Preface

As with previous editions, our goal in writing this book is to make it routine to carry out the complex calculations necessary to fully interpret regression models for categorical outcomes. Interpreting these models is complex because the models are nonlinear. Software packages that fit these models often do not provide options that make it simple to compute the quantities that are useful for interpretation; when they do provide these options, there is usually little guidance as to how to use them. In this book, we briefly describe the statistical issues involved in interpretation and then show how you can use Stata to make these computations.

While our purpose remains the same, this third edition is an almost complete rewrite of the second edition—almost every line of code in our SPost commands has been rewritten. Advances in computing and the addition of new features to Stata has expanded the possibilities for routinely applying more sophisticated methods of interpretation. As a result, ideas we noted in previous editions as good in principle are now much more straightforward to implement in practice. For example, while you could compute average marginal effects using commands discussed in previous editions, it was difficult and few people did so (ourselves included). Likewise, in previous editions, we relegated methods for dealing with nonlinearities and interactions on the right-hand side of the model to the last chapter, and our impression was that few readers took advantage of these ideas because they were comparatively difficult and error-prone to use.[1]

These limitations changed with the addition of factor variables and the `margins` command in Stata 11. It took us personally quite a while to fully absorb the potential of these powerful enhancements and decide how best to take advantage of them. Plus, Stata 13 added several features that were essential for what we wanted to do.

This third edition considers the same models as the second edition of the book. We still find these to be the most valuable models for categorical outcomes. And, as in previous editions, our discussion is limited to models for cross-sectional data. While we would like to consider models for panel data and other hierarchical data structures, doing so would at least double the size of an already long book.

1. Those who have read previous editions will note that this last chapter has been dropped entirely. In addition to covering linked variables of the right-hand side, that chapter also discussed adapting our commands to other estimation commands; however, this is now obsolete because `margins` works with most estimation commands. We also dropped the section on working effectively in Stata because Long's (2009) *Workflow of Data Analysis Using Stata* covers these topics in detail.

We note, however, that many of our SPost commands—such as `mtable`, `mgen`, and `mchange` (hereafter referred to as the `m*` commands)—are based on `margins` and can be used with any model that is supported by `margins`. This is a substantial change from our earlier `prchange`, `prgen`, `prtab`, and `prvalue` commands, which only worked with the models discussed in the book. A second major improvement is that our `m*` commands work with weights and survey estimation, because these are supported by `margins`.

SPost was originally developed using Stata 4 and Stata 5. Since then, our commands have often been enhanced to use new features in Stata. Sometimes these enhancements have led to code that was not as efficient, robust, or elegant as we would have liked. In SPost13, we rewrote much of the code, incorporated better returns, improved output, and removed obscure or obsolete features.

How to cite

Our commands are not officially part of Stata. We have written them in an effort to contribute to the community of researchers whose work involves extensive use of the models we cover. If you use our commands or other materials in published work, we ask that you cite our work in the same way that you cite other useful sources. We ask that you simply cite the book rather than providing different citations to different commands:

> Long, J. S., and J. Freese. 2014. *Regression Models for Categorical Dependent Variables in Stata*. 3rd ed. College Station, TX: Stata Press.

Thanks

Hundreds of people have contributed to our work since 2001. We cannot possibly mention them all here, but we gratefully acknowledge them for taking the time to give us their ideas. Thousands of students have taken classes using previous editions of the books, and many have given us suggestions to make the book or the commands more effective.

In writing this third edition, several people deserve to be mentioned. Ian Anson and Trent Mize tested commands and provided comments on draft chapters. Tom VanHeuvelen ran labs in two classes that used early versions of the commands, helped students work around bugs, made valuable suggestions on how to improve the commands, provided detailed comments on each chapter, was a sounding board for ideas, and was an exceptional research assistant. Rich Williams gave us many suggestions that improved the book and our commands. He has a (sometimes) valuable gift for finding bugs as well. Scott Long gratefully acknowledges the support provided by the College of Arts and Sciences at Indiana University.

People at StataCorp provided their expertise in many ways. More than this, though, we are grateful for their engagement and support of our project. Jeff Pitblado was enor-

mously helpful as we incorporated factor variables and `margins` into our commands. His advice made our code far more compact and reliable. Vince Wiggins provided valuable advice on our graphing commands and helped us understand `margins` better. Lisa Gilmore, as always, did a great job moving the book from draft to print. Most importantly, discussions with David Drukker stimulated our thinking about a new edition and, as always, asked challenging questions that made our ideas better.

Illinois and Indiana
August 2014

Jeremy Freese
Scott Long