we obtain so-called *cross-level interactions* in the reduced form

$$
\begin{aligned}
y_{ij} &= \underbrace{\gamma_{11}+\gamma_{12}w_{2j}+\gamma_{13}w_{3j}+\zeta_{1j}}_{\eta_{1j}} + \underbrace{(\gamma_{21}+\gamma_{22}w_{2j}+\gamma_{23}w_{3j}+\zeta_{2j})}_{\eta_{2j}}x_{ij}+\epsilon_{ij}\\
&= \underbrace{\gamma_{11}+\gamma_{12}w_{2j}+\gamma_{13}w_{3j}+\gamma_{21}x_{ij}+\gamma_{22}w_{2j}x_{ij}+\gamma_{23}w_{3j}x_{ij}}_{\text{fixed}} + \underbrace{\zeta_{1j}+\zeta_{2j}x_{ij}+\epsilon_{ij}}_{\text{random}}
\end{aligned}
$$

The effect of `lrt` now depends on the type of school, with γ_{22} representing the additional effect of `lrt` on `gcse` for boys' schools compared with mixed schools and γ_{23} the additional effect for girls' schools compared with mixed schools.

For estimation in `xtmixed` or `gllamm`, it is necessary to convert the two-stage formulation to the reduced form. For instance, to fit the model above using `xtmixed`, we must generate new variables representing interactions. First, we produce dummy variables for type of school:

```
. tabulate schgen, generate(w)

    schgend |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      2,169       53.44       53.44
          2 |        513       12.64       66.08
          3 |      1,377       33.92      100.00
------------+-----------------------------------
      Total |      4,059      100.00

. rename w2 boys

. rename w3 girls
```

Then we create the interaction terms:

```
. generate boys_lrt = boys*lrt

. generate girls_lrt = girls*lrt
```

Using `xtmixed`, we obtain

```
. xtmixed gcse lrt boys girls boys_lrt girls_lrt || school: lrt,
> cov(unstructured) mle

Mixed-effects ML regression                     Number of obs      =      4059
Group variable: school                          Number of groups   =        65

                                                Obs per group: min =         2
                                                               avg =      62.4
                                                               max =       198

                                                Wald chi2(5)       =    803.90
Log likelihood = -13998.825                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        gcse |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lrt |   .5712362   .0271255    21.06   0.000     .5180712    .6244012
        boys |   .8546725   1.085019     0.79   0.431    -1.271926    2.981271
       girls |   2.433411   .8433395     2.89   0.004     .7804959    4.086326
    boys_lrt |  -.0230097   .0573892    -0.40   0.688    -.1354905    .0894711
   girls_lrt |   -.029544   .0447029    -0.66   0.509    -.1171601    .0580721
       _cons |  -.9976069    .506808    -1.97   0.049    -1.990932   -.0042815
------------------------------------------------------------------------------
```

```
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Unstructured         |
                     sd(lrt) |   .1199144   .0189129      .0880277    .1633514
                   sd(_cons) |   2.797928   .2886796      2.285666    3.424997
             corr(lrt,_cons) |   .5967756   .1381306      .2614296    .8035692
-----------------------------+------------------------------------------------
                sd(Residual) |   7.441832   .0839662      7.279068    7.608236
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =   381.45   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

We see that students from girls' schools perform significantly better at the 5% level than students from mixed schools whereas students from boys' schools do not perform significantly better than students from mixed schools. The effect of `lrt` does not differ significantly between boys' schools and mixed schools or between girls' schools and mixed schools. The estimates and the parameters in the two-stage formulation are given in the last three columns of table 4.1 on page 155.

Although equivalent models can be specified using either the reduced-form (used in `xtmixed` and `gllamm`) or the two-stage formulation (used in the HLM software of Raudenbush et al. 2004), in practice model specification to some extent depends on the approach adopted. For instance, cross-level interactions are easily included using the two-stage specification in the HLM software, whereas same-level interactions must be created outside the program. Papers using this software tend to include more cross-level interactions and more random coefficients in the models (because the level-2 models look odd without residuals) than papers using for instance Stata.

Raudenbush and Bryk (2002) use slightly different notation for the two-stage formulation. For example, they use the subscript 0 instead of 1 for intercepts; see exercise 4.2.

4.10 Some warnings about random-coefficient models

It rarely makes sense to include a random slope if there is no random intercept, just like interactions between two variables usually do not make sense without including the variables themselves in ordinary regression models. Similarly, it is seldom sensible to include a random slope without including the corresponding fixed slope because it is usually strange to allow the slope to vary randomly but constrain its population mean to zero.

It may be tempting to allow many different covariates to have random slopes. However, it should be remembered that the number of parameters for the random part of the model increases rapidly with the number of random slopes because there is a variance parameter for each random effect (intercept or slope) and a covariance parameter for each pair of random effects. If there are k random slopes (plus 1 random intercept), then there are $(k+2)(k+1)/2+1$ parameters in the random part (for example, $k=3$ gives 11 parameters). Another problem is that clusters may not provide much information on cluster-specific slopes and hence on the slope variance either if the clusters are small or if x_{ij} does not vary much within clusters, or varies only in a small number of clusters. It should be noted that it does not matter if some of the clusters do not provide information on the slope variance as long as there are an adequate number of clusters that do.

It is generally not a good idea to include a random coefficient for a covariate that does not vary at a lower level than the random coefficient itself. For example, in the inner-London school data, it does not make sense to include a random slope for type of school because type of school does not vary within schools. Since we cannot estimate the effect of type of school for individual schools, it also appears impossible to estimate the variability of the effect of type of school between schools. However, as discussed in section 5.14, a level-2 random coefficient of a level-2 covariate can be used to construct a random-intercept model with a heteroskedastic random intercept.

Sometimes convergence problems will occur because the estimated covariance matrix "tries" to become nonpositive semidefinite, meaning for instance that variances try to become negative or correlations try to be greater than 1 or less than −1. All the commands in Stata force the covariance matrix to be positive semidefinite, and when parameters approach nonpermissible values, convergence can be slow or even fail. It may help to translate and rescale x_{ij} since variances and covariances are not invariant to these transformations. Often a better remedy is to simplify the model by removing some random slopes. The overall message is that random slopes should be included only if strongly suggested by the subject-matter theory related to the application.

Sometimes random-coefficient models are simply not identified. As an important example, consider balanced data with clusters of size $n_j = 2$ and with a covariate x_{ij} taking the same two values t_1 and t_2 for each cluster (an example would be the peak-expiratory-flow data, `pefr.dta`, from chapter 2). The model including a random intercept and a random slope of x_{ij} is not identified. This can be seen by considering the two distinct variances (for $i = 1$ and $i = 2$) and one covariance of the total residuals when $t_1 = 0$ and $t_2 = 1$:

$$
\begin{aligned}
\mathrm{Var}(\xi_{1j}) &= \psi_{11} + \theta \\
\mathrm{Var}(\xi_{2j}) &= \psi_{11} + 2\psi_{21} + \psi_{22} + \theta \\
\mathrm{Cov}(\xi_{1j}, \xi_{2j}) &= \psi_{11} + \psi_{21}
\end{aligned}
$$

The marginal distribution of y_{ij}, given the covariates, is completely characterized by the fixed part of the model and these three model-implied moments (two variances and a covariance). However, the three moments are determined by four parameters of the random part (ψ_{11}, ψ_{22}, ψ_{21}, and θ), so fitting the model-implied moments to the data would effectively involve solving three equations for four unknowns. The model is therefore not identified. We could identify the model by setting $\theta = 0$, which does not impose any restrictions on the covariance matrix. The original model becomes identified if the covariate x_{ij} varies also between clusters because the model-implied covariance matrix of the total residuals then differs between clusters, yielding more equations to solve for the four parameters.

4.11 Summary an

In this chapter, we have i
randomly over clusters i
vious representations of
with its own intercept a
cross-sectional study of
we considered school-sp

An important consi
terpretation of the cova
translation of the varial
terpreting the variance
formulation of random-
This formulation can b
using the more commor
pirical Bayes prediction
for individual clusters,

Random-coefficient
are considered in chap
viewed as supplementa
for noncontinuous resp

Introductory book
(2002), and Snijders a
chapters 2–4 is given i

4.12 Exercises

4.1 ❖ Inner-Lond

1. Fit the ran
2. Write dow
the mean
girl is a
hypothese
following
schools (a
schools th
3. Fit the m
whether t
sex school