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 (2023)
23, Number 3, pp. 901–904 DOI: 10.1177/1536867X231196518

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

Leonardo Grilli
Department of Statistics, Computer Science, Applications
University of Florence
Florence, Italy
leonardo.grilli@unifi.it

Carla Rampichini
Department of Statistics, Computer Science, Applications
University of Florence
Florence, Italy
carla.rampichini@unifi.it

Abstract. This article reviews *Multilevel and Longitudinal Modeling Using Stata, Fourth Edition*, by Rabe-Hesketh and Skrondal (2022, Stata Press).

Keywords: gn0095, book review, endogeneity, fixed effects, hierarchical data, mixed-effects model, random effects

1 Overview of the book

The fourth edition of *Multilevel and Longitudinal Modeling Using Stata* (MLMUS4; Rabe-Hesketh and Skrondal 2022) is the latest update of this popular book, which was published for the first time in 2005. The *Stata Journal* has already published reviews of the first and second editions (Wolfe 2006; Horton 2008). The book comprises two volumes: The first is dedicated to continuous responses, and the second to categorical responses, counts, and survival.

Nowadays, there are plenty of textbooks on multilevel modeling; most of them focus on applications and thus contain little theory. MLMUS4 achieves a good balance between theory and applications. The methods are clearly explained, and they are effectively illustrated by case studies on real datasets. The applications cover many disciplines, including medicine, economics, education, sociology, and psychology, reflecting the book's interdisciplinary nature. The exposition is rigorous, though the mathematical level is accessible to any researcher. MLMUS4 successfully implements the well-known motto that everything should be made as simple as possible but not simpler.

Most of the models presented in MLMUS4 belong to the class of generalized linear latent and mixed models, whose theory was developed by the same authors (Skrondal and Rabe-Hesketh 2004).

MLMUS4 builds a bridge between statistics and econometrics in the realm of random effects. In fact, in econometrics, random effects are exploited predominantly in longitudinal settings, prompting the development of particular methods such as panel-data instrumental-variables techniques, which are not easily accessible to researchers without backgrounds in econometrics. MLMUS4 fills the gap by introducing econometrics techniques with a language more familiar to statisticians.

Even if MLMUS4 is designed to teach how to implement multilevel modeling in Stata, it is also a good textbook for researchers planning to use different software. In fact, the excellent exposition of the methods is special in the panorama of multilevel modeling. Moreover, the applications with Stata are well separated from the theory. In addition, the Stata code and output are easy to follow even by readers unfamiliar with Stata.

The two volumes contain over 140 exercises based on over 100 real datasets. Data and Stata code for examples and exercises are available as supplementary material on the book website (https://www.stata.com/bookstore/multilevel-longitudinal-modeling-stata/). Solutions for selected exercises are also available.

2 What's new

The book is an update of the third edition (Rabe-Hesketh and Skrondal 2012), with the same chapters, although chapter 5 has been renamed. The number of pages of the two volumes increased from 974 to 1,098 pages.

The examples and exercises exploit the new commands and options of Stata 17, which entails remarkable improvements over Stata 12, the version used in the third edition. One such improvement is the Kenward–Roger degrees-of-freedom correction for more accurate inference with a small number of clusters. Note that the authors generally describe all Stata commands that can be used to fit a given model, discussing their advantages and disadvantages. This is one of the strengths of the book because the Stata procedures for multilevel and longitudinal models are spread over several suites of commands, particularly ME, XT, CM, and SEM.

Regarding the methods, the main novelty of the fourth edition is the presentation of models for longitudinal and panel data with subject-specific effects in chapter 5. The models are presented from the perspective of the type of endogeneity they deal with. This perspective clarifies the motivations underpinning the choice between fixed and random effects, which is the source of a long-lasting debate in applied research (Bell, Fairbrother, and Jones 2019). Particularly, in section 5.3 the authors present four different fixed-effects approaches that address level-two endogeneity, although the Mundlak approach is implemented by a random-effects model. Indeed, the Mundlak method, also described in section 3.7.6 for cross-sectional data, provides the same estimates for level-one parameters as the fixed-effects model.

The new section 5.4 illustrates difference-in-differences estimation for quasi-experiments and the connection with repeated-measures analysis of variance. Furthermore, the new section 5.7 is devoted to instrumental-variables estimators to deal with

level-one endogeneity. Instrumental variables and difference in differences are powerful approaches for estimating causal effects widely used in econometrics (Angrist and Pischke 2009). The authors present these methods in a language accessible to researchers of other fields.

Another novelty of the fourth edition is the description of Bayesian inference as an alternative to maximum likelihood for the estimation of binary response models with crossed random effects, where it gives a substantial computational advantage. In chapter 16, the authors briefly introduce Bayesian inference and show the implementation in this setting, highlighting the main benefits.

3 Final remarks

A suggestion for the next edition is to expand the section devoted to the Bayesian estimation of logistic models with crossed random effects to a chapter on the Bayesian estimation of multilevel models. Indeed, Bayesian inference may provide accurate uncertainty quantification in complex settings, especially for predicting random effects, compared with empirical Bayes methods that are routinely used. Those considerations are generally valid for random-effects models. Thus, a stand-alone chapter could help the reader switch to Bayesian estimation whenever convenient.

MLMUS4 is a valuable source that can be used in different ways. It is an excellent textbook on statistical modeling, where each type of model is first introduced by a motivating example, starting with the single-level case and then progressively introducing more complex random-effects structures. MLMUS4 is also a precious reference for experienced researchers wishing to deepen their understanding of specific topics. The book's interdisciplinary nature suits a broad range of applied researchers.

4 References

Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Bell, A., M. Fairbrother, and K. Jones. 2019. Fixed and random effects models: Making an informed choice. *Quality and Quantity* 53: 1051–1074. https://doi.org/10.1007/s11135-018-0802-x.

Horton, N. J. 2008. Review of Multilevel and Longitudinal Modeling Using Stata, Second Edition, by Sophia Rabe-Hesketh and Anders Skrondal. *Stata Journal* 8: 579–582. https://doi.org/10.1177/1536867X0800800411.

Rabe-Hesketh, S., and A. Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata*. 3rd ed. College Station, TX: Stata Press.

———. 2022. *Multilevel and Longitudinal Modeling Using Stata*. 4th ed. College Station, TX: Stata Press.

Skrondal, A., and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman and Hall/CRC. https://doi.org/10.1201/9780203489437.

Wolfe, R. 2006. Review of Multilevel and Longitudinal Modeling Using Stata by Rabe-Hesketh and Skrondal. *Stata Journal* 6: 138–143. https://doi.org/10.1177/1536867X0600600110.

About the authors

Leonardo Grilli, PhD, is Full Professor of Statistics at the University of Florence. His research concerns random-effects models for multilevel analysis, causal inference, latent growth curve models, mixture models, quantile regression, and statistical learning.

Carla Rampichini, PhD, is Full Professor of Statistics at the University of Florence. Her research interests relate to random-effects models for multilevel analysis, latent growth curve models, mixture models, program evaluation, causal inference, and statistical learning.