Raudenbush, S.W., & Bryk, A.S. (2002).
Hierarchical Linear Models: Applications and Data
Analysis Methods, 2nd ed. Thousand Oaks: SAGE.

4 An Illustration

Introduction

The purpose of this chapter is to illustrate the use of the techniques of estimation and hypothesis testing presented in Chapter 3. We present a series of analyses based on the models introduced in Chapter 2. For each model, we illustrate the estimation of the fixed effects and variance-covariance components, and demonstrate the use of appropriate hypothesis testing procedures for these various parameters. We also introduce through these examples some useful auxiliary description statistics that can be computed based on the maximum likelihood variance-covariance component estimates. Applications of estimation and hypothesis testing procedures for random level-1 coefficients are demonstrated in the last section of this chapter.

The examples presented below use data from a nationally representative sample of U.S. public and Catholic high schools. These data are a subsample from the 1982 High School and Beyond (HS&B) Survey,[1] and include information on 7,185 students nested within 160 schools: 90 public and 70 Catholic. Sample sizes averaged about 45 students per school.

Attention is restricted to two student-level variables: (a) the outcome, Y_{ij}, a standardized measure of math achievement; and (b) one predictor, $(\mathrm{SES})_{ij}$,

TABLE 4.1 Descriptive Statistics from American High School Data

	Variable Name	*Mean*	*sd*
Student-level variables			
Math achievement	Y_{ij}	12.75	6.88
Socioeconomic status	$(\text{SES})_{ij}$	0.00	0.78
School-level variables			
Sector	$(\text{SECTOR})_j$	0.44	0.50
School average SES	$(\text{MEAN SES})_j$	0.00	0.41

student socioeconomic status, which is a composite of parental education, parental occupation, and parental income. School-level variables include $(\text{SECTOR})_j$, an indicator variable taking on a value of one for Catholic schools and zero for public schools, and $(\text{MEAN SES})_j$, the average of the student SES values within each school. In the language introduced in Chapter 2, the level-1 units are students and the level-2 units are schools. $(\text{SES})_{ij}$ is a level-1 predictor; $(\text{SECTOR})_j$ and $(\text{MEAN SES})_j$ are level-2 predictors. Means and standard deviations of these variables are supplied in Table 4.1.

Questions motivating these analyses include the following:

1. How much do U.S. high schools vary in their mean mathematics achievement?
2. Do schools with high MEAN SES also have high math achievement?
3. Is the strength of association between student SES and math achievement similar across schools? Or is SES a more important predictor of achievement in some schools than in others?
4. How do public and Catholic schools compare in terms of mean math achievement and in terms of the strength of the SES-math achievement relationship, after we control for MEAN SES?

The One-Way ANOVA

The one-way ANOVA with random effects, described in Chapter 2, provides useful preliminary information about how much variation in the outcome lies within and between schools and about the reliability of each school's sample mean as an estimate of its true population mean.

The Model

The level-1 or student-level model is

$$Y_{ij} = \beta_{0j} + r_{ij}, \qquad [4.1]$$

where we assume $r_{ij} \sim$ independently $\mathrm{N}(0, \sigma^2)$ for $i = 1, \ldots, n_j$ students in school j, and $j = 1, \ldots, 160$ schools. We refer to σ^2 as the student-level variance. Notice that this model characterizes achievement in each school with just an intercept, β_{0j}, which in this case is the mean.

At level 2 or the school level, each school's mean math achievement, β_{0j}, is represented as a function of the grand mean, γ_{00}, plus a random error, u_{0j}:

$$\beta_{0j} = \gamma_{00} + u_{0j}, \qquad [4.2]$$

where we assume $u_{0j} \sim$ independently $\mathrm{N}(0, \tau_{00})$. We refer to τ_{00} as the school-level variance.

This yields a combined model, also often referred to as a mixed model, with fixed effect γ_{00} and random effects u_{0j} and r_{ij}:

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}. \qquad [4.3]$$

Results

Fixed Effects. From Table 4.2, the weighted least squares estimate for the grand-mean math achievement (using the estimator from Equation 3.9) is

$$\hat{\gamma}_{00} = 12.64.$$

This has a standard error of 0.24 and yields a 95% confidence interval (see Equation 3.12) of

$$12.64 \pm 1.96\,(0.24) = (12.17, 13.11).$$

Variance Components. Table 4.2 also lists restricted maximum likelihood estimates of the variance components. At the student level,

$$\widehat{\mathrm{Var}}(r_{ij}) = \hat{\sigma}^2 = 39.15.$$

TABLE 4.2 Results from the One-Way ANOVA Model

Fixed Effect		*Coefficient*	*se*	
Average school mean, γ_{00}		12.64	0.24	
Random Effect	*Variance Component*	*df*	χ^2	*p Value*
School mean, u_{0j}	8.61	159	1,660.2	.000
Level-1 effect, r_{ij}	39.15			

At the school level, τ_{00} is the variance of the true school means, β_{0j}, around the grand mean, γ_{00}. The estimated variability in these school means is

$$\widehat{\text{Var}}(\beta_{0j}) = \widehat{\text{Var}}(u_{0j}) = \hat{\tau}_{00} = 8.61.$$

To gauge the magnitude of the variation among schools in their mean achievement levels, it is useful to calculate the *plausible values range* for these means. Under the normality assumption of Equation 4.2, we would expect 95% of the school means to fall within the range:

$$\hat{\gamma}_{00} \pm 1.96(\hat{\tau}_{00})^{1/2}, \qquad [4.4]$$

which yields

$$12.64 \pm 1.96(8.61)^{1/2} = (6.89, 18.39).$$

This indicates a substantial range in average achievement levels among schools in this sample of data.

We may wish to test formally whether the estimated value of τ_{00} is significantly greater than zero. If not, it may be sensible to assume that all schools have the same mean. Formally, this hypothesis is

$$\text{H}_0\text{: } \tau_{00} = 0,$$

which may be tested using Equation 3.103. This test statistic reduces in a one-way random ANOVA model to

$$H = \sum n_j(\bar{Y}_{\cdot j} - \hat{\gamma}_{00})^2/\hat{\sigma}^2, \qquad [4.5]$$

which has a large-sample χ^2 distribution with $J - 1$ degrees of freedom under the null hypothesis. In our case, the test statistic takes on a value of 1,660.2 with 159 degrees of freedom ($J = 160$ schools). The null hypothesis is highly implausible ($p < .001$), indicating significant variation does exist among schools in their achievement.

Auxiliary Statistics. The *intraclass correlation*, which represents in this case the proportion of variance in Y between schools, is estimated by substituting the estimated variance components for their respective parameters in Equation 2.10:

$$\hat{\rho} = \hat{\tau}_{00}/(\hat{\tau}_{00} + \hat{\sigma}^2) = 8.61/(8.61 + 39.15) = 0.18, \qquad [4.6]$$

indicating that about 18% of the variance in math achievement is between schools.

Similarly, an estimator of the *reliability of the sample mean* in any school for the true school mean, β_{0j}, can also be derived by substituting the estimated variance components into Equation 3.42. That is,

$$\hat{\lambda}_j = \text{reliability}(\overline{Y}_{\cdot j}) = \hat{\tau}_{00}/\left[\hat{\tau}_{00} + (\hat{\sigma}^2/n_j)\right]. \qquad [4.7]$$

In general, the reliability of the sample mean $\overline{Y}_{\cdot j}$ will vary from school to school because the sample size, n_j, varies. However, an overall measure of the reliability can be obtained by averaging the individual school estimates:

$$\hat{\lambda} = \sum \hat{\lambda}_j/J. \qquad [4.8]$$

For the HS&B data, $\hat{\lambda} = .90$, indicating that the sample means tend to be quite reliable as indicators of the true school means.

In summary, this one-way ANOVA produces useful preliminary information in our study of math achievement in U.S. high schools. It provides an estimate of the grand mean; a partitioning of the total variation in math achievement into variation between and within schools; a *range of plausible values* for the school means and a test of the hypothesis that the variability is null; information on the degree of dependence of the observations within each school (the intraclass correlation); and a measure of the reliability of each school's sample average math achievement as an estimate of its true mean.

Regression with Means-as-Outcomes

The Model

The student-level model of Equation 4.1 remains unchanged: Student math achievement scores are viewed as varying around their school means. The school-level model of Equation 4.2 is now elaborated, however, so that each school's mean is now predicted by the MEAN SES of the school:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{MEAN SES})_j + u_{0j}, \qquad [4.9]$$

where γ_{00} is the intercept, γ_{01} is the effect of MEAN SES on β_{0j}, and we assume $u_{0j} \sim$ independently $\text{N}(0, \tau_{00})$.

Notice that the symbols u_{0j} and τ_{00} have different meanings than they had in Equation 4.2. Whereas the random variable u_{0j} had been the deviation of school j's mean from the grand mean, it now represents the residual $\beta_{0j} - \gamma_{00} - \gamma_{01}(\text{MEAN SES})_j$. Correspondingly, the variance τ_{00} is now a residual or conditional variance, that is, $\mathrm{Var}(\beta_{0j}|\mathrm{MEANSES})$, the school-level variance in β_{0j} after controlling for school MEAN SES.

Substituting Equation 4.9 into Equation 4.1 yields the combined model (or "mixed model")

$$Y_{ij} = \gamma_{00} + \gamma_{01}(\text{MEAN SES})_j + u_{0j} + r_{ij} \qquad [4.10]$$

having fixed effects γ_{00}, γ_{01}, and random effects u_{0j}, r_{ij}.

Results

Table 4.3 provides estimates and hypothesis tests for the fixed effects and the variances of the random effects.

Fixed Effects. We see a highly significant association between school MEAN SES and mean achievement ($\hat{\gamma}_{01} = 5.68, t = 16.22$). The t ratio employed for hypothesis testing of an individual fixed effect is simply the ratio of the estimated coefficient to its standard error (see Equation 3.83):

$$t = \hat{\gamma}_{01}/[\mathrm{Var}(\hat{\gamma}_{01})]^{1/2} = 5.86/0.36 = 16.22.$$

It is also possible to test the hypothesis that the grand mean is null, that is, $H_0 = \gamma_{00} = 0$, but that hypothesis is of no interest in this case.

Variance Component. The residual variance between schools, $\hat{\tau}_{00} = 2.64$, is substantially smaller than the original, $\hat{\tau}_{00} = 8.61$, estimated in the context of the random ANOVA model (see Table 4.2). A range of plausible values for school means, given that all schools have a MEAN SES of zero, is

$$\hat{\gamma}_{00} \pm 1.96(\hat{\tau}_{00})^{1/2} = 12.65 \pm 1.96(2.64)^{1/2} = (9.47, 15.83).$$

TABLE 4.3 Results from the Means-as-Outcomes Model

Fixed Effect	*Coefficient*	*se*	*t Ratio*	
Model for school means				
INTERCEPT, γ_{00}	12.65	0.15	—	
MEAN SES, γ_{01}	5.86	0.36	16.22	
Random Effect	*Variance Component*	*df*	χ^2	*p Value*
School mean, u_{0j}	2.64	158	633.52	0.000
Level-1 effect, r_{ij}	39.16			

Though this is a fairly wide range of plausible values, it is considerably smaller than the range of plausible values when MEAN SES is not held constant (Equation 4.4), which was (6.89, 18.39).

Do school achievement means vary significantly once MEAN SES is controlled? Here the null hypothesis that $\tau_{00} = 0$, where τ_{00} is now a residual variance, is tested by means of the statistic

$$\sum n_j\left[\overline{Y}_{\cdot j} - \hat{\gamma}_{00} - \hat{\gamma}_{01}(\text{MEAN SES})_j\right]^2/\hat{\sigma}^2, \qquad [4.11]$$

which, under the null hypothesis, has a χ^2 distribution with $J - 2 = 158$ degrees of freedom. In our case, the statistic has a value of 633.52, $p <$ .001, indicating that the null hypothesis is easily rejected; after controlling for MEAN SES, significant variation among school mean math achievement remains to be explained.

Auxiliary Statistics. By comparing the τ_{00} estimates across the two models, we can develop an index of the *proportion reduction in variance or "variance explained"* at level 2. In this application,

$$\text{Proportion of variance explained in } \beta_{0j}$$
$$= \frac{\hat{\tau}_{00}(\text{random ANOVA}) - \hat{\tau}_{00}(\text{MEAN SES})}{\hat{\tau}_{00}(\text{random ANOVA})}, \qquad [4.12]$$

where $\hat{\tau}_{00}(\text{random ANOVA}) = \text{Var}(\beta_{0j})$ and $\hat{\tau}_{00}(\text{MEAN SES}) = \text{Var}(\beta_{0j}|$ MEAN SES) refer to the estimates of τ_{00} under the alternative level-2 models specified by Equations 4.2 and 4.9, respectively. Note the $\hat{\tau}_{00}$(random ANOVA) provides the base in this application, because it represents the total parameter variance in the school means that is potentially explainable by alternative level-2 models for β_{0j}. The estimated proportion of variance between schools explained by the model with MEAN SES is

$$(8.61 - 2.64)/8.61 = 0.69.$$

That is, 69% of the true between-school variance in math achievement is accounted for by MEAN SES.

After removing the effect of school MEAN SES, the correlation between pairs of scores in the same school, which had been .18, is now reduced:

$$\hat{\rho} = \hat{\tau}_{00}/(\hat{\tau}_{00} + \hat{\sigma}^2)$$
$$= 2.64/(2.64 + 39.16) = .06.$$

The estimated ρ is now a *conditional intraclass correlation* and measures the degree of dependence among observations within schools that are of the same MEAN SES.

Similarly, we can calculate the reliability of the least squares residuals, $\hat{u}_{0j}$,

$$\hat{u}_{0j} = \overline{Y}_{\cdot j} - \hat{\gamma}_{00} - \hat{\gamma}_{01}(\text{MEAN SES})_j. \qquad [4.13]$$

This reliability is a *conditional reliability*, that is, the reliability with which one can discriminate among schools that are identical on MEAN SES. Substituting our new estimates of $\hat{\tau}_{00}$ and $\hat{\sigma}^2$ into Equations 4.7 and 4.8 yields an average reliability of .74. As one might expect, the reliability of the residuals is less than the reliability of the sample means.

In summary, we have learned from the means-as-outcomes model that MEAN SES is significantly positively related to mean achievement. Nonetheless, even after we hold constant, or control for, MEAN SES, schools still vary significantly in their average achievement levels.

The Random-Coefficient Model

We now consider an analysis of the SES-math achievement relationship within the 160 schools. We conceive of each school as having "its own" regression equation with an intercept and a slope, and we shall ask the following:

1. What is the average of the 160 regression equations (i.e., what are the average intercept and slope)?
2. How much do the regression equations vary from school to school? Specifically, how much do the intercepts vary and how much do the slopes vary?
3. What is the correlation between the intercepts and the slopes? (Do schools with large intercepts [e.g., high mean achievement] also have large slopes [strong relationships between SES and achievement]?)

The Model

To answer these questions, we use the random-coefficient regression model introduced in Chapter 2. Specifically, we formulate at level 1 the student-level model

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{\cdot j}) + r_{ij}. \qquad [4.14]$$

Each school's distribution of math achievement is characterized by two parameters: the intercept, β_{0j}, and the slope, β_{1j}. Because the student-level predictor is centered around its school mean, the intercept, β_{0j}, is the school-mean outcome (see Equation 2.29). Again, we assume $r_{ij} \sim$ independently $\mathrm{N}(0, \sigma^2)$, where now σ^2 is the residual variance at level 1 after controlling for student SES.

These parameters, β_{0j} and β_{1j}, vary across schools in the level-2 model as a function of a grand mean and a random error:

$$\beta_{0j} = \gamma_{00} + u_{0j}, \qquad [4.15a]$$

$$\beta_{1j} = \gamma_{10} + u_{1j}, \qquad [4.15b]$$

where

γ_{00} is the average of the school means on math achievement across the population of schools;

γ_{10} is the average SES-math regression slope across those schools;

u_{0j} is the unique increment to the intercept associated with school j; and

u_{1j} is the unique increment to the slope associated with school j.

Substituting Equation 4.15 into Equation 4.14 yields a mixed model:

$$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \overline{X}_{\cdot j}) + u_{0j} + u_{1j}(X_{ij} - \overline{X}_{\cdot j}) + r_{ij}. \qquad [4.16]$$

We assume that u_{0j} and u_{1j} are multivariate normally distributed, both with expected values of 0. We label the variances in these school effects as

$$\mathrm{Var}(u_{0j}) = \tau_{00},$$

$$\mathrm{Var}(u_{1j}) = \tau_{11}$$

and the covariance between them as

$$\mathrm{Cov}(u_{0j}, u_{1j}) = \tau_{01}.$$

Collecting these terms into a variance-covariance matrix,

$$\mathrm{Var}\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} = \mathbf{T}. \qquad [4.17]$$

Because the level-2 model is unconditional for both β_{0j} and β_{1j} (i.e., no predictors are included in Equations 4.15a and 4.15b),

$$\begin{aligned} \mathrm{Var}(u_{0j}) &= \mathrm{Var}(\beta_{0j} - \gamma_{00}) = \mathrm{Var}(\beta_{0j}), \\ \mathrm{Var}(u_{1j}) &= \mathrm{Var}(\beta_{1j} - \gamma_{10}) = \mathrm{Var}(\beta_{1j}). \end{aligned} \qquad [4.18]$$

Thus, the random-coefficient regression model provides estimates for the unconditional parameter variability in the random intercepts and slopes.

TABLE 4.4 Results from the Random-Coefficient Model

Fixed Effect		*Coefficient*	*se*	*t Ratio*
Overall mean achievement, γ_{00}		12.64	0.24	—
Mean SES-achievement slope, γ_{10}[a]		2.19	0.13	17.16
Random Effect	*Variance*	*df*	χ^2	*p Value*
School mean, u_{0j}	8.68	159	1,770.9	0.000
SES-achievement slope, u_{1j}	0.68	159	213.4	0.003
Level-1 effect, r_{ij}	36.70			

Results

Fixed Effects. Table 4.4 provides the estimates for the average regression equation within schools. Using the generalized least squares estimator of Equation 3.31 (or, equivalently, Equation 3.37), we find the average school mean

$$\hat{\gamma}_{00} = 12.64$$

and the average SES-achievement slope

$$\hat{\gamma}_{10} = 2.19.$$

The corresponding standard errors, based on Equation 3.32 (or Equation 3.38), are 0.24 and 0.13, respectively. We can use this information to formally test the null hypothesis that, *on average*, student SES is not related to math achievement within schools, that is,

$$\mathrm{H}_0: \gamma_{10} = 0.$$

Based on the test statistic of Equation 3.83:

$$t = \hat{\gamma}_{10}/(\widehat{V}_{\hat{\gamma}_{10}})^{1/2} = 2.19/0.13 = 17.16,$$

we find that, on average, student SES is significantly related ($p < .001$) to math achievement within schools.

Variance-Covariance Components. Using the general procedures for maximum likelihood estimation discussed in Chapter 3, we estimate the variance and covariance components of Equation 4.17:

$$\widehat{\mathrm{Var}}\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} \hat{\tau}_{00} & \hat{\tau}_{01} \\ \hat{\tau}_{10} & \hat{\tau}_{11} \end{bmatrix} = \begin{bmatrix} 8.68 & 0.04 \\ 0.04 & 0.68 \end{bmatrix} = \widehat{\mathbf{T}}.$$

Table 4.4 also provides the test statistics (see Equation 3.101) for the hypotheses that each of the variance components along the diagonal of $\mathbf{T}$ are null, that is,

$$\mathrm{H}_0\colon \tau_{qq} = 0 \quad \text{for } q = 0, 1.$$

Specifically, the estimated variance among the means is $\hat{\tau}_{00} = 8.68$, with a χ^2 statistic of 1,770.5, to be compared to the critical value of χ^2 with $J - 1 = 159$ degrees of freedom. We infer that highly significant differences exist among the 160 school means, a result quite similar to that encountered in the one-way ANOVA with random effects.

The estimated variance of the slopes is $\hat{\tau}_{11} = 0.68$ with a χ^2 statistic of 213.4 and 159 degrees of freedom, $p < .003$. Again, we reject the null hypothesis, in this case that $\tau_{11} = 0$, and infer that the relationship between SES and math achievement within schools does indeed vary significantly across the population of schools.

We can also now calculate a range of plausible values for both the school means and the school-specific SES-achievement slopes. Generalizing from Equation 4.4, the 95% plausible value range is

$$\hat{\gamma}_{qo} \pm 1.96(\hat{\tau}_{qq})^{1/2} \qquad [4.19]$$

for the $q = 0, \ldots, Q$ random coefficients in the level-1 model. (In this example, $Q = 1$.) The 95% plausible value range for the school means is

$$12.64 \pm 1.96(8.68)^{1/2} = (6.87, 18.41),$$

and for the SES-achievement slopes is

$$2.19 \pm 1.96(0.68)^{1/2} = (0.57, 3.81).$$

The results for the school means are similar to those previously reported for the one-way ANOVA model. In terms of the school-specific SES-achievement slopes, here, too, we find considerable variability among schools. This relationship is over seven times stronger in the most socially differentiating schools as compared to the least differentiating schools.

Auxiliary Statistics. Associated with β_{0j} and β_{1j} is also a reliability estimate (see Equation 3.59). The results are

$$\text{reliability}(\hat{\beta}_0) = 0.91$$

and

$$\text{reliability}(\hat{\beta}_1) = 0.23.$$

These indices provide answers to the question "How reliable, on average, are estimates of each school's intercept and slope based on computing the OLS regression separately for each school?" These reliabilities depend on two factors: the degree to which the true underlying parameters vary from school to school and the precision with which each school's regression equation is estimated.

The precision of estimation of the intercept (which in this application is a school mean) depends on the sample size within each school. The precision of estimation of the slope for school j depends both on the sample size and on the variability of SES within that school. Schools that are homogeneous with respect to SES will exhibit slope estimates with poor precision.

The results indicate that the intercepts are quite reliable (.91) based on an average of 50 students per school. The slope estimates are far less reliable (.23). The primary reason for the lack of reliability of the slopes is that the true slope variance across schools is much smaller than the variance of the true means. Also, the slopes are estimated with less precision than are the means because many schools are relatively homogeneous on SES.

Analogous to Equation 4.12, we can develop an index of the *proportion reduction in variance or "variance explained"* at level 1 by comparing the σ^2 estimates from these two alternative models. Notice that the estimate of the student-level variance $\hat{\sigma}^2$ is now 36.70. By comparison, the estimated variance in the one-way random ANOVA model, which did not include SES as a level-1 predictor, was 39.15. Thus,

$$\begin{aligned}&\text{Proportion variance explained at level 1}\\&= \frac{\hat{\sigma}^2(\text{random ANOVA}) - \hat{\sigma}^2(\text{SES})}{\hat{\sigma}^2(\text{random ANOVA})}\\&= \frac{39.15 - 36.70}{39.15} = .063,\end{aligned} \qquad [4.20]$$

where $\hat{\sigma}^2$(random ANOVA) and $\hat{\sigma}^2$(SES) refer to estimates of σ^2 based on the level-1 models specified by Equations 4.1 and 4.14, respectively. Note that σ^2(random ANOVA) provides the appropriate base in this application because it represents the total within-school variance that can be explained by any level-1 model.

We see that adding SES as a predictor of math achievement reduced the within-school variance by 6.3%. Hence, we can conclude that SES accounts for about 6% of the student-level variance in the outcome. When we recall that MEAN SES accounted for better than 60% of the between-school variance in the outcome, it is clear that the association between these two variables is far stronger at the school level than at the student level.

Finally, the model also produces a maximum likelihood estimate of the covariance between the intercept and the slope. When combined with the estimates of the intercept and slope variances, we can estimate the *correlation between the intercept and slope* using Equation 2.3. In this case, the correlation of slope and intercept is .02, indicating that there is little association between school means and school SES effects.

An Intercepts- and Slopes-as-Outcomes Model

Having estimated the variability of the regression equations across schools, we now seek to build an explanatory model to account for this variability. That is, we seek to understand *why* some schools have higher means than others and why in some schools the association between SES and achievement is stronger than in others.

The Model

The student-level model remains the same as in Equation 4.14. However, we now expand the school-level model to incorporate two predictors: SECTOR and MEAN SES. The resulting school-level model can be written as

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{MEAN SES})_j + \gamma_{02}(\text{SECTOR})_j + u_{0j}, \qquad [4.21a]$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{MEAN SES})_j + \gamma_{12}(\text{SECTOR})_j + u_{1j}, \qquad [4.21b]$$

where u_{0j} and u_{1j} are again multivariate normally distributed with means of zero and variance-covariance matrix $\mathbf{T}$. The elements of $\mathbf{T}$ are now residual or conditional variance-covariance components. That is, they represent residual dispersion in β_{0j} and β_{1j} after controlling for MEAN SES and SECTOR.

Combining the school-level model (Equation 4.21) and the student-level model (Equation 4.14) yields

$$\begin{aligned} Y_{ij} = {} & \gamma_{00} + \gamma_{01}(\text{MEAN SES})_j + \gamma_{02}(\text{SECTOR})_j \\ & + \gamma_{10}(X_{ij} - \overline{X}_{\cdot j}) + \gamma_{11}(\text{MEAN SES})_j(X_{ij} - \overline{X}_{\cdot j}) \\ & + \gamma_{12}(\text{SECTOR})_j(X_{ij} - \overline{X}_{\cdot j}) + u_{0j} + u_{1j}(X_{ij} - \overline{X}_{\cdot j}) + r_{ij}, \end{aligned} \qquad [4.22]$$

which illustrates that the outcome may be viewed as a function of the overall intercept (γ_{00}), the main effect of MEAN SES (γ_{01}), the main effect of SECTOR (γ_{02}), the main effect of SES (γ_{10}), and two cross-level interactions involving SECTOR with student SES (γ_{12}) and MEAN SES with student SES (γ_{11}), plus a random error

$$u_{0j} + u_{1j}(X_{ij} - \overline{X}_{\cdot j}) + r_{ij}.$$

Three kinds of questions motivate the analysis:

1. Do MEAN SES and SECTOR significantly predict the intercept? We estimate γ_{01} to study whether high-SES schools differ from low-SES schools in mean achievement (controlling for SECTOR). Similarly, we estimate γ_{02} to learn whether Catholic schools differ from public schools in terms of the mean achievement once MEAN SES is controlled.
2. Do MEAN SES and SECTOR significantly predict the within-school slopes? We estimate γ_{11} to discover whether high-SES schools differ from low-SES schools in terms of the strength of association between student SES and achievement within them (controlling for SECTOR). We estimate γ_{12} to examine whether Catholic schools differ from public schools in terms of the strength of association between student SES and achievement (controlling for MEAN SES).
3. How much variation in the intercepts and the slopes is explained by using SECTOR and MEAN SES as predictors? To answer these questions, we estimate $\text{Var}(u_{0j}) = \tau_{00}$ and $\text{Var}(u_{1j}) = \tau_{11}$ and compare these with the estimates presented above from the random-coefficient regression model.

Results

Fixed Effects. Table 4.5 displays the results. We see, first, that MEAN SES is positively related to school mean math achievement, $\hat{\gamma}_{01} = 5.33$, $t = 14.45$. Also, Catholic schools have significantly higher mean achievement than do public schools, controlling for the effect of MEAN SES, $\hat{\gamma}_{02} = 1.23$, $t = 4.00$.

With regard to the slopes, there is a tendency for schools of high MEAN SES to have larger slopes than do schools with low MEAN SES, $\hat{\gamma}_{11} = 1.03$, $t = 3.42$. Catholic schools have significantly weaker SES slopes, on average, than do public schools, $\hat{\gamma}_{12} = -1.64$, $t = -6.76$.

These results are depicted graphically in Figure 4.1. The fitted relationship between SES and math achievement is displayed for the Catholic and public sectors. Within each sector, results are displayed for (1) a high-SES school (one standard deviation above the mean), (2) a medium-SES school, and (3) a low-SES school (one standard deviation below the mean). Perhaps the most notable feature of the figure is that the within-school math-SES slopes are substantially less steep in the Catholic sector than in the public sector. This sector effect holds for schools at each level of MEAN SES. There is also

TABLE 4.5 Results from the Intercepts- and Slopes-as-Outcomes Model

Fixed Effects	*Coefficient*	*se*	*t Ratio*	
Model for school means				
INTERCEPT, γ_{00}	12.10	0.20	—	
MEAN SES, γ_{01}	5.33	0.37	14.45	
SECTOR, γ_{02}	1.23	0.31	4.00	
Model for SES-achievement slopes				
INTERCEPT, γ_{10}	2.94	0.16	—	
MEAN SES, γ_{11}	1.03	0.30	3.42	
SECTOR, γ_{12}	−1.64	0.24	−6.76	
Random Effects	*Variance Component*	*df*	χ^2	*p Value*
School mean, u_{0j}	2.38	157	605.30	0.000
SES-achievement slope, u_{1j}	0.15	157	162.31	0.369
Level-1 effect, r_{ij}	36.68			

a tendency for high-SES schools to have steeper slopes than do low-SES schools. This tendency is evident in both sectors. Main effects of MEAN SES and SECTOR are also evident. The MEAN SES effect is manifest by the solid lines that in both plots have positive slopes.

Chapter 3 discusses a procedure for testing multiparameter hypotheses regarding the fixed effects. One may wonder, for example, whether the variable SECTOR is needed in the model. Perhaps no distinction is justified between Catholic and public schools in terms of effectiveness or equity. The null hypothesis may be written as

$$\begin{aligned} H_0\colon \gamma_{02} &= 0 \\ \gamma_{12} &= 0. \end{aligned}$$

If $\gamma_{02} = 0$, Catholic and public schools do not differ in mean achievement after controlling for MEAN SES. Similarly, if $\gamma_{12} = 0$, Catholic and public schools do not differ with respect to their average SES-math achievement relationships with schools. If both null hypotheses are true, the variable SECTOR may be dropped from the model. Using Equation 3.91, we obtain a χ^2 statistic of 64.38, df $= 2$, $p < .001$, indicating that one or both of the null hypotheses is false.

Variance-Covariance Components. Recall that the results of fitting the random-coefficient model provided information about the variation and

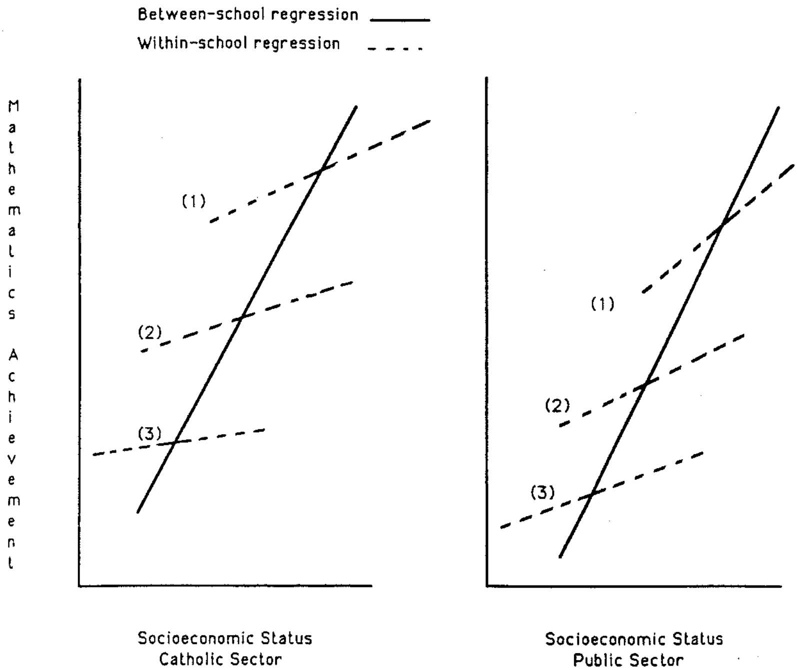

Figure 4.1. Regressions of Mathematics Achievement as a Function of Student and School SES Within Catholic and Public Sectors

NOTE: Schools 1, 2, and 3 are of high, medium, and low SES, respectively.

covariation of the intercepts and SES slopes across schools. Our interest now focuses on the residual variation and covariation of the intercepts and slopes, that is, the variation left unexplained by SECTOR and MEAN SES. The maximum likelihood point estimates are

$$\widehat{\mathrm{Var}}\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} = \begin{pmatrix} \hat{\tau}_{00} & \hat{\tau}_{01} \\ \hat{\tau}_{10} & \hat{\tau}_{11} \end{pmatrix} = \begin{pmatrix} 2.38 & 0.19 \\ 0.19 & 0.15 \end{pmatrix}. \qquad [4.23]$$

We note that both $\hat{\tau}_{00}$ and $\hat{\tau}_{11}$, the estimated variances of the intercepts and the slopes, are considerably smaller than they had been without control for SECTOR and MEAN SES. Perhaps there remains no significant residual variation in the intercepts and slopes after control for SECTOR and MEAN SES. Regarding the intercepts, the null hypothesis

$$\mathrm{H}_0\colon \tau_{00} = 0$$

is rejected, as indicated by the χ^2 statistic of 605.30, df $= J - S_q - 1 = 157, p < .001$. Thus, significant variation in the intercepts remains unexplained even after controlling for SECTOR and MEAN SES. Regarding the slopes, the null hypothesis

$$\mathrm{H}_0\colon \tau_{11} = 0$$

is retained, as indicated by the χ^2 statistic of 162.31, df $= J - S_q - 1 = 157$, $p = .369$. This test suggests that no significant variation in the slopes remains unexplained after controlling for SECTOR and MEAN SES.

Another way to examine the significance of variance and covariance components is to compare two models, one model including the variance components of interest and a second, simpler model that constrains certain components to zero. If the fit of the simpler model to the data is significantly worse than the fit of the more complex model, the simpler model is rejected as inadequately representing the variation in the data. However, if there is no significant difference in the fit of the two models, the simpler model will typically be preferred. Applying this logic to the current data, we compare a model with random intercepts and slopes to a model with only a random intercept.

Our "intercepts- and slopes-as-outcomes" model included four unique variance and covariance parameters: (a) the student-level variance, σ^2; (b) the residual variance of the school means, τ_{00}; (c) the residual variance of the SES-math slopes, τ_{11}; and (d) the residual covariance between the means and the slopes, τ_{01}. In general, the number of variance-covariance parameters estimated in a two-level model is $m(m+1)/2+1$, where m is the number of random effects in the level-2 model. In our case, $m = 2$. Chapter 3 (see Equations 3.105 through 3.107) described a likelihood-ratio test that can be used to test the composite null hypothesis

$$\mathrm{H}_0\colon \begin{pmatrix} \tau_{11} = 0 \\ \tau_{01} = 0 \end{pmatrix}.$$

One first estimates the full model with four variance-covariance parameters, and then one estimates a reduced model with just two parameters (σ^2 and τ_{00})—that is, where τ_{11} and τ_{01} have been constrained to zero. One then compares the deviance associated with the two models and asks whether the reduction in deviance associated with the more complex model is justified.

In our case, the results are as follows:

Model	*Number of Parameters*	*Deviance*
Restricted, D_0	2	46,514.0
Unrestricted, D_1	4	46,513.1

The reduction in deviance is 0.9, which is not significant when compared against the χ^2 distribution with 2 df. Hence, the simpler model seems justified. We infer that explanatory power is not significantly enhanced by specifying the residual SES-achievement slopes as random. The reduced model with β_{1j} specified as nonrandomly varying appears sufficient.

Auxiliary Statistics. Analogous to Equation 4.12, we can develop a *proportion reduction in variance or variance-explained statistic for each of the random coefficients (intercepts and slopes) from the level-*1 *model.* The variance estimates from the random-coefficient regression model estimated earlier provide the base for these statistics:

$$\text{Proportion variance explained in } \beta_{qj} = \frac{\hat{\tau}_{qq}(\text{random regression}) - \hat{\tau}_{qq}(\text{fitted model})}{\hat{\tau}_{qq}(\text{random regression})}, \qquad [4.24]$$

where $\hat{\tau}_{qq}$(random regression) denotes the qth diagonal element of $\mathbf{T}$ estimated under the random-regression model (Equations 4.14 and 4.15) and $\hat{\tau}_{qq}$(fitted model) denotes the corresponding element in the T matrix estimated under an intercepts- and slopes-as-outcomes model (in this case Equations 4.14 and 4.21).

In this application, we see a substantial reduction in variance of the school means once MEAN SES and SECTOR are controlled. Specifically, whereas the unconditional variance of intercepts had been 8.68, the residual variance is now $\hat{\tau}_{00} = 2.38$. This means that 73% of the parameter variation in mean achievement, $\text{Var}(\beta_{0j})$, has been explained by MEAN SES and SECTOR [i.e., $(8.68 - 2.38)/8.68 = 0.73$]. Similarly, the residual variance of the slopes is $\hat{\tau}_{11} = 0.15$, which, when compared to the unconditional variance of .68, implies a reduction of 78%. Clearly, most of the slope variability is associated with MEAN SES and SECTOR. Once this is controlled, only a small residual portion of variation remains unexplained. Chapter 5 provides a more detailed discussion, with some caveats, of strategies for monitoring explained variance (see "Use of Proportion Reduction in Variance Statistics").

Estimating the Level-1 Coefficients for a Particular Unit

In this chapter, we have characterized the distribution of achievement in each high school in terms of two school-specific parameters: a school's mean math achievement and a regression coefficient describing the relationship between SES and math achievement. We have viewed these level-1 coefficients as "random parameters" varying over the population of schools with some of that random variation a function of measured predictors. As mentioned in Chapter 3, we can obtain both point and interval estimates for each random level-1 coefficient. Formally, these are empirical Bayes estimators, also known as shrinkage estimators. These shrinkage estimators may be subdivided into two categories: *unconditional* and *conditional* shrinkage estimators. We illustrate each approach and compare these with the OLS estimates.

Ordinary Least Squares

The most obvious strategy for estimating the regression equation for a particular school is simply to fit a separate model to each school's data by ordinary least squares (OLS). The model for each school might simply be Equation 4.14. Recall that with SES centered around the school mean, the intercept β_{0j} is the mean outcome for that school, and the regression coefficient β_{1j} represents the expected difference in achievement per unit difference in SES within that school. OLS will produce unbiased estimates of these parameters for any school that has at least two cases. Indeed, if the errors of the model are independently normally distributed, the OLS estimates are the unique, minimum-variance, unbiased estimators of these parameters. Nonetheless, the OLS estimates for any given school may not be very accurate as we illustrate below.

Columns 1 and 2 in Table 4.6 present separate OLS estimates for 12 selected schools out of the 160 cases in the HS&B data set. These estimates were calculated for each school using Equation 3.54. Based on these estimates, we might identify case 4 as an especially good school where there is a high average level of achievement, $\hat{\beta}_{04} = 16.26$, that is distributed in an equitable social fashion, $\hat{\beta}_{14} = 0.13$.

Figure 4.2a shows the OLS estimates for all 160 schools. The intercept estimates (vertical axis) are plotted against the slope estimates (horizontal axis). Quite a few schools yield negative estimates of the SES-achievement relationship. Moreover, the apparent dispersion of the OLS estimates greatly exceeds the maximum likelihood estimate of the variance of the true slopes. Earlier, we estimated the variance of the true slope parameters to be 0.68. Yet the sample variance of the OLS slope estimates depicted in Figure 4.2a is 2.66. If we were to define effective and equitable schools as those with large means and small SES-math achievement slopes, we might identify many schools like case 4.

TABLE 4.6 Comparison of Estimated Level-1 Coefficients for a Sample of HS&B Cases

	OLS Estimates		*Empirical Bayes Estimates Unconditional Model*		*Empirical Bayes Estimates Conditional Model*				
Case	$\hat{\beta}_{0j}$	$\hat{\beta}_{1j}$	β^*_{0j}	β^*_{1j}	β^*_{0j}	β^*_{1j}	n_j	*MEAN SES*	*SECTOR*
4	16.26	0.13	15.62	2.05	16.20	1.82	20	0.53	1
15	15.98	2.15	15.74	2.19	16.01	1.85	53	0.52	1
17	18.11	0.09	17.41	1.95	17.25	3.67	29	0.69	0
22	11.14	−0.78	11.22	1.16	10.89	0.58	67	−0.62	1
27	13.40	4.10	13.32	2.53	12.95	3.02	38	−0.06	0
53	9.52	3.74	9.76	2.74	9.37	2.47	51	−0.64	0
69	11.47	6.18	11.64	2.71	11.92	3.06	25	0.08	0
75	9.06	1.65	9.28	2.01	9.30	0.70	63	−0.59	1
81	15.42	5.26	15.25	3.12	15.52	2.01	66	0.43	1
90	12.14	1.97	12.18	2.14	12.34	3.01	50	0.19	0
135	4.55	0.25	6.42	1.93	8.55	2.61	14	0.03	0
153	10.28	0.76	10.71	2.07	9.67	2.36	19	−0.59	0

Unconditional Shrinkage

In general, the accuracy of OLS regression estimates for any school depends on the sample size within the school, n_j, and the range represented in the level-1 predictor variable, X_{ij}. If n_j is small, the mean estimate, $\hat{\beta}_{0j}$, will be imprecise. If a school has a small sample or a restricted range on SES, the slope estimate, $\hat{\beta}_{1j}$, will also tend to be imprecise. The empirical Bayes (EB) estimates of each school's regression line takes into account this imprecision in the OLS estimates.

Columns 3 and 4 in Table 4.6 present the EB estimates for the 12 selected schools from HS&B. These estimates, β^*_{0j} and β^*_{1j}, are based on the unconditional level-2 model (Equations 4.14 and 4.15) and calculated using Equation 3.56 (or, equivalently, Equation 3.66). Notice that the EB estimates for case 4 differ substantially from the OLS estimates. The estimated average achievement level has dropped by 0.64 (from 16.26 to 15.62) and the SES-achievement slope has risen from 0.13 to 2.05. While the estimate of the overall achievement level for the school remains relatively high, the equity effect has disappeared. While case 4 looked superior to case 15 in terms of the OLS estimates, these two schools appear indistinguishable in terms of the empirical Bayes estimates (15.62 versus 15.74 for β^*_{0j} and 2.05 versus 2.19 for β^*_{1j}). The key factor here is the relatively small sample size of only 20 students in case 4. As a result, the OLS estimates for this unit are not very precise, and the EB estimates are shrunk toward the overall mean achievement, $\hat{\gamma}_{00} = 12.64$, and the overall mean SES-achievement slope, $\hat{\gamma}_{10} = 2.19$. Notice that this occurs for all of the cases with relatively small sample sizes (cases 17, 69, 135, and 153), with sample sizes 29, 25, 14, and 19, respectively.

Figure 4.2b displays the empirical Bayes estimates of the intercepts (vertical axis) and the math-SES slopes (horizontal axis) for all 160 schools. Notice that the empirical Bayes slope estimates are much more concentrated around the sample average than are the OLS estimates in Figure 4.2a. Unlike the collection of OLS slope estimates, none of the empirical Bayes estimates is negative. Also, the sample variance of the empirical Bayes slope estimates is only .14, much smaller than the sample variance of the OLS slope estimates (2.66). In fact, the sample variance of the empirical Bayes slope estimates is smaller than the maximum likelihood estimate of the variance of the true slopes (0.68).

We note that these results generalize:

$\mathrm{Var}(\hat{\beta}_j)$	>	$\widehat{\mathrm{Var}}(\beta_j)$	>	$\mathrm{Var}(\beta_j^*)$
observed variability in OLS estimates	>	maximum likelihood estimates of variance in level-1 coefficients, i.e. $\hat{\tau}_{00}$, $\hat{\tau}_{11}$ under Equation 4.15	>	observed variability in the empirical Bayes estimates

The fact that the empirical Bayes estimates have less variance than the estimated true variance is an expected result. In general, the shrinkage is slightly exaggerated; empirical Bayes tends to pull the estimates "too far" toward the sample average.

It is also interesting to contrast the slope shrinkage in Figure 4.2, with the results for the achievement intercepts. Recall that the intercepts are just the mean achievement in each school and are much more reliably estimated than are the slopes (.91 versus .23). Given the greater precision of the intercept estimates, we would expect that the empirical Bayes estimator would rely more heavily on this component, and less shrinkage should occur. This result is displayed in Figure 4.2 (compare the vertical axes of Figures 4.2a and 4.2b). Unlike the slopes, where the shrinkage is substantial, the difference between the OLS and the empirical Bayes estimates for the intercepts is only modest. This same pattern can be observed in the results for the 12-school subsample presented in Table 4.6.

In general, the behavior of the empirical Bayes estimator is simpler in the case of random-intercept models than in models that also have random slopes.

```
(a) Ordinary Least Squares

                -
                -                     *       *  *
            18.0+                       2   ***
                -                       2*****2 * *          *
     OLS        -                    *    ********2 *     *
    Intercept   -                      * 22 *33*424* *  **
                -                      ** *2***   6*424**
            12.0+                           22 *3   *2*2**2  *
                -                    2 *   2*22 * **2** **  2
                -                *    2*   * 22 2 * *2
                -                     *    **   **   *
                -                     *    **    *2 *
             6.0+                          **
                -                    *   *    *
                    +---------+---------+---------+---------+--
                  -6.0      -3.0      0.0       3.0       6.0

                                      OLS Slope
```

```
(b) Empirical Bayes(a)

                -
                -                            * *
            18.0+                             *4*
                -                             *3* *
     EB         -                            *483 *
  Intercept     -                            *66+2
                -                            *34+6
            12.0+                              7373
                -                           **3654**
                -                          * 3242*
                -                            *3*32
                -                            * **
             6.0+                             2*
                -                             **
                    +---------+---------+---------+---------+--
                  -6.0      -3.0      0.0       3.0       6.0

                                      EB Slope
```

Figure 4.2. Ordinary Least Squares and Empirical Bayes Intercept and Slope Estimates for 160 High Schools

NOTE: Intercept estimates are plotted on the vertical axis, slope estimates on the horizontal axis.
a. A "+" indicates the presence of more than nine observations.

For example, consider the present case, in which we focus on each school's SES-achievement slope and its intercept. The empirical Bayes estimate of each component will depend on the other component. This dependence will be strong when the maximum likelihood estimate of the correlation between the two components is large.

Conditional Shrinkage

A tool for increasing accuracy in estimating β_{0j} and β_{1j} is conditional shrinkage. Rather than pulling each OLS regression line toward the grand-mean regression line of $\hat{\gamma}_{00}$ and $\hat{\gamma}_{10}$, the OLS regression lines will now be pulled toward a *predicted value* based on the school-level model.

Unconditional shrinkage of the random-coefficient regression model (i.e., Equation 4.15) yields

$$\begin{bmatrix} \beta^*_{0j} \\ \beta^*_{1j} \end{bmatrix} = \mathbf{\Lambda}_j \begin{bmatrix} \hat{\beta}_{0j} \\ \hat{\beta}_{1j} \end{bmatrix} + (\mathbf{I} - \mathbf{\Lambda}_j) \begin{bmatrix} \hat{\gamma}_{00} \\ \hat{\gamma}_{10} \end{bmatrix}, \tag{4.25}$$

where Λ_j is based on the estimation of σ^2 and $\mathbf{T}$ for the model in Equations 4.14 and 4.15. In contrast, the intercepts- and slopes-as-outcomes model (see Equation 3.56 or, equivalently, Equation 3.66) yields conditional shrinkage toward predicted values of β_{0j} and β_{1j}. That is,

$$\begin{bmatrix} \beta^*_{0j} \\ \beta^*_{1j} \end{bmatrix} = \mathbf{\Lambda}_j \begin{bmatrix} \hat{\beta}_{0j} \\ \hat{\beta}_{1j} \end{bmatrix}$$

$$+(\mathbf{I}-\mathbf{\Lambda}_j) \begin{bmatrix} \hat{\gamma}_{00} + \hat{\gamma}_{01} \ \ (\text{MEAN SES})_j \ \ +\hat{\gamma}_{02} \ \ (\text{SECTOR})_j \\ \hat{\gamma}_{10} + \hat{\gamma}_{11} \ \ (\text{MEAN SES})_j \ \ \ \hat{\gamma}_{12} \ \ (\text{SECTOR})_j \end{bmatrix}, \tag{4.26}$$

where Λ_j is now based on the estimation of σ^2 and $\mathbf{T}$ for the model in Equations 4.14 and 4.21.

As with unconditional shrinkage, the effects of conditional shrinkage can be quite extreme when the within-group sample size, n_j, is small. (See results in columns 5 and 6 in Table 4.6.) Notice for case 135 that the OLS estimates for β_{0j} and β_{1j} were $(4.55, 0.25)$ respectively; with unconditional shrinkage, they moved to $(6.42, 1.93)$, respectively; and for conditional shrinkage they ended up at $(8.55, 2.61)$. Under conditional shrinkage, the OLS regression line for case 135 is being pulled toward a predicted value for β_{0j} and β_{1j} based on the information from this school about its social class (MEAN SES = .03) and sector (SECTOR = 0 = public). Substituting these values into the estimated equations in Table 4.5 yields predictions of

$$\hat{E}(\beta_{0j}) = 12.10 + 5.33(0.03) + 1.23(0) = 12.26,$$

$$\hat{E}(\beta_{1j}) = 2.94 + 1.03(0.03) - 1.64(0) = 2.97.$$

Notice that the empirical Bayes estimate for β_{1j}, $\beta^*_{1j} = 2.61$, entails virtual complete shrinkage from the original OLS value of 0.25 toward the predicted value of 2.97. In addition to the fact that the sample size is small for this unit ($n_{135} = 14$) and that regression slopes are less reliable, the amount of change under conditional shrinkage also depends on the precision of the prediction equations, which, in turn, are a function of the residual variances in **T**. Since $\hat{\tau}_{11} = 0.15$, the prediction equation is relatively precise in this application and the shrinkage more extensive. We note that the same factors are at work in the conditional shrinkage of $\hat{\beta}_{0j}$ toward β^*_{0j}. The proportional amount of shrinkage $\hat{\beta}_{0j}$ versus $\hat{\beta}_{1j}$ is less, however, because school means are considerably more reliable than slopes, for any fixed value of n_j. Since the original sample mean is relatively more precise, empirical Bayes gives relatively more weight to $\hat{\beta}_{0j}$ in the estimation of β^*_{0j}.

Results in Table 4.6 also illustrate the differential effect that conditional shrinkage can have on various units. Compare, for example, case 22 (a low-SES Catholic school) with case 27 (an average-SES public school). The OLS estimates suggest substantial differences between these two schools in both average achievement level (11.14 versus 13.40) and SES-achievement slopes (−0.78 versus 4.10). With unconditional shrinkage, much of the observed differences are "shrunk out." We know, however, from the results in Table 4.6 that school MEAN SES and SECTOR predict both the mean achievement level and the SES-achievement slope for a school. When we take this into account through conditional shrinkage, most of the original differences reappear: 10.89 versus 12.95 for β_{0j} and 0.58 versus 3.02 for β_{1j}.

Another interesting set of comparisons are schools 17 and 81. Both are relatively high SES schools, with case 17 being public and case 81 being Catholic. Both schools have relatively high average achievement levels ($\hat{\beta}_{0,17} = 18.11$; $\hat{\beta}_{0,81} = 15.42$), but the Catholic school has a very steep SES-achievement slope ($\hat{\beta}_{1,81} = 5.26$) as compared to the public school ($\hat{\beta}_{1,17} = 0.09$). Both of the schools are multivariate outliers in that their OLS slope estimates appear inconsistent with all the other information we have in the data set. That is, from Table 4.5 we expect high-SES public schools to have steep slopes, not their Catholic counterparts. In this instance, conditional shrinkage literally reorders the two equations. While β^*_{0j} remains higher for the public school (17.25 versus 15.52), the conditional shrinkage estimate for the SES-achievement slopes is now lower in the Catholic school than in the public (2.01 versus 3.67)!

The effects of conditional shrinkage can also be discerned through the use of OLS and empirical Bayes residuals (see Equations 3.50, 3.60; and

3.49, 3.61 respectively). The OLS residuals for the intercepts and slopes in Equation 4.21 are

$$\hat{u}_{0j} = \hat{\beta}_{0j} - \left[\hat{\gamma}_{00} + \hat{\gamma}_{01}(\text{MEAN SES})_j + \hat{\gamma}_{02}(\text{SECTOR})_j\right], \qquad [4.27a]$$

$$\hat{u}_{1j} = \hat{\beta}_{1j} - \left[\hat{\gamma}_{10} + \hat{\gamma}_{11}(\text{MEAN SES})_j + \hat{\gamma}_{12}(\text{SECTOR})_j\right]. \qquad [4.27b]$$

The corresponding empirical Bayes residuals are

$$u^*_{0j} = \beta^*_{0j} - \left[\hat{\gamma}_{00} + \hat{\gamma}_{01}(\text{MEAN SES})_j + \hat{\gamma}_{02}(\text{SECTOR})_j\right], \qquad [4.28a]$$

$$u^*_{1j} = \beta^*_{1j} - \left[\hat{\gamma}_{10} + \hat{\gamma}_{11}(\text{MEAN SES})_j + \hat{\gamma}_{12}(\text{SECTOR})_j\right]. \qquad [4.28b]$$

Figure 4.3 displays the results. The intercept residuals are plotted on the vertical axis and the slope residuals on the horizontal axis. The OLS slope residuals are highly misleading. They suggest considerable unexplained variability in the SES-achievement relationships. In contrast, the empirical Bayes residuals are tightly clumped around zero with even less dispersion than in Figure 4.2. This result is consistent with the results in Table 4.5, where 78% of the variability in β_{1j} was accounted for by MEAN SES and SECTOR.

In contrast, the empirical Bayes and OLS residuals for the intercept are much more similar. These residuals, however, are less dispersed than for the unconditional model (Figure 4.2), which is consistent with the fact that Equation 4.21a accounts for about 73% of the variance in β_{0j}.

Comparison of Interval Estimates

In addition to point estimates for the level-1 coefficients, we can also compute empirical Bayes interval estimates using Equation 3.65. We illustrate here the use of these procedures for two selected schools, cases 22 and 135 from the HS&B data, and compare results to confidence interval estimates from separate OLS regressions on each school's individual data set. The 95% confidence intervals for β_{0j} and β_{1j} under ordinary least squares, unconditional and conditional shrinkage appear in Table 4.7.

Notice that for case 22, where $n_j = 67$, the widths of the confidence intervals for school mean achievement, β_{0j}, are quite similar across all three analyses. This results from the large within-school sample size coupled with the overall high reliability for school means. In contrast, case 135, where $n_j = 14$, experiences some reduction in the width of the confidence intervals as we move from OLS to unconditional shrinkage to conditional shrinkage. The 95% confidence interval under the conditional model of Equation 4.21 is about a third smaller than obtained from ordinary least squares estimates

```
(a) Ordinary Least Squares Residuals

                 -
                 -
             6.0+                              **
                 -                     *
    OLS          -                        *        *  *  *
 Intercept       -                 *      3 2 3*3***
                 -                   **   *22263433233**    *
             0.0+                  ** 2*23323232262**     2 *
                 -                   * * 2*4*4***232* 4     *
                 -                     *  2323* *22* 2
                 -                        *  **
                 -                           * *   **
            -6.0+
                 -                      *
                 -
                 -
                   ---------+---------+---------+---------+---------+----
                          -6.0      -3.0       0.0       3.0       6.0

                                            OLS Slope
```

```
(b) Empirical Bayes Residuals(a)

                 -
                 -
             6.0+
                 -                              2
    EB           -                             2*
 Intercept       -                             6**
                 -                            2++
             0.0+                             9+72
                 -                            ++4
                 -                            54
                 -                            24
                 -
            -6.0+
                 -
                   ---------+---------+---------+---------+---------+----
                          -6.0      -3.0       0.0       3.0       6.0

                                            EB Slope
```

Figure 4.3. Ordinary Least Squares and Empirical Bayes Intercept and Slope Residuals for 160 High Schools

NOTE: Intercept residuals are plotted on the vertical axis, slope residuals on the horizontal axis.
a. A "+" indicates the presence of more than nine observations.

using only this school's data. A gain in precision has been effected by bringing to bear on this estimation problem all of the information in the data set.

These improvements in precision are even more substantial when we compare the confidence interval estimates for the SES-achievement slopes. For case 22, the empirical Bayes 95% confidence interval, based on the conditional model, is about 75% smaller than OLS; for case 135, it is about 85% smaller. Notice that in both cases the likelihood of a negative

TABLE 4.7 Comparison of 95% Confidence Interval Estimates for Random Level-1 Coefficients

	OLS		*Unconditional Shrinkage*		*Conditional Shrinkage*	
Case	β_{0j}	β_{1j}	β_{0j}	β_{1j}	β_{0j}	β_{1j}
22	(9.69, 12.59)	(−3.01, 1.45)	(9.81, 12.63)	(−0.15, 2.43)	(9.64, 12.24)	(0.24, 1.10)
135	(1.37, 7.73)	(−4.11, 4.61)	(3.65, 9.21)	(0.41, 3.45)	(6.43, 10.77)	(2.19, 3.11)

SES-achievement slope becomes trivial under the conditional model; in contrast, under OLS, negative results appear quite plausible.

Cautionary Note

The conditional shrinkage estimators will be substantially more accurate than the OLS estimators *when the level-2 model is appropriately specified.* That is, the underlying assumption of empirical Bayes conditional shrinkage is that, given the predictors in the level-2 model, the regression lines are "conditionally exchangeable." This means, in the case of Equation 4.21, that once MEAN SES and SECTOR have been taken into account, there is no reason to believe that the deviation of any school's regression line from its predicted value is larger or smaller than that of any other school. This assumption depends strongly on the validity of the level-2 model. If that model is misspecified, the empirical Bayes estimates will also be misspecified: The estimates of the γ parameters will be biased and the empirical Bayes shrinkage will lead to distortion in each group's estimated equation. We shall return to this concern in Chapter 5 when we consider the problem of estimating the effectiveness of individual organizations.

Summary of Terms Introduced in This Chapter

Plausible value range for $\boldsymbol{\beta}_q$: We can compute a range of plausible values for any random level-1 coefficient, β_q. The 95% plausible value range for β_q is $\gamma_{q0} \pm 1.96(\tau_{qq})^{1/2}$. Based on the assumption that the random effects at level 2 are normally distributed, we would expect to find values of β_{qj} within range for 95% of the level-2 units.

Proportion of variance explained at level 1: An index of the proportion reduction in variance or "variance explained" at level 1 as X predictors are entered into the level-1 model. This is computed by comparing the residual σ^2 estimate from a fitted model with the σ^2 estimate from some "base" or reference model. The reference model chosen for computing these statistics at level 1 is often the one-way random-effects ANOVA model. (See Equation 4.20.)

Proportion of variance explained at level 2 in each $\boldsymbol{\beta_q}$: An index of the proportion reduction in variance or "variance explained" in each random level-1 coefficient (intercepts and slopes) as W predictors are added to the level-2 model for any β_{qj}. This is computed by comparing the residual τ_{qq} estimate from a fitted model with the τ_{qq} estimate from some "base" or reference model. The reference model chosen for computing these statistics at level 2 is often the random-coefficient regression model. (See Equation 4.24.)

Note

1. Data files and a user's manual are available from the Office of Educational Research and Improvement, U.S. Department of Education, 55 New Jersey Avenue, Washington, DC 20208-1327.