Table 4.1: Maximum likelihood estimates for inner-London school data

Parameter	Model 1: Random intercept Est	(SE)	Model 2: Random coefficient Est	(SE)	Model 3: Rand. coefficient & level-2 covariates Est	(SE)	γ_{rr}
Fixed part							
β_1 [_cons]	0.02	(0.40)	−0.12	(0.40)	−1.00	(0.51)	γ_{11}
β_2 [lrt]	0.56	(0.01)	0.56	(0.02)	0.57	(0.03)	γ_{21}
β_3 [boys]					0.85	(1.09)	γ_{12}
β_4 [girls]					2.43	(0.84)	γ_{13}
β_5 [boys_lrt]					−0.02	(0.06)	γ_{22}
β_6 [girls_lrt]					−0.03	(0.04)	γ_{23}
Random part							
`xtmixed`							
$\sqrt{\psi_{11}}$	3.04		3.01		2.80		
$\sqrt{\psi_{22}}$			0.12		0.12		
ρ_{21}			0.50		0.60		
$\sqrt{\theta}$	7.52		7.44		7.44		
`gllamm`							
ψ_{11}	9.21		9.04				
ψ_{22}			0.01				
ψ_{21}			0.18				
θ	56.57		55.37				
Log likelihood	−14,024.80		−14,004.61		−13,998.83		

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 `xtmixed`, 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 (here abbreviated as `cov(unstructured)`) because `xtmixed` will otherwise set the covariance ψ_{21} (and the corresponding correlation) to zero by default. Maximum likelihood estimates for the random-coefficient model are then obtained using

```
. xtmixed gcse 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(1)       =    779.80
Log likelihood = -14004.613                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        gcse |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lrt |   .5567291   .0199367    27.92   0.000     .5176539    .5958043
       _cons |  -.1150841   .3978336    -0.29   0.772    -.8948236    .6646554
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Unstructured         |
                     sd(lrt) |   .1205631   .0189827      .0885508    .1641483
                   sd(_cons) |   3.007436   .3044138      2.466252    3.667375
             corr(lrt,_cons) |   .4975474   .1487416      .1572843    .7322131
-----------------------------+------------------------------------------------
                sd(Residual) |   7.440788   .0839482      7.278059    7.607157
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =   443.64   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

Because the variance option was not used, the output shows the standard deviations sd(lrt) of the slope and sd(_cons) of the intercept instead of variances, and the correlation between intercepts and slopes corr(lrt,_cons) instead of the covariance. We can obtain the covariance matrix using the postestimation command `estat recovariance`:

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

             |       lrt      _cons
-------------+----------------------
         lrt |  .0145355
       _cons |  .1804036    9.04467
```

The maximum likelihood estimates for the random-coefficient model are also given under "Model 2: Random intercept and slope" in table 4.1. We store the estimates under the name `rc` for later use:

```
. estimates store rc
```

We could also use restricted maximum likelihood estimation by specifying the `reml` option.

4.5.2 Using gllamm

Random-intercept model

We start by using `gllamm` to fit the random-intercept model:

```
. gllamm gcse lrt, i(school) adapt
number of level 1 units = 4059
number of level 2 units = 65

Condition Number = 35.786606

gllamm model

log likelihood = -14024.799

------------------------------------------------------------------------------
        gcse |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lrt |   .5633697   .0124863    45.12   0.000      .538897    .5878425
       _cons |   .0239115   .4002945     0.06   0.952    -.7606514    .8084744
------------------------------------------------------------------------------

Variance at level 1
------------------------------------------------------------------------------

  56.572669 (1.2662546)

Variances and covariances of random effects
------------------------------------------------------------------------------


***level 2 (school)

    var(1): 9.2127069 (1.8529779)
------------------------------------------------------------------------------
```

We store the estimates under the name `rig`:

```
. estimates store rig
```

Random-coefficient model

In the previous `gllamm` command for the random-intercept model, all that was required to specify the random intercept was the `i(school)` option.

To introduce a random slope ζ_{2j}, we will also need to specify the variable multiplying the random slope in (4.1), i.e., x_{ij} or `lrt`. This is done by specifying an equation for the slope:

```
. eq slope: lrt
```

We also need an equation for the variable multiplying the random intercept ζ_{1j}, and since it is an intercept, we just specify a variable equal to 1:

```
. generate cons = 1
. eq inter: cons
```

We must also add a new option, nrf(2), standing for "number of random effects is 2" (an intercept and a slope), and specify both equations, inter and slope, in the eqs() option.

To speed up estimation, we use the previous estimates for the random-intercept model as starting values for the regression coefficients and random-intercept variance and set the starting values for the additional two parameters (for the random-slope variance and the random intercept and slope covariance) to zero:

```
. matrix a = e(b)
. matrix a = (a,0,0)
```

(The order in which the parameters are given in the matrix matters, and when going from a random-intercept to a random-coefficient model, the two new parameters are always at the end.) To use the parameter matrix a as starting values, we specify the from(a) and copy options. Finally, to get good estimates as fast as possible, we use a spherical quadrature rule of degree 15 (see sec. 6.11.2), specified using ip(m) and nip(15):

```
. gllamm gcse lrt, i(school) nrf(2) eqs(inter slope) ip(m) nip(15)
> adapt from(a) copy
number of level 1 units = 4059
number of level 2 units = 65

Condition Number = 35.440522

gllamm model

log likelihood = -14004.613

------------------------------------------------------------------------------
        gcse |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lrt |    .556729    .0199969    27.84   0.000     .5175358    .5959222
       _cons |  -.1150848    .3982896    -0.29   0.773    -.8957181    .6655485
------------------------------------------------------------------------------

Variance at level 1
------------------------------------------------------------------------------

  55.365324 (1.2492818)

Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (school)

    var(1): 9.0446842 (1.8310103)
    cov(2,1): .18040306 (.06915204) cor(2,1): .49754322

    var(2): .01453559 (.00457725)
------------------------------------------------------------------------------
```

The `gllamm` output gives the maximum likelihood estimates of the within-school variance θ under `Variance at level 1` and the estimates of the elements ψ_{11}, ψ_{21}, and ψ_{22}, of the covariance matrix of the random intercept and slope under `Variances and covariances of random effects`. These estimates are also given in table 4.1.

4.6 Testing the slope variance

Before interpreting the parameter estimates, we may want to test whether the random-coefficient model "fits better" than the nested random-intercept model. Specifically, we test the null hypothesis

$$H_0\colon\ \psi_{22} = \psi_{21} = 0$$

which is equivalent to the hypothesis that the random slopes ζ_{2j} are zero. The null hypothesis lies on the boundary of the parameter space since the variance ψ_{22} must be nonnegative. Therefore, as discussed in section 2.6.2, the likelihood-ratio statistic L does not have a chi-squared distribution under the null hypothesis. Fortunately, the correct p-value for testing the slope variance can also simply be obtained by dividing the naive p-value from the likelihood-ratio test by 2.

The naive likelihood-ratio test can be performed using the `lrtest` command in Stata:

```
. lrtest rc ri
Likelihood-ratio test                                 LR chi2(2)  =     40.37
(Assumption: ri nested in rc)                         Prob > chi2 =    0.0000
Note: LR test is conservative
```

The output correctly states that the test is conservative. We can divide the p-value by 2 but this makes no difference to the conclusion that the random-intercept model is rejected in favor of the random-coefficient model.

4.7 Interpretation of estimates

The population-mean intercept and slope are estimated as −0.12 and 0.56, respectively. These estimates are similar to those for the random-intercept model (see table 4.1) and also similar to the means of the school-specific least-squares regression lines given on page 144. The estimated random-intercept standard deviation and level-1 residual standard deviation are somewhat lower than for the random-intercept model. The latter is due to a better fit of the school-specific regression lines for the random-coefficient model, which relaxes the parallel regression line restriction. The estimated covariance matrix of the intercepts and slopes is similar to the sample covariance matrix of the ordinary least-squares estimates reported on page 144.

The easiest way to interpret the estimated standard deviations of the random intercept and random slope is to form intervals within which 95% of the schools' random intercepts and slopes are expected to lie. It is important to remember that these intervals represent ranges within which 95% of the realizations of a random variable are expected to lie, a concept different from confidence intervals that are ranges within which an *unknown parameter* is believed to lie. For the intercepts, we obtain $-0.12 \pm 1.96 \times 3.01$, so 95% of schools have their intercept in the range -6.0 to 5.8. In other words, the school mean GCSE scores for children with average (`lrt=0`) LRT scores vary between -6.0 and 5.8. For the slopes, we obtain $0.56 \pm 1.96 \times 0.12$ giving an interval from 0.32 to 0.80. Ninety-five percent of schools have slopes between 0.32 and 0.80. This exercise of forming intervals is particularly important for slopes because it is useful to know whether the slopes are likely to have different signs for different schools, although that would be odd in the current example. The range from 0.32 to 0.80 is fairly wide and the regression lines of schools may cross: one school could "add more value" (produce higher mean GCSE scores for given LRT scores) than another school for students with low LRT scores and add less value than the other school for students with high LRT scores.

The estimated correlation $\widehat{\rho}_{21} = 0.50$ between intercepts and slopes means that schools with larger mean GCSE scores for students with average LRT scores than other schools also tend to have larger slopes than those other schools. This information, combined with the random-intercept and slope variances and the range of LRT scores determines how much the lines cross, something that is best explored by plotting the predicted regression lines for the schools as demonstrated in section 4.8.3.

The variance of the total residual ξ_{ij} (equal to the conditional variance of the responses y_{ij} given the covariate x_{ij}) was given in (4.2). We can estimate the corresponding standard deviation by plugging in the maximum likelihood estimates:

$$
\begin{aligned}
\sqrt{\widehat{\mathrm{Var}}(\xi_{ij}|x_{ij})} &= \sqrt{\widehat{\psi}_{11} + 2\widehat{\psi}_{21}x_{ij} + \widehat{\psi}_{22}x_{ij}^2 + \widehat{\theta}} \\
&= \sqrt{9.0447 + 2 \times 0.1804 \times x_{ij} + 0.0145 \times x_{ij}^2 + 55.3653}
\end{aligned}
$$

A graph of the estimated standard deviation of the total residual against the covariate `lrt` (x_{ij}) can be obtained using the `twoway function` command:

```
. twoway function sqrt(9.0447+2*0.1804*x+0.0145*x^2+55.3653), range(-30 30)
> xtitle(LRT) ytitle(Estimated standard deviation of total residual)
```

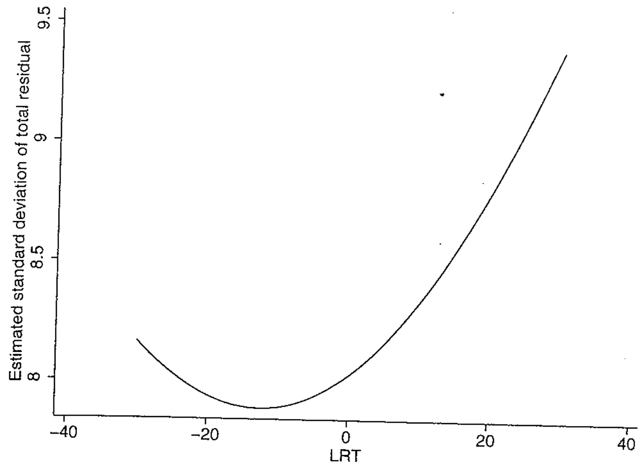

Figure 4.8: Heteroskedasticity of total residual ξ_{ij}

The estimated standard deviation of the total residual varies between just under 8 and just under 9.5.

4.8 Assigning values to the random intercepts and slopes

Having obtained estimated model parameters $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\psi}_{11}$, $\widehat{\psi}_{22}$, $\widehat{\psi}_{21}$, and $\widehat{\theta}$, we now assign values to the random intercepts and slopes, treating the estimated parameters as known (see sec. 2.9). This is useful for model visualization, residual diagnostics, and inference for individual clusters as will be demonstrated in sections 4.8.3–4.8.5.

4.8.1 Maximum likelihood estimation

Maximum likelihood estimates of the random intercepts and slopes can be obtained by first estimating the total residuals $\widehat{\xi}_{ij} = y_{ij} - (\widehat{\beta}_1 + \widehat{\beta}_2 x_{ij})$ and then fitting individual regressions of $\widehat{\xi}_{ij}$ on x_{ij} for each school by OLS using the `statsby` prefix command. We first retrieve the `xtmixed` estimates stored under `rc`, then obtain the estimated total residual and use `statsby` to produce the variables `mli` and `mls` containing the maximum likelihood estimates $\widehat{\zeta}_{1j}$ and $\widehat{\zeta}_{2j}$ of the random intercepts and slopes, respectively:

```
estimates restore rc
(results rc are active now)
predict fixed, xb
generate totres = gcse - fixed
```

```
. statsby mli=_b[_cons] mls=_b[lrt], by(school) saving(ols, replace):
> regress totres lrt
(running regress on estimation sample)
      command:  regress totres lrt
          mli:  _b[_cons]
          mls:  _b[lrt]
           by:  school
Statsby groups
---+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................
..............                                    50
. sort school
. merge school using ols
variable school does not uniquely identify observations in the master data
. drop _merge
```

Maximum likelihood estimates will not be available for schools with only one observation or for schools within which x_{ij} does not vary. There are no such schools in the dataset, but school 48 has only two observations and the maximum likelihood estimates look odd:

```
. list lrt gcse mli mls if school==48, clean noobs
       lrt      gcse       mli         mls
   -4.5541   -1.2908   -32.607   -7.458484
   -3.7276   -6.9951   -32.607   -7.458484
```

Here the fitted line goes through the two points exactly and the intercept and slope take on extreme values, a phenomenon often encountered when using maximum likelihood estimation of random effects for clusters that provide little information. In general, we therefore do not recommend using this method and suggest using empirical Bayes prediction instead.

4.8.2 Empirical Bayes prediction

As discussed for random-intercept models in section 2.9.2, empirical Bayes (EB) predictions have a smaller prediction error variance than maximum likelihood estimates due to shrinkage towards the mean. Furthermore, empirical Bayes predictions are available for schools with only one observation or only one unique value of x_{ij}.

Empirical Bayes predictions of the random intercepts ζ_{1j} and slopes ζ_{2j} can be obtained using the predict command with the reffects option after estimation with xtmixed:

```
. estimates restore rc
. predict ebs ebi, reffects
```

Here we specified the variable names ebs and ebi for the empirical Bayes predictions $\widetilde{\zeta}_{2j}$ and $\widetilde{\zeta}_{1j}$ of the slopes and intercepts. The intercept variable comes last because xtmixed treats the intercept as the last random effect, as reflected by the output. This order is consistent with Stata's convention of treating the fixed intercept as the last regression parameter in estimation commands.

To compare the empirical Bayes predictions with the maximum likelihood estimates, we list one observation per school for schools 1–9 and school 48:

```
. list school mli ebi mls ebs if pickone==1 & (school<10 | school==48), noobs
```

school	mli	ebi	mls	ebs
1	3.948386	3.749336	.1526115	.1249755
2	4.937837	4.702129	.2045584	.1647261
3	5.69259	4.79768	.0222564	.0808666
4	.1526213	.3502505	.2047173	.1271821
5	2.719524	2.462805	.1232875	.0720576
6	6.14715	5.18381	-.0213859	.0586242
7	4.100311	3.640942	-.3144541	-.1488697
8	-.1368859	-.1218861	.010678	.0068854
9	-2.2586	-1.767983	-.1555334	-.0886194
48	-32.607	-.4098185	-7.458484	-.0064854

Most of the time, the empirical Bayes predictions are closer to zero than the maximum likelihood estimates due to shrinkage as discussed for random-intercept models in section 2.9.2. However, for models with several random effects, the relationship between empirical Bayes predictions and maximum likelihood estimates is somewhat more complex than for random-intercept models. The benefit of shrinkage is apparent for school 48 where the empirical Bayes predictions appear more reasonable than the maximum likelihood estimates.

We can see shrinkage more clearly by plotting the empirical Bayes predictions against the maximum likelihood estimates and superimposing a $y = x$ line. For the random intercept, the command is

```
. twoway (scatter ebi mli if pickone==1 & school!=48, mlabel(school))
> (function y=x, range(-10 10)), xtitle(ML estimate)
> ytitle(EB prediction) legend(off)
```

and for the random slope, it is

```
. twoway (scatter ebs mls if pickone==1 & school!=48, mlabel(school))
> (function y=x, range(-0.6 0.6)), xtitle(ML estimate)
> ytitle(EB prediction) legend(off)
```

These commands produce the graphs in figure 4.9. (We excluded school 48 from the graphs because the ML estimates are so extreme.)

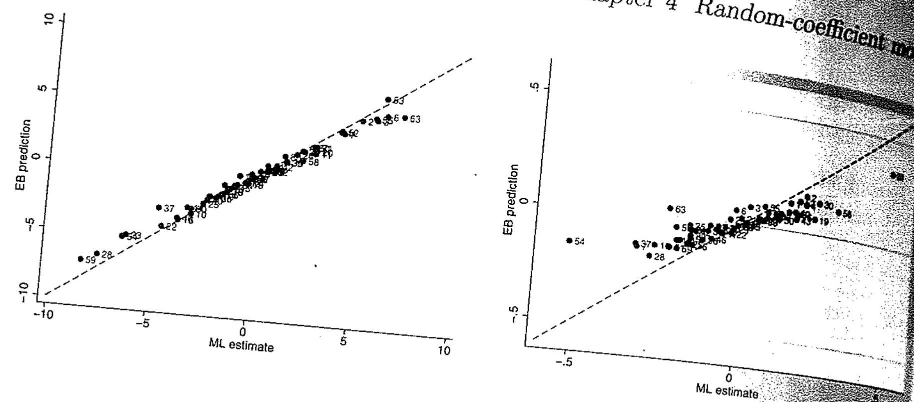

Figure 4.9: Scatterplot of empirical Bayes (EB) predictions versus maximum likelihood (ML) estimates of school-specific intercepts (left) and slopes (right) with equality shown as reference lines

For maximum likelihood estimates above the mean, the empirical Bayes prediction tends to be smaller than the maximum likelihood estimate; the reverse is true for maximum likelihood estimates below the mean.

4.8.3 Model visualization

To better understand the random-intercept and random-coefficient models, and in particular the variability implied by the random part, it is useful to produce graphs of predicted model-implied regression lines for the individual schools. This can be achieved using the `predict` command with the `fitted` option to obtain school-specific fitted regression lines, with maximum likelihood estimates substituted for the regression parameters (β_1 and β_2) and empirical Bayes predictions substituted for the random effects (ζ_{1j} for the random-intercept model and ζ_{1j} and ζ_{2j} for the random-intercept model). For instance, for the random-coefficient model, the predicted regression line for school j is

$$\widehat{y}_{ij} = \widehat{\beta}_1 + \widehat{\beta}_2 x_{ij} + \widetilde{\zeta}_{1j} + \widetilde{\zeta}_{2j} x_{ij}$$

These predictions are obtained and plotted as follows:

```
. predict murc, fitted
. sort school lrt
. twoway (line murc lrt, connect(ascending)), xtitle(LRT)
> ytitle(Empirical Bayes regression lines for model 2)
```

To obtain predictions for the random-intercept model, we must first restore the estimates stored under the name `ri`:

```
. estimates restore ri
(results ri are active now)
. predict muri, fitted
. sort school lrt
```

```
. twoway (line muri lrt, connect(ascending)), xtitle(LRT)
> ytitle(Empirical Bayes regression lines for model 1)
```

The resulting graphs of the school-specific regression lines for both the random-intercept model and the random-coefficient model are given in figure 4.10.

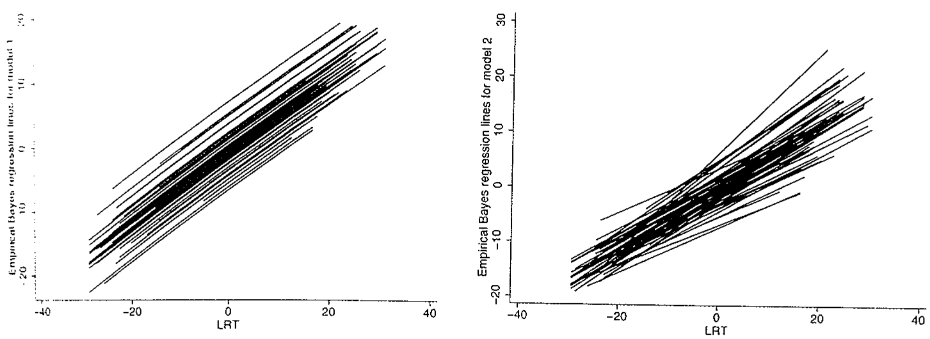

Figure 4.10: Empirical Bayes (EB) predictions of school-specific regression lines for the random-intercept model (left) and the random-intercept and random-slope model (right)

The predicted school-specific regression lines are parallel for the random-intercept model (with vertical shifts given by the $\widetilde{\zeta}_{1j}$) but are not parallel for the random-coefficient model where the slopes $\beta_2 + \widetilde{\zeta}_{2j}$ also vary across schools.

After estimation with gllamm, the syntax for obtaining school-specific regression lines would be

```
estimates restore rcg
gllapred murcg, linpred
```

4.8.4 Residual diagnostics

If the normality assumptions for the random intercepts ζ_{1j}, random slopes ζ_{2j}, and level-1 residuals ϵ_{ij} are satisfied, the corresponding empirical Bayes predictions should also have normal distributions. After estimation with xtmixed, we obtain the predicted level-1 residuals using

```
. predict res1, residuals
```

To plot the distributions of the predicted random effects, we must pick one prediction per school, and we can accomplish this using the pickone variable created earlier. We can now plot all three distributions using

```
. histogram reff1 if pickone==1, normal xtitle(Predicted random slopes)
. histogram reff2 if pickone==1, normal xtitle(Predicted random intercepts)
. histogram res1, normal xtitle(Predicted level-1 residuals)
```

The histograms in figure 4.11 and 4.12 look approximately normal although the one for the slopes is perhaps a little skewed:

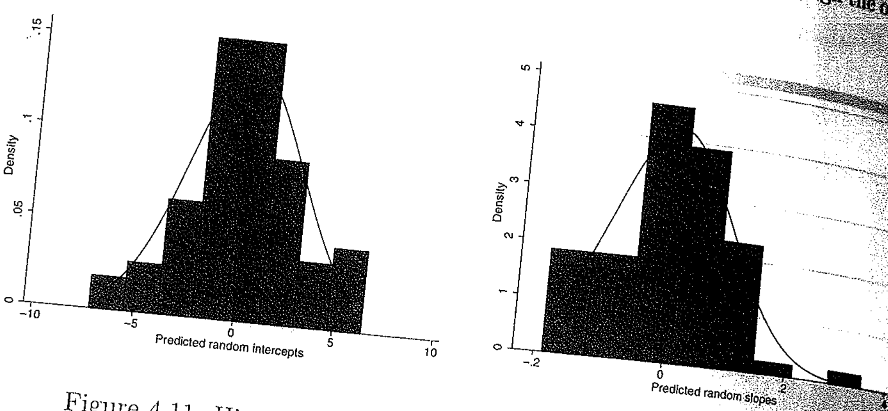

Figure 4.11: Histograms of predicted random intercepts and slopes

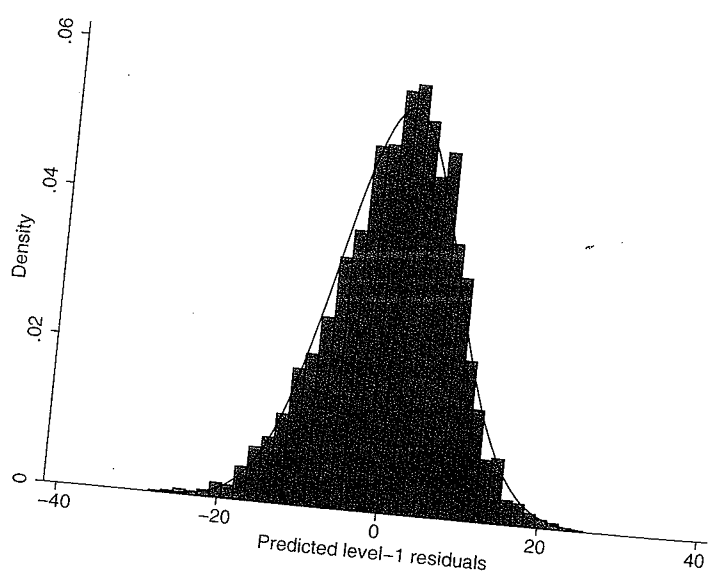

Figure 4.12: Histogram of predicted level-1 residuals

To obtain standardized level-1 residuals, use the `rstandard` option in the `predict` command after estimation using `xtmixed`. After estimation with `gllamm`, standardized level-2 residuals (random intercepts and random slopes) can be obtained using `gllapred` with the `ustd` option and standardized level-1 residuals with the `pearson` option.

4.8.5 Inferences for individual schools

The random-intercept predictions $\widetilde{\zeta}_{1j}$ can be viewed as measures of how much "value" the schools add for children with LRT scores equal to zero (the mean). Therefore, the left panel of figure 4.9 sheds some light on the research question: which schools are most effective? We could also produce similar plots for children with different values x^0 of the LRT scores:

$$\widehat{\beta}_1 + \widehat{\beta}_2 x^0 + \widetilde{\zeta}_{1j} + \widetilde{\zeta}_{2j} x^0$$

For instance, in a similar application, Goldstein et al. (2000) substitute the 10th percentile of the intake measure to compare school effectiveness for poorly performing children. It does not matter whether we add the predicted fixed part of the model since the ranking of schools is not affected by this. Unfortunately, `xtmixed` does not provide standard errors for random-effects predictions at the time of writing. We therefore restore the `gllamm` estimates

```
. estimates restore rcg
```

and use the postestimation command `gllapred` with the `u` option to obtain empirical Bayes predictions and corresponding standard errors for the random intercepts $\widetilde{\zeta}_{1j}$ and random slopes $\widetilde{\zeta}_{2j}$

```
. gllapred reff, u
(means and standard deviations will be stored in reffm1 reffs1 reffm2 reffs2)
```

The first random effect, with predictions stored in `reffm1`, is now the random intercept ζ_{1j}, whereas this was the second random effect in `xtmixed`. The predictions of the random slopes ζ_{2j} are stored in `reffm2`.

Returning to the question of comparing the schools' effectiveness for children with LRT scores equal to 0, we can plot the predicted random effects with approximate 95% confidence intervals (based on the prediction error standard deviations in `reffs1`) using the following commands:

```
. gsort +reffm1 -f
. generate rank = sum(f)
. generate labpos = reffm1 + 1.96*reffs1 + .4
. serrbar reffm1 reffs1 rank if f==1, addplot(scatter labpos rank,
> mlabel(school) msymbol(none) mlabpos(0)) scale(1.96) xtitle(Rank)
> ytitle(Prediction) legend(off)
```

(The school labels were added to the graph by superimposing a scatterplot onto the error bar plot using the `addplot()` option where the vertical positions of the labels are given by the variable `labpos`.) The resulting graph is shown in figure 4.13.

(Continued on next page)

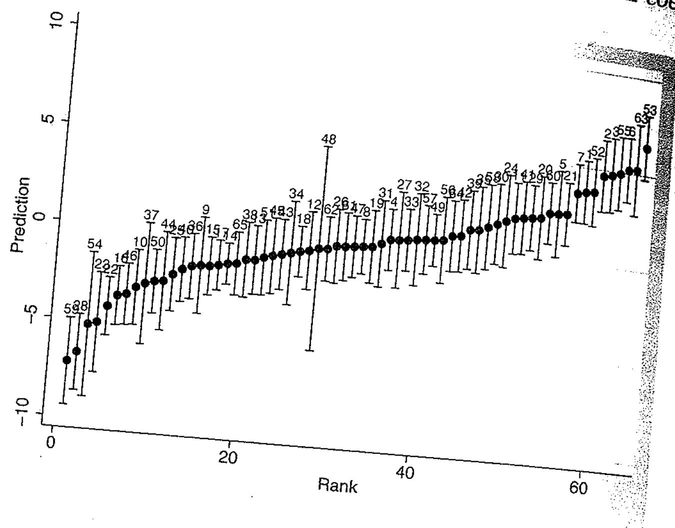

Figure 4.13: Random-intercept predictions and approximate 95% confidence intervals versus ranking (school identifiers shown on top of confidence intervals)

The interval for school 48 is particularly wide because there are only two students from this school in the dataset. It is clear from the large confidence intervals that the rankings are not precise and that perhaps only a coarse classification into poor, medium, and good schools can be justified. (To obtain confidence intervals for different values of $r^{(}$ requires posterior correlations that are produced by gllapred with the corr() option.)

4.9 Two-stage model formulation

In this section, we describe an alternative way of specifying random-coefficient models that is popular in some branches of U.S. social science, such as education (e.g., Raudenbush and Bryk 2002). As shown below, models are specified in two stages (for levels 1 and 2), necessitating a distinction between level-1 and level-2 covariates. Many people find this formulation helpful for interpreting and specifying models. Identical models can be formulated using either the approach discussed up to this point or the two-stage formulation.

To express the random-coefficient model using a two-stage formulation, the *level-1 model* is written as

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

where the intercept η_{1j} and slope η_{2j} are school-specific coefficients (the Greek letter η is pronounced eta). The *level-2 models* have these coefficients as responses

$$\begin{aligned} \eta_{1j} &= \gamma_{11} + \zeta_{1j} \\ \eta_{2j} &= \gamma_{21} + \zeta_{2j} \end{aligned} \tag{4.5}$$

Sometimes the first of these level-2 models is referred to as a "means as outcomes" or "intercepts as outcomes" model and the second as a "slopes as outcomes" model. It is assumed that, given the covariate(s), the residuals or disturbances ζ_{1j} and ζ_{2j} in the level-2 model have a bivariate normal distribution with zero mean and covariance matrix:

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix}, \qquad \psi_{21} = \psi_{12}$$

It is important to remember that the level-2 models cannot be estimated on their own because the random effects η_{1j} and η_{2j} are not observed. Instead, we must substitute the level-2 models into the level-1 model to obtain the *reduced-form* model for the observed responses y_{ij}

$$\begin{aligned} y_{ij} &= \underbrace{\gamma_{11} + \zeta_{1j}}_{\eta_{1j}} + \underbrace{(\gamma_{21} + \zeta_{2j})}_{\eta_{2j}} x_{ij} + \epsilon_{ij} \\ &= \underbrace{\gamma_{11} + \gamma_{21} x_{ij}}_{\text{fixed}} + \underbrace{\zeta_{1j} + \zeta_{2j} x_{ij} + \epsilon_{ij}}_{\text{random}} \\ &\equiv \beta_1 + \beta_2 x_{ij} + \zeta_{1j} + \zeta_{2j} x_{ij} + \epsilon_{ij} \end{aligned}$$

In the reduced form, the fixed part is usually written first followed by the random part, and we can return to our previous notation by defining $\beta_1 \equiv \gamma_{11}$ and $\beta_2 \equiv \gamma_{21}$. The above model is thus equivalent to the model in (4.1).

Any level-2 covariates (covariates that do not vary at level 1) are included in the level-2 models. For instance, we could include dummy variables for type of school, w_{2j} for boys' schools and w_{3j} for girls' schools with mixed schools as the reference category. If we include these dummy variable in the model for the random intercept

$$\eta_{1j} = \gamma_{11} + \gamma_{12} w_{2j} + \gamma_{13} w_{3j} + \zeta_{1j}$$

the reduced form becomes

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

If we also include the dummy variables for type of school in the model for the random slope

$$\eta_{2j} = \gamma_{21} + \gamma_{22} w_{2j} + \gamma_{23} w_{3j} + \zeta_{2j}$$