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 class of multilevel models for longitudinal data is known as growth-curve models.

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.

Each chapter is based on one or more research problems and real datasets. We walk trough 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 (2004). Other resources for learning Stata are listed at *http://www.stata.com*.

We have included applications from different disciplines, including medicine, economics, education, 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 statisticians.

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/mlmus/pefr, clear
```

If the datasets are in a local directory, omit the path, and type

```
. use pefr, clear
```

We have collaborated for many years and developed a general model framework called generalized linear latent and mixed models (GLLAMM) together with Andrew Pickles. The methodology is described in our recent book (Skrondal and Rabe-Hesketh 2004) and implemented in the Stata program `gllamm` (Rabe-Hesketh, Skrondal, and Pickles 2004). The current book is a hands-on introduction to using `gllamm` and other Stata commands for one important class of models within the framework: multilevel and longitudinal models. We hope that similar books on measurement and structural-equation modeling will materialize in the future.

This book makes extensive use of the Stata commands `xtmixed` and `gllamm`. The former requires Stata release 9 or later, and the latter requires that `gllamm` be installed. The easiest way of installing `gllamm` is by issuing the Stata command:

```
. ssc install gllamm
```

The `gllamm` command was developed to handle a wide array of response types, including continuous, ordered, and unordered categorical responses and counts. Due to this generality, the estimation is implemented using numerical integration, which is computationally demanding. However, for the special case of continuous responses, there are computationally efficient methods readily available, and this approach is implemented in the very useful `xtmixed` command. We will therefore focus on using `xtmixed` for

continuous responses in this book. However, for didactic reasons, we also introduce `gllamm` in this setting and occasionally use it to obtain results that are not available in `xtmixed` at the time this book was published. For dichotomous and ordinal responses and counts, there are no such computational shortcuts, and although other Stata commands (e.g., `xtlogit`) are available for simple models, we will almost exclusively use `gllamm`. This is because `gllamm` is more general and has more options for predictions and diagnostics.

Detailed descriptions of the syntax for `gllamm` and its postestimation commands `gllapred` and `gllasim` are given in the appendices. For quick and easy reference, we have placed the bare essentials in appendix A.

For readers who are new to multilevel and longitudinal modeling, the first four chapters should be read sequentially and can form the basis of an introductory course on this topic. The remaining four chapters could then be read in any order, allowing readers or course organizers to choose the topics they find most interesting. For a course on linear models, the first three chapters, the first half of chapter 7, and perhaps chapter 8 could be used.

Berkeley and London
July 2005

Sophia Rabe-Hesketh
Anders Skrondal