent Octave code for solving rational expectations models on his own (Ubuntu/Linux) web server: http://sims.princeton.edu. Octave is a free alternative for MATLAB, but Sims points out that procedures with the same names can have different effects in the two languages.

Computer algebra packages like Mathematica are Maple are now also used for fast numerical computations, and are therefore more suited for applied econometrics, but they haven't had a big impact yet. The recently developed package MathStatica for Mathematica, by Colin Rose and Murray Smith, can save applied econometricians work in the analytical derivations of complicated likelihoods.

8 Simultaneous use of different softwares

As the tables and the discussion in the previous sections illustrate, many econometric techniques can now be implemented using existing mathematical and statistical software packages. No single software can serve all purposes, which explains why more and more packages coexist and why many researchers use several products next to each other.

Thanks to the search engine Google and free specific Internet aggregators of economic and econometric research (papers, articles, books, citations, data, and software) like RePEc at www.repec.org, it is now easy to find properly documented econometric source code written for one of the main econometric softwares on the web. However, it is still difficult to assess the quality of this code if one does have access to the software for which it was originally developed. As most of these codes for academic research papers are available free of charge, authors cannot be expected to set up a helpdesk, and one has to resort to mailing lists and Internet forums, which also may be unreliable. Unsurprisingly, given the background of most econometricians, robust, high quality econometric procedures seldom come for free.

The modular structure of econometric and statistical software makes it possible to use codes outside their original environment. This helps the reproducibility required in academic econometrics. For example, Laurent and Urbain (2003) provide an interface called `M@ximize` for Ox, based on OxGauss, so that the wide range of econometric GAUSS programs available on the net can be run without a licence for GAUSS or Constrained Maximum Likelihood for GAUSS. Markus Krätzig developed a graphical user interface (GUI) for econometric modelling, JStatCom, see Krätzig (2006), which he built on top of GAUSS code and the GAUSS run time engine (GRTE) to create JMulti as a standalone program. JStatCom can also be used in combination with MATLAB and Ox. John Breslaw of Econotron software introduced Symbolic Tools which extends GAUSS and the GRTE with infinite precision computer algebra of Maple. Cameron Rookley wrote the free GTOML (GAUSStoMATLAB) scripts which translate GAUSS code into MATLAB. This requires the free powerful OO programming language Perl, see www.perl.com and www.cameronrookley.com.

Diethelm Würtz, author of Rmetrics, provided an interface in R for the G@RCH package that Laurent and Peters (2005) developed for Ox, but this still requires the availability of Ox. Many statistical packages have been ported to R, for example BRugs, which embeds OpenBUGS in R. Robert Henson (2004) introduced a MATLAB R-link with functions for calling R from within MATLAB, Bengtsson (2005) increased the communication possibilities between MATLAB and R.

Integrating codes from different applications can save time, but has its dangers. Evaluation and improvement of existing implementations for nontrivial procedures should be a constant concern, see e.g. the discussion of numerical precision of econometric packages by McCullough and Vinod (1999), which generated a series of changes in testing procedures. Note also the evaluation of random number generators (RNGs) as in McCullough (2006) and Doornik (2006). Reliability of RNGs is now extremely important as simulation based inference starts to dominate both macroeconometrics and microeconometrics. Even if the RNG is right, and expert econometric knowledge is available, there is plenty of room for undetected mistakes. The home page of the BUGS project (Bayesian inference Using Gibbs Sampling) phrases this at follows: "Independent corroboration of MCMC results is always valuable!" "MCMC is inherently less robust than analytic statistical methods. There is no in-built protection against misuse.". Even before econometric modelling starts one should apply Hendry (1980)'s "three golden rules of econometrics: test, test and test" to the freshly developed or imported software.

9 New Econometric Modelling features

Pagan and Wickens (1989) surveyed applied econometric methods twenty years ago. Four estimation methods were discussed: maximum likelihood, GMM (Generalized Method of Moments), M-estimators and non-parametric estimation and different types of inference: frequentist and Bayesian, large sample asymptotics and the bootstrap for tests in small samples. They concluded: "... when it comes to an area such as econometrics. Gone are the days when a single individual could have a detailed knowledge of all divisions of the subject. Just twenty years ago this might have been possible." and "the years since then have witnessed a fragmentation of econometrics. The biggest division has been between micro and macro econometrics." As indicated in section 2 many new data types, estimators, inference methods and diagnostic procedures have been analysed by applied econometricians since 1989. The fragmentation now also applies to the software development with dozens of procedures published on the net for the same purpose.

Although applied nonparametric econometrics has been on the rise, model based econometrics still dominates the field of applied econometrics. A key aspect that distinguishes model based econometric software is the standard availability of features for the interactive modelling cycle: models are not only easily specified and estimated, but diagnostic tests, easy respecification, and re-estimation facilities are provided in order to make the interpretation of parameter estimates and forecasts as credible as possible. Today, this requires a graphical (WIMP) interface that is sufficiently intuitive and easy to learn and remember for new users.

This recursive modelling is especially relevant for the econometric analysis of time series, where new observations become available in a natural order, with associated testing possibilities and possible adaptations of existing models. In the context of dynamic linear regression models PcGive was the first program to cater for the influential general-to-specific methodology of econometric model selection. A 'Progress' menu in PcGive simplifies the interactive model selection process. Although this feature per se has not been copied in other packages, a wide range of standard specification tests and diagnostics for estimated models has now become a crucial ingredient of every econometric software.

The model selection process can be automated. Successful automated model selection has long been available for pure Box-Jenkins time series modelling for forecasting in the AutoBox software by David Reilly and in the Census X-11-ARIMA program for seasonal adjustment of the US Census. Automated linear dynamic model selection for economic analysis, based on a wide range of robust diagnostic tests and multiple-path general-to-specific modelling is available in the PcGive procedure Autometrics, Doornik (2008).

However, also automated model selection methods, even if they encompass generalised linear models of 'Statistical Learning' as in Hastie, Tibshirani, and Friedman (2001), or fractional instead of zero-one model weights of Bayesian Model Averaging (BMA), as in Raftery, Madigan, and Hoeting (1997), still require a 'most general' adequately specified model, for which extensive tests should be available.

Stochastic simulation and bootstrap analysis of econometric models should be available as a matter of course, both for the interpretation of nonlinear models, and for associated statistical inference. James Davidson's (nonlinear) time series modelling package TSMod, reviewed by Fuertes, Izzeldin, and Murphy (2005) has this feature for all models in the package: "Bootstrap p-values for diagnostic and significance tests, using the simulation module to generate bootstrap draws." If the inference is simulation based, one also needs diagnostics on the efficacy and reliability of the associated simulation methods.

User interfaces will have to be updated. Following Google and Gretl, users will expect econometric software to deal with labels and numbers in their native language and with application menus in using their own character sets. The graphical interface will also need reconstruction as customers adapt to modern graphical interfaces. New interfaces will help to make better use of the many options that programs and procedures have, both on the user's own computer and on Internet archives. Many procedures are ineffective because they are hard to find in the current menu structures. Based on a user history the menus will 'automatically' select the best options for the user.

The market for specific econometric software is too small for one program to keep up with all recent scientific developments in econometrics, mathematics and statistics, to keep advanced knowledgeable customers interested in buying updates, and to implement lessons from Human-Computer Interaction (HCI) research to keep attracting new customers.

The presence of trends implies some predictability of future developments. The pattern that has emerged in the last twenty five years does not make it likely that new fully-fledged dedicated econometric software packages with high academic standards are going to be developed. Academic returns on high quality, robust, versatile, and well documented and supported econometric software development are low. Changing citation practices for software use, as exemplified by the JAE Data and code archive may increase these returns in the years ahead.

In this chapter I discuss over twenty years of changing software use and software development for innovative applied econometrics. An increasing range of softwares has become relevant in this period. I classify this large collection of programs and I assess the continuity of their use. Finally, I point out new directions for econometric modelling software development.

References

Anderson, R., W. Greene, B. D. McCullough, and H. D. Vinod (2008). The role of data/code archives in the future of economic research. *Journal of Economic Methodology 15*(1), 99–119.

Arellano, M. and S. Bond (1991, April). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*(2), 277–97.

Baltagi, B. H. (1999). Applied econometrics rankings: 1989-1995. *Journal of Applied Econometrics 14*(4), 423–441.

Baum, C. (2003). Facilitating applied economic research with Stata. In S. B. Nielsen (Ed.), *Programming Languages and Systems in Computational Economics and Finance*, pp. 173–179. Kluwer Academic Publishers.

Baum, C. F. (2006). *An Introduction to Modern Econometrics Using Stata*. College Station, TX, USA: Stata Press.

Bauwens, L., A. Escribano, and M. Lubrano (2007). The econometrics of industrial organization. *Journal of Applied Econometrics 22*(7), 1153–1156.

Bengtsson, H. (2005). R.matlab - local and remote Matlab connectivity in R. Preprint in Mathematical Sciences (manuscript in progress), Mathematical Statistics, Centre for Mathematical Sciences, Lund University, Sweden. [manuscript in progress].

Brillet, J.-L. (1989). Econometric modelling on microcomputers: A review of major software packages. *Journal of Applied Econometrics 4*(1), 73–92.

Brooks, C., S. P. Burke, and G. Persand (2003). Multivariate GARCH models: Software choice and estimation issues. *Journal of Applied Econometrics 18*, 725–734.

Brown, B. W., A. Monfort, and H. K. van Dijk (1993). Introduction to special issue on Econometric Inference using Simulation Techniques. *Journal of Applied Econometrics 8*, S1–S3.

Buckheit, J. and D. L. Donoho (1995). WaveLab and reproducible research. In A. Antoniadis and G. Oppenheim (Eds.), *Wavelets in Statistics*. New York, NY, USA: Springer Verlag.

Carmona, R. A. (2004). *Statistical Analysis of Financial Data in S-Plus*. New York: Springer-Verlag.

Christensen, B. J., N. D. Gupta, and J. Rust (2004). Introduction: Special issue on the econometrics of social insurance. *Journal of Applied Econometrics 19*(6), 647–648.

Cribari-Neto, F. (1997). Econometric programming environments: GAUSS, ox and S-PLUS. *Journal of Applied Econometrics 12*(1), 77–89.

Cribari-Neto, F. and S. Zarkos (1999). R: Yet another econometric programming environment. *Journal of Applied Econometrics 14*, 319–329.

Diebold, F. X. and M. W. Watson (1996). Introduction: Econometric Forecasting. *Journal of Applied Econometrics 11*(5), 453–454.

Doan, T. A. (2004). *User's Manual RATS, Version 5*. Evanston, IL, USA, www.estima.com: Estima.

Doornik, J. A. (1998). *Object-oriented Matrix Programming using Ox*. London, U.K.: Timberlake Consultants Ltd, www.oxmetrics.com.

Doornik, J. A. (2003). Object Oriented Programming in econometrics and statistics using Ox, a comparison with C++, Java and C# In S. B. Nielsen (Ed.), *Programming Languages and Systems in Computational Economics and Finance*, Dordrecht, The Netherlands, pp. 173–179. Kluwer Academic Publishers.

Doornik, J. A. (2006). The role of simulation in econometrics. In T. C. Mills and K. Patterson (Eds.), *Palgrave Handbook of Econometrics, Vol 1: Econometric Theory*, pp. 787–811. Palgrave MacMillan.

Doornik, J. A. (2007). Conversion of high-period random numbers to floating point. *ACM Transactions on Modeling and Computer Simulation (ACM TOMACS) 17*, Article 3 (January 2007).

Doornik, J. A. (2008). Autometrics. In J. L. Castle and N. Shephard (Eds.), *Festschrift in Honour of David F. Hendry*, pp. forthcoming. Oxford, U.K.: Oxford University Press.

Durlauf, S. N. and R. A. Moffitt (2003). Introduction: Special issue on empirical analysis of social interactions. *Journal of Applied Econometrics 18*(5), 499–499.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1007.

Ericsson, N. R. (1988). A review of Data-FIT: An interactive econometric modelling package for IBM-Compatible PCs. *Journal of Applied Econometrics 3*(4), 319–332.

Eviews (2004). *Eviews 5 User's Guide*. Irvine, CA. USA, www.eviews.com: Quantitative Micro Software.

Franses, P. H. and M. McAleer (2002). Financial volatility: An introduction. *Journal of Applied Econometrics 17*(5), 419–424.

Franses, P. H., H. K. van Dijk, and D. van Dijk (2005). On the dynamics of business cycle analysis: editors' introduction. *Journal of Applied Econometrics 20*(2), 147–150.

Fuertes, A.-M., M. Izzeldin, and A. Murphy (2005). A guided tour of TSMod 4.03. *Journal of Applied Econometrics 20*(5), 691–698.

GAUSS (2005). *GAUSS 7.0 User's Guide*. Maple Valley, WA, USA.: Aptech systems Inc. /Trafford Publishing, www.aptech.com.

Geweke, J., J. Rust, and H. K. Van Dijk (2000). Introduction: : Inference and Decision Making. *Journal of Applied Econometrics 15*(6), 545–546.

Goldstein, R., J. Anderson, A. Ash, B. Craig, D. Harrington, and M. Pagano (1989). Survival analysis software on MS/PC-DOS computers. *Journal of Applied Econometrics 4*(4), 393–414.

Greene, W. H. (2007). *LIMDEP 9.0 Econometric Modeling Guide*. New York, NY, USA, www.limdep.com: Econometric Software Inc.

Greene, W. H. (2008). *Econometric Analysis, Sixth Edition*. Prentice-Hall, Upper Saddle River, NJ, USA.

Hall, B. H. and C. Cummins (2005). *TSP 5.0 User's Guide*. Palo Alto, CA, USA, www.tspintl.com: TSP International.

Härdle, W. and J. Horowitz (2000). Internet-based econometric computing. *Journal of Econometrics 95*, 333–345.

Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning, Second Edition*. Springer.

Hendry, D. F. (1980). Econometrics-alchemy or science? *Economica 47*, 386–406.

Hendry, D. F. and J. A. Doornik (2000). The impact of computational tools on time-series econometrics. In T. Coppock (Ed.), *Information Technology and Scholarship Applications in the Humanities and Social Sciences*, pp. 257–269. Oxford: Oxford University Press/British Academy.

Hendry, D. F. and M. H. Pesaran (2001). Introduction to special issue in memory of John Denis Sargan, 1924-1996: Studies in empirical macroeconometrics. *Journal of Applied Econometrics 16*(3), 197–202.

Henson, R. (2004). *MATLAB R-Link*. IN: MATLAB Central, www.mathworks.com/matlabcentral.

Horowitz, J., M.-J. Lee, B. Melenberg, and A. van Soest (1998). Introduction: Application for Semiparametric Methods for Micro-Data. *Journal of Applied Econometrics 13*(5), 431–433.

Johansen, S. (1995). Identifying restrictions of linear equations with applications to simultaneous equations and cointegration. *Journal of Econometrics 69*, 111–132.

Juillard, M. (1996). Dynare : a program for the resolution and simulation of dynamic models with forward variables through the use of a relaxation algorithm. Technical report.

Kapteyn, A., N. Kieffer, and J. Rust (1995). Introduction: The Microeconometrics of Dynamic Decision Making. *Journal of Applied Econometrics 10*, S1–S7.

Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press.

Koenker, R. and A. Zeileis (2007). Reproducible econometric research (A critical review of the state of the art). Technical Report 60, Department of Statistics and Mathematics Wirtschaftsuniversität Wien.

Koopman, S. J., A. C. Harvey, J. A. Doornik, and N. Shephard (2007). *STAMP 8, Structural Time Series Analyser, Modeller and Predictor*. London, U.K.: Timberlake Consultants Press.

Koopman, S. J., N. Shephard, and J. A. Doornik (1999). Statistical algorithms for models in state space using SsfPack 2.2. *The Econometrics Journal 2*, 107–160, www.ssfpack.com.

Krätzig, M. (2006). A software framework for data analysis. *Computational Statistics and Data Analysis 53*, 618–634.

Ladiray, D. and B. Quenneville (2001). *Seasonal Adjustment with the X-11 Method*. New York: Springer-Verlag.

Laurent, S. and J.-P. Peters (2005). *G@RCH 4.0, Estimating and Forecasting ARCH Models*. London, UK.: Timberlake Consultants Press.

Laurent, S. and J.-P. Urbain (2003). Bridging the gap between Ox and Gauss using OxGauss, paper presented at the first Oxmetrics user conference, London. Technical report, Center for Econometrics and Operations Research, Louvain-la-Neuve, Belgium.

LeSage, J. P. and R. K. Pace (Eds.) (2004). *Advances in Econometrics, Volume 18: Spatial and Spatiotemporal Econometrics*, Oxford, UK. Elsevier.

Li, Q. and J. S. Racine (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter (2000). WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing 10*, 325–337.

Lütkepohl, H. and M. Krätzig (2004). The software JMulti. In H. Lütkepohl and M. Krätzig (Eds.), *Applied Time Series Econometrics*, pp. Chapter 8. Cambridge, UK, www.jmulti.de: Cambridge University Press.

Magnus, J. R. and M. S. Morgan (1997). Design of the experiment. *Journal of Applied Econometrics 12*(5), 459–465.

MATLAB (2004). *MATLAB 7.1*. Boston, MA, USA: The Mathworks Inc. www.mathworks.com.

McAleer, M. (1989). Introduction to special issue on Topics in Applied Econometrics. *Journal of Applied Econometrics 4*, S1–S3.

McCullough, B. D. (1999). Econometric software reliability: EViews, LIMDEP, SHAZAM and TSP. *Journal of Applied Econometrics 14*(2), 191–202.

McCullough, B. D. (2006). A review of TESTU01. *Journal of Applied Econometrics 21*(5), 677–682.

McCullough, B. D. and H. D. Vinod (1999). The numerical reliability of econometric software. *Journal of Economic Literature 37*, 633–665.

Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics 17*(5), 479–508.

Micah Altman, M. and M. McDonald (2001). Choosing reliable statistical software. *PS: Political Science & Politics 34*, 681–687.

Ooms, M. and J. A. Doornik (2006). Econometric software development: Past, present and future. *Statistica Neerlandica 60*, 206–224.

Pagan, A. (1994). Introduction Calibration and Econometric Research: An Overview. *Journal of Applied Econometrics 9*, S1–S10.

Pagan, A. R. and M. R. Wickens (1989, December). A survey of some recent econometric methods. *Economic Journal 99*, 962–1025.

Pesaran, M. H. and M. P. Potter (1992). Nonlinear Dynamics and Econometrics: An Introduction. *Journal of Applied Econometrics 7*, S1–S7.

Peterson, S. P. (1991). A review of Stata 2.1. *Journal of Applied Econometrics 6*(2), 207–212.

Racine, J. (2000). Review: The Cygwin tools: A GNU toolkit for Windows. *Journal of Applied Econometrics 15*(3), 331–341.

Racine, J. C. and R. Hyndman (2002). Using R to teach econometrics. *Journal of Applied Econometrics 17*, 175–189.

Raftery, A. E., D. Madigan, and J. A. Hoeting (1997). Bayesian Model Averaging for linear regression models. *Journal of the American Statistical Association 92*, 179–191.

Renfro, C. G. (Ed.) (2004a). *Computational Econometrics: Its Impact on the Development of Quantitative Economics*, Amsterdam, The Netherlands, www.iospress.com. IOS Press.

Renfro, C. G. (2004b). Econometric software: The first fifty years in perspective. *Journal of Economic and Social Measurement 29*, 9–107.

Rossi, P. E. (2006). Making R packages under Windows, a tutorial. Technical Report URL: http://faculty.chicagogsb.edu/peter.rossi/, Graduate School of Business U of Chicago IL USA.

Rossi, P. E., G. Allenby, and R. McCulloch (2005). *Bayesian Statistics and Marketing*. New York: John Wiley.

SAS (2004). *SAS/ETS (Econometrics and Time Series 9.1 User's Guide*. Cary, NC, USA, www.sas.com: SAS Publishing.

Schoenberg, R. (1997). Constrained maximum likelihood. *Computational Economics 10*, 251–266.

Shumway, R. H. and D. S. Stoffer (2006). *Time Series Analysis and Its Applications: With R Examples, 2nd edition*. Springer.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica 48*(1), 1–48.

Stone, R. (1986). Nobel Memorial Lecture 1984: The Accounts of Society. *Journal of Applied Econometrics 1*(1), 5–28.

Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Trivedi, P. K. (1997). Introduction: Econometric models of event counts. *Journal of Applied Econometrics 12*(3), 199–201.

Tsay, R. S. (2005). *Analysis of Financial Time Series, 2nd Edition*. New York: John Wiley & Sons.

Venables, W. and B. Ripley (2002). *Modern Applied Statistics with S, Fourth Edition*. New York etc.: Springer Verlag.

Whistler, D. K., K. J. White, S. D. Wong, and D. Bates (2004). *SHAZAM Version 10 User's Reference Manual*. Vancouver, Canada, www.econometrics.com: Northwest Econometrics.

Zivot, E. and J. Wang (2005). *Modeling Financial Time Series with S-PLUS, Second Edition*. New York etc.: Springer Verlag, New York etc.

Table 1: Research articles in JAE per data type per year

	95	96	97	98	99	00	01	02	03	04	05	06	07	08	total
panel data	9	5	7	8	5	8	4	4	15	21	9	17	21	7	140
time series	16	22	13	14	18	15	24	18	14	16	30	27	18	8	253
cross section	8	4	10	8	4	7	5	5	3	9	6	14	16	6	105
simulated	.	1	.	.	.	.	1	.	1	.	.	1	.	.	4
experiment					1	.	.	1	.	.	.	.	.	.	2
metadata					1	.	.	.	.	.	.	.	1	.	2
auction					1	.	.	.	1	.	.	2	.	.	4
scanner						1	.	.	.	.	.	.	1	.	2
algebra								1	.	.	.	.	.	.	1
total	33	32	30	30	30	31	34	29	34	46	45	61	57	21	513

NOTES: panel data: data with a small time series dimension and a large cross section dimension, time series: data with large time series dimension, larger than cross section dimension, cross section: cross section data without time series dimension, experiment: data from experimental economics, simulated: data from random number generator (RNG) and known data generating process (DGP), metadata: data summarising results from other articles, auction: empirical data from auctions.

Sources: Data archive JAE, www.econ.queensu.ca/jae/, JSTOR: www.jstor.org, www3.interscience.wiley.com, ISSN code JAE: 08837252.

Table 2: Softwares and Reviews JAE per software per two years, part 1: Econometric softwares

Software (old)-package	Type	88	90	92	94	96	98	00	02	04	06	08	tot	V.08	Y.	Company/Author(s)	Country
AREMOS	E	.	1	.	.	.	.	.	.	.	.	.	1	5.3	03	Global Insight	US
EasyReg	E	.	.	.	.	.	1	.	.	.	.	.	1	2007	07	H. Bierens	US
ESP	E	.	1	.	.	.	.	.	.	.	.	.	1	†	92	J.P. Cooper, O.A. Curtis	US
Eviews (MicroTSP)	E	.	1	1	.	1	.	3	1	1	1	1	10	6	07	QMS, D. Lilien	US
GAUSS - GAUSSX	E	.	.	.	.	1	.	.	.	.	.	.	1	9	08	Econotron Soft, J. Breslaw	CA
Gretl	E	.	.	.	.	.	.	.	.	1	1	1	3	1.7.5	08	A. Cottrell, R. Lucchetti	US, IT
LIMDEP	E	1	1	1	1	.	.	2	.	1	.	.	7	9	07	W. Greene	US
Microfit (DataFit)	E	2	.	.	1	.	1	.	.	.	.	.	3	4	98	B. Pesaran, M.H. Pesaran	UK
MODLER	E	.	1	.	.	.	.	.	.	.	.	.	1	10.7	08	C. Renfro	US
PCBRAP	E	.	.	1	.	.	.	.	.	.	.	.	1		02	A. Zellner	US
OxMetrics-PcGive(PcFiml)	E	1	.	1	.	2	1	.	.	1	1	.	7	12	07	J.A. Doornik, D.F. Hendry	UK
PERM	E	.	.	1	.	.	.	.	.	.	.	.	1	†	94		US
SHAZAM	E	2	.	1	.	1	.	1	.	.	.	.	5	10	08	D. Whistler, K. White	US
SORITEC	E	.	1	.	.	.	.	.	1	.	.	.	2		98	J. Sneed	US
TSP	E	.	.	.	.	1	1	1	.	.	.	.	3	5	08	B. Hall, C. Cummins	US
Autobox	ETS	1	.	.	.	.	.	.	.	.	.	.	1	6	07	Automatic Forecasting Sys.	US
Dynare	ETS	.	.	.	.	.	.	.	.	.	.	.	.	4	08	M. Juillard	FR
Forecast Master	ETS	.	1	.	.	.	.	.	.	.	.	.	1		98	Scientific Systems Company	US
GAUSS - COINT	ETS	.	.	.	.	1	.	.	.	.	.	.	1	2	94	S. Ouliaris, P.C.B. Phillips	US
GAUSS - FANPAC	ETS	.	.	.	.	.	.	1	.	1	.	.	2	2	02	R. Schoenberg	US
GAUSS- GRTE - JMulti	ETS	.	.	.	.	.	.	.	.	.	.	.	.	4.2.1	08	M. Krätzig, H. Lütkepohl	DE,IT
GAUSS - TSM	ETS	.	.	.	.	.	.	.	.	.	.	.	.	1	07	Aptech, N. Lohonen	US
MATLAB - BDS	ETS	.	.	.	.	.	.	.	1	.	.	.	1		99	L. Kanzler	DE
MTS	ETS	.	1	.	.	.	.	.	.	.	.	.	1	1	96	Automatic Forecasting Sys.	US
Ox - TSMod	ETS	.	.	.	.	.	.	.	.	.	1	.	1	4.26	08	J. Davidson	UK
RATS	ETS	.	1	.	.	.	1	.	.	1	.	.	3	7	07	Estima, T. Doan	US
RATS -CATS	ETS	.	.	.	.	.	1	.	.	.	.	.	1	2	06	H. Hansen, K. Juselius	DK
SAS - ETS	ETS	.	1	.	.	1	.	.	.	1	.	.	3	9.1	07	SAS Institute	US
SIMPC	ETS	.	.	.	.	1	.	.	.	.	.	.	1	†	94	H. Don	NL
S-PLUS- FinMetrics	ETS	.	.	.	.	.	.	.	.	2	.	.	2	3	07	E. Zivot, J. Wang	US
STAMP	ETS	.	1	.	.	.	.	.	.	.	1	.	2	8	07	S.J. Koopman, A.C. Harvey	NL,UK
TSW (TRAMO/SEATS)	ETS	.	.	.	.	.	.	.	1	.	.	.	1	1	08	G. Caporello, A. Maravall	ES
X-12 ARIMA (X-11)	ETS	.	.	.	.	.	.	.	.	.	.	.	.	0.3	07	U.S. Census, B. Monsell	US

NOTES: See also Tables 3 and 4. Softwares in alphabetical order within type categories. Review counts per two year periods, 88: 1987-1988,. . ., 08: 2007-2008. Old names in parentheses. tot: total number of reviews per software, . for unreviewed softwares, †: discontinued. V.08: Last version in June 2008, Y.: year of last update, Company/Author(s): Name of producing company/author(s). Not all authors mentioned. Type descriptions as E : Econometrics package, ETS : Econometrics Time Series package.

Table 3: Softwares and Reviews JAE per software per two years, part 2: Various packages

Software - Package	Type	88	90	92	94	96	98	00	02	04	06	08	Tot	V.08	Y.	Author	Country
GAUSS - Micro-EBA	ECS	.	.	1	.	.	.	.	.	.	.	.	1	†		J. Fowles	US
PcGive PcNaive	EMC	.	.	.	1	.	.	.	.	.	.	.	1	5	08	J.A. Doornik, D.F. Hendry	UK
BACC	EMC2	.	.	.	.	.	.	1	.	.	.	.	1	2003	03	J. Geweke, W. McCausland	US
Ox	EMPL	.	.	.	.	.	1	.	.	.	.	.	1	5.1	08	OxMetrics, J.A. Doornik	UK
OxGauss-M@ximize	EMPL	.	.	.	.	.	.	.	.	.	1	.	1	1.0	03	S. Laurent, J.P. Urbain	BE,NL
Fortran - FRONTIER	EPD	.	.	.	.	.	.	1	.	.	.	.	1	4.1	03	T. Coelli	AU
GAUSS - DPD	EPD	.	.	.	.	.	.	.	.	.	.	.	.	98	98	M. Arellano, S. Bond	ES,UK
GAUSS - ExpEnd	EPD	.	.	.	.	.	.	.	.	1	.	.	1	1	02	F. Windmeijer	UK
Maple	MCA	.	.	.	.	1	.	.	.	.	.	.	1	12	08	Maplesoft	CA
Mathematica	MCA	.	.	.	.	.	.	.	.	1	.	.	1	5.2	08	Wolfram Research	US
GNUPlot	G	.	.	.	.	.	.	.	.	.	1	.	1	4.3	08	GNUplot team	
JStatCom	GUI	.	.	.	.	.	.	.	.	.	.	.	.	2.4	08	M. Krätzig	DE
wxWidgets	GUI	.	.	.	.	.	.	.	.	.	.	.	.	2.8	08	wxWidgets project, J. Smart	
GAUSS	MPL	.	.	1	1	.	1	1	.	.	.	.	4	9	07	Aptech	US
MATLAB	MPL	.	.	.	1	.	1	.	1	.	.	.	3	7.6	08	Mathworks	US
Maxima	MPL	.	.	.	.	.	.	.	.	.	.	1	1	5.15	08	Maxima team	US
NAG	MPL	.	1	.	.	.	.	.	.	.	.	.	1	Mark 21	07	NAG Group Ltd.	UK
C++ - Newmat	MPL	.	.	.	.	1	.	.	.	.	.	.	1	10	06	R. Davies	US
Octave	MPL	.	.	.	.	.	.	2	1	.	.	.	3	3.01	08	J. Eaton	US
Scilab	MPL	.	.	.	.	.	.	1	1	.	.	.	2	4.1	07	Scilab Consortium	FR
Yorick	MPL	.	.	.	.	.	.	1	.	.	.	.	1	2.1	08	D. Munro	US
LaTeX	MWPL	.	.	.	.	.	.	.	1	.	.	.	1	2.7	07	C. Schenk	DE
MPI LAM	NMT	.	.	.	.	.	.	.	1	.	.	.	1	7.3	07	LAM/MPI team	
ParallelKnoppix	NMT	.	.	.	.	.	.	.	.	.	.	1	1	2.9	08	M. Creel	ES
C++	NPL	.	.	.	.	1	.	.	.	.	.	.	1	4.2	08	i.a. GNUcc team	
Fortran	NPL	.	.	.	.	.	.	.	.	.	.	.	.	95/2003	07	i.a. NAG fortran	UK
Debian-GNU/Linux	OS+	.	.	.	.	.	.	1	.	.	.	1	2	4.0	08	Debian team	
Cygwin	OST	.	.	.	.	.	.	1	.	.	.	.	1	1.5	08	Cygwin team	
Perl	TNPL	.	.	.	.	.	.	.	.	1	.	.	1	5.10	08	Perl team	
Python	TNPL	.	.	.	.	.	.	1	.	.	.	.	1	2.5	08	Python Team	

NOTES: See also Tables 2 and 4. Type descriptions as ECS : Econometrics Cross Section package, EMC2: Econometrics Bayesian Markov Chain Monte Carlo package, EMPL: Econometrics Matrix Programming Language, EPD : Econometrics Panel Data package, MCA: Mathematics Computer Algebra package, G: Graphics package, GUI: Graphical User Interface, MPL: Mathematical Matrix Programming Language, MWPL: Mathematical Word Processing Language, NMT: Numerical Tool (parallel computing), OS+: Operating System plus applications, OST: Operating System cross-over package, TNPL: Text processing and numerical programming language.

Table 4: Softwares and Reviews JAE per software per two years, part 3: Statistical packages

Software - package	Type	88	90	92	94	96	98	00	02	04	06	08	tot	V.08	Y.	Author	Country
SYSTAT	S	.	1	.	.	.	.	.	.	.	.	.	1	12	08	Systat Soft, L. Wilkinson	US
Math.- MathStatica	SCA	.	.	.	.	.	.	.	.	1	.	.	1	1.5	06	C. Rose, M.D. Smith,	AU
BMDP	SCS	.	1	.	.	.	.	.	.	.	.	.	1	2007	07	W. Dixon, M.B. Brown	US
GAIM	SCS	.	.	.	1	.	.	.	.	.	.	.	1	†	.	T. Almudevar, R. Tibshirani	US
NCSS	SCS	.	1	.	.	.	.	.	.	.	.	.	1	2007	07	NCSS	US
N-KERNEL	SCS	.	1	.	.	.	.	.	.	.	.	.	1	†			US
Stat/Transfer	SDT	.	.	.	.	.	.	.	.	.	.	.	.	9	07	Circle Systems Inc.	US
Excel	SG	.	.	.	.	.	.	.	.	.	.	.	.	2007	07	Microsoft	US
STATGRAPHICS	SG	.	.	.	1	.	.	.	.	.	.	.	1	XV.II	07	StatPoint Inc.	US
ViSta	SG	.	.	.	.	.	.	.	1	.	.	.	1	7.9	07	P.M. Valero-Mora, M. Friendly	CH,CA
TESTU01	SMC	.	.	.	.	.	.	.	.	.	1	.	1	1.2.1	08	R. Simard	CA
BUGS -Open BUGS	SMC2	.	.	.	.	.	.	.	.	.	.	.	.	1.4.3	07	D. Lunn, A. Thomas	UK, FI
C++ -BIOGEME	SPD	.	.	.	.	.	.	.	.	.	.	.	.	1.6	08	M. Bierlaire	CH
R	SPL	.	.	.	.	.	.	2	1	.	1	.	4	2.7	08	R team	
SC	SPL	.	.	.	1	.	.	.	.	.	.	.	1	2.03	05	T. Dusoir	FR
S-PLUS	SPL	.	.	.	1	.	1	.	.	.	.	.	2	8	07	Insightful Corp.	US
Stata	SPL	.	1	1	1	.	.	.	1	.	.	.	4	10	07	StataCorp, W. Gould	US
SST	SPL	1	.	1	.	.	.	.	.	.	.	.	2	3	04	J. Dubin	US
Xplore	SPL	.	1	.	.	.	1	.	.	.	.	.	2	4.7	07	MD*Tech, W. Härdle	DE
LISREL	SSS	.	.	.	.	.	.	.	.	1	.	.	1	8.8	06	SSI International	US
SPSS	SSS	.	.	.	.	.	.	.	.	.	.	.	.	16	07	SPSS Inc.	CA
R ts	STS	.	.	.	.	.	.	.	1	.	.	.	1	0.15	08	A. Trapletti	AT
Total reviews		7	18	10	10	11	11	20	13	14	9	5	128				
Range reviews		6	18	10	10	11	11	15	13	13	9	5	77				
Total articles		7	9	10	9	8	9	11	9	9	6	5	92				
Not reviewed													13				

NOTES: Softwares: in alphabetical order within type categories. Review counts per two year periods, 88: 1987-1988,..., 08: 2007-2008, Old names in parentheses. tot: total number of reviews per software, . for unreviewed softwares, V.08: Last version in June 2008, Y.: year of last update, Company/Author(s): Name of producing company/author(s). Not all authors mentioned. Type descriptions as S: Statistics package, SCA: Statistics Computer Algebra package, SCS : Statistical Cross Section package, SG: Statistical Graphics package, SDT: Statistical Data Transfer package, SPD : Statistical Panel Data package, SMC: Statistical Simulation (Random Number Generator testing) package SMC2: SMC2: Econometrics Bayesian Markov Chain Monte Carlo package, SPL: Statistical Programming Language. Reviews in JAE often discuss multiple packages, so that Total reviews (Tables 2, 3 and 4 exceeds Total articles. Sources: JAE in JSTOR: www.jstor.org and in www3.interscience.wiley.com, ISSN code JAE: 08837252. URLs for softwares available via the econometric links of The Econometrics Journal, www.econometriclinks.om

Table 5: Research articles in JAE with specific softwares per software per year

Software	Type	Year														
		95	96	97	98	99	00	01	02	03	04	05	06	07	08	tot
TSP	E	1	1	.	.	.	1	.	1	.	.	.	.	.	.	4
GAUSS	MPL	6	1	2	3	2	3	.	2	4	5	5	8	10	7	58
RATS	ETS	2	.	.	.	.	.	1	.	.	1	.	2	.	.	6
SAS	SPL	2	.	1	3	.	1	1	.	.	1	1	1	2	.	13
Fortran	NPL	2	2	1	2	5	.	2	.	.	1	.	1	.	.	16
C	NPL			1	.	.	.	.	.	.	1	.	.	.	.	2
LIMDEP	E				1	.	1	.	.	.	.	.	1	.	.	3
S-PLUS	SPL				1	.	.	.	.	.	.	.	2	1	.	4
MATLAB	MPL					2	.	.	2	.	.	2	5	5	1	17
SHAZAM	E						1	.	.	.	.	.	.	.	.	1
Stata	SPL						1	.	.	.	7	1	4	2	6	21
SPSS	SSS						1	.	.	.	1	.	1	.	.	3
Ox	EMPL						2	.	2	2	1	2	2	1	.	12
GAUSSX	E							1	.	.	.	.	.	.	.	1
Xplore	SPL							1	.	.	.	.	.	.	.	1
PcGive	E							2	.	.	.	.	.	.	.	2
STAMP	ETS							1	.	.	.	1	.	.	.	2
Ox G@RCH	ETS									1	.	.	1	.	.	2
Excel	SG										1	.	.	.	.	1
Eviews	E											2	1	.	.	3
R	SPL											2	.	1	.	3
BACC	EMC2												1	.	.	1
GAUSS TSM	ETS												1	.	.	1
EasyReg	E													1	.	1
BIOGEME	SPD													1	.	1
Range		5	3	4	5	3	8	7	4	3	9	8	14	9	3	25
Missing		21	28	25	21	21	22	26	22	28	30	32	37	37	8	358
Articles		33	32	30	30	30	31	34	29	34	46	45	61	57	21	513

Software ordered by time of first mentioned use in JAE article (1995.1-2008.4). Counts of research articles using the specific software. Range: number of different softwares mentioned per year. Missing: research articles NOT mentioning specific software. Articles: number of research articles per year. Type descriptions as E : Econometrics package, ECS : Econometrics Cross Section package, EMC2: Econometrics Bayesian Markov Chain Monte Carlo package, EMPL: Econometrics Matrix Programming Language, ETS : Econometrics Time Series package, NPL : Numerical Programming Language (3rd generation), SG : Statistical Graphics package, MPL : Matrix Programming Language, SPL : Statistical Programming Language, SSS : Statistical package for Social Sciences, SPD : Statistical Panel Data package Range: Sources: Data archive JAE, www.econ.queensu.ca/jae/, JSTOR: www.jstor.org, www3.interscience.wiley.com, ISSN code JAE: 08837252.

Table 6: Ordinary Least Squares in programming languages for applied econometrics

Software	type	how to	statements / commands
GAUSS	MPL	begin, declarations	not necessary
		read data	`load myX[4,3]=yX.asc;`
		select y	`y=myX[.,1] ;`
		add constant to X	`X=ones(rows(myX),1)~myX[.,2:3];`
		get b	`b=y/X;`
		show b	`print b;`
MATLAB	MPL	begin declarations	not necessary
		read data	`myX=load('-ascii','yX.asc');`
		select y	`y=myX(:,1);`
		add constant to X	`X=[ones(size(myX,1), myX) myX(:,2:3)];`
		get b	`b=X\y;`
		show b	`b`
Stata	SPL	begin, declarations	not necessary
		read data	`infile y x1 x2 using yX.mat`
		add constant to X	added by default
		get b	`regress y x1 x2`
		show b	`matrix list e(b)`
R and S-PLUS	SPL	begin, declarations	not necessary
		read data	`myX<-read.table("yX.dat", header=T)`
		add constant to X	added by default
		get b	`b=lm(y1~x1+x2,data=myX)$coefficients`
		show b	`b`
		extra line in `yX.dat`	`y1 x1 x2 \\ variable names`
Ox	EMPL	begin main program	`main()`
			`{`
		declarations	`decl myX, mX, vy, vb, vyhat;`
		read data	`myX=loadmat("yX.mat");`
		select y	`vy=myX[][0];`
		add constant to X	`mX=1~myX[][1:2];`
		get b	`olsc(vy, mX, &vb);`
		fit y: $\widehat{y} = Xb$	`vyhat=mX*vb;`
		show b	`println("b: ",vb);`
		end	`}`
		extra line in `yX.mat`	`4 3 \\ dimensions`

NOTES: MPL: Matrix programming language, SPL: Statistical programming language, EMPL: Econometric matrix programming language. Computation 3×1 regression coefficient vector b in linear model $y = X\beta + u$, where X is a 4×3 matrix starting with a column of ones. y and X are read from file. myX: $4\times$ 3 matrix with numerical data read from human readable data file yX.asc. For R we read yX.dat and for Ox we read yX.mat, starting with an extra line as indicated.