The Stata Journal (2003)
3, Number 4, pp. 440–444

Review of Maximum Likelihood Estimation with Stata by Gould, Pitblado, and Sribney

Stephen P. Jenkins
Institute for Social and Economic Research
University of Essex, UK
stephenj@essex.ac.uk

Abstract. The new book by Gould, Pitblado, and Sribney (2003) is reviewed.

Keywords: gn0009, maximum likelihood, Stata programming

1 Introduction

This is the second edition of a book originally published in 1999, when Stata 6 was current. There are significant additions, not only of Jeff Pitblado to the list of authors and the use of a larger font size (which definitely assists readability). Several changes reflect the updating of Stata itself: this edition was timed to coincide with the release of Stata 8.1, which incorporated several notable extensions to the maximum likelihood command `ml`. For example, you can now fit models to complex survey data (there are `svy` options), and there are three new optimization algorithms (Berndt–Hall–Hall–Hausman, Davidon–Fletcher–Powell, and Broyden–Fletcher–Goldfarb–Shanno) in addition to the modified Newton–Raphson algorithm, and you can switch between them. Variance estimates can be calculated using the outer product of gradients (not just via the inverse Hessian). These additions are all discussed in the new edition; so, too, is the fitting of models with linear constraints using the `constraint` command. But arguably you do not need a book to take advantage of these features: there is always the online help, for example. So, what does the book actually provide?

2 What is in the book?

Like the first edition, it is essentially an instruction manual for researchers who need maximum likelihood estimators for models that are not already prepackaged in Stata. Not so many years ago, if the ML estimation command that you wanted was not in Stata, you would most likely seek the command in another package or think about writing a special program in Gauss, C, or Fortran, or simply give up. Stata's `ml` command provides another route. Use of the command does not require special programming skills, and it offers access to many features that we have come to expect of built-in Stata commands. With such power comes complexity, however, and I doubt whether many researchers could use the `ml` command effectively relying entirely on the online help and *Reference* manual entries. You have to draw on a book like this. As a nonprogrammer, I found the first edition invaluable in helping me to write a number of different ML estimation commands. That said, I also had to consult StataCorp Technical Support on each occasion to complete my programs. The book made things substantially easier but not easy; that is the nature of the subject. The same can be said for this edition.

The book opens, as in the first edition, with a nice overview of ML estimation theory and numerical optimization methods, now extended to discuss the alternative optimization algorithms. The authors communicate the key issues with a clarity and succinctness that a number of textbook authors could aspire to. Chapter 2 is an overview of what the `ml` command does (without getting bogged down in the detail of syntax diagrams) and is a notable improvement on its earlier counterpart. Chapters 3–8 go through in detail how to use Stata to maximize your own likelihood function. Chapters 9 and 10 are about how to package your code into a do-file and write your own command using ado-files. Chapter 11 (new) details how to add survey estimation features to your estimator.

Six examples are worked through in fine detail in chapter 12. Although there are examples scattered throughout the book (and the previous edition), this is a welcome addition. As in many learning contexts, it is by consideration of a number of illustrations in detail that you begin to understand why you have to do certain things in a particular way. You are then better placed to apply and adapt them to different applications. The particular examples are the logit and probit models, Weibull and Cox continuous time hazard regression models, random effects linear regression for panel data, and seemingly unrelated regression. The seemingly unrelated regression example is new compared with the first edition and is important because it shows how the use of a concentrated likelihood function can ensure convergence in difficult maximization problems. In fact, this is the first mention of this strategy, and it would have helped if there were greater elaboration of it and its wider applicability (perhaps in chapter 1).

The book is completed by three appendices, references, and indices. The appendices are valuable reference points; they provide syntax diagrams, checklists for the different maximization methods (`lf`, `d0`, `d1`, `d2`), and complete listings of the code used to estimate the models considered in chapter 12. The subject index is rather thin and could do with more entries. For example, there is no entry for “weights”.

The most valuable addition to this edition, from my personal perspective, is section 4.5 about panel-data likelihoods. Many of `ml`’s great features are available by default only if the likelihood function takes the canonical “linear form”. In essence, this means that the overall sample likelihood is the sum of the likelihood contributions from each subject (person or firm, etc.), and there is one row in the dataset for each subject. One can then use maximization method `lf`; features such as robust variance calculations are very easy to incorporate, and speed and accuracy are hardly compromised. With panel data likelihoods, there are repeated observations (multiple rows in the dataset) for each subject and the linear form property does not hold. (These likelihoods include not only those for standard time-series cross-sectional data, but also those for discrete time hazard regression models.) One can use maximization method `d0` (with numerical derivatives and Hessian matrix calculations) relatively straightforwardly, but estimators may be slow to converge, and one cannot have robust variance estimates, survey options, or outer product of gradients variances. To get these features, and to speed up convergence, requires calculation of analytical first derivatives, implying use of method `d1` (or `d2`, which requires coding of the negative Hessian matrix calculation as well).

Coding these derivative calculations is a formidable exercise in my experience. For example, I recall spending several weeks in the summer of 2000 trying to write my command `spsurv` to use method `d1` rather than `d0`. Despite having the first edition of this book as my constant companion (and also getting helpful assistance from StataCorp Technical Support), I finally gave up. It seemed as though my model was not of the type that could be estimated in the way that was suggested. Things are different now. Section 4.5 of the current edition provides a much clearer and extended discussion of how to maximize panel-data likelihoods, and I now understand how the `mlvecsum` function may provide a route to writing a `d1` evaluator for my likelihood (though I have not yet tested this idea). Observe, too, the advent of new function `mlmatbysum` for use with `ml` when coding a `d2` evaluator. It enables calculation of various components of second derivative formulas that `mlmatsum` cannot handle. Of course, coding has to be preceded by production of the formulas for the derivatives, and this remains a complex task for many likelihoods, with plenty of scope for making errors. It is of some consolation that at least we have methods `d1debug` and `d2debug` to check analytical derivative calculations against numerically calculated ones (chapter 5).

3 What is missing from the book?

More examples would be very welcome, not only for the reasons I gave earlier, but also because some types of models are not discussed at all. For instance, there is no discussion of how to fit parametric statistical distributions to sample data. As it happens, it is relatively straightforward to do this (once you know how!); see *inter alia* my commands `smfit` and `dagumfit` for fitting the three parameter Singh–Maddala (Burr type 12) and Dagum (Burr type 3) distributions (Jenkins 1999). I only learned how to use `ml` to estimate these models after consulting Technical Support. Other candidates for examples might be a model with sample selection (for example, a Tobit model allowing for sample selection) or a double-hurdle model. In both cases, the model consists of two equations, with error terms correlated across equations.

I would have also liked to have seen more about the writing of ado-files to maximize likelihoods. For many researchers, getting estimation results is all that they are interested in, and for them the emphasis in the book will be about right; why go to the extra trouble of writing a new estimation command? However, experience tells us that there are substantial benefits to the Stata user community from the increased availability of new and well-written estimation commands. So, there is an important role that a book like this can play in supporting budding command writers. Chapter 10 is a helpful start, but it could go further. Several examples that come to mind from personal experience are how to (1) pass variable names (permanent and temporary) from the main command ado file to the evaluator subroutine that is called by `ml`, (2) write code for parsing multi-equation model commands, and (3) save items using `ereturn` in an estimation program and access them in the associated display program. I concede that some of these topics might be labeled general programming ones, but they arise commonly in the ML estimation context and so deserve treatment there.

There are other areas where readers could benefit more generally from the authors' accumulated experience and undoubted wisdom. One example is a more exhaustive and systematic discussion of numerical accuracy and precision issues to add to those scattered about in the book already (for example, why it is better to code `norm(-x)` rather than `1 - norm(x)`, and so on). Also, model parameters other than regression coefficients typically have theoretical bounds (for example, correlations lie between −1 and 1), and one should incorporate these into the estimation process. This points to the need for a more explicit and systematic cataloging of the transformations that are available, together with discussion of how to reverse the transformation post-estimation and derive the variance–covariance matrix of the relevant parameters in their original metric (`_diparm` is your friend). Extended discussion of the types of models for which likelihood maximization may prove difficult, and what to do in such cases, would be useful. Another, rather more advanced, example would be illustration of how to incorporate numerical methods such as Gauss–Hermite quadrature into user-written programs. More generally, the examples considered could be more systematically classified according to the issues that each of them raised so that budding users could more easily see which case or cases was most relevant to their problem.

4 Conclusions

If you are considering whether to buy this book, look first at *http://www.stata-press.com/books/ml.html*. This web page gives the full table of contents, and you can download for free the preface, chapter 1 (Theory and Practice), the subject index, and the datasets referred to in the book (although not the code used in the examples). Researchers who already have the first edition and who are only occasional users of `ml` might well wait for the next edition and rely on the online help and the references manuals in the interim. But if you are a regular user of `ml` or are a potential new user without the first edition, I would strongly recommend getting this edition. The book costs US$42, which is not too expensive for books of this sort, but choose your shipping option carefully. (The speediest shipping option to Europe costs only a dollar less than the book itself!)

5 References

Gould, W., J. Pitblado, and W. Sribney. 2003. *Maximum Likelihood Estimation with Stata*. 2d ed. College Station, TX: Stata Press.

Jenkins, S. P. 1999. sg106: Fitting Singh–Maddala and Dagum distributions by maximum likelihood. *Stata Technical Bulletin* 48: 19–25. In *Stata Technical Bulletin Reprints*, vol. 8, 261–268. College Station, TX: Stata Press.

About the Author

Stephen Jenkins is a professor at the Institute for Social and Economic Research, University of Essex, Colchester UK, and an Associate Editor of the *Stata Journal*. He has written a number of commands using `ml` (all available using `ssc`): `smfit`, `dagumfit`, `pgmhaz`, `spsurv`, and (with Lorenzo Cappellari) `mvprobit`.