The Stata Journal (2008)
8, Number 4, pp. 579–582

Review of Multilevel and Longitudinal Modeling Using Stata, Second Edition, by Sophia Rabe-Hesketh and Anders Skrondal

Nicholas J. Horton
Smith College
Clark Science Center
Northampton, MA
nhorton@email.smith.edu

Abstract. This article reviews *Multilevel and Longitudinal Modeling Using Stata, Second Edition*, by Sophia Rabe-Hesketh and Anders Skrondal.

Keywords: gn0043, multilevel, longitudinal, Rabe-Hesketh, Skrondal, mlmus2

1 Content

There is a continued growth in the use of multilevel regression modeling to account for settings where observations are repeated over time, clustered within some other hierarchy, or both. The second edition of *Multilevel and Longitudinal Modeling Using Stata* (MLMUS2; Rabe-Hesketh and Skrondal 2008) complements other texts that provide an applied introduction to these methods (e.g., Fitzmaurice, Laird, and Ware [2004]; Diggle et al. [2002]; Singer and Willett [2003]), with the added bonus of a particular focus on Stata.

The first edition (with 317 pages plus front matter) did an excellent job describing and illustrating multilevel, longitudinal, and clustered regression methods in the context of real-world examples and implementations within Stata, but it fell somewhat short as a comprehensive introduction to this material (Horton 2006). MLMUS2 (which weighs in at 562 pages plus 34 pages of front matter) is a superset of the previous material, adding new chapters, examples, and exercises.

A new first chapter reviews linear regression. This is a particularly welcome addition because it facilitates the use of Stata for new users in terms of linear regression methods that are assumed as a prerequisite for the book. It features an extended case study regarding gender discrimination that is analyzed with descriptive, graphical, and inferential procedures. The chapter shows off important syntax, powerful features (e.g., `testparm` and `lincom`), and sophisticated graphics. It also introduces notation that is generalized throughout the book.

An expanded part II (entitled *Two-level linear models*) takes full advantage of the additional space (approximately 70 pages' worth) to provide a gentler and more comprehensive introduction to these methods. Separate chapters have been created for variance components, random-intercept, random-coefficient, and longitudinal models.

As in other chapters, detailed case studies with real-world datasets are accompanied by the Stata code to fit models, generate functions of parameters (e.g., predicted values or functions of parameters, plus associated confidence intervals), plot predicted values, and carry out residual analysis.

Part III (*Two-level generalized linear models*) includes chapters on binary, ordinal, and count models, and a new chapter focused on discrete-time survival models. The use of overdispersed count models (important given the lack of robustness of Poisson models to model misspecification) is described in some detail.

Part IV (*Models with nested and crossed random effects*) closes the main part of the text by addressing higher-level models. Because most packages are designed for nested effects, several useful ad hoc tricks are included (for example, using superclusters or virtual level-3 variables).

A series of appendices provide the syntax for the `gllamm` (generalized linear latent and mixed models; see http://www.gllamm.org), `gllapred`, and `gllasim` commands. These commands, written by the authors, add important advanced functionality to Stata.

The book has several strengths and attractive features. The examples are compelling and have been augmented by datasets taken from a variety of research settings primarily in the medical and social sciences. These analyses are easy to replicate as a reader: all datasets remain freely available for download with the `use` command.

Optional sections (e.g., background on adaptive quadrature or simulation studies for missing-at-random missingness) are flagged to allow them to be skipped upon first reading.

New exercises, included at the end of each chapter, add to the value of the book by providing opportunities to integrate knowledge and to practice implementing and interpreting models. As an example, chapter 9 includes exercises using epilepsy, headache, police stops, patents, school absenteeism, blindness, lip cancer, and skin cancer data. Many of the exercises include software tips and specific hints about model specification. A solutions manual, including code segments, is available to instructors.

Although the `gllamm` package is used for some analyses, the book does not address all the models that are described in detail in Skrondal and Rabe-Hesketh (2004). A companion book to MLMUS2 that provides a more accessible presentation of advanced latent variable and structural-equation models would likely sell well.

A continued strength of the text is its use of the same notation as Skrondal and Rabe-Hesketh (2004). The book complements existing Stata documentation (primarily the *Longitudinal/Panel-Data Reference Manual* [StataCorp 2007]).[1]

1. The Stata web site for the *Longitudinal/Panel-Data Reference Manual* includes the table of contents, as well as freely downloadable sample entries for `xt`, `xtintro`, `xtreg`, and `xtmixed`. Visit the web site at http://www.stata.com/bookstore/xt.html.

There are only a few limitations and quibbles. Some mention of zero-inflated count models would be useful, as well as a discussion of models that directly estimate the variance–covariance matrix (akin to the `REPEATED` statement in `SAS PROC MIXED`, though to be fair these models cannot be fit in Stata 10.1 because of overparameterization in the current implementation of `xtmixed`).

MLMUS2 provides a stand-alone introduction to multilevel and longitudinal modeling, with references to other books and papers as needed (interspersed throughout the text and in "Further reading" sections). The book complements the texts described above and the authors' more advanced and more comprehensive companion volume (Skrondal and Rabe-Hesketh 2004) in a helpful way.

I envision regularly using this book in several ways. I will replace my first edition with this one and keep it, along with the *Longitudinal/Panel-Data Reference Manual* (StataCorp 2007), on my shelf as a reference. I also envision using it as a primary text for a longitudinal regression models course at the advanced undergraduate or master's level. Finally, I can imagine using it as a tutorial in regression modeling using Stata and using it as an accessible introduction to more advanced methods.

The authors have provided a well-rounded and complete approach to model-fitting and interpretation of an important family of models. Once again, they are to be commended for helping foster the appropriate use of these regression models.

2 References

Diggle, P. J., P. Heagerty, K.-Y. Liang, and S. L. Zeger. 2002. *Analysis of Longitudinal Data*. 2nd ed. Oxford: Oxford University Press.

Fitzmaurice, G. M., N. M. Laird, and J. H. Ware. 2004. *Applied Longitudinal Analysis*. Hoboken, NJ: Wiley.

Horton, N. J. 2006. Review of Rabe-Hesketh and Skrondal, "Multilevel and Longitudinal Modeling Using Stata". *American Statistician* 60: 293–294.

Rabe-Hesketh, S., and A. Skrondal. 2008. *Multilevel and Longitudinal Modeling Using Stata*. 2nd ed. College Station, TX: Stata Press.

Singer, J. D., and J. B. Willett. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford: Oxford University Press.

Skrondal, A., and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC.

StataCorp. 2007. *Stata 10 Longitudinal/Panel-Data Reference Manual*. College Station, TX: Stata Press.

About the author

Nicholas Horton is an associate professor of mathematics and statistics at Smith College, Northampton, MA. His research interests involve the development and dissemination of missing data and longitudinal regression models, applied to substance abuse and psychiatric epidemiology studies.