

# 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](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  $\blacklozenge$  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 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 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  
August 2021*

Sophia Rabe-Hesketh  
Anders Skrondal



# Contents

	List of tables	xv
	List of figures	xvii
	List of displays	xxiii
	Preface	xxv
	Multilevel and longitudinal models: When and why?	1
<b>I</b>	<b>Preliminaries</b>	<b>9</b>
<b>1</b>	<b>Review of linear regression</b>	<b>11</b>
	1.1 Introduction . . . . .	11
	1.2 Is there gender discrimination in faculty salaries? . . . . .	11
	1.3 Independent-samples t test . . . . .	12
	1.4 One-way analysis of variance . . . . .	17
	1.5 Simple linear regression . . . . .	19
	1.6 Dummy variables . . . . .	27
	1.7 Multiple linear regression . . . . .	30
	1.8 Interactions . . . . .	38
	1.9 Dummy variables for more than two groups . . . . .	44
	1.10 Other types of interactions . . . . .	50
	1.10.1 Interaction between dummy variables . . . . .	50
	1.10.2 Interaction between continuous covariates . . . . .	53
	1.11 Nonlinear effects . . . . .	54
	1.12 Residual diagnostics . . . . .	58
	1.13 ❖ Causal and noncausal interpretations of regression coefficients . .	59
	1.13.1 Regression as conditional expectation . . . . .	59

1.13.2	Regression as structural model . . . . .	60
1.14	Summary and further reading . . . . .	62
1.15	Exercises . . . . .	63
<b>II</b>	<b>Two-level models</b>	<b>73</b>
<b>2</b>	<b>Variance-components models</b>	<b>75</b>
2.1	Introduction . . . . .	75
2.2	How reliable are peak-expiratory-flow measurements? . . . . .	76
2.3	Inspecting within-subject dependence . . . . .	77
2.4	The variance-components model . . . . .	79
2.4.1	Model specification . . . . .	79
2.4.2	Path diagram . . . . .	80
2.4.3	Between-subject heterogeneity . . . . .	82
2.4.4	Within-subject dependence . . . . .	82
	Intraclass correlation . . . . .	82
	Intraclass correlation versus Pearson correlation . . . . .	84
2.5	Estimation using Stata . . . . .	85
2.5.1	Data preparation: Reshaping from wide form to long form . . . . .	86
2.5.2	Using xtreg . . . . .	87
2.5.3	Using mixed . . . . .	89
2.6	Hypothesis tests and confidence intervals . . . . .	91
2.6.1	Hypothesis test and confidence interval for the population mean . . . . .	91
2.6.2	Hypothesis test and confidence interval for the between-cluster variance . . . . .	95
	Likelihood-ratio test . . . . .	95
	❖ Score test . . . . .	96
	F test . . . . .	99
	Confidence intervals . . . . .	99
2.7	❖ Model as data-generating mechanism . . . . .	100
2.8	Fixed versus random effects . . . . .	102

2.9	Crossed versus nested effects . . . . .	104
2.10	Parameter estimation . . . . .	106
2.10.1	Model assumptions . . . . .	106
	Mean structure and covariance structure . . . . .	107
	Distributional assumptions . . . . .	108
2.10.2	Different estimation methods . . . . .	108
2.10.3	Inference for $\beta$ . . . . .	110
	Estimate and standard error: Balanced case . . . . .	110
	Estimate: Unbalanced case . . . . .	113
2.11	Assigning values to the random intercepts . . . . .	113
2.11.1	Maximum “likelihood” estimation . . . . .	114
	Implementation via OLS . . . . .	114
	Implementation via the mean total residual . . . . .	116
2.11.2	Empirical Bayes prediction . . . . .	116
2.11.3	Empirical Bayes standard errors . . . . .	122
	Posterior and comparative standard errors . . . . .	122
	Diagnostic standard errors . . . . .	123
	Accounting for uncertainty in $\hat{\beta}$ . . . . .	124
2.11.4	❖ Bayesian interpretation of REML estimation and prediction . . . . .	125
2.12	Summary and further reading . . . . .	125
2.13	Exercises . . . . .	126
<b>3</b>	<b>Random-intercept models with covariates</b>	<b>133</b>
3.1	Introduction . . . . .	133
3.2	Does smoking during pregnancy affect birthweight? . . . . .	133
3.2.1	Data structure and descriptive statistics . . . . .	135
3.3	The linear random-intercept model with covariates . . . . .	138
3.3.1	Model specification . . . . .	138
3.3.2	Model assumptions . . . . .	139
3.3.3	Mean structure . . . . .	140



3.3.4	Residual covariance structure . . . . .	141
3.3.5	Graphical illustration of random-intercept model . . . . .	141
3.4	Estimation using Stata . . . . .	144
3.4.1	Using xtreg . . . . .	144
3.4.2	Using mixed . . . . .	147
3.5	Coefficients of determination or variance explained . . . . .	148
3.6	Hypothesis tests and confidence intervals . . . . .	152
3.6.1	Hypothesis tests for individual regression coefficients . . . . .	152
3.6.2	Joint hypothesis tests for several regression coefficients . . . . .	154
3.6.3	Predicted means and confidence intervals . . . . .	156
3.6.4	Hypothesis test for random-intercept variance . . . . .	158
3.7	Between and within effects of level-1 covariates . . . . .	159
3.7.1	Between-mother effects . . . . .	159
3.7.2	Within-mother effects . . . . .	162
3.7.3	Relations among estimators . . . . .	164
3.7.4	Level-2 endogeneity and cluster-level confounding . . . . .	165
3.7.5	Conventional Hausman test . . . . .	168
3.7.6	Allowing for different within and between effects . . . . .	170
3.7.7	Robust Hausman test . . . . .	175
3.8	Fixed versus random effects revisited . . . . .	176
3.9	Assigning values to random effects: Residual diagnostics . . . . .	178
3.10	More on statistical inference . . . . .	181
3.10.1	❖ Overview of estimation methods . . . . .	181
	Pooled OLS . . . . .	181
	Feasible generalized least squares (FGLS) . . . . .	181
	ML by iterative GLS (IGLS) . . . . .	183
	ML by Newton–Raphson and Fisher scoring . . . . .	183
	ML by the expectation-maximization (EM) algorithm . . . . .	183
	REML . . . . .	184

3.10.2	Consequences of using standard regression modeling for clustered data . . . . .	185
	Purely between-cluster covariate . . . . .	185
	Purely within-cluster covariate . . . . .	186
3.10.3	❖ Power and sample-size determination . . . . .	186
	Purely between-cluster covariate . . . . .	189
	Purely within-cluster covariate . . . . .	190
3.11	Summary and further reading . . . . .	190
3.12	Exercises . . . . .	192
<b>4</b>	<b>Random-coefficient models</b>	<b>201</b>
4.1	Introduction . . . . .	201
4.2	How effective are different schools? . . . . .	201
4.3	Separate linear regressions for each school . . . . .	202
4.4	Specification and interpretation of a random-coefficient model . . .	208
	4.4.1 Specification of a random-coefficient model . . . . .	208
	4.4.2 Interpretation of the random-effects variances and covariances . . . . .	211
4.5	Estimation using mixed . . . . .	214
	4.5.1 Random-intercept model . . . . .	214
	4.5.2 Random-coefficient model . . . . .	216
4.6	Testing the slope variance . . . . .	219
4.7	Interpretation of estimates . . . . .	220
4.8	Assigning values to the random intercepts and slopes . . . . .	222
	4.8.1 Maximum “likelihood” estimation . . . . .	222
	4.8.2 Empirical Bayes prediction . . . . .	224
	4.8.3 Model visualization . . . . .	225
	4.8.4 Residual diagnostics . . . . .	227
	4.8.5 Inferences for individual schools . . . . .	229
4.9	Two-stage model formulation . . . . .	232
4.10	Some warnings about random-coefficient models . . . . .	235
	4.10.1 Meaningful specification . . . . .	235

4.10.2	Many random coefficients . . . . .	235
4.10.3	Convergence problems . . . . .	236
4.10.4	Lack of identification . . . . .	237
4.11	Summary and further reading . . . . .	238
4.12	Exercises . . . . .	239
<b>III</b>	<b>Models for longitudinal and panel data</b>	<b>247</b>
	<b>Introduction to models for longitudinal and panel data (part III)</b>	<b>249</b>
<b>5</b>	<b>Subject-specific effects, endogeneity, and unobserved confounding</b>	<b>269</b>
5.1	Introduction . . . . .	269
5.2	Random-effects approach: No endogeneity . . . . .	272
5.3	Fixed-effects approach: Level-2 endogeneity . . . . .	274
5.3.1	De-meaning and subject dummies . . . . .	274
	De-meaning . . . . .	275
	Subject dummies . . . . .	278
5.3.2	Hausman test . . . . .	279
5.3.3	Mundlak approach and robust Hausman test . . . . .	280
5.3.4	First-differencing . . . . .	282
5.4	Difference-in-differences and repeated-measures ANOVA . . . . .	284
5.4.1	Does raising the minimum wage reduce employment? . . . . .	284
5.4.2	❖ Repeated-measures ANOVA . . . . .	288
5.5	Subject-specific coefficients . . . . .	290
5.5.1	Random-coefficient model: No endogeneity . . . . .	290
5.5.2	Fixed-coefficient model: Level-2 endogeneity . . . . .	292
5.6	Hausman–Taylor: Level-2 endogeneity for level-1 and level-2 co- variates . . . . .	294
5.7	Instrumental-variable methods: Level-1 (and level-2) endogeneity . . . . .	299
5.7.1	Do deterrents decrease crime rates? . . . . .	299
5.7.2	Conventional fixed-effects approach . . . . .	300
5.7.3	Fixed-effects IV estimator . . . . .	302
5.7.4	Random-effects IV estimator . . . . .	304

5.7.5	More Hausman tests . . . . .	306
5.8	Dynamic models . . . . .	308
5.8.1	Dynamic model without subject-specific intercepts . . . . .	309
5.8.2	Dynamic model with subject-specific intercepts . . . . .	312
5.9	Missing data and dropout . . . . .	318
5.9.1	❖ Maximum likelihood estimation under MAR: A simulation . . . . .	319
5.10	Summary and further reading . . . . .	322
5.11	Exercises . . . . .	323
<b>6</b>	<b>Marginal models</b>	<b>333</b>
6.1	Introduction . . . . .	333
6.2	Mean structure . . . . .	333
6.3	Covariance structures . . . . .	334
6.3.1	Unstructured covariance matrix . . . . .	339
6.3.2	Random-intercept or compound symmetric/exchangeable structure . . . . .	343
6.3.3	Random-coefficient structure . . . . .	345
6.3.4	Autoregressive and exponential structures . . . . .	348
6.3.5	Moving-average residual structure . . . . .	351
6.3.6	Banded and Toeplitz structures . . . . .	353
6.4	Hybrid and complex marginal models . . . . .	357
6.4.1	Random effects and correlated level-1 residuals . . . . .	357
6.4.2	Heteroskedastic level-1 residuals over occasions . . . . .	358
6.4.3	Heteroskedastic level-1 residuals over groups . . . . .	360
6.4.4	Different covariance matrices over groups . . . . .	362
6.5	Comparing the fit of marginal models . . . . .	363
6.6	Generalized estimating equations (GEE) . . . . .	367
6.7	Marginal modeling with few units and many occasions . . . . .	369
6.7.1	Is a highly organized labor market beneficial for economic growth? . . . . .	370
6.7.2	Marginal modeling for long panels . . . . .	370

6.7.3	Fitting marginal models for long panels in Stata . . . . .	371
6.8	Summary and further reading . . . . .	374
6.9	Exercises . . . . .	376
<b>7</b>	<b>Growth-curve models</b>	<b>387</b>
7.1	Introduction . . . . .	387
7.2	How do children grow? . . . . .	387
7.2.1	Observed growth trajectories . . . . .	388
7.3	Models for nonlinear growth . . . . .	389
7.3.1	Polynomial models . . . . .	389
	Estimation using mixed . . . . .	390
	Predicting the mean trajectory . . . . .	393
	Predicting trajectories for individual children . . . . .	396
7.3.2	Piecewise linear models . . . . .	398
	Estimation using mixed . . . . .	399
	Predicting the mean trajectory . . . . .	402
7.4	Two-stage model formulation and cross-level interaction . . . . .	403
7.5	Heteroskedasticity . . . . .	405
7.5.1	Heteroskedasticity at level 1 . . . . .	405
7.5.2	Heteroskedasticity at level 2 . . . . .	407
7.6	How does reading improve from kindergarten through third grade? . . . . .	409
7.7	Growth-curve model as a structural equation model . . . . .	409
7.7.1	Estimation using sem . . . . .	411
7.7.2	Estimation using mixed . . . . .	417
7.8	Summary and further reading . . . . .	420
7.9	Exercises . . . . .	421
<b>IV</b>	<b>Models with nested and crossed random effects</b>	<b>429</b>
<b>8</b>	<b>Higher-level models with nested random effects</b>	<b>431</b>
8.1	Introduction . . . . .	431
8.2	Do peak-expiratory-flow measurements vary between methods within subjects? . . . . .	432

8.3	Inspecting sources of variability . . . . .	434
8.4	Three-level variance-components models . . . . .	435
8.5	Different types of intraclass correlation . . . . .	438
8.6	Estimation using mixed . . . . .	439
8.7	Empirical Bayes prediction . . . . .	442
8.8	Testing variance components . . . . .	443
8.9	Crossed versus nested random effects revisited . . . . .	445
8.10	Does nutrition affect cognitive development of Kenyan children? . .	447
8.11	Describing and plotting three-level data . . . . .	448
	8.11.1 Data structure and missing data . . . . .	448
	8.11.2 Level-1 variables . . . . .	449
	8.11.3 Level-2 variables . . . . .	451
	8.11.4 Level-3 variables . . . . .	451
	8.11.5 Plotting growth trajectories . . . . .	452
8.12	Three-level random-intercept model . . . . .	453
	8.12.1 Model specification: Reduced form . . . . .	453
	8.12.2 Model specification: Three-stage formulation . . . . .	454
	8.12.3 Estimation using mixed . . . . .	455
8.13	Three-level random-coefficient models . . . . .	458
	8.13.1 Random coefficient at the child level . . . . .	458
	Estimation using mixed . . . . .	458
	8.13.2 Random coefficient at the child and school levels . . . . .	460
	Estimation using mixed . . . . .	460
8.14	Residual diagnostics and predictions . . . . .	462
8.15	Summary and further reading . . . . .	466
8.16	Exercises . . . . .	468
<b>9</b>	<b>Crossed random effects</b>	<b>483</b>
9.1	Introduction . . . . .	483
9.2	How does investment depend on expected profit and capital stock?	484
9.3	A two-way error-components model . . . . .	485

9.3.1	Model specification . . . . .	485
9.3.2	Residual variances, covariances, and intraclass correlations .	486
	Longitudinal correlations . . . . .	486
	Cross-sectional correlations . . . . .	486
9.3.3	Estimation using mixed . . . . .	487
9.3.4	Prediction . . . . .	491
9.4	How much do primary and secondary schools affect attainment at age 16? . . . . .	493
9.5	Data structure . . . . .	494
9.6	Additive crossed random-effects model . . . . .	496
9.6.1	Specification . . . . .	496
9.6.2	Intraclass correlations . . . . .	497
9.6.3	Estimation using mixed . . . . .	497
9.7	Crossed random-effects model with random interaction . . . . .	501
9.7.1	Model specification . . . . .	501
9.7.2	Intraclass correlations . . . . .	501
9.7.3	Estimation using mixed . . . . .	502
9.7.4	Testing variance components . . . . .	505
9.7.5	Some diagnostics . . . . .	507
9.8	❖ A trick requiring fewer random effects . . . . .	509
9.9	Summary and further reading . . . . .	512
9.10	Exercises . . . . .	513
<b>A</b>	<b>Useful Stata commands</b>	<b>525</b>
	<b>References</b>	<b>527</b>
	<b>Author index</b>	<b>541</b>
	<b>Subject index</b>	<b>547</b>

# Contents

List of tables	xvii
List of figures	xix
List of displays	xxv
<b>V</b> Models for categorical responses	<b>555</b>
<b>10</b> Dichotomous or binary responses	<b>557</b>
10.1 Introduction	557
10.2 Single-level logit and probit regression models for dichotomous responses	557
10.2.1 Generalized linear model formulation	558
Labor-participation data	561
Estimation using logit	561
Estimation using glm	565
10.2.2 Latent-response formulation	566
Logistic regression	568
Probit regression	568
Estimation using probit	569
10.3 Which treatment is best for toenail infection?	571
10.4 Longitudinal data structure	571
10.5 Proportions and fitted population-averaged or marginal probabilities	573
Estimation using logit	575
10.6 Random-intercept logistic regression	577
10.6.1 Model specification	577
Reduced-form specification	577
Two-stage formulation	578



10.6.2	Model assumptions . . . . .	578
10.6.3	Estimation . . . . .	579
	Using xtlogit . . . . .	580
	Using melogit . . . . .	584
	Using gllamm . . . . .	585
10.7	Subject-specific or conditional versus population-averaged or marginal relationships . . . . .	586
10.8	Measures of dependence and heterogeneity . . . . .	590
10.8.1	Conditional or residual intraclass correlation of the latent responses . . . . .	590
10.8.2	Median odds ratio . . . . .	591
10.8.3	❖ Measures of association for observed responses at me- dian fixed part of the model . . . . .	592
10.9	Inference for random-intercept logistic models . . . . .	594
10.9.1	Tests and confidence intervals for odds ratios . . . . .	594
10.9.2	Tests of variance components . . . . .	595
10.10	Maximum likelihood estimation . . . . .	596
10.10.1	❖ Adaptive quadrature . . . . .	596
10.10.2	Some speed and accuracy considerations . . . . .	599
	Integration methods and number of quadrature points . . . . .	599
	Starting values . . . . .	601
	Using melogit and gllamm for collapsible data . . . . .	602
	Spherical quadrature in gllamm . . . . .	602
10.11	Assigning values to random effects . . . . .	603
10.11.1	Maximum “likelihood” estimation . . . . .	603
10.11.2	Empirical Bayes prediction . . . . .	604
10.11.3	Empirical Bayes modal prediction . . . . .	606
10.12	Different kinds of predicted probabilities . . . . .	608
10.12.1	Predicted population-averaged or marginal probabilities . . . . .	608
10.12.2	Predicted subject-specific probabilities . . . . .	609

	Predictions for hypothetical subjects: Conditional probabilities . . . . .	609
	Predictions for the subjects in the sample: Posterior mean probabilities . . . . .	611
10.13	Other approaches to clustered dichotomous data . . . . .	617
10.13.1	Conditional logistic regression . . . . .	617
	Estimation using clogit . . . . .	618
10.13.2	Generalized estimating equations (GEE) . . . . .	619
	Estimation using xtgee . . . . .	620
10.14	Summary and further reading . . . . .	622
10.15	Exercises . . . . .	624
<b>11</b>	<b>Ordinal responses</b>	<b>635</b>
11.1	Introduction . . . . .	635
11.2	Single-level cumulative models for ordinal responses . . . . .	635
11.2.1	Generalized linear model formulation . . . . .	636
11.2.2	Latent-response formulation . . . . .	637
11.2.3	Proportional odds . . . . .	641
11.2.4	❖ Identification . . . . .	642
11.3	Are antipsychotic drugs effective for patients with schizophrenia? . . . . .	645
11.4	Longitudinal data structure and graphs . . . . .	645
11.4.1	Longitudinal data structure . . . . .	646
11.4.2	Plotting cumulative proportions . . . . .	647
11.4.3	Plotting cumulative sample logits and transforming the time scale . . . . .	648
11.5	Single-level proportional-odds model . . . . .	650
11.5.1	Model specification . . . . .	650
	Estimation using ologit . . . . .	651
11.6	Random-intercept proportional-odds model . . . . .	654
11.6.1	Model specification . . . . .	654
	Estimation using meologit . . . . .	654
	Estimation using gllamm . . . . .	655

11.6.2	Measures of dependence and heterogeneity . . . . .	657
	Residual intraclass correlation of latent responses . . . . .	657
	Median odds ratio . . . . .	657
11.7	Random-coefficient proportional-odds model . . . . .	658
11.7.1	Model specification . . . . .	658
	Estimation using meologit . . . . .	658
	Estimation using gllamm . . . . .	660
11.8	Different kinds of predicted probabilities . . . . .	662
11.8.1	Predicted population-averaged or marginal probabilities . . . . .	662
11.8.2	Predicted subject-specific probabilities: Posterior mean . . . . .	665
11.9	Do experts differ in their grading of student essays? . . . . .	669
11.10	A random-intercept probit model with grader bias . . . . .	670
11.10.1	Model specification . . . . .	670
	Estimation using gllamm . . . . .	670
11.11	Including grader-specific measurement-error variances . . . . .	672
11.11.1	Model specification . . . . .	672
	Estimation using gllamm . . . . .	673
11.12	❖ Including grader-specific thresholds . . . . .	675
11.12.1	Model specification . . . . .	675
	Estimation using gllamm . . . . .	676
11.13	❖ Other link functions . . . . .	681
	Cumulative complementary log–log model . . . . .	681
	Continuation-ratio logit model . . . . .	681
	Adjacent-category logit model . . . . .	683
	Baseline-category logit and stereotype models . . . . .	683
11.14	Summary and further reading . . . . .	684
11.15	Exercises . . . . .	685
<b>12</b>	<b>Nominal responses and discrete choice</b> . . . . .	<b>695</b>
12.1	Introduction . . . . .	695
12.2	Single-level models for nominal responses . . . . .	696

12.2.1	Multinomial logit models . . . . .	696
	Transport data version 1 . . . . .	700
	Estimation using mlogit . . . . .	701
12.2.2	Conditional logit models with alternative-specific covariates	705
	Transport data version 2: Expanded form . . . . .	707
	Estimation using clogit . . . . .	710
	Estimation using cmclogit . . . . .	711
12.2.3	Conditional logit models with alternative- and unit-specific covariates . . . . .	712
	Estimation using clogit . . . . .	713
	Estimation using cmclogit . . . . .	714
12.3	Independence from irrelevant alternatives . . . . .	715
12.4	Utility-maximization formulation . . . . .	716
12.5	Does marketing affect choice of yogurt? . . . . .	718
12.6	Single-level conditional logit models . . . . .	721
	12.6.1 Conditional logit models with alternative-specific intercepts	721
	Estimation using clogit . . . . .	722
	Estimation using cmclogit . . . . .	723
12.7	Multilevel conditional logit models . . . . .	723
	12.7.1 Preference heterogeneity: Brand-specific random intercepts	724
	Estimation using cmxtmixlogit . . . . .	725
	Estimation using gllamm . . . . .	731
	12.7.2 Response heterogeneity: Marketing variables with ran- dom coefficients . . . . .	735
	Estimation using cmxtmixlogit . . . . .	737
	Estimation using gllamm . . . . .	739
	12.7.3 ♦ Preference and response heterogeneity . . . . .	741
	Estimation using cmxtmixlogit . . . . .	742
	Estimation using gllamm . . . . .	744
12.8	Prediction of marginal choice probabilities . . . . .	745

12.9	Prediction of random effects and household-specific choice probabilities . . . . .	747
12.10	Summary and further reading . . . . .	751
12.11	Exercises . . . . .	753
<b>VI</b>	<b>Models for counts</b>	<b>761</b>
<b>13</b>	<b>Counts</b>	<b>763</b>
13.1	Introduction . . . . .	763
13.2	What are counts? . . . . .	763
13.2.1	Counts versus proportions . . . . .	763
13.2.2	Counts as aggregated event-history data . . . . .	764
13.3	Single-level Poisson models for counts . . . . .	765
13.4	Did the German healthcare reform reduce the number of doctor visits? . . . . .	767
13.5	Longitudinal data structure . . . . .	767
13.6	Single-level Poisson regression . . . . .	768
13.6.1	Model specification . . . . .	768
	Estimation using poisson . . . . .	769
	Estimation using glm . . . . .	771
13.7	Random-intercept Poisson regression . . . . .	772
13.7.1	Model specification . . . . .	772
13.7.2	Measures of dependence and heterogeneity . . . . .	773
13.7.3	Estimation . . . . .	773
	Using xtpoisson . . . . .	773
	Using mepoisson . . . . .	775
	Using gllamm . . . . .	776
13.8	Random-coefficient Poisson regression . . . . .	778
13.8.1	Model specification . . . . .	778
	Estimation using mepoisson . . . . .	779
	Estimation using gllamm . . . . .	782
13.9	Overdispersion in single-level models . . . . .	784

13.9.1	Normally distributed random intercept . . . . .	784
	Estimation using <code>xtpoisson</code> . . . . .	785
13.9.2	Negative binomial models . . . . .	786
	Mean dispersion or NB2 . . . . .	786
	Constant dispersion or NB1 . . . . .	788
13.9.3	Quasilikelihood . . . . .	788
	Estimation using <code>glm</code> . . . . .	789
13.10	Level-1 overdispersion in two-level models . . . . .	790
13.10.1	Random-intercept Poisson model with robust standard errors . . . . .	791
	Estimation using <code>mepoisson</code> . . . . .	791
13.10.2	Three-level random-intercept model . . . . .	792
13.10.3	Negative binomial models with random intercepts . . . . .	792
	Estimation using <code>menbreg</code> . . . . .	793
13.10.4	The HHG model . . . . .	794
13.11	Other approaches to two-level count data . . . . .	794
13.11.1	Conditional Poisson regression . . . . .	794
	Estimation using <code>xtpoisson, fe</code> . . . . .	796
	Estimation using Poisson regression with dummy variables for clusters . . . . .	796
13.11.2	Conditional negative binomial regression . . . . .	797
13.11.3	Generalized estimating equations . . . . .	797
	Estimation using <code>xtgee</code> . . . . .	798
13.12	Marginal and conditional effects when responses are MAR . . . . .	799
	❖ Simulation . . . . .	799
13.13	Which Scottish counties have a high risk of lip cancer? . . . . .	803
13.14	Standardized mortality ratios . . . . .	804
13.15	Random-intercept Poisson regression . . . . .	806
13.15.1	Model specification . . . . .	806
	Estimation using <code>gllamm</code> . . . . .	807
13.15.2	Prediction of standardized mortality ratios . . . . .	808

13.16	❖ Nonparametric maximum likelihood estimation . . . . .	811
13.16.1	Specification . . . . .	811
	Estimation using gllamm . . . . .	811
13.16.2	Prediction . . . . .	816
13.17	Summary and further reading . . . . .	816
13.18	Exercises . . . . .	818
<b>VII</b>	<b>Models for survival or duration data</b>	<b>827</b>
	<b>Introduction to models for survival or duration data (part VII)</b>	<b>829</b>
<b>14</b>	<b>Discrete-time survival</b>	<b>835</b>
14.1	Introduction . . . . .	835
14.2	Single-level models for discrete-time survival data . . . . .	835
14.2.1	Discrete-time hazard and discrete-time survival . . . . .	835
	Promotions data . . . . .	835
14.2.2	Data expansion for discrete-time survival analysis . . . . .	838
14.2.3	Estimation via regression models for dichotomous responses . . . . .	840
	Estimation using logit . . . . .	842
14.2.4	Including time-constant covariates . . . . .	845
	Estimation using logit . . . . .	846
14.2.5	Including time-varying covariates . . . . .	849
	Estimation using logit . . . . .	853
14.2.6	Multiple absorbing events and competing risks . . . . .	855
	Estimation using mlogit . . . . .	858
14.2.7	Handling left-truncated data . . . . .	860
14.3	How does mother's birth history affect child mortality? . . . . .	861
14.4	Data expansion . . . . .	862
14.5	❖ Proportional hazards and interval-censoring . . . . .	864
14.6	Complementary log-log models . . . . .	865
14.6.1	Marginal baseline hazard . . . . .	866
	Estimation using cloglog . . . . .	867

14.6.2	Including covariates . . . . .	868
	Estimation using cloglog . . . . .	869
14.7	Random-intercept complementary log–log model . . . . .	872
14.7.1	Model specification . . . . .	872
	Estimation using mecloglog . . . . .	872
14.8	❖ Population-averaged or marginal vs. cluster-specific or conditional survival probabilities . . . . .	875
14.9	Summary and further reading . . . . .	879
14.10	Exercises . . . . .	880
<b>15</b>	<b>Continuous-time survival</b>	<b>887</b>
15.1	Introduction . . . . .	887
15.2	What makes marriages fail? . . . . .	888
15.3	Hazards and survival . . . . .	889
15.4	Proportional hazards models . . . . .	895
15.4.1	Piecewise exponential model . . . . .	897
	Estimation using streg . . . . .	900
	Estimation using poisson . . . . .	905
15.4.2	Cox regression model . . . . .	906
	Estimation using stcox . . . . .	907
15.4.3	Cox regression via Poisson regression for expanded data . . . . .	910
	Estimation using xtpoisson, fe . . . . .	911
15.4.4	Approximate Cox regression: Poisson regression, smooth baseline hazard . . . . .	911
	Estimation using poisson . . . . .	912
15.5	Accelerated failure-time models . . . . .	914
15.5.1	Log-normal model . . . . .	916
	Estimation using streg . . . . .	917
	Estimation using stintreg . . . . .	919
15.6	Time-varying covariates . . . . .	920
	Estimation using streg . . . . .	923
15.7	Does nitrate reduce the risk of angina pectoris? . . . . .	924



15.8	Marginal modeling . . . . .	926
15.8.1	Cox regression with occasion-specific dummy variables . . .	927
	Estimation using <code>stcox</code> . . . . .	928
15.8.2	Cox regression with occasion-specific baseline hazards . . .	929
	Estimation using <code>stcox, strata</code> . . . . .	930
15.8.3	Approximate Cox regression . . . . .	930
	Estimation using <code>poisson</code> . . . . .	932
15.9	Multilevel proportional hazards models . . . . .	933
15.9.1	Cox regression with gamma shared frailty . . . . .	934
	Estimation using <code>stcox, shared</code> . . . . .	935
15.9.2	Approximate Cox regression with log-normal shared frailty	938
	Estimation using <code>mepoisson</code> . . . . .	939
15.9.3	Approximate Cox regression with normal random intercept and coefficient . . . . .	940
	Estimation using <code>mepoisson</code> . . . . .	942
15.10	Multilevel accelerated failure-time models . . . . .	943
15.10.1	Log-normal model with gamma shared frailty . . . . .	943
	Estimation using <code>streg</code> . . . . .	944
15.10.2	Log-normal model with log-normal shared frailty . . . . .	945
	Estimation using <code>mestreg</code> . . . . .	946
15.10.3	Log-normal model with normal random intercept and random coefficient . . . . .	947
	Estimation using <code>mestreg</code> . . . . .	948
15.11	Fixed-effects approach . . . . .	949
15.11.1	Stratified Cox regression with subject-specific baseline hazards . . . . .	949
	Estimation using <code>stcox, strata</code> . . . . .	950
15.12	❖ Different approaches to recurrent-event data . . . . .	951
15.12.1	Total-time risk interval . . . . .	952
15.12.2	Counting-process risk interval . . . . .	956
15.12.3	Gap-time risk interval . . . . .	958

15.13 Summary and further reading . . . . . 959  
 15.14 Exercises . . . . . 960

**VIII Models with nested and crossed random effects 969**

**16 Models with nested and crossed random effects 971**

16.1 Introduction . . . . . 971  
 16.2 Did the Guatemalan-immunization campaign work? . . . . . 971  
 16.3 A three-level random-intercept logistic regression model . . . . . 973  
     16.3.1 Model specification . . . . . 974  
     16.3.2 Measures of dependence and heterogeneity . . . . . 974  
         Types of residual intraclass correlations of the latent  
             responses . . . . . 974  
         Types of median odds ratios . . . . . 975  
     16.3.3 Three-stage formulation . . . . . 975  
     16.3.4 Estimation . . . . . 976  
         Using melogit . . . . . 976  
         Using gllamm . . . . . 980  
 16.4 A three-level random-coefficient logistic regression model . . . . . 984  
     16.4.1 Estimation . . . . . 985  
         Using melogit . . . . . 985  
         Using gllamm . . . . . 988  
 16.5 Prediction of random effects . . . . . 991  
     16.5.1 Empirical Bayes prediction . . . . . 991  
     16.5.2 Empirical Bayes modal prediction . . . . . 993  
 16.6 Different kinds of predicted probabilities . . . . . 994  
     16.6.1 Predicted population-averaged or marginal probabilities:  
         New clusters . . . . . 994  
     16.6.2 Predicted median or conditional probabilities . . . . . 995  
     16.6.3 Predicted posterior mean probabilities: Existing clusters . . . . . 996  
 16.7 Do salamanders from different populations mate successfully? . . . . . 998  
 16.8 Crossed random-effects logistic regression . . . . . 1000

16.8.1	Setup for estimating crossed random-effects model using melogit . . . . .	1001
16.8.2	Approximate maximum likelihood estimation . . . . .	1003
	Estimation using melogit . . . . .	1003
16.8.3	Bayesian estimation . . . . .	1007
	Brief introduction to Bayesian inference . . . . .	1007
	Priors for the salamander data . . . . .	1010
	Estimation using bayes: melogit . . . . .	1011
16.8.4	Estimates compared . . . . .	1019
16.8.5	Fully Bayesian versus empirical Bayesian inference for random effects . . . . .	1020
16.9	Summary and further reading . . . . .	1026
16.10	Exercises . . . . .	1027
<b>A</b>	<b>Syntax for gllamm, eq, and gllapred: The bare essentials</b>	<b>1035</b>
<b>B</b>	<b>Syntax for gllamm</b>	<b>1041</b>
<b>C</b>	<b>Syntax for gllapred</b>	<b>1053</b>
<b>D</b>	<b>Syntax for gllasim</b>	<b>1057</b>
	<b>References</b>	<b>1061</b>
	<b>Author index</b>	<b>1077</b>
	<b>Subject index</b>	<b>1085</b>



## 4 Random-coefficient models

### 4.1 Introduction

In the previous chapter, we considered linear random-intercept models where the overall level of the response was allowed to vary between clusters after controlling for covariates. In this chapter, we include random coefficients or random slopes in addition to random intercepts, thus also allowing the effects of covariates to vary between clusters. Such models involving both random intercepts and random slopes are often called *random-coefficient models*. In longitudinal settings, where the level-1 units are occasions and the clusters are typically subjects, models with a random-coefficient of time are also referred to as growth-curve models. Such models are the topic of chapter 7.

### 4.2 How effective are different schools?

Here we analyze a dataset on inner-London schools that accompanies the MLwiN software (Rasbash et al. 2019) and is part of the data analyzed by Goldstein et al. (1993).

At age 16, students took their Graduate Certificate of Secondary Education (GCSE) exams in a number of subjects. A score was derived from the individual exam results. Such scores often form the basis for school comparisons, for instance, to allow parents to choose the best school for their child. However, schools can differ considerably in their intake achievement levels. It may be argued that what should be compared is the “value added”; that is, the difference in mean GCSE score between schools after controlling for the students’ achievement before entering the school. One such measure of prior achievement is the London Reading Test (LRT) taken by these students at age 11.

The dataset `gcse.dta` has the following variables:

- `school`: school identifier
- `student`: student identifier
- `gcse`: Graduate Certificate of Secondary Education (GCSE) score ( $z$  score, multiplied by 10)
- `lrt`: London Reading Test (LRT) score ( $z$  score, multiplied by 10)
- `girl1`: dummy variable for student being a girl (1: girl; 0: boy)
- `schgend`: type of school (1: mixed gender; 2: boys only; 3: girls only)

One purpose of the analysis is to investigate the relationship between GCSE and LRT and how this relationship varies between schools. The model can then be used to address the question of which schools appear to be most effective, taking prior achievement into account.

We read in the data by using

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

### 4.3 Separate linear regressions for each school

Before developing a model for all 65 schools combined, we consider a separate model for each school. For school  $j$ , an obvious model for the relationship between GCSE and LRT is a simple regression model,

$$y_{ij} = \beta_{1j} + \beta_{2j}x_{ij} + \epsilon_{ij}$$

where  $y_{ij}$  is the GCSE score for the  $i$ th student in school  $j$ ,  $x_{ij}$  is the corresponding LRT score,  $\beta_{1j}$  is the school-specific intercept,  $\beta_{2j}$  is the school-specific slope, and  $\epsilon_{ij}$  is a residual error term with school-specific variance  $\theta_j$ .

For school 1, OLS estimates of the intercept  $\hat{\beta}_{11}$  and the slope  $\hat{\beta}_{21}$  can be obtained using `regress`,

```
. regress gcse lrt if school==1
```

Source	SS	df	MS	Number of obs	=	73
Model	4084.89189	1	4084.89189	F(1, 71)	=	59.44
Residual	4879.35759	71	68.7233463	Prob > F	=	0.0000
				R-squared	=	0.4557
				Adj R-squared	=	0.4480
Total	8964.24948	72	124.503465	Root MSE	=	8.29

gcse	Coefficient	Std. err.	t	P> t	[95% conf. interval]
lrt	.7093406	.0920061	7.71	0.000	.5258856 .8927955
_cons	3.833302	.9822377	3.90	0.000	1.874776 5.791828

where we have selected school 1 by specifying the condition `if school==1`.

To assess whether this is a reasonable model for school 1, we can obtain the predicted (ordinary least squares) regression line for this school (with  $j = 1$ ),

$$\hat{y}_{i1} = \hat{\beta}_{11} + \hat{\beta}_{21}x_{i1}$$

by using the `predict` command with the `xb` option:

```
. predict p_gcse, xb
```

We superimpose this line on the scatterplot of the data for the school, as shown in figure 4.1.

```
. twoway (scatter gcse lrt) (line p_gcse lrt, sort) if school==1,
> xtitle(LRT) ytitle(GCSE)
```

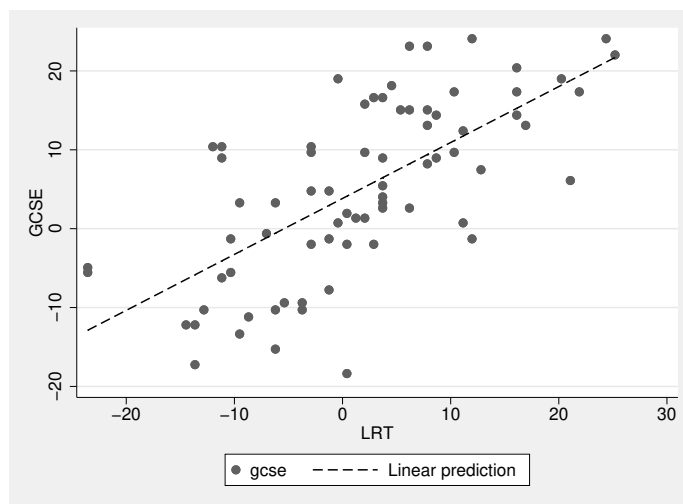


Figure 4.1: Scatterplot of `gcse` versus `lrt` for school 1 with ordinary least-squares regression line

We can also produce a *trellis graph* containing such plots for all 65 schools by using

```
. twoway (scatter gcse lrt) (lfit gcse lrt, sort lpatt(solid)),
> by(school, compact legend(off) cols(5))
> xtitle(LRT) ytitle(GCSE) ysize(3) xsize(2)
```

where `lfit` plots regression lines estimated by OLS. The resulting graph, shown in figure 4.2, suggests that the model assumptions are reasonably met. The schools appear to vary in both their intercepts and slopes.

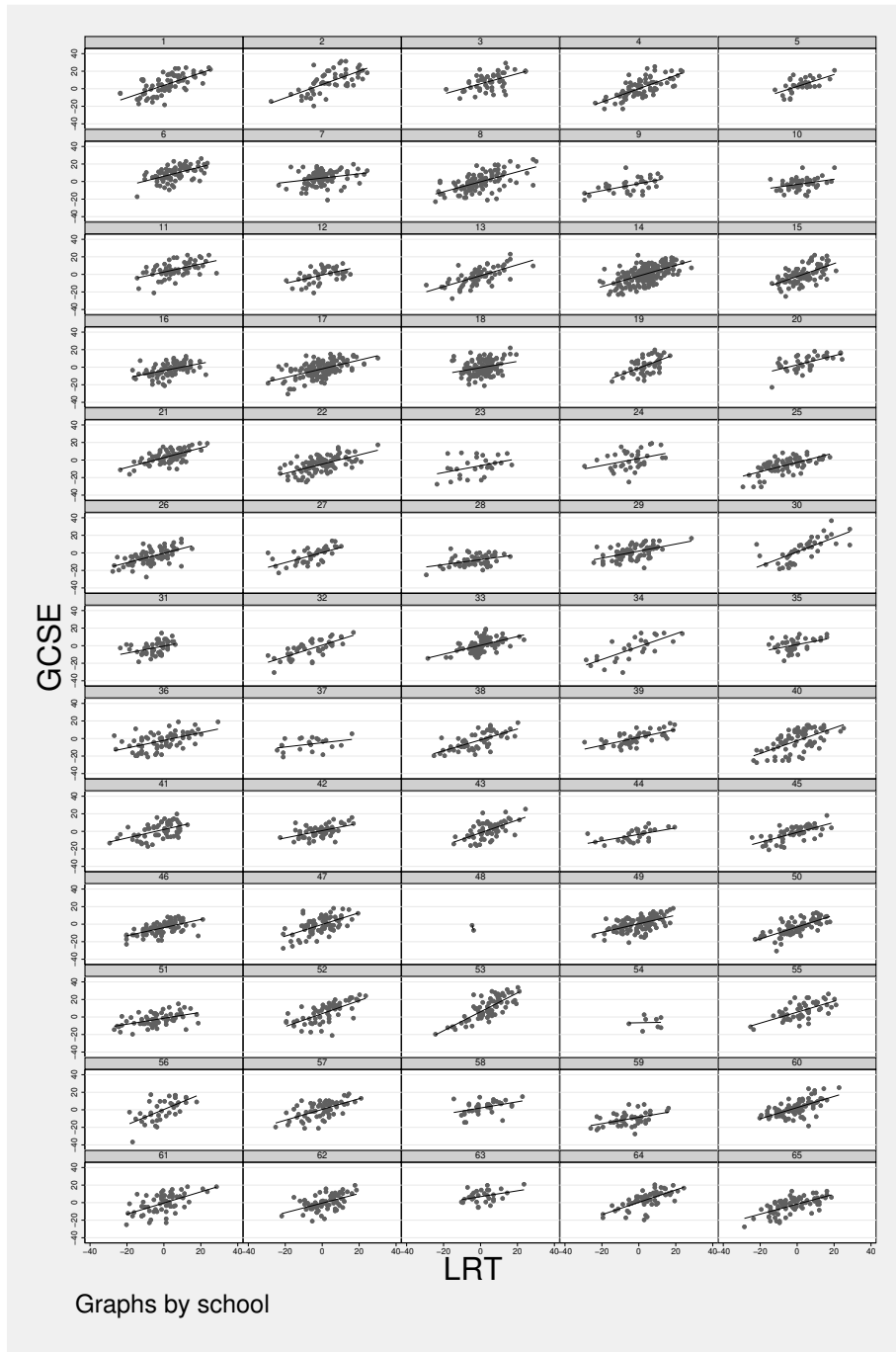


Figure 4.2: Trellis of scatterplots of `gcse` versus `lrt` with fitted regression lines for all 65 schools



We will now fit a simple linear regression model for each school, which is easily done using Stata's prefix command `statsby`. Then we will examine the variability in the estimated intercepts and slopes.

First, calculate the number of students per school by using `egen` with the `count()` function to preclude fitting lines to schools with fewer than five students below:

```
. egen num = count(gcse), by(school)
```

Then, use `statsby` to create a new dataset, `ols.dta`, in the working directory with the variables `inter` and `slope` containing OLS estimates of the intercepts (`_b[_cons]`) and slopes (`_b[lrt]`) from the command `regress gcse lrt if num>4` applied to each school:

```
. statsby inter=_b[_cons] slope=_b[lrt], by(school) saving(ols):
> regress gcse lrt if num>4
(running regress on estimation sample)
      Command: regress gcse lrt if num>4
      inter:  _b[_cons]
      slope:  _b[lrt]
      By:    school

Statsby groups
-----|-----|-----|-----|-----|-----|-----|-----
      1 | 2 | 3 | 4 | 5 |
.....|.....|.....|.....|.....|.....|.....|.....
.....|.....|.....|.....|.....|.....|.....|.....
.....|.....|.....|.....|.....|.....|.....|.....
```

The new dataset also contains the variable `school` and is sorted by `school`, making it easy to merge it into the original dataset (the “master data”) after sorting the latter by `school`:

```
. sort school
. merge m:1 school using ols
      Result              Number of obs
-----|-----|-----|-----|-----|-----|-----|-----
      Not matched              2
      from master              2  (_merge==1)
      from using              0  (_merge==2)
      Matched              4,057  (_merge==3)

. drop _merge
```

Here we have specified `m:1` in the `merge` command, which stands for “many-to-one merging” (observations for several students per school in the master data, but only one observation per school in the “using data”). We see that two of the students in the master data did not have matches in the using data (because their school, school 48, had fewer than 5 students in the data, so we did not compute OLS estimates for that school). We have deleted the variable `_merge` produced by the `merge` command to avoid error messages when we run the `merge` command in the future.

A scatterplot of the OLS estimates of the intercept and slope is produced using the following command and given in figure 4.3:

```
. twoway scatter slope inter, xtitle(Intercept) ytitle(Slope)
```

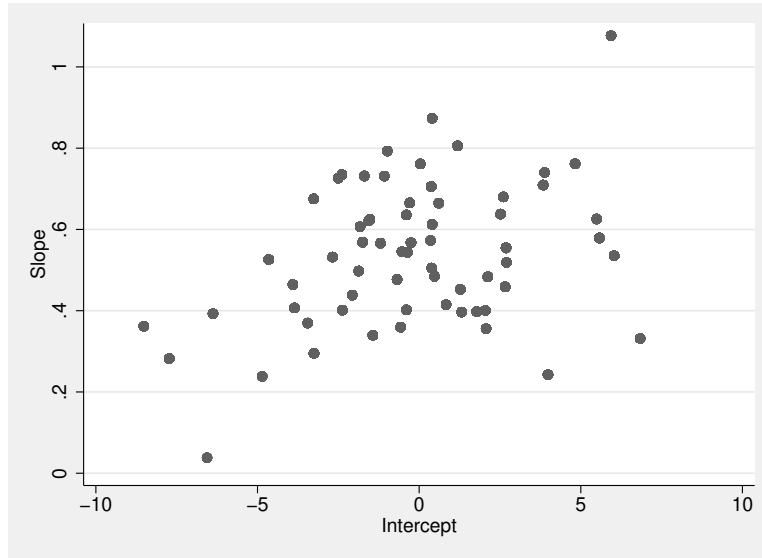


Figure 4.3: Scatterplot of estimated intercepts and slopes for all schools with at least five students

We see that there is considerable variability between the estimated intercepts and slopes of different schools. To investigate this further, we first create a dummy variable to pick out one observation per school,

```
. egen pickone = tag(school)
```

and then produce summary statistics for the schools by using the `summarize` command:

```
. summarize inter slope if pickone==1
```

Variable	Obs	Mean	Std. dev.	Min	Max
inter	64	-.1805974	3.291357	-8.519253	6.838716
slope	64	.5390514	.1766135	.0380965	1.076979

To allow comparison with the parameter estimates obtained from the random-coefficient model considered later on, we also obtain the covariance matrix of the estimated intercepts and slopes:

```
. correlate inter slope if pickone==1, covariance  
(obs=64)
```

	inter	slope
inter	10.833	
slope	.208622	.031192

The diagonal elements, 10.83 and 0.03, are the sample variances of the intercepts and slopes, respectively. The off-diagonal element, 0.21, is the sample covariance between the intercepts and slopes, equal to the correlation times the product of the intercept and slope standard deviations.

We can also obtain a *spaghetti plot* of the predicted school-specific regression lines for all schools. We first calculate the fitted values  $\hat{y}_{ij} = \hat{\beta}_{1j} + \hat{\beta}_{2j}x_{ij}$ ,

```
. generate pred = inter + slope*lrt
(2 missing values generated)
```

and sort the data so that `lrt` increases within a given school and then jumps to its lowest value for the next school in the dataset:

```
. sort school lrt
```

We then produce the plot by typing

```
. twoway (line pred lrt, connect(ascending)), xtitle(LRT)
> ytitle(Fitted regression lines)
```

The `connect(ascending)` option is used to connect points only as long as `lrt` is increasing and ensures that only data for the same school are connected. The resulting graph is shown in figure 4.4.

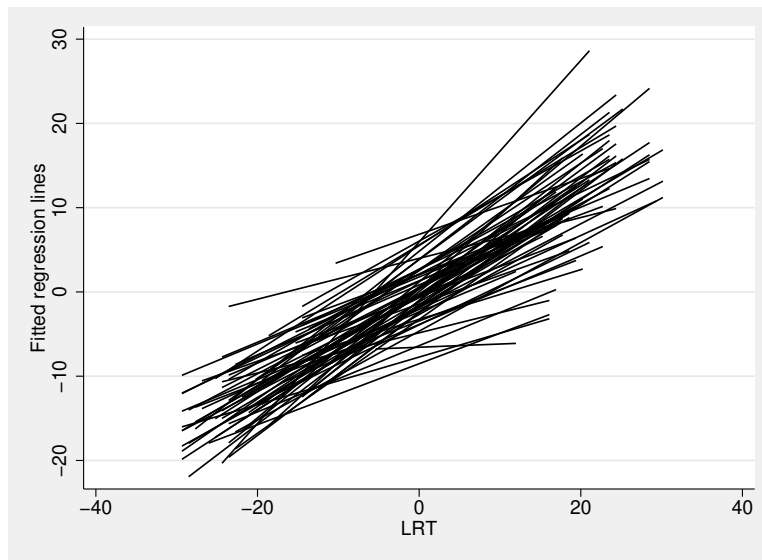


Figure 4.4: Spaghetti plot of ordinary least-squares regression lines for all schools with at least five students

## 4.4 Specification and interpretation of a random-coefficient model

### 4.4.1 Specification of a random-coefficient model

How can we develop a joint model for the relationships between `gcse` and `lrt` in all schools that allows intercepts and slopes to differ between schools?

One way would be to use dummy variables for all schools (omitting the overall intercept) and interactions between these dummy variables and `lrt` (omitting the overall slope of `lrt`). The school-specific intercepts are then the coefficients of the dummy variables and the school-specific slopes are the interaction coefficients. The only difference between the resulting model and separate regressions is that a common residual error variance  $\theta_j = \theta$  is assumed. However, this model has 130 regression coefficients! Furthermore, if the schools are viewed as a (random) sample of schools from a population of schools, we are not interested in the individual coefficients characterizing each school's regression line. Rather, we would like to estimate the mean intercept and slope as well as the (co)variability of the intercepts and slopes in the population of schools.

A parsimonious model for the relationships between `gcse` and `lrt` can be obtained by specifying a school-specific random intercept  $\zeta_{1j}$  and a school-specific random slope  $\zeta_{2j}$  for `lrt` ( $x_{ij}$ ):

$$\begin{aligned} y_{ij} &= \beta_1 + \beta_2 x_{ij} + \zeta_{1j} + \zeta_{2j} x_{ij} + \epsilon_{ij} \\ &= (\beta_1 + \zeta_{1j}) + (\beta_2 + \zeta_{2j}) x_{ij} + \epsilon_{ij} \end{aligned} \quad (4.1)$$

Here  $\zeta_{1j}$  represents the deviation of school  $j$ 's intercept from the mean intercept  $\beta_1$ , and  $\zeta_{2j}$  represents the deviation of school  $j$ 's slope from the mean slope  $\beta_2$ .

Given all covariates  $\mathbf{X}_j$  in cluster  $j$ , it is assumed that the random effects  $\zeta_{1j}$  and  $\zeta_{2j}$  have zero expectations:

$$E(\zeta_{1j} | \mathbf{X}_j) = 0$$

$$E(\zeta_{2j} | \mathbf{X}_j) = 0$$

It is also assumed that the level-1 residual  $\epsilon_{ij}$  has zero expectation, given the covariates and the random effects:

$$E(\epsilon_{ij} | \mathbf{X}_j, \zeta_{1j}, \zeta_{2j}) = 0$$

It follows from these mean-independence assumptions that the random terms  $\zeta_{1j}$ ,  $\zeta_{2j}$ , and  $\epsilon_{ij}$  are all uncorrelated with the covariate  $x_{ij}$  and with  $\bar{x}_{.j}$  and that  $\epsilon_{ij}$  is uncorrelated with both  $\zeta_{1j}$  and  $\zeta_{2j}$ . Both the intercepts  $\zeta_{1j}$  and slopes  $\zeta_{2j}$  are assumed to be uncorrelated across schools, and the level-1 residuals  $\epsilon_{ij}$  are assumed to be uncorrelated across schools and students.

An illustration of this random-coefficient model with one covariate  $x_{ij}$  for one cluster  $j$  is shown in the bottom panel of figure 4.5. A random-intercept model is shown for comparison in the top panel.

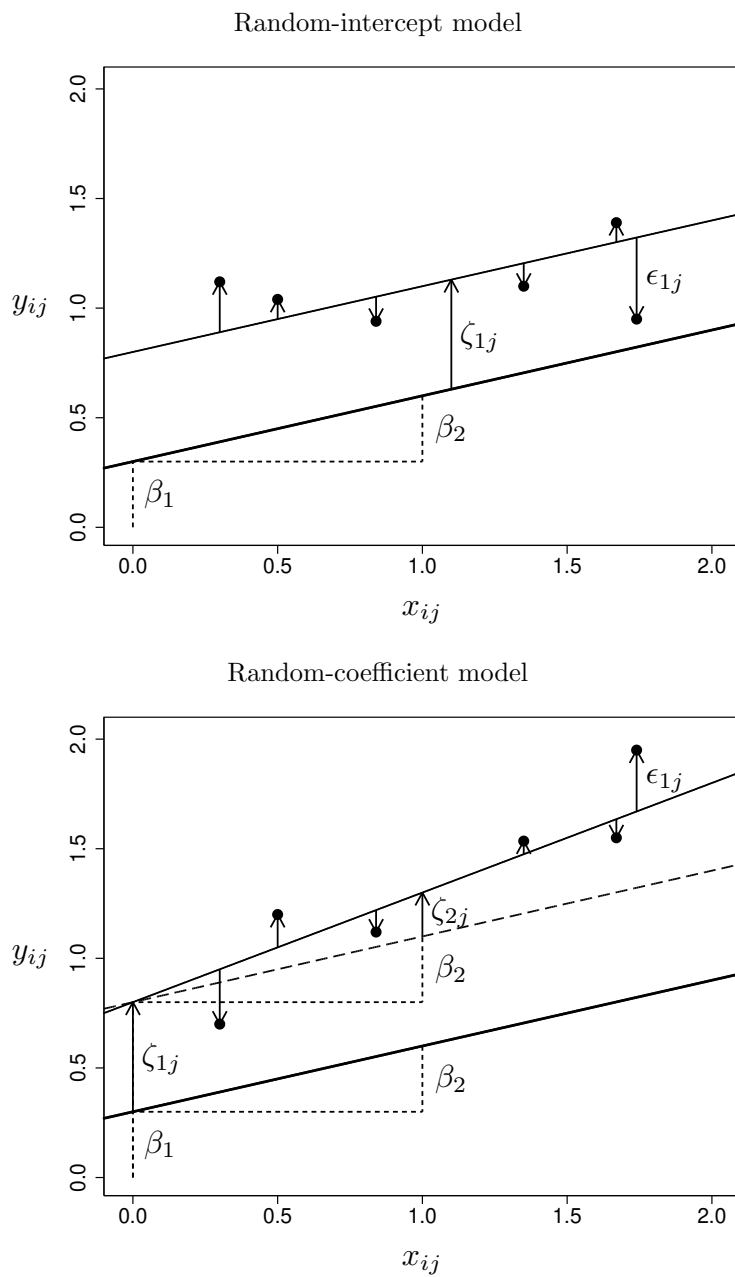


Figure 4.5: Illustration of random-intercept and random-coefficient models

In each panel, the lower bold and solid line represents the population-averaged or marginal regression line

$$E(y_{ij}|x_{ij}) = \beta_1 + \beta_2 x_{ij}$$

across all clusters. The higher and thinner solid line represents the cluster-specific regression line for cluster  $j$ . The arrows from the cluster-specific regression lines to the responses  $y_{ij}$  are the within-cluster residual error terms  $\epsilon_{ij}$  (with variance  $\theta$ ).

For the random-intercept model, the cluster-specific line is

$$E(y_{ij}|x_{ij}, \zeta_{1j}) = (\beta_1 + \zeta_{1j}) + \beta_2 x_{ij}$$

which is parallel to the population-averaged line with vertical displacement given by the random intercept  $\zeta_{1j}$ . In contrast, in the random-coefficient model, the cluster-specific or conditional regression line

$$E(y_{ij}|x_{ij}, \zeta_{1j}, \zeta_{2j}) = (\beta_1 + \zeta_{1j}) + (\beta_2 + \zeta_{2j})x_{ij}$$

is not parallel to the population-averaged line but has a greater slope because the random slope  $\zeta_{2j}$  is positive in the illustration. Here the dashed line is parallel to the population-averaged regression line and has the same intercept as cluster  $j$ . The vertical deviation between this dashed line and the line for cluster  $j$  is  $\zeta_{2j}x_{ij}$ , as shown in the diagram for  $x_{ij}=1$ . The bottom panel illustrates that the total intercept for cluster  $j$  is  $\beta_1 + \zeta_{1j}$  and the total slope is  $\beta_2 + \zeta_{2j}$ . It is clear that  $\zeta_{2j}x_{ij}$  represents an *interaction* between the clusters, treated as random, and the covariate  $x_{ij}$ .

Given  $\mathbf{X}_j$ , the random intercept and random slope have a bivariate distribution assumed to have 0 means and covariance matrix  $\Psi$ :

$$\Psi = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix} \equiv \begin{bmatrix} \text{Var}(\zeta_{1j}|\mathbf{X}_j) & \text{Cov}(\zeta_{1j}, \zeta_{2j}|\mathbf{X}_j) \\ \text{Cov}(\zeta_{2j}, \zeta_{1j}|\mathbf{X}_j) & \text{Var}(\zeta_{2j}|\mathbf{X}_j) \end{bmatrix}, \quad \psi_{21} = \psi_{12}$$

Hence, given the covariates, the variance of the random intercept is  $\psi_{11}$ , the variance of the random slope is  $\psi_{22}$ , and the covariance between the random intercept and the random slope is  $\psi_{21}$ . The correlation between the random intercept and random slope given the covariates becomes

$$\rho_{21} \equiv \text{Cor}(\zeta_{1j}, \zeta_{2j}|\mathbf{X}_j) = \frac{\psi_{21}}{\sqrt{\psi_{11}\psi_{22}}}$$

For maximum likelihood (ML) and restricted maximum likelihood (REML) estimation a normal distribution is specified for the level-1 error  $\epsilon_{ij}$  and a bivariate normal distribution for the random intercept and random slope, given  $\mathbf{X}_j$ . An example of a bivariate normal distribution with  $\psi_{11} = \psi_{22} = 4$  and  $\psi_{21} = \psi_{12} = 1$  is shown as a perspective plot in figure 4.6. Specifying a bivariate normal distribution implies that the (marginal) univariate distributions of the intercept and slope are also normal.

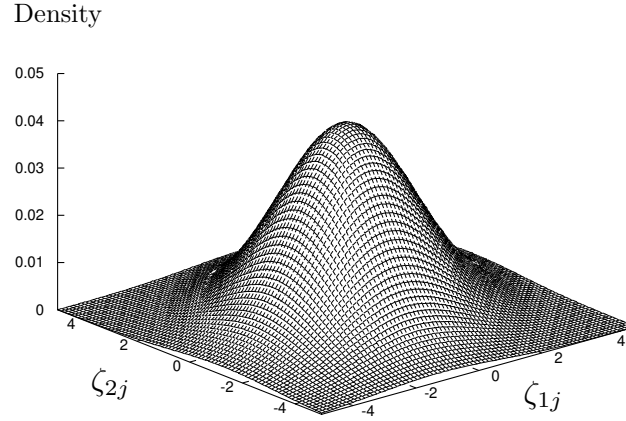


Figure 4.6: Perspective plot of bivariate normal distribution

#### 4.4.2 Interpretation of the random-effects variances and covariances

Interpreting the covariance matrix  $\Psi$  of the random effects (given the covariates  $\mathbf{X}_j$ ) is not completely straightforward.

First, the random-slope variance  $\psi_{22}$  and the covariance between random slope and intercept  $\psi_{21}$  depend not just on the scale of the response variable but also on the scale of the covariate, here `lrt`. Let the units of the response and covariate be denoted as  $u_y$  and  $u_x$ , respectively. For instance, in an application in chapter 7 that considers children's increase in weight over time,  $u_y$  is kilograms and  $u_x$  is years. The units of  $\psi_{11}$  are  $u_y^2$ , the units of  $\psi_{21}$  are  $u_y^2/u_x$ , and the units of  $\psi_{22}$  are  $u_y^2/u_x^2$ . It therefore does not make sense to compare the magnitude of random-intercept and random-slope variances.

Another issue is that the total residual variance is no longer constant as in random-intercept models. The total residual is now

$$\xi_{ij} \equiv \zeta_{1j} + \zeta_{2j}x_{ij} + \epsilon_{ij}$$

and the conditional variance of the responses given the covariate, or the conditional variance of the total residual, is

$$\text{Var}(y_{ij}|\mathbf{X}_j) = \text{Var}(\xi_{ij}|\mathbf{X}_j) = \psi_{11} + 2\psi_{21}x_{ij} + \psi_{22}x_{ij}^2 + \theta \quad (4.2)$$

This variance is a (quadratic) function of the covariate  $x_{ij}$ , and the total residual is therefore *heteroskedastic*. The conditional covariance for two students  $i$  and  $i'$  with covariate values  $x_{ij}$  and  $x_{i'j}$  in the same school  $j$  is

$$\begin{aligned}\text{Cov}(y_{ij}, y_{i'j} | \mathbf{X}_j) &= \text{Cov}(\xi_{ij}, \xi_{i'j} | \mathbf{X}_j) \\ &= \psi_{11} + \psi_{21}x_{ij} + \psi_{21}x_{i'j} + \psi_{22}x_{ij}x_{i'j}\end{aligned}\quad (4.3)$$

and the conditional intraclass correlation becomes

$$\text{Cor}(y_{ij}, y_{i'j} | \mathbf{X}_j) = \frac{\text{Cov}(\xi_{ij}, \xi_{i'j} | \mathbf{X}_j)}{\sqrt{\text{Var}(\xi_{ij} | \mathbf{X}_j)\text{Var}(\xi_{i'j} | \mathbf{X}_j)}}$$

where we can plug in the covariance from (4.3) and the variances from (4.2). When  $x_{ij} = x_{i'j} = 0$ , the expression for the intraclass correlation is the same as for the random-intercept model and represents the correlation of the total residuals (from the overall mean regression line) for two students in the same school who both have `lrt` scores equal to 0 (the mean in this case). However, for pairs of students  $i$  and  $i'$  in the same school  $j$  with other values of `lrt`, the intraclass correlation is a complicated function of `lrt` ( $x_{ij}$  and  $x_{i'j}$ ).

Due to the heteroskedastic total residual variance, it is not straightforward to define coefficients of determination—such as  $R^2$ ,  $R_2^2$ , and  $R_1^2$ , discussed in section 3.5—for random-coefficient models. Snijders and Bosker (2012, 114) suggest removing the random coefficient(s) for the purpose of calculating the coefficient of determination because this will usually yield values that are close to correct (see their section 7.2.2 for how to obtain the correct version).

Finally, interpreting the parameters  $\psi_{11}$  and  $\psi_{21}$  can be difficult because their values depend on the translation of the covariate or, in other words, on how much we add or subtract from the covariate. Adding a constant to `lrt` and refitting the model would result in different estimates of  $\psi_{11}$  and  $\psi_{21}$  (see also exercise 4.9). This is because the intercept variance is the variability in the vertical positions of school-specific regression lines where `lrt`=0 (the position where `lrt`=0 changes when `lrt` is translated) and the covariance or correlation is the tendency for regression lines that are higher up where `lrt`=0 to have higher slopes. This lack of invariance of  $\psi_{11}$  and  $\psi_{21}$  to translation of the covariate  $x_{ij}$  is illustrated in figure 4.7. Here identical cluster-specific regression lines are shown in the two panels, but the covariate  $x'_{ij} = x_{ij} - 3.5$  in the lower panel is translated relative to the covariate  $x_{ij}$  in the upper panel. The intercepts are the intersections of the regression lines with the vertical lines at 0. Clearly these intercepts vary more in the upper panel than the lower panel, whereas the correlation between intercepts and slopes is negative in the upper panel and positive in the lower panel.



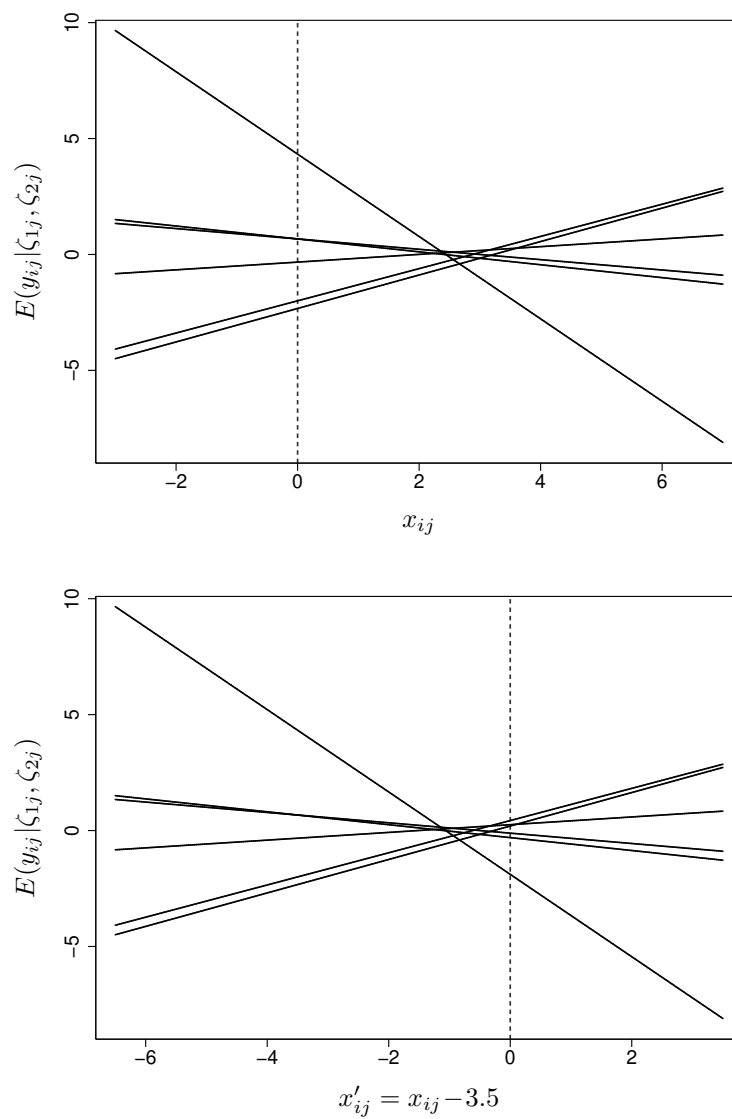


Figure 4.7: Cluster-specific regression lines for random-coefficient model, illustrating lack of invariance under translation of covariate (*Source*: Skrondal and Rabe-Hesketh 2004)

To make  $\psi_{11}$  and  $\psi_{21}$  interpretable, it makes sense to translate  $x_{ij}$  so that the value  $x_{ij} = 0$  is a useful reference point in some way. Typical choices are either mean centering (as for `lrm`) or, if  $x_{ij}$  is time, as in growth-curve models, defining 0 to be the initial time in some sense. Because the magnitude and interpretation of  $\psi_{21}$  depend on the location (or translation) of  $x_{ij}$ , which is often arbitrary, it generally does not make sense to set  $\psi_{21}$  to 0 by specifying uncorrelated intercepts and slopes.

A useful way of interpreting the magnitudes of the estimated variances  $\hat{\psi}_{11}$  and  $\hat{\psi}_{22}$  is by constructing intervals that contain the intercepts and slopes of 95% of clusters in the population (treating estimates as known parameters). Assuming that the intercepts and slopes are normally distributed with means  $\hat{\beta}_1$  and  $\hat{\beta}_2$  and variances  $\hat{\psi}_{11}$  and  $\hat{\psi}_{22}$ , these intervals are  $\hat{\beta}_1 \pm 1.96 \sqrt{\hat{\psi}_{11}}$  and  $\hat{\beta}_2 \pm 1.96 \sqrt{\hat{\psi}_{22}}$ . To aid interpretation of the random part of the model, it is also useful to produce plots of predicted school-specific regression lines, as discussed in section 4.8.3.

## 4.5 Estimation using mixed

The `mixed` command can be used to fit linear random-coefficient models by ML or REML. (`xtreg` can only fit two-level random-intercept models.)

### 4.5.1 Random-intercept model

We first consider a random-intercept model discussed in the previous chapter:

$$y_{ij} = (\beta_1 + \zeta_{1j}) + \beta_2 x_{ij} + \epsilon_{ij}$$

This model is a special case of the random-coefficient model in (4.1) with  $\zeta_{2j} = 0$  or, equivalently, with zero random-slope variance and zero random-intercept and random-slope covariance,  $\psi_{22} = \psi_{21} = 0$ .

ML estimates for the random-intercept model can be obtained using `mixed` with the `mle` option (the default), and we also use the `vce(robust)` option for robust standard errors:

```
. mixed gcse lrt || school:, mle stddeviations vce(robust)
Mixed-effects regression      Number of obs   =    4,059
Group variable: school        Number of groups =     65
                               Obs per group:
                               min =         2
                               avg =        62.4
                               max =        198
                               Wald chi2(1)   =    852.73
Log pseudolikelihood = -14024.799          Prob > chi2    =    0.0000
                               (Std. err. adjusted for 65 clusters in school)
```

gcse	Robust		z	P> z	[95% conf. interval]	
	Coefficient	std. err.				
lrt	.5633697	.0192925	29.20	0.000	.5255572	.6011823
_cons	.0238706	.4050143	0.06	0.953	-.7699428	.8176841

Random-effects parameters	Estimate	Robust std. err.	[95% conf. interval]	
school: Identity				
sd(_cons)	3.035269	.3154741	2.475863	3.72107
sd(Residual)	7.521481	.1306016	7.269813	7.781861

To allow later comparison with random-coefficient models via likelihood-ratio tests, we store these estimates by using

```
. estimates store ri
```

The random-intercept model assumes that the school-specific regression lines are parallel. The common coefficient or slope  $\beta_2$  of `lrt`, shared by all schools, is estimated as 0.56 and the mean intercept as 0.02. Schools vary in their intercepts with an estimated standard deviation of 3.04. Within the schools, the estimated residual standard deviation around the school-specific regression lines is 7.52. The within-school correlation, after controlling for `lrt`, is therefore estimated as

$$\hat{\rho} = \frac{\hat{\psi}_{11}}{\hat{\psi}_{11} + \hat{\theta}} = \frac{3.035^2}{3.035^2 + 7.521^2} = 0.14$$

We could obtain this within-school correlation by typing `estat icc`.

The ML estimates for the random-intercept model are also given under “Random intercept” in table 4.1.

Table 4.1: Maximum likelihood estimates for inner-London-schools data with robust standard errors

Parameter	Random intercept		Random coefficient		Rand. coefficient & level-2 covariates		$\gamma_{xx}$
	Est	(SE)	Est	(SE)	Est	(SE)	
Fixed part							
$\beta_1$ [_cons]	0.02	(0.41)	-0.12	(0.40)	-1.00	(0.55)	$\gamma_{11}$
$\beta_2$ [lrt]	0.56	(0.02)	0.56	(0.02)	0.57	(0.02)	$\gamma_{21}$
$\beta_3$ [boys]					0.85	(0.96)	$\gamma_{12}$
$\beta_4$ [girls]					2.43	(0.84)	$\gamma_{13}$
$\beta_5$ [boys_lrt]					-0.02	(0.05)	$\gamma_{22}$
$\beta_6$ [girls_lrt]					-0.03	(0.05)	$\gamma_{23}$
Random part							
$\sqrt{\psi_{11}}$	3.04		3.01		2.80		
$\sqrt{\psi_{22}}$			0.12		0.12		
$\rho_{21}$			0.50		0.60		
$\sqrt{\theta}$	7.52		7.44		7.44		
Log likelihood	-14,024.80		-14,004.61		-13,998.83		

### 4.5.2 Random-coefficient model

We now relax the assumption that the school-specific regression lines are parallel by introducing random school-specific slopes  $\beta_2 + \zeta_{2j}$  of `lrt`:

$$y_{ij} = (\beta_1 + \zeta_{1j}) + (\beta_2 + \zeta_{2j})x_{ij} + \epsilon_{ij}$$

To introduce a random slope for `lrt` using `mixed`, we simply add that variable name in the specification of the random part, replacing `school:` with `school: lrt`. We must also specify the `covariance(unstructured)` option because `mixed` will otherwise set the covariance  $\psi_{21}$  (and the corresponding correlation) to 0 by default. ML estimates for the random-coefficient model are then obtained using

```
. mixed gcse lrt || school: lrt, covariance(unstructured) mle stddeviations
> vce(robust)
Mixed-effects regression                Number of obs   =    4,059
Group variable: school                  Number of groups =     65
                                         Obs per group:
                                         min =         2
                                         avg =        62.4
                                         max =        198
                                         Wald chi2(1)   =    767.80
Log pseudolikelihood = -14004.613      Prob > chi2    =    0.0000
                                         (Std. err. adjusted for 65 clusters in school)
```

gcse	Robust		z	P> z	[95% conf. interval]	
	Coefficient	std. err.				
lrt	.556729	.0200919	27.71	0.000	.5173496	.5961084
_cons	-.115085	.4009294	-0.29	0.774	-.9008922	.6707222

Random-effects parameters	Estimate	Robust std. err.	[95% conf. interval]	
school: Unstructured				
sd(lrt)	.1205646	.0236268	.0821128	.1770224
sd(_cons)	3.007444	.3134589	2.451765	3.689065
corr(lrt,_cons)	.4975415	.1751841	.0894796	.7625783
sd(Residual)	7.440787	.1251535	7.19949	7.690172

Because the `stddeviations` option was used, the output shows the standard deviations, `sd(lrt)`, of the slope and `sd(_cons)` of the intercept instead of variances. It also shows the correlation between intercepts and slopes, `corr(lrt,_cons)`, instead of the covariance. We can obtain the estimated covariance matrix either by replaying the estimation results without the `stddeviations` option (or with the `variance` option),

```
mixed, variance
```

or by using the postestimation command `estat recovariance`:

```
. estat recovariance
Random-effects covariance matrix for level school
```

	lrt	_cons
lrt	.0145358	
_cons	.1804042	9.04472

The ML estimates for the random-coefficient model were also given under “Random coefficient” in table 4.1. We store the estimates under the name `rc` for later use:

```
. estimates store rc
```

We can also obtain the model-implied residual standard deviations and correlations among the GCSE scores for students in a particular school by using the `estat`

`wcorrelation` command. For schools that have many students in the data, the correlation matrix is too large to display without wrapping, so we choose school 54, which has 8 students in the data, for illustration. First, we sort the data in ascending order of `lrt` within `school` and list the values of `lrt` because they will affect both the standard deviations and correlations, as shown in equations (4.3) and (4.2):

```
. sort school lrt
. list school lrt if school==54, clean noobs
      school      lrt
      -----
      54    -5.3806
      54     2.058
      54     2.8845
      54     3.711
      54     9.4967
      54    10.323
      54    11.976
      54    11.976
```

Now, we obtain the estimated residual standard deviations and correlations for school 54:

```
. estat wcorrelation, at(school=54)
Standard deviations and correlations for school = 54:
Standard deviations:
      obs |      1      2      3      4      5      6      7      8
      ----|-----
      sd | 7.930  8.076  8.098  8.121  8.315  8.348  8.415  8.415
Correlations:
      obs |      1      2      3      4      5      6      7      8
      ----|-----
      1 | 1.000
      2 | 0.129  1.000
      3 | 0.130  0.153  1.000
      4 | 0.131  0.155  0.158  1.000
      5 | 0.137  0.170  0.173  0.177  1.000
      6 | 0.138  0.172  0.175  0.179  0.202  1.000
      7 | 0.139  0.176  0.179  0.183  0.208  0.212  1.000
      8 | 0.139  0.176  0.179  0.183  0.208  0.212  0.218  1.000
```

The standard deviations increase with increasing `lrt`. To interpret the pattern of the correlations, we can look down the columns, which corresponds to holding the `lrt` for one student constant and looking at the correlations as the `lrt` of the other student increases. We see that the corresponding correlations increase.

Here we used ML estimation. REML estimation should be used instead when the number of clusters is small ( $J - q < 42$ , see display 2.1) and this method is requested by specifying the `reml` option.

