The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from Sage Publishing via telephone 805-499-9774 (U.S. customers) or 44-(0)20-73248701 (International), email journals@sagepub.com, or online at

https://journals.sagepub.com/home/stj

The *Stata Journal* (ISSN 1536-867X) is a quarterly journal of Stata Press and is published by Sage Publishing in association with StataCorp LLC. Stata, Stata, Stata Press, Mata, Mata, and NetCourse are registered trademarks of StataCorp LLC.

The Stata Journal (2025)
25, Number 3, pp. 677–684 DOI: 10.1177/1536867X251365502

Review of Maximum Likelihood Estimation with Stata, Fifth Edition, by Pitblado, Poi, and Gould

Richard Williams
Department of Sociology
University of Notre Dame
Notre Dame, IN
rwilliam@nd.edu
https://orcid.org/0000-0001-9974-8307

Abstract. I review *Maximum Likelihood Estimation with Stata, Fifth Edition*, by Pitblado, Poi, and Gould (2024 [Stata Press]).

Keywords: gn0104, book review, maximum likelihood, Stata programming

1 Introduction

I had mixed feelings when I was asked to review the fifth edition of *Maximum Likelihood Estimation with Stata* by Pitblado, Poi, and Gould. On the one hand, the third edition (released in 2006) was absolutely invaluable to me when I wrote my `gologit2` (Williams 2006) and `oglm` (Williams 2010) commands. On the other hand, I found parts of that book overwhelming. I did not have the mathematical skills or at least the perseverance needed for some of the more advanced features. But once I decided to work through the book anyway, I also found that I did not need those advanced features to accomplish what I wanted.

My impression of the fifth edition is very similar. Some (including me) may well find parts of the book intimidating. But even those with less advanced statistical skills can find the book extremely useful, and those with more advanced skills may find that this is exactly the book they need.

2 Who is the book for?

As the preface to the book notes, "*Maximum Likelihood Estimation with Stata, Fifth Edition* is written for researchers in all disciplines who need to compute maximum likelihood estimators that are not available as prepackaged routines." This first sentence alone may be enough for most potential readers to decide whether to read further. Many users are perfectly happy with Stata's built-in features occasionally supplemented by the community-contributed commands that are easily available from Statistical Software Components and elsewhere. If they are not happy, they know that StataCorp is very receptive to suggestions and may provide what they want in the near future.

Although less obviously, the first sentence also identifies other people that probably will not be interested in the book. The book is targeted at users who want to write their own *estimation* commands, which are for statistical procedures like `logit` and `probit` and `regress`. Estimation commands typically allow postestimation commands like `predict`, `margins`, and `test`. Readers may want to write Stata commands for many repetitive tasks, but they do not need to program new statistical procedures; they just need a program that makes using existing procedures easier. Therefore, many initially interested readers may soon realize that other Stata books on programming, such as Kit Baum's 2016 *An Introduction to Stata Programming, Second Edition*, are much better fits to their needs.

At least some, however, will be pleased to learn that they do not have to stick only with what Stata and other users choose to provide. In more than a few cases, users are willing and able to take on those challenges themselves. This group is the main target audience for the book.

There is one other potential audience: owners of earlier editions. Is the new version worth upgrading to? My answer would be an emphatic yes. Stata has improved greatly since the 2006 third edition that I used. Some of those changes were reflected in the 2010 fourth edition, but even this edition is now 15 years old. The syntax and examples in the fifth edition are extensive and up to date. Discussions of Mata and the new `mlexp` command were either brief or nonexistent in earlier editions. There are many features in the old editions that I never used, and I would want to have the most current edition if I did decide to try them. You might save a little money by sticking with an older edition, but the additional time you would have to spend getting up to date on things that are succinctly covered in the new edition would make its purchase well worth it.

The preface then further explains what skills you need to get the most out of the book. The reader should be familiar with Stata, but only relatively basic programming skills are required (and even those without such skills can probably learn them from the book fairly easily). I think this may make mastery of the book sound easier than it is, but at the same time, I do not think the book is as difficult as it may first seem.

In short, who is the book for? My best guess is that most Stata users will have little or no interest or need for the book. But for those who wish to push the frontiers, who wish to use statistical techniques not already available to them, or who perhaps want to substantially improve routines that are already there, this book may be invaluable. I will add that motivated programmers have a great chance of being successful even if much of the book seems too advanced for them.

3 Organization of the book

Chapter 1 begins with a discussion of likelihood theory. For many, this will be a review. For others, all the equations may be a bit daunting. But this chapter already contains much more information than many users will ever need, such as discussions of the Newton–Raphson algorithm and numerical second derivatives. Those who already find

themselves feeling a bit lost will likely find that the detailed discussions and numerous examples contained throughout the book (complete with sample Stata code) make them feel much more comfortable.

Chapter 2 discusses the relatively new `mlexp` command, which "allows you to specify a likelihood function directly in a dialog box or at the command line and obtain maximum likelihood (ML) estimates with a minimum of effort" (p. 29). The book notes that the command does not work with all likelihood functions, but when it does work, "its simplicity cannot be matched" (p. 29).

Like much of the rest of the book, the chapter gives excellent examples with Stata code using familiar techniques, for example, `regress` and `probit`. The examples not only show you how `mlexp` works but also help you understand how maximum likelihood concepts apply to techniques you are already familiar with. If commands like `regress` and `probit` did not already exist, a user could easily replicate them with straightforward `mlexp` commands.

Chapter 3 provides an introduction to the `ml` command, which can be used to write programs that implement more complicated maximum-likelihood estimators. For many, deriving the log-likelihood function for your model may be the most difficult task. Once you do that, however, the process is often relatively straightforward. More detailed examples for `probit` and `regress` are offered. Furthermore, you will feel that you have already read much of this chapter before because the syntax for `ml` shares so many similarities with other Stata estimation commands. For example, you will see how to use robust standard errors, perform weighted estimation, and use the `svy` prefix. For pedagogical purposes, the chapter uses the general form (`gf0`) evaluator, but later chapters show methods that are generally superior.

Chapter 4 provides a much more detailed description of `ml` programming and shows how to overcome limitations noted. Particularly critical is the discussion of likelihood-evaluator methods in section 4.3. Again, the discussion may seem a bit daunting to many, but this point on page 63 is worth highlighting:

> Use method `lf` [linear form] when the likelihood function meets the linear-form restrictions and you do not wish to program first or second derivatives. *This is the most popular method among users and is often the only method one ever needs.* [Emphasis added.]

I think that last sentence is one of the most critical in the book. Most users do not need much of the advanced material presented. If they want to make their programs faster, they may want to program first or second derivatives or possibly use Mata (discussed in chapters 15 and 16). But for most people, perfectly adequate `ml` commands can be written using relatively simple methods. (Exceptions are discussed in chapter 7.)

The chapter again provides excellent examples. It reviews popular options, many of which will seem familiar because they are also popular with other Stata commands. Section 4.6 provides a brief overview of all the steps you should follow to write and debug your own likelihood functions, but these points are discussed more later.

Chapter 5 discusses method `lf` in more detail and says, "Use this method whenever possible. Method `lf` is fast, accurate, and easy to use" (p. 79). The main emphasis is on providing very helpful examples, which now include a discussion of the Weibull model. Not to be overlooked is section 5.3, which explains why it is so important to generate temporary variables using double precision. It further tells you what problems you can safely ignore because method `lf` takes care of them for you. It also shows, despite what the term `lf` (linear form) might seem to imply, that nonlinear specifications are possible. It closes with yet another powerful argument for using method `lf`, noting that it often has substantial speed advantages over other alternatives.

Depending on their needs and interests, many readers can skim or even skip many of the next few chapters. Chapter 6 covers additional `lf` evaluators `lf0`, `lf1`, and `lf2`. These methods let you specify analytic first and second derivatives. You have to do things like program the Hessian matrix and scores. Program execution may be faster but may or may not be worth the additional programming effort and statistical expertise that are required. As usual, anyone who wants to use the more advanced `lf` methods is given several examples as guidance.

Chapter 7 discusses methods `d0`, `d1`, and `d2`. Depending on their goals, readers may find that they can skip this chapter completely or must read it very closely. When a `d` evaluator and an `lf` evaluator can both fit the same model, the `lf` evaluator is quicker and simpler. So why use `d` evaluators at all? As page 130 explains,

> Method `d0` does not assume that the overall log likelihood ln L is the sum of the log likelihoods from each observation. Some models—such as the Cox proportional hazards model, conditional logistic regression, and panel-data models—produce log likelihoods for *groups* of observations and so do not lend themselves to methods that assume the likelihood function satisfies the linear-form restrictions. [Emphasis added.]

I have never needed to program anything using the method `d` evaluators. I consider myself lucky because these methods are obviously much more complicated than anything I have attempted. For those who have both the need and the determination to write such programs, chapter 7's in-depth discussion and detailed examples should be invaluable resources.

Chapters 8, 9, 10, and 11 all deal with programming issues and displaying results. Chapter 8 shows how to use `ml check` to debug programs. Chapter 9 discusses setting initial values. Good start values can speed up execution or even be critical to get the program to converge at all. Chapter 10 discusses interactive maximization and offers some tips for dealing with programs that are not converging. Chapter 11 shows how to graph convergence and display the output. If the section on displaying output seems familiar, it is because it shows mostly the same syntax that you use for displaying output from any Stata command. For example, you can specify the desired significance level for confidence intervals, how coefficients are formatted, and whether coefficients are exponentiated.

Chapters 12, 13, and 14 talk about writing do-files and ado-files. I thought these chapters were very well written, but readers who have made it this far would probably know most of this already because it is so similar to Stata's own documentation and other books on Stata programming. Still, this material may not be familiar to everyone, and the parts that are unique to `ml` programming are especially helpful.

I would advise all programmers to pay very close attention to section 14.2, *Writing your own predict command*. As explained below, I think this section is underdeveloped, but it still provides critical information. A good `predict` command can make your program far more useful. It can generate several variables that are functions of the model estimates and the data. For example, you might want to compute predicted probabilities for each case. A good `predict` command can also help enhance other postestimation commands, such as `margins`. Predict commands are not that hard to write, and the value they add to your estimation commands is usually well worth it.

Chapters 15 and 16 discuss Mata-based likelihood evaluators. Those who are already familiar with Mata will likely be especially interested in these chapters. As the book notes, many complicated statistical models are easily programmed in Mata. The chapters show how maximum likelihood evaluators can be programmed in Mata while using `ml` commands to control the process. The example programs are very good and illustrate how to replicate results presented earlier in the book. Those who have already written `ml` programs may want to consider rewriting parts of their programs using Mata to see whether execution becomes quicker. Those just starting their programs may find it easier to use Mata in the first place.

Finally, chapter 17 provides several additional examples, while appendix E lists the Stata code for those examples. Appendixes A, B, and C provide the syntax for key commands. Appendix D provides concise checklists to remember to do when programming. Once someone has read the book, they may find that the appendixes provide most of the reference material they need when doing their own programming.

4 Strengths of the book

As with most Stata Press publications, the examples are some of the most appealing parts of the book. I personally learn best by doing. Seeing how something is actually programmed in Stata helps me not only provide a template for my own programs but also understand the underlying statistical principles.

I like how the book offers something for those with basic skills and needs and for those who are willing and able to take on more demanding tasks. For me, the book became far more accessible once I realized that I did not need to understand everything! But even advanced statisticians and programmers will benefit greatly by seeing examples of how the `d` evaluators work.

I liked how familiar much of the book already seemed. If you are familiar with the syntax for commands like `logit` and `regress`, then you are already familiar with much of the syntax needed for `ml` programming. One of Stata's great strengths is its internal logic and consistency, which makes it easier to take up new tasks like `ml` programming.

I like the emphasis on Mata. I have tried reworking some of my own routines using Mata and may release them if I find that Mata yields significant runtime improvements. Those who use Mata more than I do will probably be especially attracted to the capabilities that it offers.

5 Areas for improvement

There are a few things that I wish had been done better and would probably not have required that much additional work.

I liked how the book clarified that many readers will not need everything in it. But holding the reader's hand even more would have been nice. Sections that are essential for almost every programmer could have been highlighted as such, while sections that are likely to be needed only for the most advanced tasks could have been noted as optional or recommended.

While the examples and sample code are excellent, they mostly show alternative ways to do things that Stata already does. Brief discussions of popular community-contributed estimation commands might have helped emphasize the point of just how useful the methods in the book have been and could be.

There were various places where I thought the book could have provided additional practical tips and suggestions. For example, the book notes in section 10.3 that using the `difficult` option sometimes helps with convergence problems but also points out that it often just makes matters worse. If `difficult` often does not work, what exactly is the programmer supposed to do to address convergence problems? The book does not say.

Long (1997) and others make several suggestions that could have been easily added to this section of the book. Even advanced computers do not have infinite precision. Rescaling variables (for example, measuring income in thousands of dollars rather than dollars) often makes convergence more likely and can have the added advantage of making coefficients easier to interpret (for example, a coefficient displayed as 0.000000 that is highly statistically significant can be pretty baffling to make sense of; with variable rescaling, the coefficient might instead appear as a more sensible value like 0.003275.) Long also notes that many estimation tasks encounter problems simply because the data have not been properly cleaned or because variables have not been computed correctly. In short, before spending a great deal of time trying to modify a troublesome program, you should first ensure that the problem is not due to something else, such as flaws in the data.

Incidentally, one of my own favorite suggestions for identifying problems with programs is to run the `update` and `ado update` commands. If you are lucky, some bug in Stata may have already been fixed. If not, at least all your software is up to date. I suspect that many Stata users do not even realize that software updates are possible.

Finally, and most critically, I am surprised at how little attention postestimation commands receive. The `margins` command is one of the most important additions to Stata since the fourth edition was published in 2010, yet the book hardly mentions it. As Williams (2012) and others argue, the `margins` command can make the substantive importance of results from an estimation command much clearer. For example, in a logistic regression, a coefficient of 0.357 for variable `female` (coded 1 if the respondent is female, 0 otherwise) means that other things being equal, females are more likely than others to experience the event. But how much more likely? Is the difference large enough to be substantively interesting? With the `margins` command, you might be able to determine that on average, women are 12 percentage points more likely to experience the event. You might further find that the gender gap is very small at age 20 and substantially larger by age 80. Most people would probably find that such results are much more interesting than the model coefficients.

But the `margins` command needs some help from the estimation commands to reach its full potential. For example, the `predict` command might need to be able to compute the predicted probabilities of the event occurring for each case.

While I wrote about the utility of the `margins` command in 2012, I was (properly) chastised because my own commands did not fully support it. With the help of Kerry Kammire from StataCorp, I rewrote my `gologit2` and `oglm` commands to work well with `margins`. It was not easy, but the effort was well worth it. Life might have been easier for both me and Kerry if I had a well-written discussion on how to optimize `ml` programs to work with postestimation commands. I am disappointed that the otherwise excellent fifth edition still does not include such a discussion. If I could add one section or chapter to the book, it would be on how to make your own commands take optimal advantage of `margins` and other postestimation commands available in Stata.

6 Conclusion

If you are perfectly happy with Stata as is, this book will probably be of little interest to you. Spend your money instead on one of the many other excellent books available from Stata Press.

If you do find Stata lacking in some statistical areas but find the book daunting, do not be intimidated. The book itself admits that what most people really need is not that difficult. If you focus on what is most critical and skim through the rest, you will probably be fine.

Finally, if you have in-depth mathematical training and advanced statistical needs, this is the book for you. Indeed, it might be the only book for you because I am not even sure that there is a good alternative. You already understand the statistical theory. This book shows you how to translate that theory into practice.

7 References

Baum, C. F. 2016. *An Introduction to Stata Programming*. 2nd ed. College Station, TX: Stata Press.

Gould, W., J. Pitblado, and B. Poi. 2010. *Maximum Likelihood Estimation with Stata*. 4th ed. College Station, TX: Stata Press.

Gould, W., J. Pitblado, and W. Sribney. 2006. *Maximum Likelihood Estimation with Stata*. 3rd ed. College Station, TX: Stata Press.

Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage.

Pitblado, J., B. Poi, and W. Gould. 2024. *Maximum Likelihood Estimation with Stata*. 5th ed. College Station, TX: Stata Press.

Williams, R. 2006. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal* 6: 58–82. https://doi.org/10.1177/1536867X0600600104.

———. 2010. Fitting heterogeneous choice models with oglm. *Stata Journal* 10: 540–567. https://doi.org/10.1177/1536867X1101000402.

———. 2012. Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata Journal* 12: 308–331. https://doi.org/10.1177/1536867X1201200209.

About the author

Richard Williams is a full professor and a former chairman of the Department of Sociology at the University of Notre Dame. His work has appeared in the *American Sociological Review*, *Social Forces*, *Social Problems*, *Demography*, *Cityscape*, *Stata Journal*, and *Sociological Methods and Research*. He is a coeditor of the recently published 10-volume *Sage Research Methods Foundations*. Recent work by Williams has looked at issues involving the analysis of categorical data. He notes that assumptions of heterogeneity and proportionality are often violated in commonly used logit and ordered logit models. He discusses how heterogeneous choice models and generalized ordered logit models provide potential solutions to these problems. He is the author of the Stata commands `gologit2` and `oglm`, which make the estimation of these models possible. In his work, he also shows how the use of adjusted predictions and marginal effects can be major aids in interpreting the results from categorical models. He won the 2015 *Stata Journal* Editors' Prize for his work in these areas.