Journal of Educational and Behavioral Statistics
Winter 1998, Vol. 24, No. 4, pp. 323–355

Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models

Judith D. Singer
Harvard University

Keywords: *hierarchical linear model, longitudinal analysis, measurement of change, mixed model, multilevel model, school effects analysis*

SAS PROC MIXED is a flexible program suitable for fitting multilevel models, hierarchical linear models, and individual growth models. Its position as an integrated program within the SAS statistical package makes it an ideal choice for empirical researchers and applied statisticians seeking to do data reduction, management, and analysis within a single statistical package. Because the program was developed from the perspective of a "mixed" statistical model with both random and fixed effects, its syntax and programming logic may appear unfamiliar to users in education and the social and behavioral sciences who tend to express these models as multilevel or hierarchical models. The purpose of this paper is to help users familiar with fitting multilevel models using other statistical packages (e.g., HLM, MLwiN, MIXREG) add SAS PROC MIXED to their array of analytic options. The paper is written as a step-by-step tutorial that shows how to fit the two most common multilevel models: (a) school effects models, designed for data on individuals nested within naturally occurring hierarchies (e.g., students within classes); and (b) individual growth models, designed for exploring longitudinal data (on individuals) over time. The conclusion discusses how these ideas can be extended straighforwardly to the case of three level models. An appendix presents general strategies for working with multilevel data in SAS and for creating data sets at several levels.

As multilevel models, hierarchical models and individual growth models increase in popularity, the need for credible and flexible software that can be used to fit them to data increases. In their 1994 review of the five major software programs that were then currently available, Kreft, de Leeuw and van der Leeden (1994) found that only one (BMDP-5V) was integrated into a multipurpose statistical package. The remaining four required users to conduct preliminary data reduction and data processing in a different package before outputting data files to the specialized packages for analysis. Although the last few years have seen improvements in the front-ends of the two most popular packages—

Thanks are due to Russ Wolfinger of SAS Institute who read and commented upon a previous version of this paper and to three anonymous reviewers.

HLM (Bryk, Raudenbush, & Congdon, 1996) and MLwiN (Prosser, Rasbash, & Goldstein, 1996)—many users have sought the inclusion of routines for fitting multilevel models into the major statistical packages themselves.

In 1992, SAS Institute introduced one such routine—PROC MIXED—into their large menu of offerings. In subsequent releases, SAS has updated and expanded the models and options available as part of PROC MIXED to the point that it is now a reasonable choice for researchers fitting many types of multilevel models. Although the documentation for PROC MIXED is complex (SAS Institute, 1992, 1996), and the defaults are often not appropriate for many models (Latour, Latour, & Wolfinger, 1994; Littell, Milliken, Stroup, & Wolfinger, 1996), the abilty to do data reduction, management, and analysis in a single software package makes this routine particularly attractive to a wide range of researchers.

Because PROC MIXED was developed from a distinctly different perspective than that employed by most statisticians and empirical researchers in the educational, social, and behavioral sciences, its syntax and programming logic may appear unusual to people in these fields (Ferron, 1997). Unlike HLM and MLwiN, which were written with the kinds of models used by social scientists in mind, PROC MIXED was written by agricultural and physical scientists seeking a generalization of the standard linear model that allows for both *fixed* and *random* effects (McLean, Sanders, & Stroup, 1991). Although it is not immediately obvious based upon the documentation provided by SAS, it is indeed the case that by properly specifying the mixed model, a data analyst may fit a variety of specific instances of the multilevel models, hierarchical models, and individual growth models that have become so popular in educational and behavioral research (Kreft, 1995; Hox & Kreft, 1994).

The purpose of this paper is to show educational and behavioral statisticians and researchers how they can use PROC MIXED to fit many common types of multilevel models. Rather than try to cover a broad array of models (without providing sufficient depth for the reader to understand the logic behind the syntax), I focus on two of the most common models: (a) *school effects* models, designed for data on individuals nested within naturally occurring hierarchies (e.g., students within classes, children within families, teachers within schools); and (b) *individual growth* models, designed for exploring longitudinal data (on individuals) over time. In addition, because the use of PROC MIXED does not obviate the need for substantial data processing in preparation for analysis, in the appendix I present general strategies for working with multilevel data in SAS and for creating data sets at several levels.

Multilevel models can be expressed in at least three different ways: (a) by writing separate equations at multiple levels; (b) by writing separate equations at multiple levels and then substituting in to arrive at a single equation; and (c) by writing a single equation that specifies the multiple sources of variation. Bryk and Raudenbush (1992) specify the model for each level separately, and their software program (HLM) never requires you to substitute back to derive a single equation specification. Goldstein (1995) expresses the multilevel model directly using a single equation, and his software program, MLwiN, works from that single level representation. PROC MIXED also requires that you provide a single level representation. For pedagogic reasons, in this paper I take the middle ground, initially writing the model at multiple levels (kept here to two) and then substituting in to arrive at a single equation representation.

To use this paper effectively, a basic understanding of the ideas behind multilevel modeling, hierarchical modeling, and individual growth modeling is helpful. Both Bryk and Raudenbush (1992) and Hox (1995) provide excellent introductions to these topics. In particular, the reader must understand: (a) the difference between a *fixed* effect and a *random* effect; (b) the notion of multiple levels within a hierarchy; (c) the notion that the error variance-covariance matrix can take on different structures; and (d) that centering can be a helpful way of parameterizing models so that the results are more easily interpreted.

This article does not substitute for the comprehensive documentation available through SAS, including the general PROC MIXED documentation (SAS Institute, 1992, 1996), *Getting Started with PROC MIXED* (Latour, Latour, & Wolfinger, 1994), and *The SAS System for Mixed Models* (Littell et al., 1996). My goal is simply to provide a bridge to users already familiar with multilevel modeling because the SAS documentation is thin in this regard. I have found that PROC MIXED's flexibility has led many an unsuspecting user to write a program, obtain results, and have no idea what model has been fit. The goal for the user, then, is to specify the model and to learn the syntax necessary for ensuring that this is the model being fit to the data.

Two-Level School Effects Models

I begin by presenting an overview of strategies for using PROC MIXED to fit classic two-level *school effects* models. By two-level school effects models, I am referring to situations in which you have data at two levels within an organizational hierarchy—such as students within classes or classes within schools—and you would like to examine the behavior of a level-1 outcome as a function of both level-1 and level-2 predictors.

To achieve some continuity with presentations of these models available elsewhere, I use the High School and Beyond data example that Bryk and Raudenbush (1992) include in the 1996 version of HLM for Windows (Bryk et al., 1996). Readers unfamiliar with this example should consult Chapter Five of Bryk and Raudenbush (1992) for a fuller description. The data set consists of information for 7,185 students in 160 schools (with anywhere from 14 to 67 students per school). The student-level (level-1) outcome is MATHACH. The student level (level-1) covariate is SES. There are two school-level (level-2) covariates. One is an aggregate of student level characteristics (MEANSES); the other is a school-level variable (SECTOR). MEANSES and SES are centered at the grand mean (they have means of 0). SECTOR, a dummy variable, is coded 0 and 1.

HLM (Bryk, Raudenbush, & Congdon, 1996) and MLwiN (Prosser, Rasbash, & Goldstein, 1996)—many users have sought the inclusion of routines for fitting multilevel models into the major statistical packages themselves.

In 1992, SAS Institute introduced one such routine—PROC MIXED—into their large menu of offerings. In subsequent releases, SAS has updated and expanded the models and options available as part of PROC MIXED to the point that it is now a reasonable choice for researchers fitting many types of multilevel models. Although the documentation for PROC MIXED is complex (SAS Institute, 1992, 1996), and the defaults are often not appropriate for many models (Latour, Latour, & Wolfinger, 1994; Littell, Milliken, Stroup, & Wolfinger, 1996), the abilty to do data reduction, management, and analysis in a single software package makes this routine particularly attractive to a wide range of researchers.

Because PROC MIXED was developed from a distinctly different perspective than that employed by most statisticians and empirical researchers in the educational, social, and behavioral sciences, its syntax and programming logic may appear unusual to people in these fields (Ferron, 1997). Unlike HLM and MLwiN, which were written with the kinds of models used by social scientists in mind, PROC MIXED was written by agricultural and physical scientists seeking a generalization of the standard linear model that allows for both *fixed* and *random* effects (McLean, Sanders, & Stroup, 1991). Although it is not immediately obvious based upon the documentation provided by SAS, it is indeed the case that by properly specifying the mixed model, a data analyst may fit a variety of specific instances of the multilevel models, hierarchical models, and individual growth models that have become so popular in educational and behavioral research (Kreft, 1995; Hox & Kreft, 1994).

The purpose of this paper is to show educational and behavioral statisticians and researchers how they can use PROC MIXED to fit many common types of multilevel models. Rather than try to cover a broad array of models (without providing sufficient depth for the reader to understand the logic behind the syntax), I focus on two of the most common models: (a) *school effects* models, designed for data on individuals nested within naturally occurring hierarchies (e.g., students within classes, children within families, teachers within schools); and (b) *individual growth* models, designed for exploring longitudinal data (on individuals) over time. In addition, because the use of PROC MIXED does not obviate the need for substantial data processing in preparation for analysis, in the appendix I present general strategies for working with multilevel data in SAS and for creating data sets at several levels.

Multilevel models can be expressed in at least three different ways: (a) by writing separate equations at multiple levels; (b) by writing separate equations at multiple levels and then substituting in to arrive at a single equation; and (c) by writing a single equation that specifies the multiple sources of variation. Bryk and Raudenbush (1992) specify the model for each level separately, and their software program (HLM) never requires you to substitute back to derive a single equation specification. Goldstein (1995) expresses the multilevel model directly using a single equation, and his software program, MLwiN, works from that single level representation. PROC MIXED also requires that you provide a single level representation. For pedagogic reasons, in this paper I take the middle ground, initially writing the model at multiple levels (kept here to two) and then substituting in to arrive at a single equation representation.

To use this paper effectively, a basic understanding of the ideas behind multilevel modeling, hierarchical modeling, and individual growth modeling is helpful. Both Bryk and Raudenbush (1992) and Hox (1995) provide excellent introductions to these topics. In particular, the reader must understand: (a) the difference between a *fixed* effect and a *random* effect; (b) the notion of multiple levels within a hierarchy; (c) the notion that the error variance-covariance matrix can take on different structures; and (d) that centering can be a helpful way of parameterizing models so that the results are more easily interpreted.

This article does not substitute for the comprehensive documentation available through SAS, including the general PROC MIXED documentation (SAS Institute, 1992, 1996), *Getting Started with PROC MIXED* (Latour, Latour, & Wolfinger, 1994), and *The SAS System for Mixed Models* (Littell et al., 1996). My goal is simply to provide a bridge to users already familiar with multilevel modeling because the SAS documentation is thin in this regard. I have found that PROC MIXED's flexibility has led many an unsuspecting user to write a program, obtain results, and have no idea what model has been fit. The goal for the user, then, is to specify the model and to learn the syntax necessary for ensuring that this is the model being fit to the data.

Two-Level School Effects Models

I begin by presenting an overview of strategies for using PROC MIXED to fit classic two-level *school effects* models. By two-level school effects models, I am referring to situations in which you have data at two levels within an organizational hierarchy—such as students within classes or classes within schools—and you would like to examine the behavior of a level-1 outcome as a function of both level-1 and level-2 predictors.

To achieve some continuity with presentations of these models available elsewhere, I use the High School and Beyond data example that Bryk and Raudenbush (1992) include in the 1996 version of HLM for Windows (Bryk et al., 1996). Readers unfamiliar with this example should consult Chapter Five of Bryk and Raudenbush (1992) for a fuller description. The data set consists of information for 7,185 students in 160 schools (with anywhere from 14 to 67 students per school). The student-level (level-1) outcome is MATHACH. The student level (level-1) covariate is SES. There are two school-level (level-2) covariates. One is an aggregate of student level characteristics (MEANSES); the other is a school-level variable (SECTOR). MEANSES and SES are centered at the grand mean (they have means of 0). SECTOR, a dummy variable, is coded 0 and 1.

I begin by fitting an unconditional means model, examining variation in MATHACH across schools. I then sequentially examine the effects of a school-level (level-2) predictor (MEANSES) and a student level (level-1) predictor (student SES). Having examined each type of predictor separately, I conclude this section of the paper by combining both types of predictors into a single model. Wherever possible, I use the notation used by Bryk and Raudenbush (1992). Readers more familiar with Goldstein's (1995) notation will need to make periodic translations.

Unconditional Means Model

The unconditional means model can be viewed as a one-way random effects ANOVA model. Although there are several different ways to write this model, one common approach expresses the outcome, Y_{ij}, as a linear combination of a grand mean μ, a series of deviations from that grand mean (the α_j) and a random error associated with the i^{th} student in the j^{th} school (r_{ij}):

$$Y_{ij} = \mu + \alpha_j + r_{ij} \quad \text{where}$$
$$\alpha_j \sim iid\ N(0,\tau_{00}) \quad \text{and} \quad r_{ij} \sim iid\ N(0,\sigma^2) \tag{1}$$

This model has one fixed effect (μ) and two variance components—one representing the variation between school means (τ_{00}) and the other representing the variation among students within schools (σ^2). You can fit this model in PROC MIXED quite easily using the following syntax:

```
proc mixed;
 class school;
 model mathach = ;
 random school;
```

Rather than parameterize the model this way, however, consider an alternative approach—a two-level approach—that generalizes more easily to more complex models. This strategy expresses the student-level outcome Y_{ij} using a pair of linked models: one at the student level (level-1) and another at the school-level (level-2). At level 1, we express a student's outcome as the sum of an *intercept* for the student's school (β_{0j}) and a random error (r_{ij}) associated with the i^{th} student in the j^{th} school:

$$Y_{ij} = \beta_{0j} + r_{ij} \quad \text{where } r_{ij} \sim N(0,\sigma^2) \tag{2a}$$

At level 2 (the school level), we express the school level intercepts as the sum of an overall mean (γ_{00}) and a series of random deviations from that mean (u_{0j}):

$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad \text{where } u_{0j} \sim N(0,\tau_{00}) \tag{2b}$$

Substituting (2b) into (2a) yields the multilevel model:

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \quad \text{where}$$
$$u_{0j} \sim N(0,\tau_{00}) \quad \text{and} \quad r_{ij} \sim N(0,\sigma^2) \tag{3}$$

Notice the direct equivalence between the model in (1) and the model in (3). The grand mean μ is now represented by γ_{00}, the effect of school (the α_j) is now represented by u_{0j}, and the residual associated with the i^{th} student in the j^{th} school remains r_{ij}. This model can be partitioned into two parts: a fixed part, which contains the single effect γ_{00} (for the overall intercept) and a random part, which contains two random effects (for the intercept u_{0j} and for the within-school residual r_{ij}). We fit this model to data to estimate both the fixed effect γ_{00} (which tells us about the average MATHACH score in the population) and the two random effects, τ_{00} (which tells us about the variability in school means) and σ^2 (which tells us about the variability in MATHACH within schools).

Although it may not be immediately obvious, the model in (3) postulates that the variance and covariance components take on a particular form. First, because we have not indicated otherwise, we are assuming that the r_{ij} and the u_{0j} are independent. Second, if we combine the variance components for the two random effects together into a single matrix, we would find a highly structured block diagonal matrix. For example, if there were three students in each class, we would have:

$$\begin{pmatrix}
\tau_{00}+\sigma^2 & \tau_{00} & \tau_{00} & 0\,0\,0 & 0 & 0 & 0 \\
\tau_{00} & \tau_{00}+\sigma^2 & \tau_{00} & 0\,0\,0 & 0 & 0 & 0 \\
\tau_{00} & \tau_{00} & \tau_{00}+\sigma^2 & 0\,0\,0 & 0 & 0 & 0 \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & 0 & 0\,0\,0 & \tau_{00}+\sigma^2 & \tau_{00} & \tau_{00} \\
0 & 0 & 0 & 0\,0\,0 & \tau_{00} & \tau_{00}+\sigma^2 & \tau_{00} \\
0 & 0 & 0 & 0\,0\,0 & \tau_{00} & \tau_{00} & \tau_{00}+\sigma^2
\end{pmatrix} \tag{4}$$

If the number of students per class varied, the size of each of these submatrices would also vary, although they would still have this common structure. The variance in MATHACH for any given student is assumed to be $\tau_{00} + \sigma^2$. This structure is known as *compound symmetry*. The covariance of MATHACH scores for any two students in a single class is τ_{00}. The covariance of MATHACH scores for any two students in different classes is 0.

The representation of the multilevel model in (3) leads to an alternative specification of the unconditional means model in PROC MIXED. The syntax is:

```
proc mixed noclprint covtest;
 class school;
 model mathach = /solution;
 random intercept/sub=school;
```

After invoking the procedure and identifying any categorical variables (using the CLASS statement), the MODEL statement specifies the fixed effects and the RANDOM statement specifies the random effects. Let's examine this syntax in detail by focusing on its two major parts: the structural part (the first two lines) and the modeling part (the second two lines).

Structural specification. The NOCLPRINT option on the PROC MIXED statement prevents the printing of the CLASS level information giving the numbers of schools involved in the analysis. The first time you run the program, you might not want to include this option to ensure that all relevant groups are included in the analysis. The COVTEST option on the PROC MIXED statement tells SAS that you would like hypothesis tests for the variance and covariance components (described below). This option is not necessary if you are running a version of SAS prior to 6.12. The CLASS statement indicates that SCHOOL is a classification variable whose values do not contain quantitative information.

Model specification. You use the MODEL statement to indicate fixed effects and the RANDOM statement to indicate random effects. The MODEL statement here may appear odd because it seems as if it has no predictors. In reality, it has one implied predictor, the vector 1, which represents the intercept. The /SOLUTION option asks SAS to print the estimates for the fixed effects. PROC MIXED, like HLM, includes an intercept by default. Other programs, such as MLwiN and Hedeker's MIXREG (Hedeker & Gibbons, 1996) require you to specify the intercept explicitly. If you would like to fit a model without an intercept, however, it is very easy: just add the option /NOINT to the model statement.

The RANDOM statement is crucial and its specification is usually the trickiest part about fitting mixed models. By default, there is always at least one random effect, here the lowest-level (within-school) residual r_{ij}. (This is similar to the default random effect in a typical regression model, representing the error term.) By specifying the intercept on this RANDOM statement, we are indicating the presence of a second random effect—that the INTERCEPT in the MODEL statement (which is not explicitly present but implied) should be treated not only as a fixed effect (represented by γ_{00}) but also as a RANDOM effect (represented by τ_{00}). The SUB= option on the RANDOM statement specifies the multilevel structure, indicating how the level-1 units are divided into level-2 units. Here, the subgroups are designated by the classification variable SCHOOL. Without this statement, the model fit would not be that in (3) above, but would rather be $Y_{ij} = \gamma_{00} + r_{ij}$. In other words, the variance component representing the effect of school (for the u_{0j} which has variance τ_{00}) would be omitted.

The results of fitting this model are presented below. For comparison, examine the equivalent model fit using HLM (Bryk and Raudenbush 1992; pp. 62–66).

REML Estimation Iteration History

Iteration	Evaluations	Objective	Criterion
0	1	34899.608417	
1	2	33913.503461	0.00000109
2	1	33913.484655	0.00000000

Convergence criteria met.

Covariance Parameter Estimates (REML)

Cov Parm	Ratio	Estimate	Std Error	Z	PR > |Z|
INTERCEPT	0.21992178	8.60965741	1.07782320	7.99	0.0001
Residual	1.00000000	39.14872611	0.66065147	59.26	0.0001

Model Fitting Information for MATHACH

Description	Value
Observations	7185.000
REML Log Likelihood	-23558.4
Akaike's Information Criterion	-23560.4
Schwarz's Bayesian Criterion	-23567.3
-2 REML Log Likelihood	47116.79

Solution for Fixed Effects

Parameter	Estimate	Std Error	DDF	T	PR > |T|
INTERCEPT	12.63698083	0.24433777	159	51.72	0.0001

Interpreting the output of fitting an unconditional means model. First notice that the model converged quickly. PROC MIXED is a very efficient program making it particularly nice for fitting of a wide range of models. (Of course, as models become more complex, they can take a while to converge. Imbalance can also increase the computational time.)

The next section presents the *Covariance Parameter Estimates*. These are estimates for the random effects portion of the model. In this case, we find that the estimated value of $\tau_{00} = 8.6096$ and the estimated value of $\sigma^2 = 39.1487$. (Differences between these estimates and those presented in Bryk & Raudenbush, 1992, are due to the computational improvements between the two packages. The differences between HLM 4.0 for Windows results and these results are much smaller). Hypothesis tests presented in this section indicate that both variance components are significantly different from 0 (although these tests may not be very reliable)[1]. These estimates suggest that schools do differ in their average MATHACH scores and that there is even more variation among students within schools. (The variance component within school is nearly five times the size of the variance component between schools).

Another way of thinking about the sources of variation in MATHACH is to estimate the intraclass correlation, ρ. This is equivalent to expressing the variance-covariance matrix in (4) in correlation form, with 1's on the diagonal and ρ on the appropriate off-diagonal elements. We estimate ρ, which tells us what portion of the total variance occurs between schools, as:

$$\hat{\rho} = \frac{\hat{\tau}_{00}}{\hat{\tau}_{00}+\hat{\sigma}^2} = \frac{8.6096}{8.6096 + 39.1487} = .18$$

This tells us that there is a fair bit of clustering of MATHACH within school. This suggests that an OLS analysis of these data would likely yield misleading results.[2]

The next section presents information that can be useful for comparing the goodness of fit of multiple models with the same fixed effects but different random effects.[3] The two criteria likely to be the most helpful are the AIC (Akaike's Information Criterion) and the SBC (Schwarz's Bayesian Criterion). Models that fit better will have values of these statistics that are larger. (Note that when these values are negative, as they are here, lower numbers in absolute values are preferred.) Both penalize the log-likelihood for the number of parameters estimated, with the SBC taking a higher penalty for increased complexity. Without a model against which we can compare these statistics they are not very useful. As you fit models with different specifications for the random effects, as I do later in this paper, changes in these statistics help assess differences in goodness of fit (also see Littell et al., 1996).

The last section presents parameter estimates for the fixed effects. As there is only one fixed effect, the intercept, the estimate of 12.64 tells us the average school-level math achievement score in this sample of schools. (Note, this is not the same as the average student level achievement score.)

Including Effects of School Level (level 2) Predictors

The unconditional means model provides a baseline against which we can compare more complex models. We begin with the inclusion of one level-2 variable, MEANSES, which indicates the average SES of the children within the school. Remember that MEANSES has a mean of 0 (it is centered about the grand mean), which facilitates interpretation of the intercept term γ_{00}. Thus, our first conditional model, in which MATHACH is expressed as a function of school-level SES can be written as:

$$Y_{ij} = \beta_{0j} + r_{ij} \quad \text{and} \quad \beta_{0j} = \gamma_{00} + \gamma_{01}\, MEANSES_j + u_{0j}$$
$$\text{where } r_{ij} \sim N(0,\sigma^2) \text{ and } u_{0j} \sim N(0,\tau_{00})$$

Substituting the level-2 equation into the level-1 equation yields:

$$Y_{ij} = [\gamma_{00} + \gamma_{01}\, \text{MEANSES}_j] + [u_{0j} + r_{ij}] \qquad (5)$$

To emphasize that this combined model is the sum of two parts—a fixed part and a random part—I have separated the two components using brackets [ ]. The two terms in the first bracket represents the fixed part, consisting of the two gamma terms. The two terms in the second bracket represent the random part, consisting of the u_{0j} (which represents variation in intercepts between schools) and the r_{ij} (which represents variation within schools). As before, we estimate these random effects through their respective variance components, τ_{00} and σ^2.

Notice that the only difference between the conditional model in (5) and the unconditional model in (3) is the addition of an extra fixed effect for MEANSES. Therefore, the MODEL statement, which identifies the fixed effects, must change to incorporate the predictor. The RANDOM statement, which identifies the random effects remains the same.

We fit the model in (5) using the following code:

```
proc mixed noclprint covtest;
 class school;
 model mathach = meanses/solution ddfm=bw;
 random intercept/sub=school;
```

Notice that nothing has changed except for the MODEL statement, which now includes the additional fixed effect for MEANSES. Here, for simplicity, I restrict attention to a single level-2 variable. Additional school level predictors can be included as fixed effects by appending the variable names to the MODEL statement. The other change is the option /DDFM=BW. This option asks SAS to use the "between/within" method for computing the denominator degrees of freedom for tests of fixed effects. Further details on this option are given in Littell et al. (1996) and SAS Institute (1996, pp. 565–566).

Here is the output:

REML Estimation Iteration History

Iteration	Evaluations	Objective	Criterion
0	1	33999.764766	
1	2	33759.813934	0.00000000

Convergence criteria met.

Covariance Parameter Estimates (REML)

Cov Parm	Ratio	Estimate	Std Error	Z	Pr > |Z|
INTERCEPT	0.06730855	2.63565706	0.40364376	6.53	0.0001
Residual	1.00000000	39.15783186	0.66081390	59.26	0.0001

Model Fitting Information for MATHACH

Description	Value
Observations	7185.000
REML Log Likelihood	-23480.6
Akaike's Information Criterion	-23482.6
Schwarz's Bayesian Criterion	-23489.5
-2 REML Log Likelihood	46961.28

Solution for Fixed Effects

Parameter	Estimate	Std Error	DDF	T	Pr > |T|
INTERCEPT	12.64945599	0.14921620	158	84.77	0.0001
MEANSES	5.86349698	0.36130302	158	16.23	0.0001

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
MEANSES	1	158	263.37	0.0001

Interpreting the output with a single level-2 predictor. Because there are now fixed effects (other than the INTERCEPT) to be estimated, the output includes an additional section presenting relevant hypothesis tests for the fixed effects. This Tests of Fixed Effects section can be helpful when you include a CLASSification variable [a variable that you want represented as multiple dummies] as a fixed effect and you would like a pooled test across all the levels of that variable. If you would like to suppress this additional section (as we do in subsequent illustrative programs in this paper) simply add the option NOTEST to the MODEL statement.

Fixed effects information. The term for the INTERCEPT, 12.65, estimates γ_{00}, the school mean math achievement when the remaining predictors (here, just MEANSES) are 0. Because MEANSES is centered at the grand mean (with a mean of 0), γ_{00} is the estimated MATHACH in a school of "average MEANSES." The term for MEANSES, 5.86, provides our estimate of the other fixed effect, γ_{01}, and tells us about the relationship between math achievement and MEANSES. Schools that differ by 1 point in MEANSES differ by 5.86 points in MATHACH. Its standard error of 0.36 yields an observed t-statistic of 16.22 ($p < .0001$), which indicates that we reject the null hypothesis that there is no relationship between a school's SES and the math achievement scores of its students.

Covariance parameter estimates. These tell us about the random effects. We now estimate τ_{00} to be 2.65 and σ^2 to be 39.16. Although we have used the same symbols in models (3) and (5) to represent these variance components, note that they have very different meanings. In the previous model, there were no predictors, so these were unconditional components. Having added a predictor, these are now conditional components. Notice that the conditional component for the variance within school (the residual component representing σ^2) has remained virtually unchanged (going from 39.15 to 39.16). The variance component representing variation between schools, however, has diminished markedly (going from 8.61 to 2.64). This tells us that the predictor MEANSES explains a large portion of the school-to-school variation in mean math achievement.

One way of measuring how much of the variation in school means is explained by MEANSES is to compute how much the variance component for this term (τ_{00}) has diminished between the two models. As discussed by Bryk & Raudenbush (1992, p. 65), we compute this as (8.61 − 2.65)/8.61, which yields .69, or 69%. We interpret this by saying that 69% of the explainable variation in school mean math achievement scores is explained by MEANSES. (Note that this is not the same as a traditional R^2 statistic. This percentage only talks about the fraction of explainable variation that is explained. If the amount of variation between schools is small, we might be explaining a large amount of very little! For further discussion, see Snijders and Bosker, 1994.)

Having explained 69% of the explainable variation, you might also want to know whether there is still any variation in school means remaining to be explained. The output provides two windows on this question. The first is the test for the residual variance component for intercepts, which rejects the null that τ_{00} is 0 with a z-statistic of 6.53 ($p < .0001$). Although this test is not very reliable, it suggests that even after including MEANSES, there is additional explainable variation present. The second window is to compute the *residual* intraclass correlation, the intraclass correlation among schools "of comparable *SES*." Once again, we estimate the intraclass correlation as that fraction of the sum of both variance components that occurs at the school level (i.e., 2.63/[2.63 + 39.16]), which is 0.06. We can view this residual intraclass correlation as a partial correlation, which tells us about the similarity in math achievement among students within schools after controlling for the effect of MEANSES.

Including Effects of Student-Level (level-1) Predictors

I illustrate the effect of including student level predictors by initially examining a model with only one student-level predictor (SES). To ease interpretation, and to focus on those features of the procedure unique to the inclusion of level-1 predictors, I exclude level-2 predictors in this formulation. After reviewing the steps necessary for including level-1 predictors, I fit a combined model.

Begin by thinking about what the model to include a student level predictor might look like. One simple model might be:

$$\begin{aligned} Y_{ij} &= \beta_{0j} + \beta_{1j} SES_{ij} + r_{ij}, \\ \beta_{0j} &= \gamma_{00} + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + u_{1j}, \end{aligned} \qquad (6)$$

$$\text{where } r_{ij} \sim N(0,\sigma^2) \text{ and } \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \right]$$

This model differs from the simple unconditional model in (3) in three important ways. First, we have included a single level-1 predictor, SES. Second, having included this additional fixed effect, we have also included an additional random effect. Thus, not only are we stipulating that a student's math achievement score is related to his or her SES, but also that the relationship between SES can vary across schools. (If we did not want to allow this slope coefficient to vary across schools, we could have "fixed" it by eliminating the term u_{1j} from the equation for the slope β_{1j}.) Third, having allowed the intercepts and slopes to vary across schools, we now have a larger tau matrix to represent the random effects across schools. Not only are there elements representing the *variance* components for both the intercept and slope, there is also a *covariance* component, representing the correlation between intercepts and slopes (τ_{10}).

Although this model can be fit easily in PROC MIXED, I have chosen not to present the code for doing so because of an issue about interpretation arising from the model parameterization. Consider the interpretation of the betas in equation (6). Across the full sample, SES has a mean of 0. (It is grand-mean centered.) Therefore, β_{0j} represents the average math achievement for a student of average SES across the full sample. It does not represent the average math achievement for students in school j (controlling for SES). As we add predictors to our model, we would like to see how these conditional school means relate to these other predictors. (Consider, for example, the interpretation of the effects of MEANSES presented in the previous section.) To render the parameters more interpretable, and to lead to a model in which we have both level-1 and level-2 predictors, we can rescale SES to be centered about its *school mean*, by computing $CSES_{ij} = SES_{ij} - MEANSES_j$. Unlike some specialized software programs (e.g., HLM) which ask whether you want to center variables, the data analyst must be proactive when using PROC MIXED. Given the misconceptions and misunderstandings surrounding the rationale behind centering and the effects of the different forms of centering (Kreft, de Leeuw, & Aiken, 1995), some might argue that this provision (or lack thereof) is an advantage of this program.

Let us therefore consider the following model representing the effect of a level-1 predictor:

$$\begin{aligned} Y_{ij} &= \beta_{0j} + \beta_{1j}(SES_{ij} - \bar{SES}_j) + r_{ij}, \\ \beta_{0j} &= \gamma_{00} + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + u_{1j}, \end{aligned} \qquad \text{(7a)}$$

$$\text{where } r_{ij} \sim N(0,\sigma^2) \text{ and } \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right]$$

which can be rewritten as:

$$\begin{aligned} Y_{ij} &= \gamma_{00} + u_{0j} + (\gamma_{10} + u_{1j})(SES_{ij} - \bar{SES}_j) + r_{ij} \\ &= [\gamma_{00} + \gamma_{10}(SES_{ij} - \bar{SES}_j)] \quad + \quad [u_{0j} + u_{1j}(SES_{ij} - \bar{SES}_j) + r_{ij}] \end{aligned} \qquad \text{(7b)}$$

with the assumptions as specified in (7a). This model has two fixed effects (an intercept and a slope for centered SES) and three random effects: for the intercepts (registered by the u_{0j}), for the slopes (registered by the u_{1j}), and for the students within schools (registered by the r_{ij}).

We write the PROC MIXED code for fitting this model by specifying the fixed effects on the MODEL statement and the random effects on the RANDOM statement as:

```
proc mixed noclprint covtest noitprint;
 class school;
 model mathach = cses/solution ddfm=bw notest;
 random intercept cses/sub=school type=un;
```

The NOITPRINT option on the PROC statement tells SAS not to print the iteration history (done here to save space). The MODEL statement includes the fixed effect for CSES, the centered SES variable. (Remember that the intercept

is included by default; the notest option suppresses the printing of additional hypothesis tests for the fixed effects.) Notice that the RANDOM statement has changed quite a bit from its simpler specification. Now there are two random effects—one for the INTERCEPT and one for the CSES slope. (Remember that the third random effect, for the r_{ij}, which represents the variation within-school across students, is included by default.) In addition, we have added an option specifying the *structure* of the variance-covariance matrix for the intercepts and slopes. The structure specified, UN, indicates an *unstructured* specification, which allows all three parameters to be determined by the data. This specification is common in school effects analyses. In many other multilevel analyses, you may want to try alternative specifications. I discuss this further when describing methods for fitting individual growth models. In addition to the general PROC MIXED documentation, this topic is also addressed in Wolfinger (1996) and Murray and Wolfinger (1994).

The output from this procedure is:

Covariance Parameter Estimates (REML)

Cov Parm		Ratio	Estimate	Std Error	Z	Pr > |Z|
INTERCEPT	UN(1,1)	0.23642291	8.67686615	1.07855368	8.04	0.0001
	UN(2,1)	0.00138287	0.05075209	0.40619222	0.12	0.9006
	UN(2,2)	0.01890945	0.69398853	0.28078887	2.47	0.0135
Residual		1.00000000	36.70061535	0.62575113	58.65	0.0001

Model Fitting Information for MATHACH

Description	Value
Observations	7185.000
REML Log Likelihood	-23357.1
Akaike's Information Criterion	-23361.1
Schwarz's Bayesian Criterion	-23374.9
-2 REML Log Likelihood	46714.24
Null Model LRT Chi-Square	1065.704
Null Model LRT DF	3.0000
Null Model LRT P-Value	0.0000

Solution for Fixed Effects

Parameter	Estimate	Std Error	DDF	T	Pr > |T|
INTERCEPT	12.64934611	0.24445234	159	51.75	0.0001
CSES	2.19319235	0.12825918	7024	17.10	0.0001

Interpreting the output from models with level-1 predictors. Focus first on the fixed effects. The estimate for γ_{00} (12.65) indicates that the estimated average school mean math achievement score, controlling for student SES, is 12.65. The estimate for γ_{01} (2.19) indicates that the estimated average slope representing the relationship between student SES and math achievement is 2.19. The standard errors for both these parameter estimates are very small, resulting in large t-statistics and low p-values. We conclude that, on average, there is a statistically significant relationship between student SES and math achievement scores.

The *covariance parameter estimates* tell us how much these intercepts and slopes vary across schools. Although SAS presents these estimated variance-covariance components in list form, we may rewrite the first three elements in the list as:

$$\begin{pmatrix} \hat{\tau}_{00} & \hat{\tau}_{01} \\ \hat{\tau}_{10} & \hat{\tau}_{11} \end{pmatrix} = \begin{pmatrix} 8.68 & 0.05 \\ 0.05 & 0.69 \end{pmatrix}$$

So, 8.68 tells us about the variability in intercepts, 0.69 tells us about the variability in slopes, and 0.05 tells us about the covariance between intercepts and slopes. Estimated standard errors and tests of the null hypotheses that each of these components is 0 are given in the remaining columns of the list. What do we see? First, that the intercepts are very variable; in other words, schools do differ in average math achievement levels even after controlling for the effects of student SES. Second, that the slopes are also variable (variance component of .69). We reject the null that this variance component = 0 with $p = .0135$. Third, there is little correlation between intercepts and slopes (covariance component 0.05, $p = .9006$). In other words, there is no evidence that the effects of student SES on math achievement differ depending upon the average math achievement in the school.

How much of the within school variance in math achievement is explained by student SES? Just as we compared the variance component for τ_{00} in the unconditional and conditional models (presented in the previous two sections), so, too, can we compare the estimates for σ^2 for the unconditional and conditional models. Returning to the output on page 7 we find an unconditional estimate of 39.15. Here we have a conditional estimate of 36.70. Inclusion of student level SES has therefore explained (39.15–36.70)/39.15 = 0.06, or 6% of the explainable variation within schools. Comparatively speaking, then, school SES explains much more of the variation in school level math achievment than does student SES explain the within-school variation in student level achievement. When interpreting these results, however, the previously mentioned cautions about the term "explained" variation in the context of multilevel models remain, and even escalate. Interested readers should consult Snijders and Bosker (1994) for a fuller discussion of this issue.

Including Both Level-1 and Level-2 Predictors

Having separately specified models with either just level-1 predictors or level-2 predictors, we can now consider models which contain variables of both types. Although simplicity would have us fit a model with just the effects of student SES and school SES, to achieve parallelism with Bryk and Raudenbush (1992), we also add in the effects of a second school level variable, SECTOR, coded as 0 for public schools and 1 for Catholic schools.

Begin by thinking about how you would want to specify the model to be fit. I strongly advise you to write the model out, interpreting each of the parameters, before writing code to fit the model. As models get more complex, it is not

always obvious how to parameterize the model so that the output can be used directly to answer your research question. In the previous section, for example, we saw the gains that come from centering student SES within school only after writing out a model in which student SES was not centered. I find it helpful to write separate models at the two levels and then combine them together to yield the single level representation required for PROC MIXED.

Consider the following model:

$$\begin{aligned} Y_{ij} &= \beta_{0j} + \beta_{1j}(SES_{ij} - \bar{SES}_j) + r_{ij}, \\ \beta_{0j} &= \gamma_{00} + \gamma_{01}MEANSES_j + \gamma_{02}SECTOR_j + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}MEANSES_j + \gamma_{12}SECTOR_j + u_{1j}, \\ \text{where } r_{ij} &\sim N(0,\sigma^2) \text{ and } \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right] \end{aligned} \tag{8a}$$

Notice the similarities between this model (which includes both level-1 and level-2 predictors) and the previous model (eq 7) that included only a level-1 predictor. The level-1 part of the model remains the same (because there is just the one level-1 predictor), but each part of the level-2 part of the model now has two additional fixed effects. The number of random effects remains the same. The number of random effects may be increased if an additional level-1 variable is added to the model.

We can combine the level-1 and level-2 equations together to yield:

$$\begin{aligned} Y_{ij} = {} & \gamma_{00} + \gamma_{01}MEANSES_j + \gamma_{02}SECTOR_j + \gamma_{10}(SES_{ij} - \bar{SES}_j) \\ & + \gamma_{11}MEANSES_j(SES_{ij} - \bar{SES}_j) + \gamma_{12}SECTOR_j(SES_{ij} - \bar{SES}_j) \\ & + u_{0j} + u_{1j}(SES_{ij} - \bar{SES}_j) + r_{ij} \end{aligned} \tag{8b}$$

Having written out a combined equation, we can now write the requisite PROC MIXED code. Each fixed effect on the first two lines of the equation in 8b must appear in the MODEL statement (because this is where fixed effects are indicated) and each random effect (on the last line of equation 8b) must appear in the RANDOM statement. By default, SAS includes an intercept as a fixed effect on the MODEL statement and a within-group random effect (for the r_{ij}) on the RANDOM statement. Interaction terms may be easily specified in the MODEL statement by using an asterisk (*) between the relevant variables. The code:

```
proc mixed noclprint covtest noitprint;
 class school;
 model mathach = meanses sector cses meanses*cses
                 sector*cses/solution ddfm=bw notest;
 random intercept cses/type=un sub=school;
```

yields the output:

Covariance Parameter Estimates (REML)

Cov Parm		Ratio	Estimate	Std Error	Z	Pr > |Z|
INTERCEPT	UN(1,1)	0.06485969	2.38172336	0.37171728	6.41	0.0001
	UN(2,1)	0.00524422	0.19257382	0.20451479	0.94	0.3464
	UN(2,2)	0.00276060	0.10137258	0.21381009	0.47	0.6354
Residual		1.00000000	36.72116429	0.62613331	58.65	0.0001

Model Fitting Information for MATHACH

Description	Value
Observations	7185.000
REML Log Likelihood	-23251.8
Akaike's Information Criterion	-23255.8
Schwarz's Bayesian Criterion	-23269.6
-2 REML Log Likelihood	46503.67
Null Model LRT Chi-Square	220.5683
Null Model LRT DF	3.0000
Null Model LRT P-Value	0.0000

Solution for Fixed Effects

Parameter	Estimate	Std Error	DDF	T	PR > |T|
INTERCEPT	12.11358496	0.19880323	157	60.93	0.0001
CSES	2.93876223	0.15509265	7022	18.95	0.0001
MEANSES	5.33911631	0.36929107	157	14.46	0.0001
SECTOR	1.21667252	0.30637896	157	3.97	0.0001
CSES*MEANSES	1.03887054	0.29890063	7022	3.48	0.0005
CSES*SECTOR	-1.64258263	0.23979107	7022	-6.85	0.0001

Interpreting the output of fitting models with both level-1 and level-2 predictors. Begin with the fixed effects. All are significantly different from 0 ($p < .001$). As SECTOR is a dummy variable indicating whether the school is a public school or a Catholic school, it can be helpful to rewrite a pair of fitted models, one for each sector, by substituting in the values of 0 and 1 for SECTOR:

Public: MATHACH = 12.11 + 5.34 MEANSES + 2.94 CSES
+ 1.03 MEANSES*CSES

Catholic: MATHACH = 13.33 + 5.34 MEANSES + 1.30 CSES
+ 1.03 MEANSES*CSES

The main effect of SECTOR tells us that the intercepts in these two models are significantly different. The interaction between CSES and MEANSES tells us that the slopes for CSES differ depending upon the MEANSES of the school; the interaction between CSES and SECTOR tells us that the slopes for CSES are significantly different in the two sectors. (I should note that I tested to see whether there was a two-way interaction between MEANSES and SECTOR and a three way interaction between MEANSES, CSES, and SECTOR and found none.)

We could use these equations to graph the results of the multilevel model (as done with these data by Bryk & Raudenbush, 1992, p. 73). Because the variable MEANSES has a grand mean of 0, and CSES is centered at its school mean, the six parameter estimates have easy and direct interpretations. The average public school math achievement score is 12.11; the average Catholic school score is 13.33. At average values of student and school SES, these means are significantly different. Student and school level SES are associated with math achievement in both sectors, although the magnitude of the student effect differs across sectors. There is also an interaction between student and school SES. In both sectors, the slope for student SES is higher in schools with higher mean SES levels.

The findings with respect to the random effects are a bit different. The variance component for intercepts (τ_{00}) remains significantly different from 0, suggesting that there is additional variation in school mean achievement levels that is not explained by these three factors and their interactions. Were this an actual analysis, you would interpret this finding as reason to believe that there are additional school level factors that might explain the variation in school means.

The variance component for slopes, in contrast, is very small (.10), and the null hypothesis that the slopes do not differ across schools cannot be rejected ($p = .64$). Similarly, the component representing the covariance between intercepts and slopes is also small (.19) and we cannot reject the null hypothesis that it, too, is 0 ($p = .35$).

These findings suggest that a simpler model, in which the intercepts vary across schools but the slopes do not may provide a reasonable fit to these data. We would fit such a model as follows:

```
proc mixed noclprint covtest noitprint;
 class school;
 model mathach = meanses sector cses meanses*cses
                 sector*cses/solution ddfm=bw notest;
 random intercept/sub=school;
```

Notice that the fixed portion of the model (on the MODEL statement) has remained unchanged. The random portion, however, is now simpler, involving only random intercepts, not slopes. This simplification, which leaves us with only one explicit random effect, allows us to drop the TYPE = UN option from the random statement. Because the fixed portion of the model is unchanged, we can now use the goodness-of-fit statistics to compare the two models. Fitting this model to the data, we find:

	AIC	SBC	−2LL
random intercepts and slopes	−23,255.8	−23,269.6	46,503.67
random intercepts	−23,254.4	−23,261.3	46,504.79

Recalling that we want larger values of the AIC and SBC, it appears that a model in which we do not treat the slopes as random (the second model) provides a better fit. Both the AIC and SBC are larger with this more restricted model, and the change in the −2LL is only 1.12. An approximate test of the null hypothesis that this change is 0 is given by comparing the differences in the −2LL's to a χ^2 distribution, here on two degrees of freedom (to correspond to the two additional parameters).[4] This conclusion is identical to that reached by Bryk and Raudenbush (1992, p. 76), in their analysis of these data.

Individual Growth Models

There are several ways you can fit individual growth models in PROC MIXED. One approach uses a RANDOM statement (as illustrated in the school-effects analyses). An alternative approach uses a REPEATED statement (which mimics classical repeated measures analysis of variance). Although experienced users tend to prefer the latter approach, I begin with the former approach, because it is easier to see the parallels between the school effects models presented in the previous section and the individual growth models under discusssion here. I then turn to the use of the REPEATED statement.

An Unconditional Linear Growth Model

Let's begin with a simple two-level model, in which the level-1 model is a linear individual growth model, and the level-2 model expresses variation in parameters from the growth model as random effects unrelated to any person-level covariates. By convention (and to facilitate extension to a 3-level model in which individuals within groups are tracked over time), we represent the parameters in the level-1 (within person) model using π and the parameters in the level-2 (between-person) model using β. Thus, we may write the level-1 and level-2 models as:

$$\begin{aligned} Y_{ij} &= \pi_{0j} + \pi_{1j}(TIME)_{ij} + r_{ij}, \quad \text{where } r_{ij} \sim N(0,\sigma^2) \\ &\qquad\qquad \text{and} \\ \pi_{0j} &= \beta_{00} + u_{0j}, \\ \pi_{1j} &= \beta_{10} + u_{1j}, \quad \text{where } \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right] \end{aligned} \tag{9a}$$

which can be written in combined form as:

$$Y_{ij} = [\beta_{00} + \beta_{10}TIME_{ij}] \quad + \quad [u_{0j} + u_{1j}TIME_{ij} + r_{ij}] \tag{9b}$$

Notice the direct parallels with the model used for the school effects analysis in (7b). As before, the multilevel model is expressed as the sum of two parts: a fixed part, which contains two fixed effects (for the intercept and for the effect of TIME) and a random part, which contains three random effects (for the intercept, the TIME slope, and the within person residual r_{ij}). Notice that this formulation treats both the intercept and slope as random effects (although this assumption can be changed), and that there are no level-2 covariates (this, too, can be changed).

To fit this model, you must first create a person-period data set in which each individual has one record for every time-period that he or she is observed. (SAS code to create person-period data sets is presented in the appendix.) With this data set, the syntax to fit the individual growth model using PROC MIXED looks quite similar to that for fitting a school-effects analysis with a single level-1 predictor and random intercepts and slopes:


```
proc mixed noclprint covtest;
 class id;
 model y = time/solution ddfm=bw notest;
 random intercept time/subject=id type=un;
```

Notice the similarity between this code and that for the school effects analysis. The CLASS variable has changed from SCHOOL to ID to indicate that the data represent multiple observations over time for individuals. This CLASS variable is used on the RANDOM statement to indicate that when the random effects are specified, we want to allow both intercepts and slopes to vary across people.

The MODEL statement indicates what type of growth model is to be fit. Be sure to consider a range of alternative options in specifying the growth model, and be careful about coding the variable TIME. As we saw in the school effects analysis, the interpretation of the intercept differed depending on how the within person variable (student SES, in our example) was expressed. Similarly, the intercept in the growth model can be specified in such a way that it represents *initial status* (by coding TIME = 0 for the first wave of data), *average status* (by centering TIME) or even *final status* (by coding time using negative numbers and letting 0 represent the last wave). If there are three or more waves of data, models allowing for curvilinear growth might be considered.

The RANDOM statement indicates the random effects that you want to include in your model. As with the school effects analysis, this is probably the most difficult statement to write correctly. By default, there is one random effect in the model, for the r_{ij}, representing variation within persons. To fit an individual growth model, two *additional sources of variation* need to be included: in the INTERCEPTs and in the *slopes* for TIME. The options after the / indicate how to structure the variance-covariance matrix representing these sources of variation.

- *The SUBJECT=option* (alias for SUB in school effects analyses) indicates that the data set is composed of a set of different "subjects." Subjects are assumed to be independent of each other; hence, the SUBJECT=ID command indicates that the variance covariance matrix for the random effects is to be block diagonal, with identical blocks.
- *The TYPE=option* specifies the structure of these diagonal blocks. Specifying TYPE=UN indicates that you would like to treat the variance-covariance matrix for the intercepts and slopes as unstructured, with a separate variance (or covariance) component for each of the elements. The unstructured option indicates that you would not like to place any structure on the variances for intercepts and variances for slopes (they can be different, which is usually essential as they are not likely to be identical!) and that you would not like to impose any structure on the covariance between these two either.

I illustrate the results of this analysis using the data presented in Willett (1988) on the growth in opposite naming task on four occasions for 35 individuals. TIME is coded 0, 1, 2, and 3, so that the intercept estimates the (true) value of opposite-naming skill at occasion 0 (initial status) and the slope estimates the

rate of change in (true) opposite-naming skill across occasions. This judicious coding of TIME (the level-1 predictor) is done for the same reason student level SES was centered in the school effects analysis presented earlier. By choosing an appropriate scale for TIME, the parameters of the within-person growth model become interesting in their own right, making the modeling of them as a function of between-person covariates a vehicle for answering research questions about inter-individual differences in growth. The results of fitting the model are:

REML Estimation Iteration History

Iteration	Evaluations	Objective	Criterion
0	1	1134.0992383	
1	1	1013.1957046	0.00000000

Convergence criteria met.

Covariance Parameter Estimates (REML)

Cov Parm		Ratio	Estimate	Std Error	Z	Pr > |Z|
INTERCEPT	UN(1,1)	7.51691915	1198.7767899	318.38096701	3.77	0.0002
	UN(2,1)	-1.12402041	-179.2555630	88.96341625	-2.01	0.0439
	UN(2,2)	0.83021660	132.40057143	40.21069632	3.29	0.0010
Residual		1.00000000	159.47714286	26.95655716	5.92	0.0001

Model Fitting Information for Y

Description	Value
Observations	140.0000
REML Log Likelihood	-633.411
Akaike's Information Criterion	-637.411
Schwarz's Bayesian Criterion	-643.266
-2 REML Log Likelihood	1266.823
Null Model LRT Chi-Square	120.9035
Null Model LRT DF	3.0000
Null Model LRT P-Value	0.0000

Solution for Fixed Effects

Parameter	Estimate	Std Error	DDF	T	Pr > |T|
INTERCEPT	164.37428571	6.11884861	34	26.86	0.0001
TIME	26.96000000	2.16660366	104	12.44	0.0001

Interpreting the output from an unconditional individual growth model. Notice that PROC MIXED converged in just two iterations, the minimum amount of time necessary for convergence to be evaluated. This rapid convergence results from the perfectly balanced data set. In other analyses, especially those with missing data, unbalanced data, or high degrees of collinearity, convergence is unlikely to be so rapid.

Focus first on the estimates of the fixed effects. Because this is an individual growth model with no level-2 covariates, they can be interpreted in the usual way: $\beta_{00} = 164.37$ is our estimate of the average *intercept* across persons (the average value of Y when TIME=0) and $\beta_{10} = 26.96$ is our estimate of the

average *slope* across persons. Hence, the average person began with a score of 164 and gained 27 points per testing occasion. Standard errors and tests can also be interpreted in the usual ways. We reject the null hypotheses that either of these parameters are 0 in the population.

Focus next on the random effects. We may write the estimates for the first three variance-covariance components in matrix form as:

$$\begin{pmatrix} \hat{\tau}_{00} & \hat{\tau}_{01} \\ \hat{\tau}_{10} & \hat{\tau}_{11} \end{pmatrix} = \begin{pmatrix} 1198.78 & -179.26 \\ -179.26 & 132.40 \end{pmatrix}$$

and we also conclude that the estimated value of σ^2 is 159.48. In addition to these estimates, SAS also produces standard errors for these estimates, and hypothesis tests of the null hypotheses that these population variances (and covariances) are 0. Here, you can see that all the tests reject, including those for the terms we are most interested in: for τ_{00} and for τ_{11}, which tell us that there is variation in both the intercepts and slopes that potentially could be explained by a level 2 (person-level) covariate.

The output presents several goodness of fit statistics that can be used to evaluate this model, and to compare the goodness of fit for this model with that of other (nested) models. In addition to indicating the number of observations, we are presented with the actual REML log-likelihood, and -2RLL. Please consult the SAS manual for details on these statistics.

A Linear Growth Model With a Person-Level Covariate

Having fit an unconditional growth model, we may now consider a model in which we explore whether variation in intercepts and slopes is related to a covariate. Begin by considering the model:

$$Y_{ij} = \pi_{0j} + \pi_{1j}(TIME)_{ij} + r_{ij}, \quad \text{where } r_{ij} \sim N(0,\sigma^2)$$

and

$$\begin{aligned} \pi_{0j} &= \beta_{00} + \beta_{01}COVAR_j + u_{0j}, \\ \pi_{1j} &= \beta_{10} + \beta_{11}COVAR_j + u_{1j}, \end{aligned} \quad \text{where } \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right]$$

Were we to fit this model, the interpretation of the fixed effects for β_{00} and β_{10} would be based upon conceiving of a case in which the value of COVAR was 0. As this covariate never even approaches 0, this parameterization of the level-2 model is not the most useful. So instead, we center the covariate at its grand mean, and consider the model:

$$Y_{ij} = \pi_{0j} + \pi_{1j}(TIME)_{ij} + r_{ij}, \quad \text{where } r_{ij} \sim N(0,\sigma^2)$$

and

$$\begin{aligned} \pi_{0j} &= \beta_{00} + \beta_{01}(COVAR_j - \bar{COVAR}) + u_{0j}, \\ \pi_{1j} &= \beta_{10} + \beta_{11}(COVAR_j - \bar{COVAR}) + u_{1j}, \end{aligned} \quad \text{where } \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right] \tag{10a}$$

Now, the interpretation of these fixed effects is far more straightforward. β_{00} represents the average intercept in the individual growth model and β_{10} represents the average slope.

Substituting the level-2 models into the level-1 model yields the combined representation that most closely resembles the statements needed to use PROC MIXED:

$$Y_{ij} = \beta_{00} + \beta_{10}(TIME)_{ij} + \beta_{01}(COVAR_j - \bar{COVAR}) + \beta_{11}(COVAR_j - \bar{COVAR})(TIME)_{ij} + u_{0j} + u_{1j}(TIME)_{ij} + r_{ij} \quad (10b)$$

Letting CCOVAR represent the centered covariate, we fit this model by writing:

```
proc mixed noclprint covtest;
 class id;
 model y = time ccovar time*ccovar/s ddfm=bw notest;
 random intercept time /type=un sub=id gcorr;
```

Notice the similarity between this syntax and the school effects model. Notice, too, that we have added the option GCORR to the RANDOM statement, which tells SAS to print the estimated correlation matrix amongst the random effects (see below). Fitting this model we find:

G Correlation Matrix

Parameter	Subject	Row	COL1	COL2
INTERCEPT	ID 1	1	1.00000000	-0.48945185
TIME	ID 1	2	-0.48945185	1.00000000

Covariance Parameter Estimates (REML)

Cov Parm		Ratio	Estimate	Std Error	Z	Pr > |Z|
INTERCEPT	UN(1,1)	7.75291483	1236.4127057	332.40217831	3.72	0.0002
	UN(2,1)	-1.11760998	-178.2332472	85.42977775	-2.09	0.0370
	UN(2,2)	0.67250510	107.24919114	34.67670438	3.09	0.0020
Residual		1.00000000	159.47714286	26.95655716	5.92	0.0001

Model Fitting Information for Y

Description	Value
Observations	140.0000
REML Log Likelihood	-630.142
Akaike's Information Criterion	-634.142
Schwarz's Bayesian Criterion	-639.968
-2 REML Log Likelihood	1260.285
Null Model LRT Chi-Square	120.7249
Null Model LRT DF	3.0000
Null Model LRT P-Value	0.0000

Solution for Fixed Effects

Parameter	Estimate	Std Error	DDF	T	PR > |T|
INTERCEPT	164.37428571	6.20609540	33	26.49	0.0001
TIME	26.96000002	1.99388078	103	13.52	0.0001
CCOVAR	-0.11355272	0.50401189	33	-0.23	0.8231
TIME*CCOVAR	0.43285774	0.16192784	103	2.67	0.0087

Interpreting the output from a linear growth models with a person-level covariate. First examine the fixed effects. Because we grand mean centered our level-2 covariate, the estimates for the INTERCEPT and for TIME (i.e., for β_{00} and β_{10}) are identical to what they were in the unconditional model estimated in the previous section and the interpretation is similar as well. The only difference now is that we add the phrase "controlling for the covariate" to the interpretation.

The coefficients for the centered covariate and its interaction with time are new. The coefficient for CCOVAR (−0.11) captures the relationship between the covariate and initial status. As the standard error is over four times larger than the estimate itself, we conclude that there is no relationship between initial status and the covariate. With respect to the growth rates, however, we do find an effect of the covariate. The parameter estimate of .43 indicates that individuals who differ by 1.0 with respect to the covariate have growth rates that differ by 0.43.

The estimate for σ^2 has remained unchanged at 159.47. But the estimates for the variance-covariance matrix for the slopes have changed to:

$$\begin{pmatrix} \hat{\tau}_{00} & \hat{\tau}_{01} \\ \hat{\tau}_{10} & \hat{\tau}_{11} \end{pmatrix} = \begin{pmatrix} 1236.41 & -178.23 \\ -178.23 & 107.25 \end{pmatrix}$$

Comparing these estimates to those from the unconditional model (in the previous section), we see that when it comes to estimating initial status, inclusion of the covariate did not help at all (it did not reduce the size of the variance component for intercepts). Indeed, the variance component actually increased slightly! But inclusion of the covariate did improve the fit of the growth rates. The variance component for growth rates went from 132.40 to 107.25. Computing (132.40−107.25)/132.40=0.19, we find a 19% reduction. In other words, the covariate accounts for 19% of the explainable variation in growth rates.

Exploring the Structure of Variance Covariance Matrix Within Persons

The classic growth models fit in the previous two sections place a common, but sometimes unrealistic, assumption on the behavior of the r_{ij}, the within-person residuals over time. Were we to fit a model in which only the intercepts vary across persons, we would be assuming a compound symmetric error covariance matrix for each person. When we fit a model in which the slopes vary as well, we introduce heteroscedasticity into this error covariance matrix (which can be seen through the inclusion of the effect of TIME in the random portion of the model in equation 9b).

How realistic are such assumptions? One of the strengths of PROC MIXED is that it allows the user to compare different structures for the error covariance matrix. Instead of the intercepts and slopes as outcomes model in (9a), consider the following simpler model for observations over time:

$$
\begin{aligned}
&Y_{ij} = \pi_{0j} + \pi_{1j}(TIME)_{ij} + r_{ij}, \quad \text{where } r_{ij} \sim N(0,\Sigma) \\
&\pi_{0j} = \beta_{00}, \\
&\pi_{1j} = \beta_{10}.
\end{aligned} \tag{11a}
$$

In this model, the intercepts and growth rates are assumed to be constant across people. But the model introduces a different type of complexity: the residual observations within persons (after controlling for the linear effect of TIME) are correlated through the within-person error variance-covariance matrix Σ. By considering alternative structures for Σ (that ideally derive from theory), and by comparing the goodness of fit of resulting models, the user can determine what type of structure is most appropriate for the data at hand.

Many different types of error-covariance structures are possible. If there are only three waves of data, it is worth exploring only a few of these possibilities because there is so little data for each person. With additional observations per person (in this example we have four), additional structures for the Σ matrix (called the R matrix in the language of PROC MIXED) are possible. The interested reader is referred to the *SAS System for Mixed Models* (Littell et al., 1996; pp. 92–102), the PROC MIXED documentation in *Getting Started with PROC MIXED* (Latour et al., 1994; pp. 57–58) and the helpful paper by Wolfinger (1996) devoted entirely to this topic.

The structure of the within-person error covariance matrix is specified using a REPEATED statement. To fit the model in (11a) under the assumption that Σ is compound symmetric we write:

```
proc mixed noclprint covtest noitprint;
 class id wave;
 model y = time/s notest;
 repeated wave/type=cs subject=id r;
```

Notice that I have added a second CLASS variable (WAVE) to indicate the time structured nature of the data within person and I have used WAVE on the REPEATED statement. WAVE differs from TIME in that WAVE is treated as a series of dummies, whereas TIME is treated as a continuous variable to yield the growth model. The variable specified on the REPEATED statement must be categorical (although it need not be equal interval). The TYPE=option is crucial, for it specifies the form of the within-person variance-covariance matrix. In addition to the compound symmetry specification (CS) shown here, other possibilities include UN (for unstructured) and AR(1) for autoregressive with a lag of one. The SUBJECT=ID tells SAS that there are to be separate blocks of this matrix, one for each subject. The R option asks SAS to print the R matrix.

Here is the output from the procedure run with a compound symmetry assumption:

Covariance Parameter Estimates (REML)

Cov Parm	Ratio	Estimate	Std Error	Z	Pr > |Z|
DIAG CS	2.40703008	904.80538381	242.59019002	3.73	0.0002
Residual	1.00000000	375.90115385	52.12811095	7.21	0.0001

Model Fitting Information for Y

Description	Value
Observations	140.0000
REML Log Likelihood	-650.170
Akaike's Information Criterion	-652.170
Schwarz's Bayesian Criterion	-655.097
-2 REML Log Likelihood	1300.340
Null Model LRT Chi-Square	87.3867
Null Model LRT DF	1.0000
Null Model LRT P-Value	0.0000

Solution for Fixed Effects

Parameter	Estimate	Std Error	DDF	T	Pr > |T|
INTERCEPT	164.37428571	5.77664310	34	28.45	0.0001
TIME	26.96000000	1.46560793	104	18.40	0.0001

The *SAS System for Mixed Models* presents a nice discussion of how to compare error structures (Littell et al., 1994; pp. 92–102). The idea is to compare goodness of fit statistics for different error structures, determining which one seems to fit best. As a point of comparison, consider selected results presented below obtained when two additional error structures were posited: Autoregressive (1) and totally unstructured.

Assumption	N parameters	AIC	SBC	−2RLL
Compound Symmetry	2	−652.17	−655.10	1300.34
AR (1)	2	−636.73	−641.66	1273.47
Unstructured	10	−641.71	−656.35	1263.42

Recall that we prefer models in which the AIC and SBC are larger and the −2RLL is smaller. Although the totally unstructured Σ yields the best value of −2RLL, it does so at the price of many parameters. The estimated variance covariance matrix from this model is:

$$
\begin{pmatrix}
\hat{\sigma}^2_{11} & \hat{\sigma}_{12} & \hat{\sigma}_{13} & \hat{\sigma}_{14} \\
\hat{\sigma}_{21} & \hat{\sigma}^2_{22} & \hat{\sigma}_{23} & \hat{\sigma}_{24} \\
\hat{\sigma}_{31} & \hat{\sigma}_{32} & \hat{\sigma}^2_{33} & \hat{\sigma}_{34} \\
\hat{\sigma}_{41} & \hat{\sigma}_{42} & \hat{\sigma}_{43} & \hat{\sigma}^2_{44}
\end{pmatrix}
=
\begin{pmatrix}
1308 & 977 & 921 & 564 \\
977 & 1120 & 1018 & 856 \\
921 & 1018 & 1289 & 1081 \\
564 & 856 & 1081 & 1415
\end{pmatrix}
$$

Notice the structure of this matrix—the variances along the diagonal are fairly similar, and the off diagonal elements decrease as they represent covariances between errors further spaced in time. This type of structure is exactly that specified by the lagged autoregressive structure, which is why it is not surprising

that the AR(1) model yields values of AIC and SBC that appear superior. With only two parameters, this approach estimates Σ to be:

$$\begin{pmatrix} 1324 & 1092 & 901 & 743 \\ 1092 & 1324 & 1092 & 901 \\ 901 & 1092 & 1324 & 1092 \\ 743 & 901 & 1092 & 1324 \end{pmatrix}$$

which is quite similar to the unstructured estimate, but this requires only two parameters, σ^2 and ρ. Based on these analyses, we would conclude that the AR(1) structure provides a better fit to the data. Were this an actual analysis, however, we would also consider alternative structures before stopping at this conclusion.

Having established the method for specifying the structure of the within-person error covariance matrix, we may now consider what happens when we combine this specification with the intercepts and slopes as outcomes specification considered earlier. We allow the intercepts and slopes to vary across people by writing:

```
proc mixed noclprint covtest noitprint;
 class id wave;
 model y = time ccovar time*ccovar/s ddfm=bw notest;
 random intercept time /type=un sub=id g;
 repeated wave/type=ar(1) subject=id r;
```

which yields the following output:

R Matrix for ID 1

Row	COL1	COL2	COL3	COL4
1	141.36668313	-19.36313770	2.65218857	-0.36327295
2	-19.36313770	141.36668313	-19.36313770	2.65218857
3	2.65218857	-19.36313770	141.36668313	-19.36313770
4	-0.36327295	2.65218857	-19.36313770	141.36668313

G Matrix

Parameter	Subject	Row	COL1	COL2
INTERCEPT	ID 1	1	1258.0957499	-182.4126739
TIME	ID 1	2	-182.4126739	110.94230682

Covariance Parameter Estimates (REML)

Cov Parm		Ratio	Estimate	Std Error	Z	Pr > |Z|
INTERCEPT	UN(1,1)	8.89952089	1258.0957499	333.24588494	3.78	0.0002
	UN(2,1)	-1.29035123	-182.4126739	84.55201948	-2.16	0.0310
	UN(2,2)	0.78478397	110.94230682	34.52985960	3.21	0.0013
WAVE	AR(1)	-0.00096891	-0.13697101	0.25888610	-0.53	0.5968
Residual		1.00000000	141.36668313	36.34493926	3.89	0.0001

Model Fitting Information for Y

Description	Value
Observations	140.0000
REML Log Likelihood	-630.022
Akaike's Information Criterion	-635.022
Schwarz's Bayesian Criterion	-642.304
-2 REML Log Likelihood	1260.045
Null Model LRT Chi-Square	120.9650
Null Model LRT DF	4.0000
Null Model LRT P-Value	0.0000

Solution for Fixed Effects

Parameter	Estimate	Std Error	DDF	T	Pr > |T|
INTERCEPT	164.42273345	6.19898567	33	26.52	0.0001
TIME	26.90816754	1.97745500	103	13.61	0.0001
CCOVAR	-0.12338407	0.50343450	33	-0.25	0.8079
TIME*CCOVAR	0.43573062	0.16059386	103	2.71	0.0078

When interpreting this output, it is useful to compare it to the simpler models, which included random effects for the intercepts and slopes, but which imposed no additional structure on the error covariance matrix (beyond the heteroscedastic structure of the intercepts and slopes as outcomes model). When we make these comparisons, all signs point towards the conclusion that we do not need to add the extra complexity of the autoregressive error structure, once the covariate has been taken into account. I emphasize this last phrase because the error covariance structure within persons describes the behavior of the errors—in other words, what remains after removing the other fixed and random effects in the model. In this instance, and in many others, the autoregressive structure is no longer needed after other fixed and random effects are taken into account.

What evidence am I using to reach this conclusion? First, consider the covariance estimate for the autoregressive parameter. We are unable to reject the null hypothesis that this estimate, -0.13, could have been obtained from a population in which the true value of the parameter were 0. In other words, there is little supporting evidence to increase the complexity of Σ by adding off-diagonal elements. Second, when comparing the two models that include the covariate and its interaction with time, differing only in the inclusion of the autoregressive parameter, the -2RLL statistic improves only trivially, from -630.14 without this assumption to -630.02 with this assumption. This improvement is so small that the AIC and SBC, which both penalize for the additional parameter, actually get worse. Therefore, despite the fact that there appears to be an autoregressive error structure when the covariate is not included and the slopes are not treated as random, the need for this additional structure disappears when these features are added to the model.

As this example shows, a range of models can be fit to the same data. Experienced data analysts know that selecting among competing models can be tricky, especially when the number of observations per person is relatively small.

Were we conducting this analysis to reach substantive conclusions about the relationship between the outcome and predictors, we would fit several additional models to these data, including one with an AR(1) error covariance matrix and random intercepts. Readers interested in learning more about specifying the error covariance matrix and comparing results across models should consult Van Leeuwen (1997), Goldstein, Healy, and Rasbash (1994), and Wolfinger (1993, 1996).

Conclusion

Statistical software does not a statistician make. That said, without software, few statisticians and even fewer empirical researchers would fit the kinds of sophisticated statistical models being promulgated today. The availability of flexible integrated software for fitting multilevel models holds the possibility that larger numbers of users will be able to fit reasonable statistical models to their data. Of course, as software becomes easier to use, we face the danger that statistical programming will substitute for clear statistical thinking and model development. Readers of the 1995 special issue of the *Journal of Educational and Behavioral Statistics* entitled *Hierarchical Linear Models: Problems and Prospects* (Kreft, 1995) were reminded that no piece of software will resolve the challenging statistical issues underlying decisions about model specification with complex data structures. Yet readers of this special issue were also reminded that without software, few users would fit the models we would like to see applied in education and the behavioral sciences.

The ideas presented in this paper can be easily extended to three-level (and higher-level models). In the case of "school-effects" analyses, the user must specify multiple RANDOM statements, with appropriate nesting specifications given in the SUB= option. For example, if you have data on students within teachers within schools, you could fit an unconditional means model with the syntax:

```
proc mixed noclprint covtest;
 class teacher school;
 model mathach = /solution;
 random intercept/sub=school;
 random intercept/sub=teacher(school);
```

Note that we do not have to include the option/TYPE=UN on either of the RANDOM statements because each specifies only one random effect (for the intercept). Were we to include additional random effects on either statement (that is, if we were to move beyond an unconditional means model) we would need to add this option to the appropriate line.

In the case of longitudinal analyses that track individuals who are nested within groups, the specifications in the school-effects analysis portion of this paper can be combined with the specification in the individual growth models section. For exemple, if you have longitudinal data on students nested within teachers, you can fit a three-level individual growth model with the syntax:


```
proc mixed noclprint covtest;
 class student teacher;
 model mathach = time/solution ddfm=bw;
 random intercept time/type=un sub=teacher;
 random intercept time/type=un sub=student (teacher);
```

Note that we have now specified the option /TYPE=UN on both random statements to ensure that estimation of the variance-covariance matrix is totally unconstrained.

Many other options are available to the user interested in fitting more complex mixed models. Heterogeneity in the error variance-covariance matrix can be introduced using the GROUP option on the RANDOM statement. Sampling-based Bayesian analysis can be conducted using a PRIOR statement that permits a variety of distributional specifications for the variance components parameters' prior density. SAS also provides two macros—GLMMIX and NLINMIX—that can be used for fitting generalized linear mixed models and nonlinear mixed models that do not involve the normal continuous outcomes treated here. Further details concerning all these extensions are found in Littell et al. (1996) and SAS Institute (1996).

PROC MIXED does not substitute for the excellent stand-alone multilevel software programs that are constantly being updated to fit an ever increasing array of models. Its integration into SAS, one of the most widely used integrated statistical packages, is what makes it an attractive option for many users. Because it was not designed with multilevel models in mind, the user seeking to use the program is likely best served by writing out the particular model to be fit and then identifying the appropriate syntax. Experience suggests that proceeding directly to PROC MIXED syntax is likely to produce output that is not what the user intended. But with these caveats in mind, I believe that PROC MIXED represents a valuable addition to the statistical toolkit for fitting multilevel models, hierarchical models, and individual growth models.

Notes

[1] 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. Although many other multilevel programs use the same approach to testing variance components (e.g., MLwiN and MIXREG), SAS has responded to this caution by actually dropping this section of output from the default PROC MIXED specification (in versions 6.12 and higher). That is why we needed to specify the option COVTEST on the PROC MIXED statement. An alternative approach is to compare models using familiar likelihood ratio chi-square tests that compare full and reduced models. Further details on this straightforward alternative are given in SAS Institute (1996) pages 598–599.

[2] There is another way of thinking about these variance components that you should be thinking about whenever you fit an unconditional model. The variance component for schools, here 8.6096, places an effective ceiling on the amount of variation in school means that will ever be explainable by a school level (level-2) factor. By including school

level factors (as we will do in the next section), we hope to reduce the size of this variance component, indicating that we have explained part of the explainable variation.

[3] If you want to compare models with different fixed effects you must specify METHOD=ML and use the IC option. This is because SAS uses Restricted Maximum Likelihood (REML) (also known as residual maximum likelihood) as the default method of estimation. For further discussion of the differences between these methods of estimation, and the consequences of these differences, consult Longford (1993) or Diggle, Liang and Zeger (1994).

[4] For further information about the accuracy of these tests, see Littell et al., 1996, p. 4.

Appendix: Creating SAS data sets for use in PROC MIXED

SAS data sets containing multilevel data can be organized in one of two ways: (a) *multiple record data sets*, in which each level-2 unit has multiple records, one per level-1 unit; and (b) *multiple variable data sets*, in which each level-2 unit has one record and multiple variables are used to record either the multiple occasions of measurement or the multiple members of a group (the level-1 data). In a multiple record file, the data set for 35 people with 4 occasions of measurement would have 140 records, one per person, per occasion. In a multiple variable file, the same data set would have only 35 records; 4 variables would be used to denote the individual's score on each measurement occasion.

To use PROC MIXED, you need a multiple record data set. It must contain all the variables you want to analyze (regardless of the level at which they are measured) at the lowest level possible. In this appendix, I describe how you can create this data set from a variety of existing data arrangements. The example uses the data for indivdual growth modeling analyzed in the text (Willett, 1988). By selecting from the code presented, you should be able to create whatever variables needed for multilevel analysis.

Reading in Multiple Record Files

Most data you will encounter will arrive as a multiple record file. You will have data on multiple teachers within a school, multiple students in a class, multiple children within a family, or multiple observations on individuals over time. Each observation must have an ID that identifies the group (or other level-2 unit) to whom each level-1 record belongs. An example of the level-1 data file for the growth modeling example is shown below. There are four variables: the ID in cols 1–2, the TIME of measurement in col 3, the SCORE in cols 4–6, and the COVAR in cols 7–9.

```
010205 37
011217 37
012268 37
013302 37
020219 23
021243 23
022279 23
023302 23
.......
350166 10
351197 10
352203 10
353233 10
```


The code below provides two ways of reading this multiple record file. Data step one reads it in as a multiple record file; data step two reads it in as a multiple variable file. Although it is most common to read the data in as a multiple record file, it is sometimes convenient to have the data organized as multiple variable file (especially when thinking about creating aggregate variables or computing means for centering). If you decide that this is the case for you, the second data step could be used as well, providing you with a level-2 data file. PROC SUMMARY can also be used in this regard.

Converting From Multiple Record Files to Multiple Variable Files (and vice versa)

Once you have created either type of SAS data set, it is relatively easy to convert from one data structure to the other. Data step three converts a multiple record SAS data set (data=one) into a multiple variable data set. The resultant data set is identical to the data set created in data step two. Note that I have changed the names of the array variables from T and SCORE to TVAR and SCOREVAR because the arrays T and SCORE had already been defined in a previous data step.

Data step four converts a multiple variable data set into a multiple record data set. This data step completes the cycle, enabling you to go from one form to another with ease. The resultant data set (data=four) is identical to data set one. The important idea for PROC MIXED users is that you can easily go from one data form to the other through the careful use of data steps.

Code for manipulating multilevel data sets

***Reading in as a multiple record file*;**

```
data one;
 infile test;
 input id 1-2 t 3 score 4-6 covar 7-9;
```

***Reading in as multiple variable file. *;**

```
data two:
 infile test eof=stop;
 array t[4] t1-t4;
 array score [4] score1-score4;
 do i=1 to 4 while (id=nextid);
  input id 1-2 t[i] 3 score [i] 4-6 covar 7-9;
  input nextid 1-2 @@;
 end;
 drop nextid i;
 stop: output;
```

***Converting a multiple record SAS data set into a multiple variable SAS data set. *;**

```
data three;
 array tvar[4] t1-t4;
 array scorevar [4] score1-score4;
 do i=1 to 4 until (last.id);
  set one;
  by id;
  tvar [i]=t;
  scorevar [i]=score;
 end;
 drop i t score;
```

***Converting a multiple variable data set into a multiple record data set. *;**

```
data four;
 set three;
 array tvar [4] t1-t4;
 array scorevar [4] score1-score4;
 do i=1 to 4;
  t = tvar[i];
  score = scorevar [i];
  output;
 end;
 drop i t1-t4 score1-score4;
```

References

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Bryk, A. S., Raudenbush, S. W., & Congdon, R. T. (1996). *HLM: Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs.* Chicago, IL: Scientific Software International, Inc.

Diggle, P. J., Liang, K-Y., & Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Oxford: Oxford University Press.

Ferron, J. (1997). Moving between hierarchical modeling notations. *Journal of Educational and Behavioral Statistics, 22,* 119–123.

Goldstein, H. (1995). *Multilevel Statistical Models,* 2nd ed. New York: Halstead Press.

Goldstein, H., Healy, M. J. R., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine, 13,* 1643–1655.

Hedeker, D. & Gibbons, R. D. (1996). MIXREG: a computer program for mixed-effects regression analysis with autocorrelated errors. *Computer Methods and Programs in Biomedicine, 49,* 229–252.

Hox, J. J. (1995). *Applied Multilevel Analysis.* Amsterdam: TT-Publikaties.

Hox, J. J., & Kreft, I. G. G. (Eds.). (1994). Multilevel analysis methods [Special issue]. *Sociological Methods and Research, 22*(3).

Kreft, I. G. G. (Ed.). (1995). Hierarchical linear models: Problems and prospects [Special issue]. *Journal of Educational and Behavioral Statistics. 20*(2).

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30*(1), 1–21.

Kreft, I. G. G., de Leeuw, J., & van der Leeden, R. (1994). Review of five multilevel analysis programs: BMDP-5V, GENMOD, HLM, ML3, and VARCL. *The American Statistician, 48,* 324–335.

Latour, D., Latour, K., & Wolfinger, R. D. (1994). *Getting started with PROC MIXED.* Cary, NC: SAS Institute, Inc. L

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS System for Mixed Models.* Cary, NC: SAS Institute, Inc.

Longford, N. T. (1993). *Random coefficient models.* NY: Oxford University Press.

McLean, R. A., Sanders, W. L., & Stroup, W. W. (1991). A unified approach to mixed linear models. *The American Statistician, 45,* 54–64.

Murray, D. M., & Wolfinger, R. D. (1994). Analysis issues in the evaluation of community trials: Progress towards solutions in SAS/STAT MIXED. *Journal of Community Psychology, Special Issue,* 140–154.

Prosser, R., Rasbash, J. & Goldstein, H. (1996). *MLn User's Guide.* London: Institute of Education.

SAS Institute (1992). *SAS Technical Report P–229, SAS/STAT Software: Changes and Enhancements, Cary, NC: Author.*

SAS Institute (1996). *SAS/STAT Software: Changes and Enhancements through Release 6.11,* Cary, NC: Author.

Snijders, T. A. B., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods and Research, 22*(3), 342–363.

VanLeeuwen, D. M. (1997). A note on the covariance structure in a linear model. *The American Statistician, 51*(2), 140–144.

Willett, J. B. (1988). Questions and answers in the measurement of change. In E. Rothkopf (Ed.), *Review of research in education (1988–89) (pp. 345–422). Washington, DC: American Educational Research Association.*

Wolfinger, R. D. (1993). Covariance structure selection in general mixed models. *Communications in Statistics-Simulations, 22*(4), 1079–1106.

Wolfinger, R. D. (1996). Heterogeneous variance-covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics, 1*(2), 205–230.

Author

JUDITH D. SINGER is Professor of Quantitative Methods, Graduate School of Education, Harvard University, Larsen Hall, 7th Floor, Cambridge, MA 02138; judith_singer@harvard.edu. She specializes in methods for the analysis of longitudinal and multilevel data, and the design of research.

Received June 23, 1997
Revision received October 21, 1997
Accepted November 17, 1997