



REVIEW OF STATA 7

STANISLAV KOLENIKOV*

Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260, USA and Centre for Economic and Financial Research, Moscow, Russia

1. GENERAL INTRODUCTION

Stata 7 is a general-purpose statistical package that does all of the textbook statistical analyses and has a number of procedures found only in highly specialized software.¹ Unlike most commercial packages aimed at making it possible for any Windows user to produce smart-looking graphs and tables, Stata is aimed primarily at researchers who understand the statistical tools they are using. The applications of statistics that are covered best by Stata are econometrics, social sciences, and biostatistics. The Stata tools for the latter two categories are contingency tables, stratified and clustered survey data analysis (useful also for health and labour economists), and survival data analysis (useful also in studies of duration, such as studies of the lengths of unemployment or poverty spells).

There are several things that I like about Stata. First of all, it is a very good package for doing applied research, with lots of everyday estimation and testing techniques, as well as convenient data-handling tools. The unified syntax makes all these things easy to learn and use. It is also a rapidly developing package, with excellent module extension and file sharing capabilities built into it. Stata features a great support environment, both from the developers and from advanced users of the package. The best example of the informal support network formed around Stata is the Statistical Software Components archive (SSC-IDEAS, part of RePEc) that contains several hundred user-written programs. For many users, it is also important that the academic and student prices can be quite low.

2. SPECIFICATIONS AND INSTALLATION

Stata has long been available in two versions (Intercooled and Small) for all popular platforms: Windows 9x/NT/2000, MacOS, and many Unixes (Digital Unix, HP-UX, AIX, Linux for Intel

*Correspondence to: S. Kolenikov, Department of Statistics, 117 New West, Cameron Avenue, University of North Carolina, Chapel Hill, NC, 27599-3260, USA. E-mail: skolenik@unc.edu

¹ Stata software is developed and distributed by Stata Corporation, 4905 Lakeway Drive, College Station, Texas 77845, USA. Tel. 1-979-696-4600. Fax. 1-979-696-4601. E-mail: stata@stata.com. Corporate website: <http://www.stata.com>. Worldwide distributors list: <http://www.stata.com/info/worldwide/>

and Power Mac platforms, Solaris; only the Intercooled version is available for Unix). The Small version has various technical limitations, such as a smaller number of observations/variables and smaller memory, and it is intended to be used only for undergraduate teaching. The prices of Intercooled Stata for different categories of users may differ quite substantially, with the lowest student licence price of \$55, academic prices in the range between \$100 and \$500, depending on options and eligibility, and a commercial price approaching \$1000 (for the first copy of Stata).²

The versions of Stata for Windows, Unix, and Macintosh are identical, and cross-platform compatibility is one of the long-standing commitments of Stata Corp. It is really nice to be able to sketch your data analysis at an outdated laptop with a 1% sample of the original data and then upload your data file and the program to a multiprocessor server to run it with the full data set. The reverse side of the coin is that not too much attention is given to the drag-and-drop operations one might get used to under Windows. This especially concerns graphics. However, users can copy and paste Stata output, including graphics, to their favourite word processors.

Stata is a big system in terms of capabilities, but unlike other big systems that are distributed on several CD-ROMs and require all recent Microsoft libraries, Stata is quite tolerable on a 486DX/16MB system under Linux or Windows 95. This makes Stata great for budget applications and institutions, as well as for the Third World countries, where research institutions might not have the funds to purchase the newest hardware.

Stata is shipped on a CD-ROM with all installation files for all platforms. The licensing information is sent along with the documentation. When installed, Stata 7 occupies around 10Mb of disk space. Installation of updates and additional components would take another 2–5 megabytes. Stata datasets are organized rather efficiently: the compression ratio achieved by zip archivers is about two, as compared to 10 or 15 for Excel or SAS files.

Stata installs itself in a single directory (for example, C:\STATA or /usr/local/bin/stata) without writing much to the system directories. This, however, makes Stata rather susceptible to unlicensed copying. When Stata is installed, it requests the licensing information at the first launch. The license information is platform-specific, and it also distinguishes between single user and network Stata. As a promotional offer, there is an option for a free monthly trial license.³ The network version tracks the number of users currently logged on to Stata, although each user may run several processes.

I found a couple of faults with the installation program for Windows. First, it hung when it ran out of free space on the drive where it tried to install Stata to. Second, it did not tell me how to run Stata when it was finally installed.

Stata is a very stable program. I do not remember Stata ever crashing. If there is a run-time error, Stata displays an error message with a non-zero return code and goes back to the command prompt. Users can stop calculation by pressing Ctrl-Break in Windows or Ctrl-C in Unix. It might appear non-responding at times, but when Stata receives the Break sequence, it needs to complete the operation being performed, delete all unnecessary temporary objects, and restore the status prior to last command issued, so the response to the Break command can at times be quite slow, especially when Stata uses virtual memory under Windows.

² Prices are subject to change; see <http://www.stata.com/info/order/> for up-to-date information. The prices for multiple-user network licences are also worth checking, as they involve decreasing prices per marginal user.

³ See <http://www.stata.com/postoffice/borrow.html>

3. CONCEPTS

There are three main conceptual features underlying the operation of Stata statistical software.

First, Stata is an observation-oriented package, in the way that Gauss is a matrix language and S-Plus is an object environment. The code is optimized in such a way that all operations are implicitly performed on all observations (unless a subsample is explicitly chosen), and those operations are very fast. This is due to the fact that Stata stores the whole data set in the RAM. It can also use virtual memory, so there is no formal upper limit to the number of observations in the data set. Virtual memory works well under Unix, but using it under Windows results in continual swapping that reduces the computation speed by approximately three orders of magnitude.

Second, Stata is command prompt driven, which means that all programming capabilities are just the same as the interactive ones, and the user does not have to think much converting interactive sessions into programs. Simple Stata programs are sequences of commands one would issue with the keyboard in an interactive regime. They usually have an extension `.do` (thus called do-files in Stata jargon). Stata can be launched in batch mode, that is, ordered to execute the specified do-file, write the output into the log-file with the same name, and then exit. Another nice use of do-files is the specification of the startup profile as a command line option where the user can specify keyboard shortcuts, add paths to her programs, add menus to Stata, or request updates from Stata web site. Stata 7 also has handy ways of converting the log of the interactive work into a program.

The above does not mean, however, that Stata has no menus. It does have some basic things like file operations, preferences, window management, and help. Users can add their own menus, thus effectively converting Stata into a mouse-driven package. In fact, Stata Corp. offers StataQuest, an addition that adds a lot of menus to help navigating through numerous statistical and data management tools. It can be downloaded from the Stata website and installed from inside Stata.⁴

Third, Stata is modular by its nature, resembling Matlab in that much of Stata is written in Stata code. In fact, Stata has two types of commands. The basic data, graphics, and matrix operations, elementary statistical analysis, and the command syntax parser are implemented in the main executable file as built-in commands, and everything else, including all advanced statistical modules, are (interpretable) ASCII text file programs. The latter are referred to as ado-files (Automatic DO-files), and they are usually stored in several places under the `stata/ado` directory. The development of Stata is shifting towards increased use of ado files. Currently, of about 1100 commands in the base release (including undocumented internal ones), only about 200 are implemented as kernel commands.

The 'modularity' has several implications. Stata may be easily updated, and complicated Stata programs can be started by picking the code accumulated previously for similar problems. The modular character of the language makes it natural to write easily shareable programs, providing an invaluable externality to the research community. As a result, even though Stata is produced by a commercial firm, Stata *capabilities* are produced to a large degree by an 'Open Source'-like effort of researchers. As a part of this, several user group meetings have been held around the world (seven in the UK, two in Spain, one in Netherlands, one in the USA).

A lot of user-written extensions to Stata come from the Stata mailing list and from the SSC-IDEAS (Statistical Software Components) RePEc archive maintained by Christopher Baum, Boston College. The 'officially recognized' user-written additions are published in *Stata Technical Bulletin (STB)*, a bi-monthly publication with a format close to that of the *Stata Reference manual*. It is also

⁴ See <http://www.stata.com/support/quest/>

run by Stata users rather than Stata Corp. itself. The editor of *STB* is H. Joseph Newton (Texas A&M University). *STB* hard copy publication is accompanied by free Stata modules downloadable from the Stata website and a couple of mirrors, at Harvard and in Portugal. The concept of the *STB* was being revised towards a refereed scientific journal with more essential statistical content at the time when the review was written.

Stata has a special file description format that makes it possible to search the Internet (including the SSC-IDEAS archive) for Stata components, so that users can find and get additions by typing a few commands or clicking a couple of menu entries. The willingness of Stata Corp. to provide these facilities in their programming language must be applauded. See also Section 6 on the subject of on-line resources.

4. SYNTAX AND USAGE

Most Stata commands have the following format:

```
[by variable list:] command [variable list] [if condition] [in range] [using file name]
[[weights]], [options]
```

where the square brackets denote that the argument of a command is optional (but the brackets around weights are required if the weights are specified). The first *variable list* is the list of grouping variables (so that the following command is sequentially applied to the subsets identified by the unique values of those variables, e.g. panels). The second *variable list* is the list of variables to which the command is to be applied (e.g. the dependent variable and regressor list for a regression model). The *condition* and the *range* select the appropriate subsample (e.g. *if year < 1990*). The *file name* is used for file operations (load into memory, write to disk, import from or export to a text file). The *weights* specify the sample weights of the observations used for the command. The *options* contain whatever else Stata would need to know to run the program: the choice of estimation method (e.g., fixed or random effects for panel data, full ML or two-step for Heckman selection model), the values of numeric tuning parameters, requests for more detailed output or for an alternative estimated covariance matrix to be used, etc. Some additional effort may be required for estimation of systems of equations, especially for specifying the (cross-equation) constraints.

After installation, Stata can easily be kept up to date. Updates, which include bug fixes and/or new options for a few commands, are made available two or three times a month. The updates are made available over the web and, even better, Stata itself can find them and install them. Installation times are usually minimal, since the files transferred turn out to be small (being Stata ado-files, which are generally no more than a few kbytes). Once every two months or so, the executable itself is also updated (which Stata also handles automatically), and the size of that file is just over 1.5 Mbytes.

The issue of numerical accuracy is addressed in a Stata website publication; see <http://www.stata.com/support/cert/>. Stata passed most of the NIST StRD tests, except one for regression with highly collinear regressors. It just dropped some of the regressors, which is recognized as a legitimate response for this type of problem. The built in (pseudo-)random number generator (32-bit KISS algorithm with a period of about $2^{126} \approx 10^{38}$) was also checked with the DIEHARD test and considered adequate.

5. DATA MANAGEMENT, MATRIX OPERATIONS, AND GRAPHICS

Stata supports its own binary format of data, although it can be taught to read in a text file with almost any weird formatting one can think of, like several lines per observation. Stata offers optionally (and with an additional charge) the most recent version of Stat/Transfer along with Stata installation files, so users would have no difficulties converting their data between all popular formats supported by Stat/Transfer. Another popular data format converter that can deal with Stata files is DBMS/COPY, and a number of programs can open Stata data files directly.

Stata supports file operations over the Internet: the user can simply type the URL of a file just as she would type the local path to it. Instructors may find this especially useful. Datasets (or do-files, or anything else) can be stored on their home pages and they can tell their students to access the materials inside Stata from their home page.

Stata has versatile commands for dealing with data. It has a comprehensive set of statistical, mathematical, string and other functions, although relatively more exotic things like Bessel functions are not implemented. The operations can be restricted to certain subsamples of the data. It is also possible to combine different files, e.g. by some index variables, and doing some special transformations of the data set, like transposing, stacking variables, converting the variables from `income92, income93, . . .` \times one observation per object format ('wide' format, in Stata jargon) into `income + time variables` \times many observations per object format ('long' format).

Dealing with dates in Stata is not always trivial. The dates are stored as integers (with the value of 1 assigned to 1 January, 1960, or January 1960, or 1960, depending on the frequency of the data, as specified by the user), not as an additional special format. The user would need to convert the string dates into those integers to take advantage of lags, leads, difference operators, and special time series commands. Stata provides formats to make the integer date recording scheme readable, but if the things are not done properly, the date might appear as, say, 14900 instead of 17 Oct 2000, which would make most users feel they made an error.

While Stata's programming capabilities allow you to add new commands to Stata, they do not allow you to add new functions, such as `sqrt()`, that can be used in any context. If one wrote a module to compute Bessel functions, one could use the module to assign the values for a new variable `bx` as `egen bx=bessel(n,x)` (where `egen` is the special Stata command⁵ to deal with those user-written functions), but one could not use the modifier `if bessel(x,n)` on the end of any Stata command.

Although matrices are not the main objects of Stata, most matrix operations that make sense for computational statistics can be performed, starting from simple arithmetic operations and concatenation down to Kronecker products, sweeps, and inverse matrices. Various versions of matrix products and accumulated matrices relevant to statistical analyses are also available. All the basic computational linear algebra stuff is covered, including Cholesky decomposition, singular value decompositions, and eigenvalue problems. At the time this review was written, one could not save matrices as files for further use, but new utilities for this operation were being developed. The maximum size of the matrix can be set by the user, but it cannot exceed 800 by 800.

Graphics is one of the most controversial issues in Stata. Given the Stata Corp. commitment to provide cross-platform compatible products and formats, Stata does not provide graphs that can be treated as flexible OLE objects in Windows environment, like MS Excel graphs, let

⁵ Read: Extensions to GENerate, where `generate` is the way to create new variables in simple settings. The modules for `egen` should follow some special syntax.

alone Java-based S-Plus graphical applets. There are no built-in 3D graphs, either. Still, most 2D capabilities are sufficient to create graphs acceptable for scientific publications. Graphs can be easily converted into (Encapsulated) PostScript or Windows Metafiles formats for further use in \TeX or Word documents. Stata Corp. was developing a new graphics engine at the time this review was written.

6. DOCUMENTATION AND ON-LINE RESOURCES

There are two sorts of documentation available for Stata: built-in (on-line) help and manuals. On-line help includes all the commands in the base distribution of Stata and explanations of the basic concepts of Stata (like what can be done with the keyboard, how to use the search system, and how to update Stata over the Net). More detailed explanations are available with the manuals. Stata 7 comes with a 386-page *User's Guide*, a 225-page *Graphics Manual*, a 368-page *Programming Manual*, a 180-page *Getting Started Manual*, and a 2239-page 4 volume *Reference Manual* set. The entries of the latter include more examples than the on-line help, along with technical details such as methods and formulas. In fact, they can serve as a good introduction to the statistical concepts and topics related to the commands described, and there are 5 to 10 references to the literature in the end of each entry to suggest further reading. A cheaper *Stata Reference Manual Extract* (620 pp.) is available for budget installations and applications complying with the 80/20 Pareto rule (here, 20% of the documentation entries are sufficient for 80% of problems). There have been many user suggestions to distribute the manuals in PDF format, and there are some signs of acknowledgement of this initiative by Stata Corp.

The web site of the company (<http://www.stata.com>) provides a lot of additional information. The Capabilities section fully duplicates the on-line help. There is a small on-line bookstore where users can order not only additional sets of Stata manuals, but also a number of useful books ranging from Davidson and MacKinnon (1993), Greene (2000), and Johnston and DiNardo (1997) to Mooney and Duval (1993) and McCullagh and Nelder (1989). A number of recently published books use Stata as the primary software. These include Deaton (1997), Hamilton (1997), Hardin and Hilbe (2001), Long (1997), and Rabe-Hesketh and Everitt (2000). Users can also find errata list for the manuals, register for the on-line Stata courses, download the official updates and additions to Stata from STB, or subscribe to the Stata mailing list.

Another great on-line Stata resource is the Statistical Software Components Archive SSC-IDEAS⁶ maintained by Christopher Baum at Boston College. This archive probably contains more end-user Stata commands (available free of charge) than the Stata distribution itself! Coupled with the search and indexing opportunities of the RePEc system, SSC-IDEAS makes searching for statistical routines easy and simple.

7. SUPPORT

There are several ways to obtain support for problems related to Stata, and I have never had any problems doing so. The 'official' way to get support is to send a message to the technical support group at tech-support@stata.com. The technicians from Stata Corp. generally respond within a few hours.

⁶ <http://ideas.uqam.ca/ideas/data/bocbocode.html>

There is also a great mailing list, where both Stata developers and experienced Stata users can give advice for most Stata problems (or at least refer to the relevant software if Stata does not do the required job).⁷ In fact, the founder and president of Stata participates on the list, and it is not uncommon for him to respond.⁸ I personally benefited quite a lot from just reading the statistical discussions on the list. Many of the list participants have expressed the point of view that Stata is fast, inexpensive and handy, even though there are fields where it can be outperformed by specialty packages.

For users more used to traditional support tools, a hot line (and a toll-free 1-800 number for users in USA/Canada) is available, as well as support by fax or by mail.

8. PRINCIPAL ECONOMETRIC TOOLS

Stata has a rich set of statistical tools, of which many are used in everyday econometric practice. An example of a short Stata session, which illustrates some of these tools, may be found in the JAE Data Archive.⁹

One of the most widely used commands is `regress`. It estimates OLS regressions and presents results in the form of a regression table. There are lots of variants of regressions implemented in Stata, including regression with White and Newey–West covariance matrix estimators, robust regression with Huber and biweight loss functions, regression with instrumental variables (2SLS and 3SLS for simultaneous equations), and some other regression-based techniques. The implementation of GLS is not always trivial, however. Stata has matrix commands that allow for sandwiches with the GLS matrix in the middle, but I am not aware of a command that directly provides GLS estimation. Also worth mentioning is the versatile stepwise regression tool that searches for the ‘best’ set of regressors.

Stata has a rich set of tools for regression diagnostics, including measures of influence (studentized residuals, DFBETAs, DFFITS, COVRATIO) and overall ‘regression quality’ measures (VIFs, heteroscedasticity, non-linearity (RESET) and serial correlation (Durbin–Watson) tests), as well as a number of graphical displays that help to visualize regression results (added variable, component plus residual, residual vs. fitted values, leverage vs. squared residuals plots). Not all of them are routinely used in econometric practice despite their high practical value.

The discrete and limited dependent variable commands include most of the things one could imagine in this area, except for such rarities as ordered binary response models for panel data. Readily available are standard logit and probit models, Heckman sample selection models (with ML and two-stage options), treatment effects models, tobit regression, the bivariate probit model, ordered logit and probit models, the McFadden conditional (fixed effect) logit model, the complementary log-log (Gompertz) model, and generalized linear models.

The set of commands that deal with panel data is also exhaustive: random and fixed effect (within and between) linear regression, including an option for autocorrelated errors; GLS models with various patterns of within and across panel correlations; Prais–Winsten regression; instrumental

⁷ User opinions on Stata, as well as comparisons to other statistical packages and references to the statistical methodology discussions, are selected occasionally from the list archive and made available at <http://www.komkon.org/~tacik/stata/opinions.html>.

⁸ An actual quote from one of his letters on the list to characterize the attitude to users: ‘If we have made the wrong decision in pushing this problem to the side, we can certainly reconsider. Let me know.’

⁹ <http://www.econ.queensu.ca/jae>

variables and two-stage least squares; Arellano–Bond linear dynamic panel data estimator; fixed and random effect logit and random effect probit and gompit models; panel Poisson, negative binomial, and some other generalized linear models. Specification tests (LM test for random effect, Hausman test for fixed effect) are available. The panel data can also be converted into a special form that allows the use of the `regress` command with all its capabilities of regression diagnostics.

Time series capabilities were seriously introduced with the release of Stata 6 in January 1999 and are sufficient for most intermediate level time series analyses. The canned routines include correlograms, ACF and PACF; general ML-based ARIMA and conditional heteroscedasticity (ARCH, GARCH, asymmetric, threshold, power and non-linear ARCH) models; white noise (Bartlett and portmanteau) tests and unit root (ADF, Phillips–Perron) tests. The ARCH-type and ARIMA models allow for lists of regressors, so the interpretation of these commands by Stata is that the residual process follows ARCH-type or ARIMA functional form, with the mean process given by the list of regressors. If no regressors are specified, then of course the model is applied to that variable. Programs performing vector autoregressions are not a part of the official Stata, but rather are implemented as user-written additions available from SSC-IDEAS. Converting the data set to the format understood as time series by Stata may be somewhat cumbersome, and getting used to the date formats and conversion options will take some time.

Some users might be interested in analysing duration data like unemployment or poverty spells. Stata's tools for this type of analysis are comprehensive (Cox proportional hazard model; parametric survival models with exponential, Weibull, Gompertz, lognormal, loglogistic, gamma families; various data management and graphical routines), but converting the data with the observed times to the survival time data set would require even more effort than in the time series case.

One of the greatest advantages of Stata is its maximum likelihood estimator (Gould and Sribney, 1999). Users can write their own likelihood functions and let Stata maximize them. There are several versions of the procedure. In the simplest i.i.d. case, the user just codes the 'observation-wise' likelihood function; Stata then performs summation (excluding missing observations if needed, or accounting for weights if those are specified) and maximizes the function numerically. More difficult cases can include computation of the full likelihood function (e.g. in the panel data with intrinsic dependencies). Finally, to increase computational efficiency, the user can code analytical first and second derivatives. The maximization algorithm seems to be quite involved and includes random search at the early stages, then some unidirectional optimization, and then some combination of Newton–Raphson and steepest ascent algorithms with diagnostics of concavity.

Stata has unified syntax for hypothesis testing after all estimation commands mentioned above. It can test simple linear restrictions, combine them into F - or χ^2 -tests, or perform non-linear hypothesis testing based on the delta method.

One of Stata's main weaknesses as an econometric package is the lack of a readily available GMM estimator, even though Stata already has a very good maximizer routines implemented in the maximum likelihood estimator. On the other hand, Stata has a number of estimators built on the GMM principle, such as the Arellano–Bond estimator for dynamic panel data models.

9. GETTING TO KNOW STATA

As with all software packages, learning Stata requires some initial impetus. The very first steps in Stata can be observed with its built-in tutorials (just type `tutorial` when you launch Stata for the first time in your life). The set of tutorials is rather limited, however. There is an introduction

to Stata, and tutorials on tables, graphics, regression, ANOVA, logit, survival and factor models, and data input. Several toy data sets are supplied with Stata to illustrate linear regression analysis and diagnostics, survival analysis and ANOVA, as well as tabulation and graphic capabilities.

A good way to learn Stata is through the 'NetCourses' offered by Stata Corp. These are several-week courses taught over the Internet, with lectures, homeworks, and the opportunity to get feedback from the instructors. The Introduction to Stata, Stata Programming, and Advanced Programming courses are given twice a year. Other courses offered include Survival Analysis and ML Estimation.

Some references to Stata resources may be found in the supplementary material to this review in the JAE Data Archive.

The author's experience shows that it takes about two weeks to get acquainted with the main commands and start writing your own do-files to automate basic data analysis; then about several months plus a couple of NetCourses to get used to programming sequences of commands for your data management or estimation purposes; and then the rest of your life to get to know all the capabilities of Stata through your own programming and participation in the mailing list. This pace was mainly achieved through self-education, with Stata being launched at startup on my office PC every morning and being used for at least a couple of hours per day.

10. COMPARISON WITH PREVIOUS VERSIONS

Stata is developing at a very steep rate, and several initial comments had to be removed from this review during the weeks after it was prepared, as new capabilities were introduced. For instance, one of my initial intentions was to mention the lack of VAR routines, but that niche has been filled by the `vecar` routine by Christopher Baum, Boston College.

There have been far too many new features to mention since Stata 3.1 was reviewed by Ferrall (1994). The main new features of Stata 6 (released in early 1999) were introduction of a new set of time series concepts, operators, and routines; improvements in the ML optimizer; new programming syntax that allowed explicit parsing of the command line; Internet capabilities; and a number of new statistical routines.

In Stata 7 (released in December 2000), a new GUI was developed for the Unix version of the software, and some incompatibilities between the Windows and Unix versions were fixed. Some users found it a nuisance that Stata previously did not calculate some quantiles and/or tail probabilities of statistical distributions with double precision. This was also fixed in Stata 7. Among other new features of Stata 7 were: a set of procedures for cluster analysis, long variable names, and a new output engine (SMCL: Stata Markup and Control Language). Also, Stata seems to improve its speed from one release to another by some 5–10%.

Prospective developments that were outlined by Gould ('Report to users at the 1st North American Stata User Group Meeting', unpublished, 2001) at the North America User Group Meeting, included: LM tests for omitted variables; HTML and TeX formatted log files; new graphics; Big Stata with many restrictions weakened (most importantly, the one on the matrix size); connection of Stata over the Internet.

11. FINAL REMARKS

I strongly recommend Stata for research, especially in cross-sectional, panel, and survey data. Teaching Stata might be difficult at the first stages, as you would probably have to explain all

commands and some of their details to the students if they have no knowledge of Stata. Learning Stata is an investment, not necessarily with an immediate return, but, like all investments in human capital, it pays off handsomely in the end. In a couple of months time, you would probably do things faster and more efficiently than with a mouse-driven package.

Stata is not the best package for:

- Presentation graphics (although publication graphics is OK);
- Intricate macro and financial time series analysis;
- GMM or such relatively exotic methods as neural nets, simulated annealing, CARTs, or genetic algorithms.

However, it is certainly great at:

- Pre-programmed long sequences of simple commands (which might be developed in previous interactive work);
- Panel and survey data analysis;
- Discrete and limited dependent variable analysis;
- Survival and duration analysis;
- Maximum likelihood estimation;
- Regression analysis and regression diagnostics;
- Data management and conversion;
- Desktop scientific and statistical calculations.

ACKNOWLEDGEMENTS

I am grateful to Stata Corp. for the version 7 release for review purposes and to William Gould, Nick J. Cox, Christopher Baum, and James MacKinnon, the Software Review Editor, for their most helpful comments and suggestions.

REFERENCES

- Davidson R, MacKinnon JG. 1993. *Estimation and Inference in Econometrics*. Oxford University Press: Oxford.
- Deaton A. 1997. *The Analysis of Household Surveys*. Johns Hopkins University Press: Baltimore, MD.
- Ferrall C. 1994. A review of Stata 3.1. *Journal of Applied Econometrics* **9**: 469–477.
- Gould W, Sribney W. 1999. *Maximum Likelihood Estimation with Stata*. Stata Press: College Station, TX.
- Greene WH. 2000. *Econometric Analysis*. Prentice Hall: Englewood Cliffs, NJ.
- Hamilton L. 1997. *Statistics with Stata 5*. Duxbury Press: New York.
- Hardin J, Hilbe J. 2001. *Generalized Linear Models and Extensions*. Stata Press: College Station, TX.
- Johnston J, DiNardo J. 1997. *Econometric Methods*. McGraw-Hill: New York.
- Long SJ. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications: Beverly Hills, CA.
- McCullagh P, Nelder JA. 1989. *Generalized Linear Models*, (2nd edn). Chapman and Hall: London.
- Mooney CZ, Duval RD. 1993. *Bootstrapping: A Nonparametric Approach to Statistical Inference*. Sage Publications: Beverly Hills, CA.
- Rabe-Hesketh S, Everitt B. 2000. *A Handbook of Statistical Analyses using Stata*, (2nd edn). Chapman and Hall: London.
- Stata Corp. 2001. *Stata Statistical Software: Release 7.0*. Stata Press: College Station, TX.