Preface

This book is about applied multilevel and longitudinal modeling. Other terms for multilevel models include hierarchical models, random-effects or random-coefficient models, mixed-effects models, or simply mixed models. Longitudinal data are also referred to as panel data, repeated measures, or cross-sectional time series. A popular type of multilevel model for longitudinal data is the growth-curve model.

The common theme of this book is regression modeling when data are clustered in some way. In cross-sectional settings, students may be nested in schools, people in neighborhoods, employees in firms, or twins in twin-pairs. Longitudinal data are by definition clustered because multiple observations over time are nested within units, typically subjects.

Such clustered designs often provide rich information on processes operating at different levels, for instance, people's characteristics interacting with institutional characteristics. Importantly, the standard assumption of independent observations is likely to be violated because of dependence among observations within the same cluster. The multilevel and longitudinal methods discussed in this book extend conventional regression to handle such dependence and exploit the richness of the data.

Volume 1 is on multilevel and longitudinal modeling of continuous responses using linear models. The volume consists of four parts: I. Preliminaries (a review of linear regression modeling, preparing the reader for the rest of the book), II. Two-level models, III. Models for longitudinal and panel data, and IV. Models with nested and crossed random effects. For readers who are new to multilevel and longitudinal modeling, the chapters in part II should be read sequentially and can form the basis of an introductory course on this topic. A one-semester course on multilevel and longitudinal modeling can be based on most of the chapters in volume 1 plus chapter 10 on binary or dichotomous responses from volume 2. For this purpose, we have made chapter 10 freely downloadable from https://www.stata-press.com/books/mlmus4_ch10.pdf.

Volume 2 is on multilevel and longitudinal modeling of categorical responses, counts, and survival data. This volume also consists of four parts: I. Categorical responses (binary or dichotomous responses, ordinal responses, and nominal responses or discrete choice), II. Counts, III. Survival (in both discrete and continuous time), and IV. Models with nested and crossed random effects. Each chapter starts by introducing models for nonclustered data (for example, logistic and Poisson regression) and then extends the models for clustered data by introducing random effects, leading to generalized linear mixed models. Subsequently, alternatives such as generalized estimating equations (GEE) and fixed-effects approaches are discussed. Chapter 10 on binary or dichotomous

responses is a core chapter of this volume and should be read before embarking on the other chapters. It is also a good idea to read chapter 14 on discrete-time survival before reading chapter 15 on continuous-time survival.

Our emphasis is on explaining the models and their assumptions, applying the methods to real data, and interpreting results. Many of the issues are conceptually demanding but do not require that you understand complex mathematics. Therefore, wherever possible, we introduce ideas through examples and graphical illustrations, keeping the technical descriptions as simple as possible. Some sections that go beyond an introductory course on multilevel and longitudinal modeling are tagged with the ❖ symbol. Derivations that can be skipped by the reader are given in displays. For an advanced treatment, placing multilevel modeling within a general latent-variable framework, we refer the reader to Skrondal and Rabe-Hesketh (2004), which uses the same notation as this book.

This book shows how all the analyses described can be performed using Stata. There are many advantages of using a general-purpose statistical package such as Stata. First, for those already familiar with Stata, it is convenient not having to learn a new standalone package. Second, conducting multilevel analysis within a powerful package has the advantage that it allows complex data manipulation to be performed, alternative estimation methods to be used, and publication-quality graphics to be produced, all without having to switch packages. Finally, Stata is a natural choice for multilevel and longitudinal modeling because it has gradually become perhaps the most powerful general-purpose statistics package for such models.

Each chapter is based on one or more research problems and real datasets. After describing the models, we walk through the analysis using Stata, pausing to address statistical issues that need further explanation. Do-files for each chapter can be downloaded from https://www.stata-press.com/data/mlmus4.html. Some readers may find it useful to perform the analyses while reading the book.

Stata can be used either via a graphical user interface (GUI) or through commands. We recommend using commands interactively—or preferably in do-files—for serious analysis in Stata. For this reason, and because the GUI is fairly self-explanatory, we use commands exclusively in this book. However, the GUI can be useful for learning the Stata syntax. Generally, we use the `typewriter font` to refer to Stata commands, syntax, and variables. A "dot" prompt followed by a command indicates that you can type verbatim what is displayed after the dot (in context) to replicate the results in the book. Some readers may find it useful to intersperse reading with running these commands. We encourage readers to write do-files for solving the data analysis exercises because this is standard practice for professional data analysis.

The commands used for data manipulation and graphics are explained to some extent, but the purpose of this book is not to teach Stata from scratch. For a basic introduction to Stata, we refer the reader to Acock (2018). Other books and further resources for learning Stata are listed at the Stata website.

If you are new to Stata, we recommend running all the commands given in chapter 1 of volume 1. A list of commands that are particularly useful for manipulating, describing, and plotting multilevel and longitudinal data is given in the appendix of volume 1. Examples using these and other commands can easily be found by referring to the "commands" entry in the subject index.

We have included applications from a wide range of disciplines, including medicine, economics, education, sociology, and psychology. The interdisciplinary nature of this book is also reflected in the choice of models and topics covered. If a chapter is primarily based on an application from one discipline, we try to balance this by including exercises with real data from other disciplines. The two volumes contain over 140 exercises based on over 100 different real datasets. Exercises for which solutions are available to readers are marked with Solutions, and the solutions can be downloaded from https://www.stata-press.com/books/mlmus4-answers.html. Instructors can obtain solutions to all exercises from Stata Press.

All datasets used in this book are freely available for download; for details, see https://www.stata-press.com/data/mlmus4.html. These datasets can be downloaded into a local directory on your computer. Alternatively, individual datasets can be loaded directly into net-aware Stata by specifying the complete URL. For example,

```
. use https://www.stata-press.com/data/mlmus4/pefr
```

If you have stored the datasets in the working directory, omit the path and just type

```
. use pefr
```

We will generally describe all Stata commands that can be used to fit a given model, discussing their advantages and disadvantages. An exception to this rule is that we do not discuss our own `gllamm` command in volume 1 (see the `gllamm` companion, downloadable from http://www.gllamm.org, for how to fit the models of volume 1 in `gllamm`). In volume 1, we extensively use the Stata commands `xtreg` and `mixed`, and we introduce several more specialized commands for longitudinal modeling, such as `xthtaylor`, `xtivreg`, and `xtabond`. The `sem` command for structural equation modeling is used for growth-curve modeling.

In volume 2, we use Stata's `xt` and `me` commands for different response types. For example, we use `xtlogit` and `melogit` for binary responses, `meologit` for ordinal responses, `xtpoisson` and `mepoisson` for counts, and `mestreg` for multilevel continuous-time survival modeling with shared frailties. In chapter 12 on nominal responses, we use Stata's new `cm` (for "choice model") suite of commands, such as `cmxtmixlogit`. `gllamm` is also used throughout volume 2. We also discuss commands for marginal models and fixed-effects models, such as `xtgee` and `clogit`. The online reference manuals available through the `help` command within Stata provide detailed information on all the official Stata commands for multilevel and longitudinal modeling.

The `nolog` option has been used to suppress the iteration logs showing the progress of the log likelihood. This option is not shown in the command line because we do not recommend it to users; we are using it only to save space.

We assume that readers have a good knowledge of linear regression modeling, in particular, the use and interpretation of dummy variables and interactions. However, the first chapter in volume 1 reviews linear regression and can serve as a refresher.

Errata for different editions and printings of the book can be downloaded from https://www.stata-press.com/books/errata/mlmus4.html, and answers to exercises can be downloaded from https://www.stata-press.com/books/mlmus4-answers.html.

In this fourth edition, we have thoroughly revised all chapters and updated the Stata syntax for release 17. Major additions in volume 1 include the Kenward–Roger degree-of-freedom correction for improved inference with a small number of clusters, difference-in-difference estimation for quasi experiments, and instrumental-variables estimation to handle level-1 endogeneity. In volume 2, we now introduce Bayesian estimation for crossed-effects models and extensively use several new commands (since the third edition), including `meologit`, `cmxtmixlogit`, `mestreg`, and `menbreg`.

Berkeley and Oslo Sophia Rabe-Hesketh
August 2021 Anders Skrondal