Preface

This is a book 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 the 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 since 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 due to 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.

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. Wherever possible, we therefore introduce ideas through examples and graphical illustrations, keeping the technical descriptions as simple as possible, often confining formulas to subsections that can be skipped. Some sections that go beyond an introductory course on multilevel and longitudinal modeling are tagged with the symbol ❖ . For an advanced and comprehensive treatment, we refer 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 stand-alone 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 since 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. We walk through the analysis using Stata, pausing when statistical issues arise that need further explanation. 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, this book exclusively uses commands. However, the GUI can be useful for learning the Stata syntax. Generally, we use the typewriter font `command` 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.

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 basic introductions to Stata, we refer to Kohler and Kreuter (2005) or Rabe-Hesketh and Everitt (2007). Other books and further resources for learning Stata are listed at

http://www.stata.com

We have included applications from a wide range of disciplines, including medicine, economics, education, sociology, and psychology. The interdisciplinary nature of the 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. We encourage users to write do-files for solving the data analysis exercises since this is standard practice for professional data analysis.

All datasets used in this book are freely available for download from

http://www.gllamm.org/books

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 http://www.stata-press.com/data/mlmus2/pefr
```

If you have stored the datasets in a local directory, omit the path, and type

```
. use pefr
```

We will generally describe all Stata commands that can be used for a given problem, discussing their advantages and disadvantages. We make extensive use of the Stata commands `xtreg`, `xtmixed`, `xtlogit`, `xtpoisson`, `xtmelogit`, `xtmepoisson`, and our own command `gllamm`. Using `xtmixed` requires Stata release 9 or later, `xtmelogit` and `xtmepoisson` require release 10 or later, and `gllamm` requires that the program be installed. If you are connected to the Internet, the easiest way of installing `gllamm` is by issuing the Stata command:

```
. ssc install gllamm
```

Since `gllamm` is not part of official Stata and therefore only mentioned but not described in the Stata manuals, we include detailed descriptions of the syntax for `gllamm` and its postestimation commands `gllapred` and `gllasim` in the appendices. For quick and easy reference, we have placed the bare essentials in appendix A. We refer to the *Stata Longitudinal/Panel-Data Reference Manual* (StataCorp 2007) for detailed information on all the official Stata commands for multilevel and longitudinal modeling.

We assume that readers have a good knowledge of linear regression modeling, in particular the use and interpretation of dummy variables and interactions. However, we have included a new first chapter in this edition which reviews linear regression and can serve as a refresher.

The book consists of four parts: I Preliminaries, II Two-level linear models, III Two-level generalized linear models, and IV Models with nested and crossed random effects. Part I is a review of linear regression modeling and prepares the reader for the rest of the book. 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. The remaining chapters can then be read in any order, except that chapter 6 should be read before chapter 7 and chapter 10 before chapter 11. A course on linear models could cover part II as well as the first half of chapter 10, and perhaps chapter 11, in part IV.

Errata for different editions and printings of the book can be downloaded from http://www.stata-press.com/books/errata/mlmus2.html, and answers to exercises are available for instructors; see http://www.stata-press.com/books/mlmus2-answers.html.

The second edition of the book includes 3 new chapters, comprehensive updates for Stata 10, 38 new exercises and 27 new datasets. All chapters of the previous edition have been substantially revised.

Berkeley and London   Sophia Rabe-Hesketh
December 2007   Anders Skrondal