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 over 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 over 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, random-coefficient models are also referred to as growth-curve models. Such models are discussed in chapter 5.

4.2 How effective are different schools?

We start by analyzing a dataset on inner-London schools that accompanies the MLwiN software (Rasbash et al. 2005) 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, and it may be argued that what should be compared is the "value added"; the difference in mean GCSE score between schools after controlling for 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)
- girl: dummy variable for child 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 are most effective, taking intake achievement into account.

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 ith student in school j, x_{ij} is the corresponding LRT score, β_{1j} the school-specific intercept, β_{2j} the school-specific slope, and ϵ_{ij} is a residual error term with school-specific variance θ_j.

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

```
. use http://www.stata-press.com/data/mlmus2/gcse
. regress gcse lrt if school==1
      Source |       SS       df       MS              Number of obs =      73
-------------+------------------------------           F(  1,    71) =   59.44
       Model |  4084.89209     1  4084.89209           Prob > F      =  0.0000
    Residual |  4879.35761    71  68.7233466           R-squared     =  0.4557
-------------+------------------------------           Adj R-squared =  0.4480
       Total |  8964.24969    72  124.503468           Root MSE      =    8.29

------------------------------------------------------------------------------
        gcse |      Coef.   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
------------------------------------------------------------------------------
```

To assess if this is a reasonable model, we can obtain a scatterplot of the data for school 1 with the corresponding regression line superimposed:

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

The resulting graph in figure 4.1 suggests that the model assumptions are reasonably met.

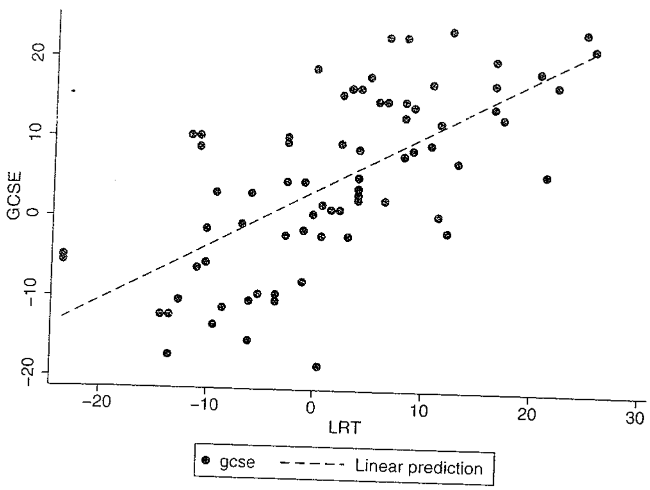

Figure 4.1: Scatterplot of gcse versus lrt for school 1 with regression line

We will now fit a simple linear regression model for each school, which is easily done using Stata's prefix command statsby, and examine the variability in the estimated intercepts and slopes. We first calculate the number of students per school to preclude fitting lines to schools with fewer than 5 students:

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

We then use statsby to create a new dataset ols.dta in the local directory with the variables inter and slope containing estimates of the intercepts (_b[_cons]) and slopes (_b[lrt]) from the command regress gcse lrt if num>4 applied to each school (as well as containing the variable 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
..................................................    50
.........
```

We can merge the estimates inter and slope into the gcse dataset using the merge command (after sorting the "master data" by school; the "using data" created by statsby are already sorted by school):

```
. sort school

. merge school using ols
variable school does not uniquely identify observations in the master data

. drop _merge
```

Here we have deleted the variable _merge produced by the merge command to avoid error messages when we want to run the merge command in the future.

A scatterplot is produced using the command

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

and given in figure 4.2. We see that there is considerable variability between the 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, whereas the off-diagonal element 0.21 is the sample covariance between the intercepts and slopes, the correlation times the product of the intercept and slope standard deviations.

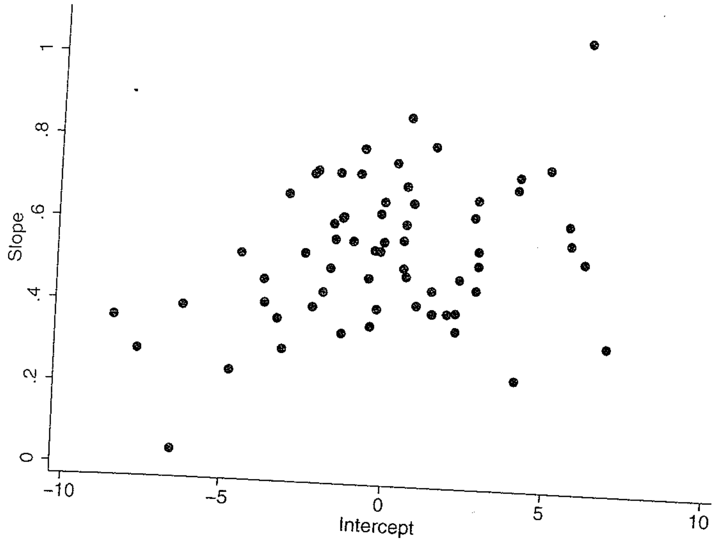

Figure 4.2: Scatterplot of intercepts and slopes for all schools with at least 5 students

We can also plot the predicted school-specific regression lines by first calculating the fitted values $\widehat{y}_{ij} = \widehat{\beta}_{1j} + \widehat{\beta}_{2j} x_{ij}$:

```
. generate pred = inter + slope*lrt
(2 missing values generated)
. sort school lrt
. twoway (line pred lrt, connect(ascending)), xtitle(LRT)
> ytitle(Fitted regression lines)
```

To produce the graph, we first sorted the data so that `lrt` increases within a given school and then jumps to its lowest value for the next school in the dataset. 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 4.3.

(*Continued on next page*)

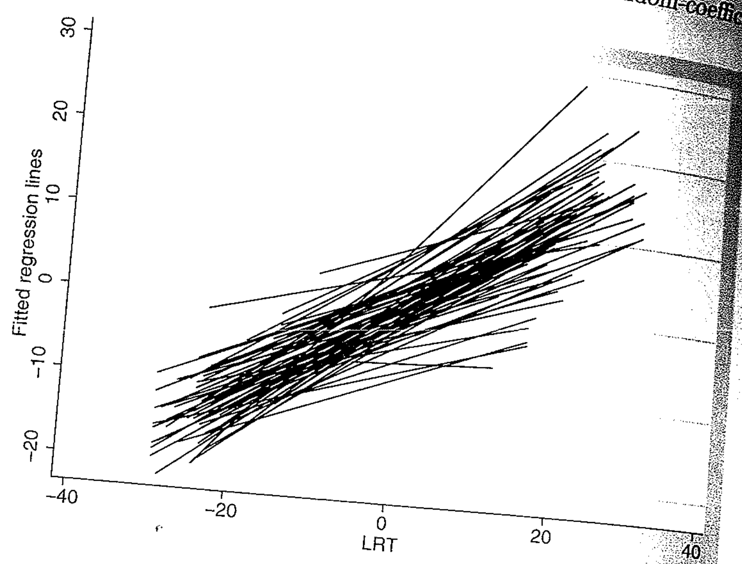

Figure 4.3: Least-squares regression lines for all schools with at least 5 students

4.4 Specification and interpretation of a random-coefficient model

4.4.1 Specification of random-coefficient model

How can we develop a joint model for the relationships between gcse and lrt in all schools?

One way would be to use dummy variables for all schools (omitting the overall constant) to estimate school-specific intercepts and interactions between these dummy variables and lrt to estimate school-specific slopes. 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 ζ_{1j} and a school-specific random slope ζ_{2j} for lrt (x_{ij}):

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

We assume that the covariate x_{ij} is exogenous with $E(\zeta_{1j}|x_{ij}) = 0$, $E(\zeta_{2j}|x_{ij}) = 0$, and $E(\epsilon_{ij}|x_{ij}, \zeta_{1j}, \zeta_{2j}) = 0$. Then ζ_{1j} represents the deviation of school j's intercept from the mean intercept β_1, and ζ_{2j} represents the deviation of school j's slope from the mean slope β_2. It follows from the zero expectations that all three random terms are uncorrelated with x_{ij} and that ϵ_{ij} is uncorrelated with both ζ_{1j} and ζ_{2j}. Both the intercepts ζ_{1j} and slopes ζ_{2j} are independent across schools and the level-1 residuals ϵ_{ij} are independent across schools and students.

An illustration of this random-coefficient model with one covariate x_{ij} for a school j is shown in the bottom panel of figure 4.4. A random-intercept model is shown for comparison in the top panel. In each panel, the lower, bold, solid line represents the population-averaged regression line

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

across all schools. The thinner solid line represents the school-specific regression line for school j. For the random-intercept model, this 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 ζ_{1j}. In contrast, in the random-coefficient model, the school-specific 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 ζ_{2j} is positive in the illustration. Here the dashed line is parallel to the population-averaged regression line and has the same intercept as school j. The vertical deviation between this dashed line and the line for school j is $\zeta_{2j}x_{ij}$, as shown in the diagram for $x_{ij}=1$. The bottom panel illustrates that the total intercept for school j is $\beta_1+\zeta_{1j}$ and the total slope is $\beta_2+\zeta_{2j}$. The arrows from the school-specific regression lines to the responses y_{ij} are the within-school residual error terms ϵ_{ij} (with variance θ). It is clear that $\zeta_j x_{ij}$ represents an *interaction* between the clusters, treated as random, and the covariate x_{ij}.

(*Continued on next page*)

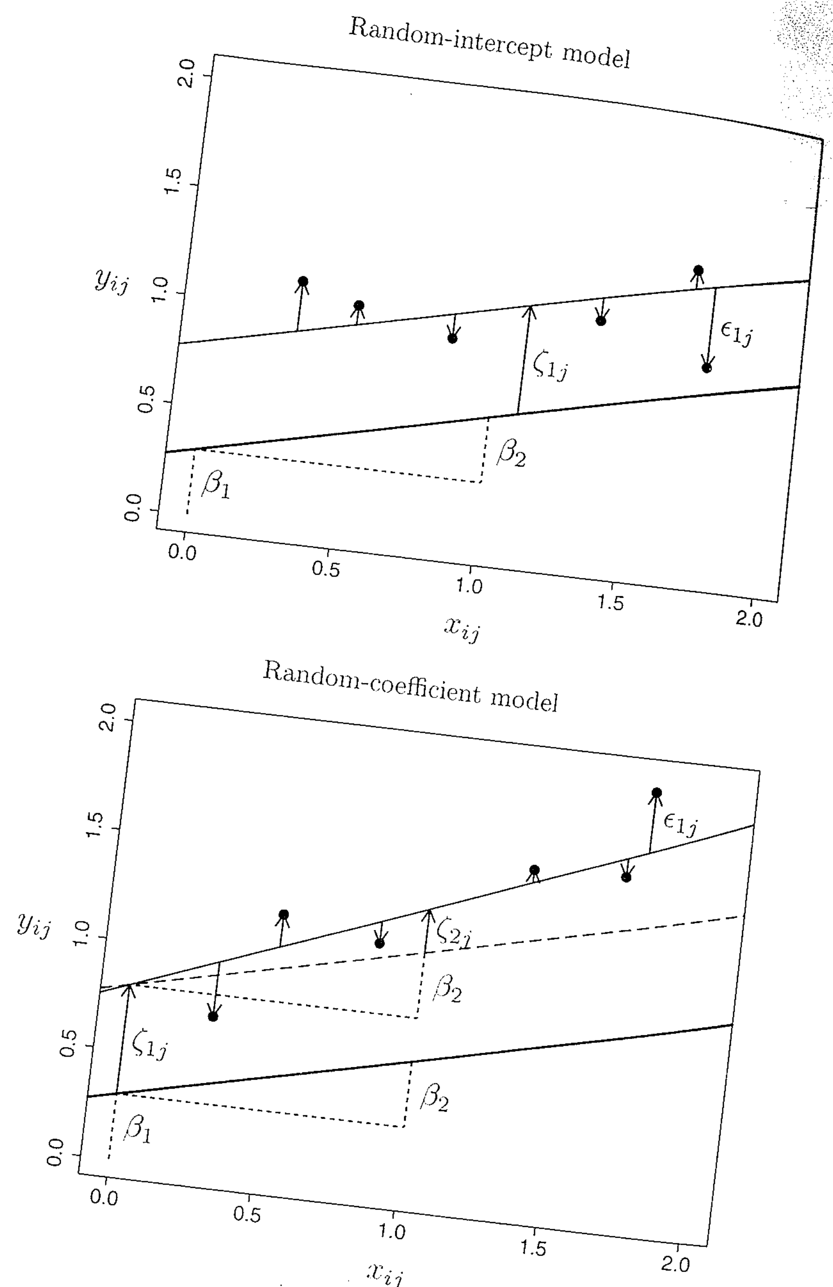

Figure 4.4: Illustration of random-intercept and random-coefficient models

We will, as is usually done, assume that, given x_{ij}, the random intercept and random slope have a bivariate normal distribution with zero mean and covariance matrix:

$$\Psi = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix} \equiv \begin{bmatrix} \mathrm{Var}(\zeta_{1j}|x_{ij}) & \mathrm{Cov}(\zeta_{1j}, \zeta_{2j}|x_{ij}) \\ \mathrm{Cov}(\zeta_{2j}, \zeta_{1j}|x_{ij}) & \mathrm{Var}(\zeta_{2j}|x_{ij}) \end{bmatrix}, \qquad \psi_{21} = \psi_{12}$$

The correlation between the random intercept and slope becomes

$$\rho_{21} = \frac{\psi_{21}}{\sqrt{\psi_{11}\psi_{22}}}$$

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 4.5 and as a contour plot in figure 4.6. Specifying a bivariate normal distribution implies that the (marginal) univariate distributions of the intercept and slope are also normal.

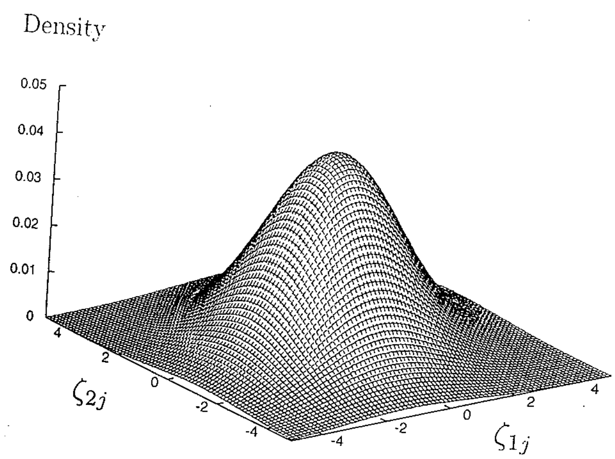

Figure 4.5: Perspective plot of bivariate normal distribution

(*Continued on next page*)

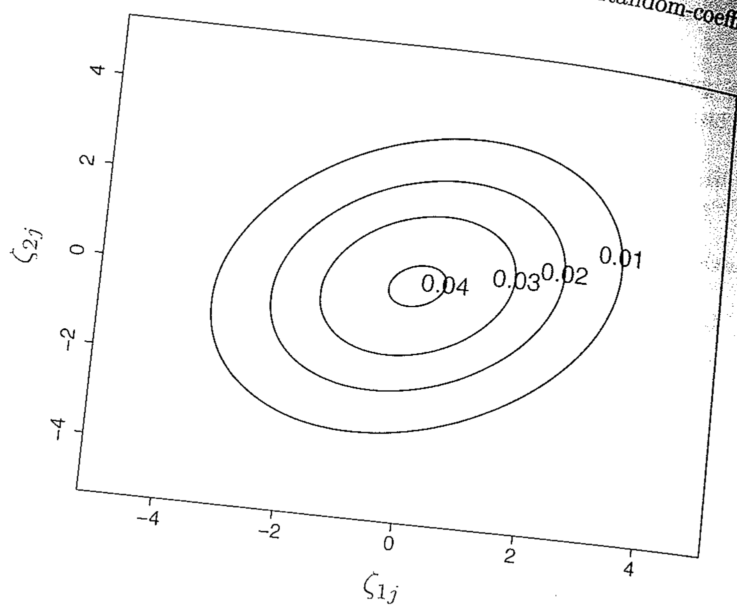

Figure 4.6: Contour plot of bivariate normal distribution

4.4.2 Interpretation of the random-effects variances and covariances

Interpreting the covariance matrix Ψ of the random effects is not straightforward. First, the random-slope variance ψ_{22} and the covariance between random slope and intercept ψ_{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 explanatory variable be denoted as u_y and u_x, respectively. For instance, in the application considered in the next chapter on children's increase in weight, u_y is kilograms, and u_x is years. The units of ψ_{11} are u_y^2, the units of ψ_{21} are u_y^2/u_x, and the units of ψ_{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.

A second difficulty 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 variance of the total residual, is

$$\mathrm{Var}(y_{ij}|x_{ij}) = \mathrm{Var}(\xi_{ij}|x_{ij}) = \psi_{11} + 2\psi_{21}x_{ij} + \psi_{22}x_{ij}^2 + \theta \qquad (4.2)$$

This variance depends on the value of the covariate x_{ij} and the total residual is therefore *heteroskedastic*. The conditional covariance for two students i and i' in the same school j is

$$
\begin{aligned}
\mathrm{Cov}(y_{ij}, y_{i'j}|x_{ij}, x_{i'j}) &= \mathrm{Cov}(\xi_{ij}, \xi_{i'j}|x_{ij}, x_{i'j}) \\
&= \psi_{11} + \psi_{21}x_{ij} + \psi_{21}x_{i'j} + \psi_{22}x_{ij}x_{i'j} \qquad (4.3)
\end{aligned}
$$

and the conditional intraclass correlation becomes

$$
\mathrm{Cor}(y_{ij}, y_{i'j}|x_{ij}, x_{i'j}) = \frac{\mathrm{Cov}(\xi_{ij}, \xi_{i'j}|x_{ij}, x_{i'j})}{\sqrt{\mathrm{Var}(\xi_{ij}|x_{ij})\mathrm{Var}(\xi_{i'j}|x_{i'j})}}
$$

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 residuals (from the overall mean regression line) for two students in the same school who both have `lrt` scores equal to 0 (the mean). However, for other pairs of students in the same school, the intraclass correlation is a complicated function of 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.

Finally, interpreting the parameters ψ_{11} and ψ_{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 ψ_{11} and ψ_{21} (see also exercise 4.7). This is because the intercept variance is the variability in the vertical positions of school-specific regression lines where `lrt`=0 (which 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 ψ_{11} and ψ_{21} to translation of the explanatory variable x_{ij} is illustrated in figure 4.7. Here identical cluster-specific regression lines are shown in the two panels, but with the explanatory variable $x'_{ij} = x_{ij} - 3.5$ in the lower panel translated relative to the explanatory variable x_{ij} in the upper panel. The intercepts are the intersections of the regression lines with the vertical line at zero. 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.

(*Continued on next page*)

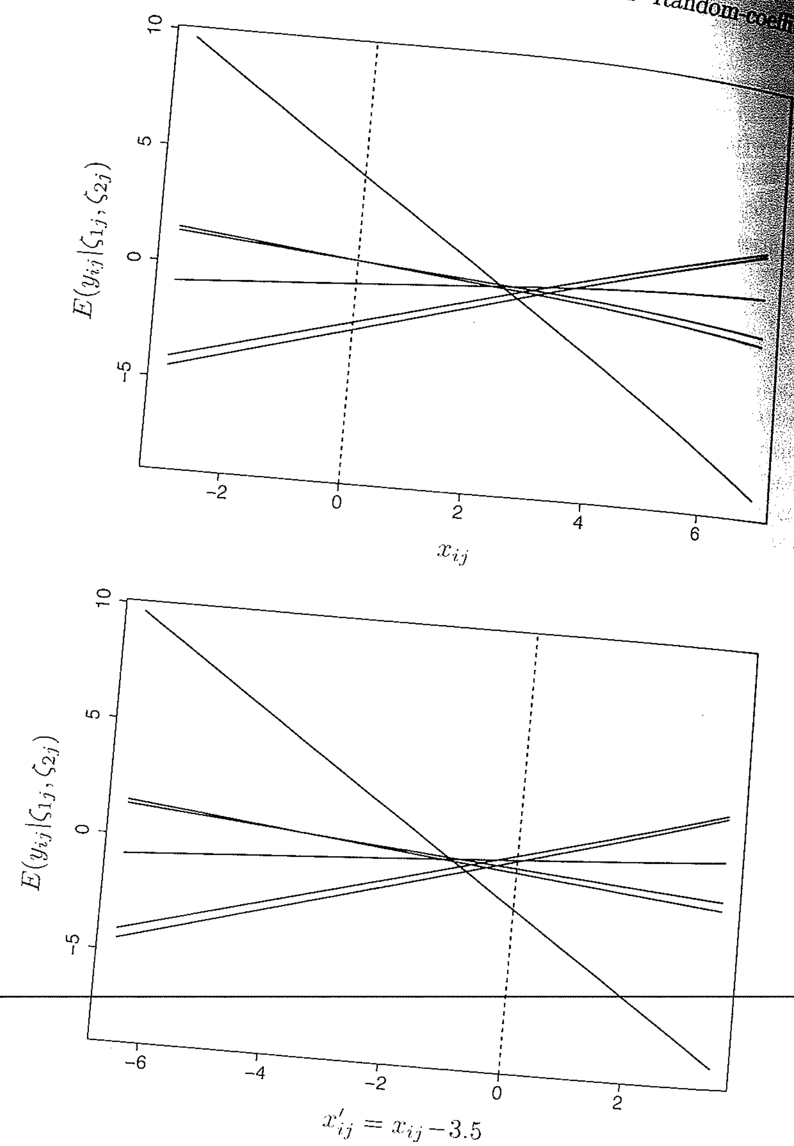

Figure 4.7: Cluster-specific regression lines for random-coefficient model, illustrating lack of invariance under translation of explanatory variable (*Source*: Skrondal and Rabe-Hesketh 2004)

To make ψ_{11} and ψ_{21} interpretable, it makes sense to translate x_{ij} so that $x_{ij} = 0$ is meaningful in some way. Typical choices are either mean centering (as for `lrt`), or, if x_{ij} is time as in the next chapter, defining 0 to be the "initial time" in some sense. Because the magnitude and interpretation of ψ_{21} depend on the translation of x_{ij}, which is often arbitrary, it generally does not make sense to set ψ_{21} to 0 by specifying uncorrelated intercepts and slopes.

A useful way of interpreting the magnitude of the estimated variances $\widehat{\psi}_{11}$ and $\widehat{\psi}_{22}$ is by considering the intervals $\widehat{\beta}_1 \pm 1.96\sqrt{\widehat{\psi}_{11}}$ and $\widehat{\beta}_2 \pm 1.96\sqrt{\widehat{\psi}_{22}}$ that contain about 95% of the intercepts and slopes in the population, respectively. To aid interpretation of the random part of the model, it is also useful to produce plots of school-specific regression lines as discussed in section 4.8.3.

4.5 Estimation using Stata

We will describe two Stata commands for linear random-coefficient models, `xtmixed` and `gllamm`. In general, we recommend using `xtmixed` rather than `gllamm` for linear random-coefficient models because it is computationally more efficient and sometimes more accurate. However, there are certain diagnostics available using the `gllapred` command for `gllamm` that are, at the time of writing this book, not provided by the `predict` command for `xtmixed`.

4.5.1 Using xtmixed

Random-intercept model

We first consider the more familiar random-intercept model

$$y_{ij} = (\beta_1 + \zeta_{1j}) + \beta_2 x_{ij} + \epsilon_{ij} \tag{4.4}$$

discussed in the previous chapter. 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 slope covariance, $\psi_{22} = \psi_{21} = 0$.

Maximum likelihood estimates for the random-intercept model can be obtained using `xtmixed` with the `mle` option:

(Continued on next page)

```
. xtmixed gcse lrt || school:, 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(1)       =   2042.57
Log likelihood = -14024.799                     Prob > chi2        =    0.0000

        gcse       Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

         lrt    .5633697    .0124654    45.19   0.000     .5389381    .5878014
       _cons    .0238706    .4002258     0.06   0.952    -.7605576    .8082987

  Random-effects Parameters      Estimate   Std. Err.     [95% Conf. Interval]

school: Identity
                   sd(_cons)     3.035271   .3052516      2.492262    3.69659

                sd(Residual)     7.521481   .0841759      7.358295    7.688285

LR test vs. linear regression: chibar2(01) =   403.27 Prob >= chibar2 = 0.0000
```

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

```
. estimates store ri
```

The random-intercept model assumes that the school-specific regression lines are parallel. The common coefficient or slope β_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

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

The maximum likelihood estimates for the random-intercept model are also given under "Model 1: Random intercept" in table 4.1.