Estimating Multilevel Models using SPSS, Stata, and SAS

Jeremy J. Albright

May 11, 2007

Multilevel data are pervasive in the social sciences. Students may be nested within schools, voters within districts, or workers within firms, to name a few examples. Statistical methods that explicitly take into account hierarchically structured data have gained popularity in recent years, and there now exist several special-purpose statistical programs designed specifically for estimating multilevel models (e.g. HLM, MLwiN). In addition, the increasing use of of multilevel models — also known as hierarchical linear and mixed effects models — has led general purpose packages such as SPSS, Stata, and SAS to introduce their own procedures for handling nested data.

Nonetheless, researchers may face two challenges when attempting to determine the appropriate syntax for estimating multilevel/mixed models using general purpose software. First, many users from the social sciences come to multilevel modeling with a background in regression models, whereas much of the software documentation utilizes examples from experimental disciplines [due to the fact that multilevel modeling methodology evolved out of ANOVA methods for analyzing experiments with random effects (Searle, Casella, and McCulloch, 1992)]. Second, notation for multilevel models is often inconsistent across disciplines (Ferron 1997).

The purpose of this document is to demonstrate how to estimate multilevel models using SPSS, Stata, and SAS. It first seeks to clarify the vocabulary of multilevel models by defining what is meant by fixed effects, random effects, and variance components. It then compares the model building notation frequently employed in applications from the social sciences with the more general matrix notation found in much of the software documentation. The syntax for centering variables and estimating multilevel models is then presented for each package.

1 Vocabulary of Mixed and Multilevel Models

Models for multilevel data have developed out of methods for analyzing experiments with random effects. Thus it is important for those interested in using hierarchical linear models to have a minimal understanding of the language experimental researchers use to differentiate between effects considered to be random or fixed. In an ideal experiment, the researcher is interested in whether the presence or absence of one factor affects scores on an outcome variable.[1] Does a particular pill reduce cholesterol more than a placebo? Can behavioral modification reduce a particular phobia better than psychoanalysis or no treatment? The factors in these experiments are said to be fixed "because the same, fixed levels would be included in replications of the study" (Maxwell and Delaney, pg. 469). That is, the researcher is only interested in the exact categories of the factor that appear in the experiment. The typical model for a one-factor experiment is:

$$y_{ij} = \mu + \alpha_j + e_{ij} \tag{1}$$

where the score on the dependent variable for individual i is equal to the grand mean of the sample (μ), the effect α of receiving treatment j, and an individual error term e_{ij}. In general, some kind of constraint is put on the alpha values, such as that they sum to zero, so that the model is identified. In addition, it is assumed that the errors are independent and normally distributed with constant variance.

In some experiments, however, a particular factor may not be fixed and perfectly replicable across experiments. Instead, the distinct categories present in the experiment represent a random sample from a larger population. For example, different nurses may administer

[1] In the parlance of experiments, a *factor* is a categorical variable. The term *covariate* refers to continuous independent variables.

an experimental drug to subjects. Usually the effect of a specific nurse is not of theoretical interest, but the researcher will want to control for the possibility that an independent caregiver effect is present beyond the fixed drug effect being investigated. In such cases the researcher may add a term to control for the random effect:

$$y_{ij} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + e_{ij} \tag{2}$$

where β represents the effect of the kth level of the random effect, and $\alpha\beta$ represents the interaction between the random and fixed effects. A model that contains only fixed effects and no random effects, such as equation 1, is known as a *fixed effects model.* One that includes only random effects and no fixed effects is termed a *random effects model.* Equation 2 is actually an example of a *mixed effects model* because it contains both random and fixed effects.

While the notation in equation 2 for the random effect is the same as for the fixed effect (that is, both are denoted by subscripted Greek letters), an important difference exists in the tests for the drug and nurse factors. For the fixed effect, the researcher is interested in only those levels included in the experiment, and the null hypothesis is that there are no differences in the means of each treatment group:

$$H_0: \quad \mu_1 = \mu_2 = ... = \mu_j$$
$$H_1: \quad \mu_j \neq \mu_{j'}$$

For the random effect in the drug example, the researcher is not interested in the particular nurses per se but instead wishes to generalize about the potential effects of drawing different nurses from the larger population. The null hypothesis for the random effect is therefore that

that its variance is equal to zero:

$$H_0: \ \ \sigma^2_\beta = 0$$
$$H_1: \ \ \sigma^2_\beta > 0$$

The estimated variance is known as a *variance component*, and estimation of these is an essential step in mixed effects models.

Oftentimes in experimental settings, the random effects are nuisances that must be controlled for. In the above example, the effect of the drug was the primary interest, whereas the nurse factor was potentially confounding but theoretically uninteresting. It is nonetheless necessary to include the relevant random effects in the model or otherwise run the risk of making false inferences about the fixed effect (and any fixed/random effect interaction). In other applications, particularly for the types of multi-level models discussed below, the random effects are indeed of substantive interest. A researcher comparing test scores of students across schools may be interested in a school effect, even if it is only possible to sample a limited number of districts.

The reason to review random effects in the context of experiments is that methods for handling multilevel data are actually special cases of mixed effects models. Hox and Kreft (1994) make the connection clearly:

> "An effect in ANOVA is said to be fixed when inferences are to be made only about the treatments actually included. An effect is random when the treatment groups are sampled from a population of treatment groups and inferences are to be made to the population of which these treatments are a sample. Random effects need random effects ANOVA models (Hays 1973). Multilevel models assume a hierarchically structured population, with random sampling of both groups and individuals within groups. Consequently, multilevel analysis models must incorporate random effects" (pgs. 285-286).

For scholars coming from non-experimental disciplines (i.e. those that rely more heavily on regression models than analysis of variance), it may be difficult to make sense of the documentation for statistical applications capable of estimating mixed models. Political scientists or sociologists, for example, come to utilize mixed models because they recognize that hierarchically structured data violate standard linear regression assumptions. However, because mixed models developed out of methods for evaluating experiments, much of the documentation for packages like SPSS, SAS, and Stata is made up of experimental examples. Hence it is important to recognize the connection between random effects ANOVA and hierarchical linear models.

Note that the motivation for utilizing mixed models for multilevel data does not rest in the different number of observations at each level, as any model including a dummy variable involves nesting (e.g. survey respondents are nested within gender). The justification instead lies in the fact that the errors within each randomly sampled level-2 unit are likely correlated, necessitating the estimation of a random effects model. Once the researcher has accounted for error non-independence it is possible to make more accurate inferences about the fixed effects of interest.

2 Notation for Mixed and Multilevel Models

Even if one is comfortable distinguishing between fixed and random effects, additional confusion may emerge when trying to make sense of the notation used to describe multilevel models. In non-experimental disciplines, researchers tend to use the notation of Raudenbush and Bryk (2002) that explicitly models the nested structure of the data. Unfortunately his approach can be rather messy, and software documentation typically relies instead on matrix notation. Both approaches are detailed in this section.

In the archetypical cross-sectional example, a researcher is interested in predicting test performance as a function of student-level and school-level characteristics. Using the model-building notation, an empty (i.e. lacking predictors) student-level model is specified first:

$$Y_{ij} = \beta_{0j} + r_{ij} \tag{3}$$

The outcome variable Y for individual i nested in school j is equal to the average outcome in unit j plus an individual-level error r_{ij}. Because there may also be an effect that is common to all students within the same school, it is necessary to add a school-level error term. This is done by specifying a separate equation for the intercept:

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{4}$$

where γ_{00} is the average outcome for the population and u_{0j} is a school-specific effect. Combining equations 3 and 4 yields:

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \tag{5}$$

Denoting the variance of r_{ij} as σ^2 and the variance of u_{0j} as τ_{oo}, the percentage of observed variation in the dependent variable attributable to school-level characteristics is found by dividing τ_{00} by the total variance:

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \tag{6}$$

Here ρ is referred to as the *intraclass correlation coefficient.* The percentage of variance attributable to student-level traits is easily found according to $1 - \rho$.

A researcher who has found a significant variance component for τ_{00} may wish to incorporate macro level variables in an attempt to account for some of this variation. For example, the average socioeconomic status of students in a district may impact the expected test performance of a school, or average test performance may differ between private and public institutions. These possibilities can be modeled by adding the school-level variables to the intercept equation,

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{MEANSES}_j) + \gamma_{02}(\text{SECTOR}_j) + u_{0j} \tag{7}$$

and substituting 7 into equation 3.

Additionally, the researcher may wish to include student-level covariates. A student's personal socioeconomic status may affect his or her test performance independent of the school's average SES score. Thus equation 3 would become:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{SES}_{ij}) + r_{ij} \tag{8}$$

If the researcher wishes to treat student SES as a random effect (that is, the researcher feels the effect of a student's SES status varies between schools), he can do so by specifying an equation for the slope in the same manner as was previously done with the intercept equation:

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{9}$$

Finally, it is possible that the effect of a level-1 variable changes across scores on a level-2 variable. The effect of a student's SES status may be less important in a private rather than a public school, or a student's individual SES status may be more important in schools with higher average SES scores. To test these possibilities, one can add the MEANSES and SECTOR variables to equation 9.

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{MEANSES}_j) + \gamma_{12}(\text{SECTOR}_j) + u_{1j} \tag{10}$$

A random-intercept and random-slope model including level-2 covariates and cross-level interactions is obtained by substituting equations 7 and 10 into 8:

$$\begin{aligned} Y_{ij} &= \gamma_{00} + \gamma_{01}(\text{MEANSES}_j) + \gamma_{02}(\text{SECTOR}_j) + \gamma_{10}(\text{SES}_{ij}) \\ &+ \gamma_{11}(\text{MEANSES}_j * \text{SES}_{ij}) + \gamma_{12}(\text{SECTOR}_j * \text{SES}_{ij}) \\ &+ u_{0j} + u_{1j}(\text{SES}_{ij}) + r_{ij} \end{aligned} \tag{11}$$

This approach of building a multilevel model through the specification and combination of different level-1 and level-2 models makes clear the nested structure of the data. However, it is long and messy, and what is more, it is inconsistent with the notation used in much of the documentation for general statistical packages. Instead of the step-by-step approach taken above, the pithier, and more general, matrix notation is often used:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\epsilon} \tag{12}$$

Here $\boldsymbol{y}$ is an n x 1 vector of responses, $\mathbf{X}$ is an n x p matrix containing the fixed effects regressors, $\boldsymbol{\beta}$ is a p x 1 vector of fixed-effects parameters, $\mathbf{Z}$ is an n x q matrix of random effects regressors, $\mathbf{u}$ is a q x 1 vector of random effects, and $\boldsymbol{\epsilon}$ is an n x 1 vector of errors. The relationship between equations 12 and 11 is clearest when, in the step-by-step approach, the fixed effects are grouped together in the first part of the right-hand-side of the equation and the random effects are grouped together in the second part.

$$\begin{aligned} Y_{ij} &= \underbrace{\gamma_{00} + \gamma_{01}(\text{MEANSES}_j) + \gamma_{02}(\text{SECTOR}_j) + \gamma_{10}(\text{SES}_{ij})}_{\text{Fixed Effects}} \\ &+ \underbrace{\gamma_{11}(\text{MEANSES}_j * \text{SES}_{ij}) + \gamma_{12}(\text{SECTOR}_j * \text{SES}_{ij})}_{\text{Fixed Effects}} \\ &+ \underbrace{u_{0j} + u_{1j}(\text{SES}_{ij})}_{\text{Random Effects}} + r_{ij} \end{aligned} \tag{13}$$

Note that it is possible for a variable to appear as both a fixed effect and a random effect (appearing in both $\mathbf{X}$ and $\mathbf{Z}$ from 12). In this example, estimating 13 would yield both fixed effect and random effect estimates for the student-level SES variable. The fixed effect would refer to the overall expected effect of a student's socioeconomic status on test scores; the random effect gives information on whether or not this effect differs between schools.

3 Estimation

An essential step to estimating multilevel models is the estimation of variance components. Up until the 1970s, the literature on variance component estimation focused on using ANOVA techniques that derived from the work of Fisher [adapted to unbalanced data by Henderson (1953)]. Since the 1970s, Full and Restricted Maximum Likelihood estimation (FML and REML, respectively) have become the preferred methods. ML approaches have several advantages, including the ability to handle unbalanced data without some of the pathologies of ANOVA methods (i.e. lack of uniqueness, negative variance estimates). Both FML and REML produce identical fixed effects estimates. The latter, however, takes into account the degrees of freedom from the fixed effects and thus produces variance components estimates that are less biased. One downside to REML is that the likelihood ratio test cannot be used to compare two models with different fixed effects specifications. In small samples with balanced data, REML is generally preferable to ML because it is unbiased. In large samples, however, differences between estimates are neglible (Snijders and Bosker 1999). Thus, in most applications, "the question of which method to use remains a matter of personal taste" (StataCorp 2005, pg. 188).

The remainder of this document provides syntax for estimating multilevel models using SPSS, Stata, and SAS. The data analyzed will be the High School and Beyond (HSB) dataset that accompanies the HLM package (Raudenbush et al. 2005). Each section will show how to estimate the empty model, a random intercept model, and a random slope model from the student performance example outlined above. The dependent variable is scores on a math achievement scale. Note that whereas HLM requires two separate data files (one corresponding to each level), SPSS, Stata, and SAS rely on only a single file. The level-2 observations are common to each case within the same macro-unit, so that if there are 50 students in one school the corresponding school-level score appears 50 times.. Each program also requires an id variable identifying the group membership of each individual. The results presented below are based on REML estimation, the default in each package.

4 SPSS

This section closely follows Peugh and Enders (2005). It demonstrates how to group-mean center level-1 covariates and estimate multilevel models using SPSS syntax. Note that it is also possible to use the **Mixed Models** option under the **Analyze** pull-down menu (see Norusis 2005, pgs. 197-246). However, length considerations limit the examples here to syntax. The SPSS syntax editor can be accessed by going to **File** → **New** → **Syntax**.

In the HSB data file, the student-level SES variable is in its original metric (a standardized scale with a mean of zero). Oftentimes researchers dealing with hierarchically structured data wish to center a level-1 variable around the mean of all cases within the same level-2 group in order to facilitate interpretation of the intercept. To group-mean center a variable in SPSS, first use the `AGGREGATE` command to estimate mean SES scores by school. In this example, the syntax would be:

```
AGGREGATE OUTFILE=sesmeans.sav
/BREAK=id
/meanses=MEAN(ses) .
```

The `OUTFILE` statement specifies that the means are written out to the file sesmeans.sav in the working directory. The `BREAK` subcommand specifies the groups within which to estimate means. The final line names the variable containing the school means `meanses`.

Next, the group means are sorted and merged with the original data using the `SORT CASES` and `MATCHFILES` commands. The centered variables are then created using the `COMPUTE` command.[2] The syntax for these steps would be:

```
SORT CASES BY id .
MATCH FILES
/TABLE=sesmeans.sav
/FILE=*
/BY id .
COMPUTE centses = ses - meanses .
EXECUTE .
```

The subcommands for `MATCH FILES` ask SPSS to take the data file saved using the

[2]To grand mean center a variable in SPSS requires only a single line of syntax. For example, `COMPUTE newvar = oldvar - mean(oldvar).`

`AGGREGATE` command and merge it with the working data (denoted by *). The matching variable is the school ID.

With the data prepared, the next step is to estimate the models of interest. The following syntax corresponds to the empty model (5):

```
MIXED mathach
/PRINT = SOLUTION TESTCOV
/FIXED = INTERCEPT
/RANDOM = INTERCEPT | SUBJECT(id) .
```

The command for estimating multilevel models is `MIXED`, followed immediately by the dependent variable. `PRINT = SOLUTION` requests that SPSS reports the fixed effects estimates and standard errors. `FIXED` and `RANDOM` specify which variables to treat as fixed and random effects, respectively. The `SUBJECT` option following the vertical line | identifies the grouping variable, in this case school ID.

The fixed and random effect estimates for this and subsequent models are displayed in Table 1. The intercept in the empty model is equal to the overall average math achievement score, which for this sample is 12.637. The variance component corresponding to the random intercept is 8.614; for the level-1 error it is 39.1483. Including the `TESTCOV` subcommand requested that SPSS report Wald-Z significance tests for the variance components, equal to the estimate divided by its standard error. In this example, the value of the Wald-Z statistic is 6.254, which is significant ($p<.001$). Note, however, that these tests should not be taken as conclusive. Singer (1998, pg. 351) writes,

> "the validity of these tests has been called into question both because they rely on large sample approximations (not useful with the small sample sizes often analyzed using multilevel models) and because variance components are known to have skewed (and bounded) sampling distributions that render normal approximations such as these questionable."

A more thorough test would thus estimate a second model constraining the variance component to equal zero and compare the two models using a likelihood ratio test.

The two variance components can be used to partition the variance across levels according to equation 6 above. The intraclass correlation coefficient for this example is equal to $\frac{8.614}{8.614+39.1483} = .1804$, meaning that roughly 18% of the variance is attributable to school traits. Because the intraclass correlation coefficient shows a fair amount of variation across schools, model 2 adds two school-level variables. These variables are `sector`, defining whether a school is private or public, and `meanses`, which is the average student socioeconomic status in the school. The SPSS syntax to estimate this model is:

```
MIXED mathach WITH meanses sector
/PRINT = SOLUTION TESTCOV
/FIXED = INTERCEPT meanses sector
/RANDOM = INTERCEPT | SUBJECT(id) .
```

The results, displayed in the second column of Table 1, show that `meanses` and `sector` significantly affect a school's average math achievement score. The intercept, representing the expected math achievement score for a student in a public school with average SES, is equal to 12.1283. A one unit increase in average SES raises the expected school mean by 5.5334. Private schools have expected math achievement scores 1.2254 units higher than public schools. The variance component corresponding to the random intercept has decreased to 2.3140, demonstrating that the inclusion of the two school-level variables has explained much of the level-2 variation. However, the estimate is still more than twice the size of its standard error, suggesting that there remains a significant amount of unexplained school-level variance (though the same caution about over-interpreting this test still applies).

A final model introduces a student-level covariate, the group-mean centered SES variable (`centses`). Because it is possible that the effect of socioeconomic status may vary across schools, SES is treated as a random effect. In addition, `sector` and `meanses` are included to model the slope on the student-level SES variable. Modeling the slope of a random effect is the same as specifying a cross-level interaction, which can be specified in the `FIXED` subcommand as in the following syntax:

```
MIXED mathach WITH meanses sector centses
/PRINT = SOLUTION TESTCOV
```

```
/FIXED = INTERCEPT meanses sector centses meanses*centses sector*centses
/RANDOM = INTERCEPT centses | SUBJECT(id) COVTYPE(UN) .
```

One important change over the previous models is the addition of the `COV(UN)` option, which specifies a structure for the level-2 covariance matrix. Only a single school-level variance component was estimated in the previous two models, thus it was unnecessary to deal with covariances. When there is more than one level-2 variance component, SPSS will assume a particular covariance structure. In many cross-sectional applications of multilevel models, the researcher does not wish to put any constraints on this covariance matrix. Thus the `UN` in the `COV` option specifies an *unstructured* matrix. In other contexts, the researcher may wish to specify a first-order autoregressive (AR1), compound symmetry (CS), identity (ID), or other structure. These alternatives are more restrictive but may sometimes be appropriate.

The results from this final model appear in the last column of Table 1. The fixed effects are all significant. Given the inclusion of the group-mean centered SES variable, the intercept is interpreted as the expected math achievement in a public school with average SES levels for a student at his or her school's average SES. In this model, the expected outcome is 12.1279. Because there are interactions in the model, the marginal fixed effects of each variable will depend on the value of the other variable(s) involved in the interaction. The marginal effect of a one-unit change in a student's SES score on math achievement depends on whether a school is public or private as well as on the school's average SES score. For a public school (where `sector=0`), the marginal effect of a one-unit change in the group-mean centered student SES variable is equal to $\frac{\partial Y}{\partial CENTSES} = \gamma_{10} + \gamma_{11}(MEANSES) = 2.945041 + 1.039232(MEANSES)$. For a private school (where `sector=1`), the marginal effect of a one-unit change in student SES is equal to $\frac{\partial Y}{\partial CENTSES} = \gamma_{10} + \gamma_{11}(MEANSES) + \gamma_{12} = 2.945041 + 1.039232(MEANSES) - 1.642674$. When cross-level interactions are present, graphical means may be appropriate for exploring the contingent nature of marginal effects in greater detail (Raudenbush & Bryk 2002; Brambor, Clark, and Golder 2006). Here the simplest interpretation is that the effect of student-level SES is significantly higher in wealthier schools and significantly lower in private schools.

The variance component for the random intercept continues to be significant, suggesting that there remains some variation in average school performance not accounted for by the variables in the model. The variance component for the random slope, however, is not significant. Thus the researcher may be justified in estimating an alternative model that constrains this variance component to equal zero.

[Table 1 about here.]

5 Stata

This section discusses how to center variables and estimate multilevel models using Stata. A fuller treatment is available in Rabe-Hesketh and Skrondal (2005) and in the Stata documentation. Since release 9, Stata includes the command `.xtmixed` to estimate multilevel models. The `.xt` prefix signifies that the command belongs to the larger class of commands used to estimate models for longitudinal data. This reflects the fact that panel data can be thought of as multilevel data in which observations at multiple time points are nested within an individual. However, the command is appropriate for mixed model estimation in general, including cross-sectional applications.

In the HSB data file, the student-level SES variable is in its original metric (a standardized scale with a mean of zero). Oftentimes researchers dealing with hierarchically structured data wish to center a level-1 variable around the mean of all cases within the same level-2 group. Group-mean centering can be accomplished by using one of two commands in Stata. The first is to use the `.collapse` command, which creates a new data file consisting of summary statistics for groups (analogous to the `AGGREGATE` command in SPSS). Following the `.collapse` command, the desired summary statistic is listed in parentheses followed by a list of variables for which the corresponding statistic(s) will be estimated. The resulting data file can then be sorted, saved, and merged with the original data. The `.generate` command completes the creation of a group-centered variable.

With the original data file (hsb.dta) as the working data, the full syntax to group-mean center the SES variable is:

```
.collapse (mean) meanses=ses, by(id)
.sort id
.save sesmeans, replace
.use hsb.dta, replace
.sort id
.merge id using sesmeans
.gen centses=ses-meanses
```

Here `meanses=ses` tells Stata to name the summary variable `meanses`. The `.merge` command combines two data files.

Alternatively, the `.statsby` command can also estimate and temporarily store summary statistics for groups. The syntax for this approach is:

```
.statsby meanses=r(mean), by(id) saving(sesmeans, replace):  summarize ses
.sort id
.merge id using sesmeans
```

The `.statsby` command stores temporary variables created by a particular command for the groups specified in the `by` option. In this example, Stata saves the group means as the variable `sesmeans` in a new data file. By default, the new data file replaces the working data. The `saving` option instead saves the data to a file, here named sesmeans, in the working directory. The `.sort` and `.merge` commands sort the working data and merge it with the newly created file.[3]

The syntax for estimating multilevel models in Stata begins with the `.xtmixed` command followed by the dependent variable and a list of independent variables. The last independent variable is followed by double vertical lines ||, after which the grouping variable and random effects are specified. `.xtmixed` will automatically specify the intercept to be random. A list of variables whose slopes are to be treated as random follows the colon. Note that, by default, Stata reports variance components as standard deviations (equal to the square root of the variance components). To get Stata to report variances instead, add the `var` option.

[3]To grand mean center in Stata requires two commands. First, `.quietly summarize oldvar`. Then, `.gen newvar=oldvar-r(mean)`.

The syntax for the empty model is the following:

```
.xtmixed mathach || id:  , var
```

The results are displayed in Table 2. The average test score across schools, reflected in the intercept term, is 12.63697. The variance component corresponding to the random intercept is 8.61403. Because this estimate is substantially larger than its standard error, there appears to be significant variation in school means.

The two variance components can be used to partition the variance across levels. The intraclass correlation coefficient is equal to $\frac{8.61403}{39.14832+8.61403} = 18.04$, meaning that roughly 18% of the variance is attributable to the school-level.

In order to explain some of the school-level variance in math achievement scores it is possible to incorporate school-level predictors into the model. For example, the socioeconomic status of the typical student, or the school's status as public or private, may influence test performance. The Stata syntax for adding these variables to the model is:

```
.xtmixed mathach meanses sector || id:  , var
```

The intercept, which now corresponds to the expected math achievement score in a public school with average SES scores, is 12.12824. Moving to a private school bumps the expected score by 1.2254 points. In addition, a one-unit increase in the average SES score is associated with an expected increased in math achievement of 5.3328. These estimates are all significant.

The variance component corresponding to the random intercept has decreased to 2.313986, reflecting the fact that the inclusion of the level-2 variables has accounted for some of the variance in the dependent variable. Nonetheless, the estimate is still more than twice the size of its standard error, suggesting that there remains variance unaccounted for.

A final model introduces the student socioeconomic status variable. Because it is possible that the effect of individual SES status varies across schools, this slope is treated as random. In addition, a school's average SES score and its sector (public or private) may interact with student-level SES, accounting for some of the variance in the slope. In order to include these cross-level interactions in the model, however, it is necessary to first explicitly create the interaction variables in Stata:

```
.gen ses_mses=meanses*centses
.gen ses_sect=sector*centses
```

When estimating more than one random effect, the researcher must also be concerned with the covariances among the level-2 variance components. As with SPSS, in Stata it is necessary to add an option specifying that the covariance matrix for the random effects is unstructured (the default is to assume all covariances are zero). The syntax for estimating the random-slope model is thus:

```
.xtmixed mathach meanses sector centses ses_mses ses_sect || id: centses,
var cov(un)
```

The results are displayed in the final column of Table 2. The intercept is 12.12793, which here is the expected math achievement score in a public school with average SES scores for a student at his or her school's average SES level. Because there are interactions in the model, the marginal fixed effects of each variable now depend on the value of the other variable(s) involved in the interaction. The marginal effect of a one-unit change in student's SES on math achievement will depend on whether a school is public or private as well as on the average SES score for the school. For a public school (where `sector`=0), the marginal effect of a one-unit change in the group-mean centered SES variable is equal to $\frac{\partial Y}{\partial CENTSES} = \gamma_{10} + \gamma_{11}(MEANSES) = 2.94504 + 1.039237(MEANSES)$. For a private school (where `sector`=1), the marginal effect of a one-unit change in student SES is equal to $\frac{\partial Y}{\partial CENTSES} = \gamma_{10} + \gamma_{11}(MEANSES) + \gamma_{12} = 2.94504 + 1.039237(MEANSES) - 1.642675$. When cross-level interactions are present, graphical means may be appropriate for exploring the contingent nature of marginal effects in greater detail. Here the simplest interpretation of the interaction coefficients is that the effect of student-level SES is significantly higher in wealthier schools and significantly lower in private schools.

The variance component for the random intercept is 2.379597, which is still large relative to its standard error of 0.3714584. Thus there remains some school-level variance unaccounted for in the model. The variance component corresponding to the slope, however, is quite small relative to its standard error. This suggests that the researcher may be justified in constraining the effect to be fixed.

By default, Stata does not report model fit statistics such as the AIC or BIC. These can be requested, however, by using the postestimation command `.estat ic`. This displays the log-likelihood, which can be converted to Deviance according to the formula $-2*$ *log likelihood*. It also displays the AIC and BIC statistics in smaller-is-better form. Comparing both the AIC and BIC statistics in Table 2 it is clear that the final model is preferable to the first two models.

[Table 2 about here.]

6 SAS

This section follows Singer (1998); a thorough treatment is available from Littell et al. (2006). The SAS procedure for estimating multilevel models is PROC MIXED.

In the HSB data file the student-level SES variable is in its original metric (a standardized scale with a mean of zero). Oftentimes, the researcher will prefer to center a variable around the mean of all observations within the same group. Group-mean centering in SAS is accomplished using the SQL procedure. The following commands create a new data file, HSB2, in the Work library that includes two additional variables: the group means for the SES variable (saved as the variable `sesmeans`) and the group-mean centered SES variable `cses`.[4]

```
PROC SQL;
CREATE TABLE hsb2 AS
SELECT *, mean(ses) as meanses,
ses-mean(ses) AS cses
FROM hsb
GROUP BY id;
QUIT;
```

[4] Grand-mean centering also uses `PROC SQL`. Excluding the `GROUP BY` statement causes the `mean(ses)` function to estimate the grand mean for the `ses` variable. The `ses-mean(ses)` statement then creates the grand-mean centered variable.

The syntax for estimating the empty model is the following:

```
PROC MIXED COVTEST DATA=hsb2;
CLASS id;
MODEL mathach = /SOLUTION;
RANDOM intercept/SUBJECT=id
RUN;
```

The `COVTEST` option requests hypothesis tests for the random effects. The `CLASS` statement identifies `id` as a categorical variable. The `MODEL` statement defines the model, which in this case does not include any predictor variables, and the `SOLUTION` option asks SAS to print the fixed effects estimates in the output. The next statement, `RANDOM`, identifies the elements of the model to be specified as random effects. The `SUBJECT=id` option identifies `id` to be the grouping variable.

The results are displayed in Table 3. The average math achievement score across all schools is 12.6370. The variance component corresponding to the random intercept is 8.6097, which has a corresponding standard error of 1.0778. Because this estimate is more than twice the size of its standard error, there is evidence of significant variation in average test scores across schools (though see the SPSS section for a caution on over-interpreting this test).

It is possible to partition the variance in the dependent variable across levels according to the ratio of the school-level variance component to the total variance. In this example, the ratio is $\frac{8.6097}{8.6097+39.1487} = .1802761$, meaning that roughly 18% of the variance is attributable to school characteristics.

In order to explain some of the school-level variation in math achievement scores it is possible to incorporate school-level predictors into the model. For example, the average socioeconomic status of a school's students may affect performance. In addition, whether a school is public or private may also make a difference. The SAS program for a model with two school level predictors is the following:

```
PROC MIXED COVTEST DATA=hsb2;
CLASS id;
MODEL mathach = meanses sector /SOLUTION;
RANDOM intercept/SUBJECT=id;
```

```
RUN;
```

The `MODEL` statement now includes the two school-level predictors following the equals sign. Nothing else is changed from the previous program.

The results are displayed in the second column of Table 3. The intercept is 12.1282, which now corresponds to the expected math achievement score for a student in a public school at that school's average SES level. A one-unit increase in the school's average SES score is associated with a 5.3328-unit increase in expected math achievement, and moving from a public to a private school is associated with an expected improvement of 1.2254. These estimates are all significant.

The variance component corresponding to the random intercept has now dropped to 2.3139, demonstrating that the inclusion of the average SES and school sector variables explains a good deal of the school-level variance. Still, the estimate remains more than twice the size of its standard error of 0.3700, suggesting that some of the school-level variance remains unexplained.

A final model adds a student-level covariate, the group-mean centered SES variable. Because it is possible that the effect of a student's SES may vary across schools, the final model treats the slope as random. Additionally, because the slope may vary according to school-level characteristics such as average SES and sector (private versus public), the final model also incorporates cross-level interactions.

The syntax for this last model is the following:

```
PROC MIXED COVTEST DATA=hsb2;
CLASS id;
MODEL mathach = meanses sector cses meanses*cses sector*cses/solution;
RANDOM intercept cses / TYPE=UN SUB=id;
RUN;
```

The `MODEL` statement adds the `cses` variable along with the cross-level interactions between `cses`at the student level and `sector` and `meanses`at the school level. `CSES` is also added to the `RANDOM` statement. The `TYPE=UN` option specifies an unstructured covariance matrix for the random effects.

The results are displayed in the final column of Table 3. The intercept of 12.1279 now refers to the expected math achievement score in a public school with average SES scores for a student at his or her school's average SES level. Because there are interactions in the model, the marginal fixed effects of each variable depend on the value of the other variable(s) involved in the interaction. The marginal effect of a one-unit change in a student's SES score on math achievement will depend on whether a school is public or private as well as on the school's average SES score. For a public school (where `sector`=0), the marginal effect of a one-unit change in the group-mean centered student SES variable is equal to $\frac{\partial Y}{\partial CENTSES} = \gamma_{10} + \gamma_{11}(MEANSES) = 2.9450 + 1.0392(MEANSES)$. For a private school (where `sector`=1), the marginal effect of a one-unit change in a student's SES is equal to $\frac{\partial Y}{\partial CENTSES} = \gamma_{10} + \gamma_{11}(MEANSES) + \gamma_{12} = 2.9450 + 1.0392(MEANSES) - 1.6427$. When cross-level interactions are present, graphical means may be appropriate for exploring the contingent nature of marginal effects in greater detail. Here the simplest interpretation of the interaction coefficients is that the effect of student-level SES is significantly higher in wealthier schools and significantly lower in private schools.

The variance component corresponding to the random intercept is 2.3794, which remains much larger than its standard error of .3714. Thus there is most likely additional school-level variation unaccounted for in the model. The variance component for the random slope is smaller than its standard error, however, suggesting that the model picks up most of the variance in this slope that exists across schools.

[Table 3 about here.]

References

[1] Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding interaction models: Improving emprical analysis. *Political Analysis*, 14, 63-82.

[2] Ferron, J. (1997). Moving between hierarchical modeling notations. *Journal of Educational and Behavioral Statistics*, 22, 119-12.

[3] Hayes, W. L. (1973). *Statistics for the Social Sciences*. New York: Holt, Rinehart, & Winston

[4] Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics*, 9, 226-252.

[5] Hox, J. J. (1994). Multilevel analysis methods. *Sociological Methods & Research*, 22, 283-299.

[6] Littell, R. C., Milliken, G. A., Stroup, W. A., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS for Mixed Models, Second Edition*. Cary, NC: SAS Institute.

[7] Maxwell, S. E. & Delaney, H. D. (2004). *Designing Experiments and Analyzing Data: A Model Comparison Perspective, Second Edition*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

[8] Norusis, M. J. (2005). *SPSS 14.0 Advanced Statistical Procedures Companion*. Upper Saddle, NJ: Prentice Hall.

[9] Peugh, J. L. & Enders, C. K. (2005). Using the SPSS mixed procedure to fit cross-sectional and longitudinal multilevel models. *Educational and Psychological Measurement*, 65, 714-741.

[10] Raudenbush, S.W. & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition*. Newbury Park, CA: Sage.

[11] Raudenbush, S., Bryk, A., Cheong, Y. F., & Congdon, R. (2005). *HLM 6: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International.

[12] Rabe-Hesketh, S. & Skrondal A. (2005). *Multilevel and Longitudinal Modeling using Stata*. College Station, TX: StataCorp LP.

[13] Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance Components*. New York: Wiley.

[14] Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 24, 323-355.

[15] Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage.

[16] StataCorp. (2005). *Stata Longitudinal/Panel Data Reference Manual, Release 9*. College Station, TX: StataCorp LP.

Table 1: Results from SPSS

Fixed Effects	Model 1	Model 2	Model 3
Intercept γ_{00}	12.636974*	12.128236*	12.127931*
	(0.244394)	(0.199193)	(0.199290)
MEANSES γ_{01}		5.332838*	5.3329*
		(0.368623)	(0.369164)
SECTOR γ_{02}		1.225400*	1.226579*
		(0.3058)	(0.306269)
CENTSES γ_{10}			2.945041*
			(0.155599)
MEANSES*CENTSES γ_{11}			1.039232*
			(0.298894)
SECTOR*CENTSES γ_{12}			−1.642674*
			(0.239778)
Random Effects	**Model 1**	**Model 2**	**Model 3**
Intercept τ_{00}	8.614025*	2.313989*	2.379579*
	(1.078804)	(0.370011)	(0.371456)
CENTSES τ_{11}			0.101216
			(0.213792)
Residual σ^2	39.148322*	39.161358*	36.721155*
	(0.660645)	(0.660933)	(0.626134)
Model Fit Statistics	**Model 1**	**Model 2**	**Model 3**
Deviance	47116.8	46946.5	46714.24
AIC	47122.79	46956.5	46726.23
BIC	47143.43	46990.9	46767.51

Table 2: Results from Stata

Fixed Effects	Model 1	Model 2	Model 3
Intercept γ_{00}	12.63697^*	12.12824^*	12.12793^*
	(0.2443937)	(0.1992)	(0.1992901)
MEANSES γ_{01}		5.332838^*	5.332876^*
		(0.3686225)	(0.3691648)
SECTOR γ_{02}		1.2254^*	1.226578^*
		(0.3058)	(0.3062703)
CENTSES γ_{10}			2.94504^*
			(0.1555962)
MEANSES*CENTSES γ_{11}			1.039237^*
			(0.2988895)
SECTOR*CENTSES γ_{12}			-1.642675^*
			(0.2397734)
Random Effects	**Model 1**	**Model 2**	**Model 3**
Intercept τ_{00}	8.614034^*	2.313986^*	2.379597^*
	(1.078805)	(0.3700)	(0.3714584)
CENTSES τ_{11}			0.1012
			(0.2138)
Residual σ^2	39.14832	39.16136	36.7212
	(0.6606446)	(0.6609331)	(0.660)
Model Fit Statistics	**Model 1**	**Model 2**	**Model 3**
Deviance	47116.8	46946.5	46503.7
AIC	47120.8	46950.5	46511.7
BIC	47126.9	46956.6	46524.0

Table 3: Results from SAS

Fixed Effects	Model 1	Model 2	Model 3
Intercept γ_{00}	12.6370*	12.1282*	12.1279*
	(0.2443)	(0.1992)	(0.1993)
MEANSES γ_{01}		5.3328*	5.3329*
		(0.3686)	(0.3692)
SECTOR γ_{02}		1.2254*	1.2266*
		(0.3058)	(0.3063)
CENTSES γ_{10}			2.9450*
			(0.1556)
MEANSES*CENTSES γ_{11}			1.0392*
			(0.2989)
SECTOR*CENTSES γ_{12}			-1.6427*
			(0.2398)
Random Effects	**Model 1**	**Model 2**	**Model 3**
Intercept τ_{00}	8.6097*	2.3139*	2.3794*
	(1.0778)	(0.3700)	(0.3714)
CENTSES τ_{11}			0.1012
			(0.2138)
Residual σ^2	39.1487	39.1614	36.7212
	(0.6607)	(0.6609)	(0.6261)
Model Fit Statistics	**Model 1**	**Model 2**	**Model 3**
Deviance	47116.8	46946.5	46503.7
AIC	47120.8	46950.5	46511.7
BIC	47126.9	46956.6	46524.0