Microeconometrics Using Stata

Revised Edition

A. COLIN CAMERON
PRAVIN K. TRIVEDI

Stata Press

Contents

Preface to the Revised Edition

Microeconometrics Using Stata, published in December 2008, was written for Stata 10.1. The book incorporated version 10.1 additions to Stata 10.0, most notably, the new random-number generators.

In this revised edition, we present other additions to Stata 10 that appear for the first time in Stata 11. With few exceptions, we present these additions in a way that reproduces the results given in the first edition.

First, we introduce the new construct of factor variables. These provide a simple way to specify models with sets of indicator variables formed from a categorical variable and to specify models with interactions. Factor variables replace the `xi` prefix command. See especially section 1.3.4 and the end of section 2.4.7.

Second, we describe the new `margins` command for prediction and for computation of marginal effects in regression models. The `margins` command with options including the `dydx()` option replaces the Stata `mfx` command and the user-written `margeff` command. Additionally, the `margins` command when used in conjunction with factor variables can simplify computation of marginal effects in models with interactions. See sections 10.5 and 10.6, especially subsections 10.5.7 and 10.6.5. Throughout this revised edition, notably, in chapters 14–17, we replace `mfx` and `margeff` with the `margins` command.

In the first edition, we most often calculated the marginal effect at the mean (MEM), rather than the average marginal effect (AME), because the `mfx` command did not compute the AME. The new `margins` command can compute both the MEM and the AME. In this revised edition, we have endeavored to replicate the results given in the first edition. For that reason, we continue to most frequently calculate the MEM, though in practice, the AME is usually preferred.

Third, we describe the new `gmm` command for generalized method of moments and nonlinear instrumental-variables estimation. See sections 10.3.8 and 17.5.2.

Fourth, we present some minor changes that need to be made to the existing `ml` command when the `d1` and `d2` methods are used. These changes arise because the `ml` command is now a front-end to the new Mata `moptimize()` function. We also present the new `lf0`, `lf1`, and `lf2` methods. See section 11.6. The Mata `optimize()` v evaluator has been renamed to gf evaluator; see section 11.7.

Preface to the First Edition

This book explains how an econometrics computer package, Stata, can be used to perform regression analysis of cross-section and panel data. The term microeconometrics is used in the book title because the applications are to economics-related data and because the coverage includes methods such as instrumental-variables regression that are emphasized more in economics than in some other areas of applied statistics. However, many issues, models, and methodologies discussed in this book are also relevant to other social sciences.

The main audience is graduate students and researchers. For them, this book can be used as an adjunct to our own *Microeconometrics: Methods and Applications* (Cameron and Trivedi 2005), as well as to other graduate-level texts such as Greene (2008) and Wooldridge (2002). By comparison to these books, we present little theory and instead emphasize practical aspects of implementation using Stata. More advanced topics we cover include quantile regression, weak instruments, nonlinear optimization, bootstrap methods, nonlinear panel-data methods, and Stata's matrix programming language, Mata.

At the same time, the book provides introductions to topics such as ordinary least-squares regression, instrumental-variables estimation, and logit and probit models so that it is suitable for use in an undergraduate econometrics class, as a complement to an appropriate undergraduate-level text. The following table suggests sections of the book for an introductory class, with the caveat that in places formulas are provided using matrix algebra.

Stata basics	Chapter 1.1–1.4
Data management	Chapter 2.1–2.4, 2.6
OLS	Chapter 3.1–3.6
Simulation	Chapter 4.6–4.7
GLS (heteroskedasticity)	Chapter 5.3
Instrumental variables	Chapter 6.2–6.3
Linear panel data	Chapter 8
Logit and probit models	Chapter 14.1–14.4
Tobit model	Chapter 16.1–16.3

Although we provide considerable detail on Stata, the treatment is by no means complete. In particular, we introduce various Stata commands but avoid detailed listing and description of commands as they are already well documented in the Stata manuals

and online help. Typically, we provide a pointer and a brief discussion and often an example.

As much as possible, we provide template code that can be adapted to other problems. Keep in mind that to shorten output for this book, our examples use many fewer regressors than necessary for serious research. Our code often suppresses intermediate output that is important in actual research, because of extensive use of command `quietly` and options `nolog`, `nodots`, and `noheader`. And we minimize the use of graphs compared with typical use in exploratory data analysis.

We have used Stata 10, including Stata updates.[1] Instructions on how to obtain the datasets and the do-files used in this book are available on the Stata Press web site at http://www.stata-press.com/data/mus.html. Any corrections to the book will be documented at http://www.stata-press.com/books/mus.html.

We have learned a lot of econometrics, in addition to learning Stata, during this project. Indeed, we feel strongly that an effective learning tool for econometrics is hands-on learning by opening a Stata dataset and seeing the effect of using different methods and variations on the methods, such as using robust standard errors rather than default standard errors. This method is beneficial at all levels of ability in econometrics. Indeed, an efficient way of familiarizing yourself with Stata's leading features might be to execute the commands in a relevant chapter on your own dataset.

We thank the many people who have assisted us in preparing this book. The project grew out of our 2005 book, and we thank Scott Parris for his expert handling of that book. Juan Du, Qian Li, and Abhijit Ramalingam carefully read many of the book chapters. Discussions with John Daniels, Oscar Jorda, Guido Kuersteiner, and Doug Miller were particularly helpful. We thank Deirdre Patterson for her excellent editing and Lisa Gilmore for managing the LaTeX formatting and production of this book. Most especially, we thank David Drukker for his extensive input and encouragement at all stages of this project, including a thorough reading and critique of the final draft, which led to many improvements in both the econometrics and Stata components of this book. Finally, we thank our respective families for making the inevitable sacrifices as we worked to bring this multiyear project to completion.

Davis, CA
Bloomington, IN
October 2008

A. Colin Cameron
Pravin K. Trivedi

1. To see whether you have the latest update, type `update query`. For those with earlier versions of Stata, some key changes are the following: Stata 9 introduced the matrix programming language, Mata. The syntax for Stata 10 uses the `vce(robust)` option rather than the `robust` option to obtain robust standard errors. A mid-2008 update of version 10 introduced new random-number functions, such as `runiform()` and `rnormal()`.