

Paper SD-14

An Intermediate Primer to Estimating Linear Multilevel Models using SAS[®] PROC MIXED

Bethany A. Bell, Whitney Smiley, Mihaela Ene, Phillip R. Sherlock, Jr., Genine L. Blue
University of South Carolina

ABSTRACT

This paper expands upon Bell et al.'s (2013) "A Multilevel Model Primer Using SAS[®] PROC MIXED" in which we presented an overview of estimating two and three-level linear models via PROC MIXED. However, in our earlier paper, we, for the most part, relied on simple options available in PROC MIXED. In this paper, we present a more advanced look at common PROC MIXED options used in the analysis of social and behavioral science data, as well introduce users to two different SAS macros previously developed for use with PROC MIXED: one to examine model fit (MIXED_FIT; Ene, Smiley, & Bell, 2012) and the other to examine distributional assumptions (MIXED_DX; Bell et al., 2010). Specific statistical options presented in the current paper include (a) PROC MIXED statement options for estimating statistical significance of variance estimates (COVTEST, including problems with using this option) and estimation methods (METHOD =), (b) MODEL statement option for degrees of freedom estimation (DDFM =), and (c) RANDOM statement option for specifying the variance/covariance structure to be used (TYPE =). Given the importance of examining model fit, we also present methods for estimating changes in model fit through an illustration of the SAS macro MIXED_FIT. Likewise, the SAS macro MIXED_DX is introduced to remind users to examine distributional assumptions associated with two-level linear models, including normality and homogeneity of level-1 and level-2 residuals. To maintain continuity with the 2013 introductory PROC MIXED paper, thus, providing users with a set of comprehensive guides for estimating multilevel models using PROC MIXED, we use the same real world data sources that we used in our earlier primer paper (Bell et al., 2013).

Keywords: MULTILEVEL MODELING, PROC MIXED, MODEL FIT, STATISTICAL ASSUMPTIONS

INTRODUCTION

Hierarchically organized data are commonplace in educational, clinical, and other research settings. Thus, it is imperative that researchers account for each level of nesting in any data set. Luckily, SAS has several procedures that allow researchers to estimate multilevel models, thus, accounting for the nested structure of the data, when necessary. When dealing with linear outcomes, PROC MIXED can be used to account for numerous levels of nesting within hierarchically organized data. Whereas the MIXED procedure makes estimating these types of models appear simple, as with most procedures in SAS, there are several options that researchers must consider when specifying his/her model. At the 2013 SAS Global Forum in San Francisco, Bell, Ene, Smiley, and Schoeneberger presented the logic behind multilevel models as well as some basic demonstrations on how to use PROC MIXED to estimate two- and three-level organizational models as well as two-level growth models. The live-streamed presentation can be found here: http://www.livestream.com/sasglobalforumpapers/video?clipId=pla_4e1c2a5f-9217-4e2f-af58-63caf9ac29c9&utm_source=library&utm_medium=ui-thumb

Although their paper was quite informative for the beginning PROC MIXED user, their examples did not include details about option selections specified on their PROC MIXED, MODEL, and RANDOM statements. For example, although they show users how to request models be estimated using maximum likelihood through the METHOD = ML option on the PROC MIXED statement, they did not include a detailed discussion of the differences between METHOD = ML and METHOD = REML. Hence, the need for the current paper; by building from the examples in Bell et al. (2013), the current paper presents users detailed discussions and illustrations about (a) PROC MIXED statement options for estimating statistical significance of variance estimates (COVTEST, including problems with using this option) and estimation methods (METHOD =), (b) MODEL statement option for degrees of freedom estimation (DDFM =), and (c) RANDOM statement option for specifying the variance/covariance structure to be used (TYPE =). In addition, given the importance of examining model fit we also present methods for estimating changes in model fit through an illustration of the SAS macro MIXED_FIT (Ene, Smiley, & Bell, 2012). Likewise, the SAS macro MIXED_DX (Bell et al., 2010) is used to show users how to examine distributional assumptions associated with two-level linear models, including normality and homogeneity of level-1 and level-2 residuals.

PROC MIXED STATEMENT OPTIONS

COVTEST OPTION

In the original primer article by Bell, Ene, Smiley, and Schoeneberger (2013) provide examples for estimating a 2-level organization model as well as a 2-level growth model. The syntax for both of these models contained the COVTEST as a statement option such as the one below, which is an option used to generate hypothesis testing output for the variance and covariance components (aka random effects). This is a necessary statement in PROC MIXED if significance testing is to be conducted on the model's random effects (i.e., if a researcher wants to see *p*-values associated with the random effect parameter estimates, this option must be included on the PROC MIXED statement).

As an example, syntax and output from a data set presented in Bell et al. (2013) is provided here. In this example there is a 2-level organization model with students nested within schools. In this example, standardized mathematics achievement (ma_z) is being modeled as a function of students' limited English proficiency (LEP) status and exceptional child (EC) status at level-1 and schools' treatment status (treatment) and the proportion of students receiving free and reduced lunch (schl_frl_c) at level-2. Additionally LEP and EC are provided on the random statement to allow these two student characteristics to randomly vary across schools. Notice COVTEST is included as a statement option.

```
PROC MIXED data=lib.SESUG13 COVTEST noclprint method= ML;
class schoolid;
model ma_z = LEP EC treatment schl_frl_c/solution ddfm = SATTERTHWAITTE;
random intercept LEP EC/ sub=schoolid type=vc;
```

The code above would generate the output in Figure 1. Notice that in addition to the parameter estimate and standard error, using the COVTEST option provides a Z value column as well as a *p*-value associated with that Z value. If this option is not included on the PROC MIXED command, then the Z and *p*-values would not be included in the output.

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	schoolid	0.2831	0.07387	3.83	<.0001
LEP	schoolid	0.05115	0.01611	3.18	0.0007
EC	schoolid	0.08612	0.02588	3.33	0.0004
Residual		0.6738	0.006681	100.85	<.0001

Figure 1. Random effect output using the COVTEST option

Whereas the *p*-values for the variance components are appealing and would suggest that the Intercept, LEP, and EC all significantly vary across schools, interpretation of the results from the COVTEST option can be misleading. Because variance components are known to be skewed and bounded, a sampling distribution that relies on a normal approximation such as the Z value used in the COVTEST option is not acceptable (Singer, 1998). In fact, starting with SAS version 6.12 and higher the COVTEST output was dropped as automatic output. The only exception in which sole use of the COVTEST option is encouraged is in cases in which very large sample sizes are used. However, when used in conjunction with other methods for examining the statistical significance of variance and covariances, the results from the COVTEST option can serve as the starting point for determining the significance of model variance components.

In addition to requesting the COVTEST option with the PROC MIXED statement and examining the resulting *p*-values as a first step for determining the statistical significance of a model's variance components, researchers are encouraged to examine changes in model fit between a model without random slopes and a model with random slopes. Specifically, when models are nested (i.e., model 2 is an extension of model 1) then the likelihood ratio test (LRT) can be used to assess model fit (Snijders & Bosker, 2012). The likelihood ratio test calculates the change in deviance (-2LL) between the smaller model and the larger model and statistically tests the change using a χ^2 distribution, with degrees of freedom equal to the change in number of parameters from the smaller model to the larger model. If the two models are not nested then changes in model fit between the two models needs to be examined using AIC and BIC. Moreover, when examining changes in model fit in which the models differ in their random effects, such as is the case when using changes in model fit to determine the statistical significance of model

variance components, models need to be estimated using restricted maximum likelihood (REML). This is different than the presentation of examining changes in model fit presented in Bell et al. (2013) – in those examples, the models differed in their fixed effects, thus, they had to be estimated using maximum likelihood (ML). More about these estimation methods is presented in the section below.

Thus, the proper way to assess significance of variance components is a two-step process – first, you can examine the hypothesis test results from the COVTEST option in PROC MIXED; if any of the variance or covariance estimates have p-values less than alpha, you proceed with examining changes in model fit using via the LRT (if models are nested) or AIC and BIC (if models are not nested).

Using the same data we used in the example above with the COVTEST option, below is an explanation of how you can examine the significance of the two random slopes through a model building process. Let's assume that we have already gone through the model building process to determine the best fitting model for the fixed effects (see Bell et al., 2013 for details on this approach) and the best fitting fixed effect model is one that predicts the standardized math achievement scores from two student-level variables (LEP and EC) and two school-level variables (treatment and schl_frl_c). Now, we want to determine if LEP and EC significantly vary across schools – i.e., we want to estimate a model that contains random slopes for the two student-level variables. Then, we will compare the model fit of the full model (i.e., the model with random slopes) to the model fit of the reduced model (i.e., the model without random slopes). And, remember, because we are comparing nested models that differ in their variance components, we have to estimate our models using REML and not ML. For this example, I could estimate the reduced and full model and then compare the changes in model fit by hand, but, instead, I use the MIXED_FIT macro (Ene, Smiley, & Bell, 2012) to demonstrate how models can be compared to determine if variance and covariance parameters are statistically significant or not.

MIXED_FIT uses information from ODS tables generated from PROC MIXED to produce numeric output consisting of a table that includes the likelihood ratio test (LRT), change in Akaike Information Criterion (AIC), change in Bayesian Information Criterion (BIC), and change in Pseudo- R^2 (See Ene et al, 2013 for more information on the MIXED_FIT macro, including macro code as well as an overview of nested vs. non-nested model comparison). Since we are using nested models to assess if the two variance components are significant, we will want to look at the LRT and p-value associated with the LRT. Remember, that the results from the COVTEST option alone suggests that both LEP and EC significantly vary across schools, but, given the inherent bias in the Z distribution used to generate those p-values, we now need to verify that inclusion of the random slopes significantly improves model fit.

The SAS syntax and call to the MIXED_FIT macro are below:

```
Title 'Model 1: Level-1 Model Random Intercept Only';
PROC MIXED data=lib.SESUG13 COVTEST noclprint method= REML;
class schoolid;
model ma_z = lep ec treatment schl_frl_c/solution ddfm = SATTERTHWAITE;
random intercept/ sub=schoolid type=vc;
ods output Fitstatistics=FS_Model_2 SolutionF=SF_Model_2;
run;
```

```
Title 'Model 2: Level-1 and Level-2 Model Random Intercept and Slopes';
PROC MIXED data=lib.SESUG13 COVTEST noclprint method= REML;
class schoolid;
model ma_z = lep ec treatment schl_frl_c/solution ddfm = SATTERTHWAITE;
random intercept lep ec/ sub=schoolid type=vc;
ods output Fitstatistics=FS_Model_3 SolutionF=SF_Model_3;
run;
```

```
%Mixed_Fit(fullmodel=FS_Model_3,redmodel=FS_Model_2,DFfull=SF_Model_3,DFred=SF_Model_)
; run;
```

Once the above syntax is executed and the models are estimated, we used the MIXED_FIT macro to obtain the model fit statistics between the model with random slopes and the model without random slopes. More details on using MIXED_FIT are provided in (Ene, Smiley, & Bell, 2012).

Figure 2 contains the output from MIXED_FIT; we can see the p -value for the LRT is greater than .05. Thus, the full model with LEP and EC on the RANDOM statement fits significantly better than the reduced model with a random intercept but no random slopes. These results suggest that the LEP and EC slopes significantly vary across schools. Had the p -value been less than .05, we would conclude that there was not enough evidence to suggest these random slopes significantly varied across schools. The raw output values that were used in MIXED_FIT to calculate the LRT were $-2LL = 50181.3$ for the full model and $-2LL = 50474.2$ for the reduced model.

Change in AIC	Change in BIC	Change in -2LL	mixture method p -value
Reduced-Full	Reduced-Full	Reduced-Full	for nested models
Useful with non-nested models	Useful with non-nested models	for nested models only	for nested models only
288.918	286.050	292.918	0.0000

Figure 2. MIXED_FIT output for testing significance of LEP and EC random slopes

METHOD OPTION

Researchers also need to consider another statement option related to how the model should be estimated. On the statement line, there is also a METHOD = option. In Bell et al.'s (2013) primer paper, we only included the option for maximum likelihood estimation (ML). However, analysts have the option of estimating models in PROC MIXED using restricted maximum likelihood (REML). Which estimation method to use (ML vs. REML) primarily depends on the focus of the research questions being examined. More specifically, the difference between ML and REML estimation occurs in how the variance components are estimated. With ML, the fixed effects and variance components are estimated simultaneously, which has been shown to yield biased variance component estimates. REML, on the other hand, estimates the variance components and fixed effects separately; by first estimating the variance components and then estimating the fixed effects, REML allows the maximization of a likelihood function based just on the variance components, which eliminates the bias problem (Raudenbush & Bryk, 2002; Snijders & Boskers, 2012). When the level-2 sample size is large, the difference in the variance component estimates will be trivial, however, smaller level-2 sample sizes will yield more bias in estimates generated using ML.

Whereas REML sounds like it should be the dominant option, especially since it is the default estimation method in PROC MIXED, ML continues to be used quite frequently given two important advantages that it offers analysts: it is computationally easier and it allows researchers to compare nested models that differ in fixed effects using the LRT. When REML is used to estimate models, the LRT can only be used to compare nested models that differ in variance components. Whereas REML is useful for examining the significance of variance components as illustrated above, it is not useful for the overall model building process that is typically used in multilevel research applications. Thus, which method should be used depends on the substantive interest of the researcher.

If a researcher is only interested in the fixed effects and does not include any random slopes in the model (i.e., the only random effect is the intercept), he/she can proceed with the model building process [see Bell et al. (2013) for more details on this approach], using ML since the nested models would only differ in fixed effects. Or, if a researcher is primarily interested in variance components and does not plan to utilize a model building process to help identify the best fitting model in terms of fixed effects (i.e., a researcher simply estimates a single model containing all fixed effects) then REML would be the preferred estimation method, as it would allow the researcher to examine changes in model fit as variance components are added to the model. When researchers are interested in both fixed and random effects, a two-step process is typically used in which first the best fitting fixed effect model is determined through a model building process and use of the LRT to test for improved model fit – these models would be estimated using ML. Then, once the best fitting and most parsimonious fixed effect model is identified, researchers can re-estimate the model using REML – this newly estimated model will serve as the variance component base or reduced model (i.e., this first REML model would only contain a random intercept). Then, the variance component model building process can be used, with each subsequent model estimated using REML.

Below are two examples of SAS syntax for estimating our two-level model with students nested in schools; once with METHOD = ML and the other with METHOD = REML. Note that if no METHOD = option is specified, the default estimation is REML. Given the size of our sample, differences in the variance components from these two models would be trivial, therefore, no output is provided. Instead, the main focus of this section of the paper has been to not only illustrate how to request the two estimation methods in PROC MIXED, but, also, to provide a summary of the nuances between these methods and how important it is for researchers to select the most appropriate estimation method, given the focus of his/her research study.

```

PROC MIXED data=lib.SESUG13 noclprint method= ML;
class schoolid;
model ma_z = lep ec treatment schl_frl_c/solution ddfm = SATTERTHWAITE;
random intercept lep ec/ sub=schoolid type=vc;

PROC MIXED data=lib.SESUG13 noclprint method= REML;
class schoolid;
model ma_z = lep ec treatment schl_frl_c/solution ddfm = SATTERTHWAITE;
random intercept lep ec/ sub=schoolid type=vc;

```

MODEL STATEMENT OPTIONS

The MODEL statement is used to specify the fixed effects to be included in the model. The DDFM= option allows the user to specify which degrees of freedom (DF) to use when estimating the model. Degrees of freedom provide an approximate sample size based on the number of individuals and predictors at level one and the number of groups and predictors at level 2. This value is necessary for determining the significance of the parameter estimates at each level, which are based on t- distributions. In addition, during the process of model building, degrees of freedom are necessary to use the chi-square distribution which is used to determine the significance of changes to the -2LL values between nested models. Thus, it is important for researchers to carefully consider the different degrees of freedom methods available in PROC MIXED and to select the best method for a given research scenario.

SAS® provides five degrees of freedom methods to use within the PROC MIXED statement: residual (RESIDUAL), between-within (BETWITHIN), containment (CONTAIN), Satterthwaite (SATTERTHWAITE), and Kenward-Roger (KENWARDROGER). Both the Satterthwaite and Kenward-Roger DF methods are well-suited for most multilevel models as they both are intended to be used in models with unbalanced designs and both can handle complex covariance structures. Moreover, the Kenward-Roger DF method also adjusts for small sample bias, so it is ideal to use when estimating models with relatively small samples. Below is a brief summary of each of the five DF methods available in PROC MIXED, followed by examples on how to use the DDFM= option within PROC MIXED for two-level organizational models and two-level growth models.

The RESIDUAL degrees of freedom estimation method was the default option in previous versions of PROC MIXED. This method for specifying DF only gives the correct DFs when the level-1 errors are independent and identically distributed and when there are no level-2 errors (i.e., a situation in which multilevel modeling is not needed. There is no reason to ever use this DF method when using PROC MIXED.

The BETWITHIN degrees of freedom estimation method is more applicable to repeated measures (longitudinal) models than to organizational models. This method works by partitioning the residual degrees of freedom into between-participants and within-participants. Therefore, effects that change within an individual are assigned within-participants DF and effects that do not change within an individual are assigned the between-participants DF.

The CONTAIN degrees of freedom estimation method is the default method in PROC MIXED when a random statement is used. This option can lead to exact degrees of freedom when the design is balanced and when the level-1 errors are independent and identically distributed. This choice becomes more questionable, however, as the design becomes less balanced or when a complex error structure is needed at level-1.

The SATTERTHWAITE method approximates the degrees of freedom, and is designed for use with unbalanced designs and more complex covariance structures. Given the reality of most models being unbalanced, this method is most applicable for researchers working with hierarchical data.

The KENWARDROGER method also approximates the degrees of freedom, and again is designed for use with unbalanced designs and complex covariance structures. By adjusting the covariance matrix, this method also adjusts for small-sample bias. Based on previous research that shows that using the Kenward-Roger DF with small sample sizes yields unbiased parameter and standard error estimates, we recommend this method be used for models containing less than 30 level-1 units and 30 level-2 units (Bell, Morgan, Schoeneberger, Kromrey, & Ferron, 2012).

INVESTIGATING DDFM OPTIONS FOR TWO-LEVEL ORGANIZATIONAL MODELS

We will use the same two-level organizational data and syntax as we used in the previous organizational model example – students nested in schools, predicting math achievement from two student-level variables and two school-level variables. We will not present an organizational model using the BETWEEN DDFM as it is not an appropriate method for organizational models.

The PROC MIXED syntax for using the RESIDUAL degrees of freedom method is below followed by the Fixed Effects output in Figure 3.

```
PROC MIXED data=lib.SESUG13 COVTEST noclprint;
class schoolid;
model ma_z = lep ec treatment schl_frl_c/solution ddfm = RESIDUAL;
random intercept lep ec/ sub=schoolid type=vc;
run;
```

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	0.02681	0.07254	2E4	0.37	0.7117
lep	-0.09375	0.04513	2E4	-2.08	0.0378
ec	-0.4432	0.05842	2E4	-7.59	<.0001
treatment	0.09139	0.1864	2E4	0.49	0.6239
schl_frl_c	-1.8327	0.3359	2E4	-5.46	<.0001

Figure 3. Fixed effect output using DDFM = RESIDUAL

The syntax for using the CONTAIN degrees of freedom method is below followed by the Fixed Effects output in Figure 4.

```
PROC MIXED data=lib.SESUG13 COVTEST noclprint;
class schoolid;
model ma_z = lep ec treatment schl_frl_c/solution ddfm = CONTAIN;
random intercept lep ec/ sub=schoolid type=vc;
run;
```

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	0.02681	0.07254	29	0.37	0.7144
lep	-0.09375	0.04513	30	-2.08	0.0464
ec	-0.4432	0.05842	30	-7.59	<.0001
treatment	0.09139	0.1864	2E4	0.49	0.6239
schl_frl_c	-1.8327	0.3359	2E4	-5.46	<.0001

Figure 4. Fixed effect output using DDFM = CONTAIN

The syntax for using the SATTERTHWAITE degrees of freedom method is below followed by the Fixed Effects output in Figure 5.

```
PROC MIXED data=lib.SESUG13 COVTEST noclprint;
class schoolid;
model ma_z = lep ec treatment schl_frl_c/solution ddfm = SATTERTHWAITE;
random intercept lep ec/ sub=schoolid type=vc;
run;.
```

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	0.02681	0.07254	26	0.37	0.7147
lep	-0.09375	0.04513	28.7	-2.08	0.0468
ec	-0.4432	0.05842	29.9	-7.59	<.0001
treatment	0.09139	0.1864	25.9	0.49	0.6280
schl_frl_c	-1.8327	0.3359	25.7	-5.46	<.0001

Figure 5. Fixed effect output using DDFM = SATTERTHWAITE

The syntax for using the KENWARDROGER degrees of freedom method is below followed by the Fixed Effects output in Figure 6.

```
PROC MIXED data=lib.SESUG13 COVTEST noclprint;
class schoolid;
model ma_z = lep ec treatment schl_frl_c/solution ddfm = KENWARDROGER;
random intercept lep ec/ sub=schoolid type=vc;
run;
```

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	0.02681	0.07255	26	0.37	0.7148
lep	-0.09375	0.04518	28.7	-2.07	0.0471
ec	-0.4432	0.05848	29.9	-7.58	<.0001
treatment	0.09139	0.1864	25.9	0.49	0.6280
schl_frl_c	-1.8327	0.3359	25.7	-5.46	<.0001

Figure 6. Fixed effect output using DDFM = KENWARDROGER

Table 1 compares the degrees of freedom and t-statistic results, respectively, from the four models presented above using the different DDFM= options within the MODEL statement of the PROC MIXED command. The parameter estimates and standard errors were identical for each of the methods used, thus, the *t*-values were also identical across models. However, the degrees of freedom changed between options, resulting in varying *p*-values. Although the differences across DF methods in our example are not very large and the results of the hypothesis test for each fixed effect are the same for each DF method, as sample sizes become smaller and more unbalanced, the variation in these values will be more pronounced. Again, given the mathematical advantages and empirical evidence of its good performance with small sample sizes (Bell et al., 2012), we recommend users consider using the Kenward-Roger DF estimation method as their default when estimating linear models in PROC MIXED.

Table 1. Comparison of Resulting Statistics from Two-level Organizational Model with four DDFM= options

	RESIDUAL	CONTAIN	SATTERTHWAITE	KENWARDROGER
Intercept	.37(2E4), <i>p</i> =.7177	.37(29), <i>p</i> =.7144	.37(26), <i>p</i> =.7147	.37(26), <i>p</i> =.7148
lep	-2.08(2E4), <i>p</i> =.0378	-2.08(30), <i>p</i> =.0464	-2.08(28.7), <i>p</i> =.0468	-2.07(28.7), <i>p</i> =.0471
ec	-7.59(2E4), <i>p</i> <.0001	-7.59(30), <i>p</i> <.0001	-7.59(29.9), <i>p</i> <.0001	-7.58(29.9), <i>p</i> <.0001
treatment	.49(2E4), <i>p</i> =.6239	.49(2E4), <i>p</i> =.6239	.49(25.9), <i>p</i> =.6280	.49(25.9), <i>p</i> =.6280
schl_frl_c	-5.46(2E4), <i>p</i> <.0001	-5.46(2E4), <i>p</i> <.0001	-5.46(25.7), <i>p</i> <.0001	-5.46(25.7), <i>p</i> <.0001

INVESTIGATING DDFM OPTIONS FOR TWO-LEVEL GROWTH MODELS

The example two-level growth model that will be used to examine the degrees of freedom methods comes from Bell et al.'s (2013) SAS Global Forum presentation "A Multilevel Model Primer Using SAS® PROC MIXED." This model investigates how children's reading achievement changes throughout the first years of school and if a child's sex is related to reading achievement in kindergarten. Using a sample of 120 children pulled from the Early Childhood Longitudinal Study-Kindergarten (ECLS-K; Early Childhood Longitudinal Study – Kindergarten, n.d.), this model has time (level-1) nested within students (level-2) and specifies a child's sex as a binary level-2 predictor variable. Given that the BETWITHIN degrees of freedom method are designed for repeated measures modeling, for this section, we do not present examples of all five DF options. Instead, we have included examples of using the BETWITHIN option and the KENWARDROGER option, given that it tends to yield reliable estimates across a variety of sample conditions.

The syntax for using the BETWITHIN degrees of freedom method is below followed by the Fixed Effects output in Figure 7.

```
proc mixed COVTEST noclprint data = lib.eclsDDFM1 method=ml;
class childid;
model reading = sex timemos/solution ddfm = BETWITHIN;
random intercept timemos/sub=childid type=VC;
RUN;
```

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	18.6770	1.1976	118	15.59	<.0001
sex	2.1814	1.6650	118	1.31	0.1927
TIMEMOS	1.5837	0.04337	355	36.52	<.0001

Figure 7. Fixed effect output using DDFM = BETWITHIN

The syntax for using the KENWARDROGER degrees of freedom method is below followed by the Fixed Effects output in Figure 8.

```
proc mixed COVTEST noclprint data = lib.eclsDDFM1 method=ml;
class childid;
model reading = sex timemos/solution ddfm = KENWARDROGER;
random intercept timemos/sub=childid type=VC;
RUN;
```

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	18.6770	1.1985	123	15.58	<.0001
sex	2.1814	1.6675	117	1.31	0.1934
TIMEMOS	1.5837	0.04337	127	36.51	<.0001

Figure 8. Fixed effect output using DDFM = KENWARDROGER

Table 2 compares the results generated for the t -statistic, degrees of freedom, and p -value using the five DDFM= options within the MODEL statement of the PROC MIXED command. The parameter estimates and standard errors were identical for the two methods used, thus, the t -values were also identical across models. However, the degrees of freedom changed between options. As shown in Table 2, the main difference between the BETWITHIN and KENWARDROGER degrees of freedom methods occurs in how the DF are estimated at level-1. This example clearly shows how the each time point in the growth model is accounted for when using the BETWITHIN DF method (i.e., instead of "seeing" the sample as 120 observations the BETWITHIN method considers each time point for each student as an observation). Obviously, with a large t -value of 36.52, both DF estimates will still yield p -values less

than .05, however, if the sample was smaller and a smaller difference existed between time points, it is plausible that the results of the hypothesis test could differ between these two methods.

Table 2. Comparison of Resulting Statistics from Two-level Growth Model with two DDFM= options

	BETWITHIN	KENWARDROGER
Intercept	15.59(118), <i>p</i> <.0001	15.59(123), <i>p</i> <.0001
sex	1.31(118), <i>p</i> =.1927	1.31(117), <i>p</i> =.1927
TIMEMOS	36.52(355), <i>p</i> <.0001	36.52(127), <i>p</i> <.0001

RANDOM STATEMENT OPTIONS

The RANDOM statement in PROC MIXED is used to specify what parameters are allowed to vary across level-2 units. As explained in Bell et al. (2013), this statement is the key element of what differentiates multilevel linear models from single-level OLS models. Specifically, this is the part of the PROC MIXED code where users specify for the intercept to be modeled as varying across level-2 units. Likewise, this is also the statement in PROC MIXED where users specify, which, if any, level-1 variables have random slopes that vary across level-2 units. An important element of the RANDOM statement is specifying what covariance structure SAS should use when estimating the G (i.e., the covariance matrix of level-2 errors) and R (i.e., the covariance matrix of level-1 errors) matrices. This is accomplished through the TYPE = option. Although there are oodles of variance/covariance matrices that can be specified through the TYPE = option, in this section we present three commonly used options: VC, UN, and AR(1).

TYPE=VC (variance components) is the default structure for both G and R covariance structures in PROC MIXED. This structure is relatively simple in terms of estimation; covariance matrix of level-2 errors (G-matrix) is estimated to have separate variances but no covariance and the covariance matrix of level-1 errors (R-matrix) are modeled to have a common variance with no covariance. When the VC structure is used, the number of variance components that are estimated equals the number of variables included on the RANDOM statement plus the variance component of level-1 errors (i.e., the residual variance component).

Below is the SAS syntax and covariance parameter output for our two-level organizational model example using TYPE = VC to estimate the covariance structures.

```
PROC MIXED data=lib.SESUG13 COVTEST noclprint method= REML;
class schoolid;
model ma_z = lep ec treatment schl_frl_c/solution ddfm = SATTERTHWAIT;
random intercept lep ec/ sub=schoolid type=vc;
run;
```

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	schoolid	0.08277	0.02351	3.52	0.0002
lep	schoolid	0.05120	0.01608	3.18	0.0007
ec	schoolid	0.08698	0.02611	3.33	0.0004
Residual		0.6738	0.006681	100.85	<.0001

Figure 9. Covariance output using TYPE = VC

TYPE = UN (unstructured) specifies a general covariance matrix. With the unstructured estimation, the G-matrix is estimated to contain different variances for the intercept and slope(s), as well as covariances between these parameters. In generating a more complex G-matrix than TYPE = VC, the unstructured covariance matrix also estimates many more parameters than the variance component option. Specifically, TYPE = UN will generate [t

$(t+1)/2]$ variance and covariance parameter estimates where t = the number of variables listed on the RANDOM statement. In addition, the covariance output will also contain the variance of the level-1 errors (i.e., the residual variance estimate) that is automatically generated. Thus, as shown in the example below, simply estimating a random intercept and two random slopes will yield seven variance estimates in our two-level organizational model.

The SAS syntax and corresponding covariance parameter output for our two-level organizational model example using TYPE = UN to estimate the covariance structures is below.

```
PROC MIXED data=lib.SESUG13 COVTEST noclprint method= REML;
class schoolid;
model ma_z = lep ec treatment schl_frl_c/solution ddfm = SATTERTHWAITE;
random intercept lep ec/ sub=schoolid type=un;
run;
```

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	schoolid	0.09781	0.03907	2.50	0.0062
UN(2,1)	schoolid	-0.02182	0.02294	-0.95	0.3414
UN(2,2)	schoolid	0.05011	0.01571	3.19	0.0007
UN(3,1)	schoolid	-0.04213	0.04437	-0.95	0.3423
UN(3,2)	schoolid	0.03205	0.01549	2.07	0.0386
UN(3,3)	schoolid	0.08886	0.02782	3.19	0.0007
Residual		0.6738	0.006681	100.85	<.0001

Figure 10. Covariance output using TYPE = UN

At first, the interpretation of the TYPE = UN output can seem a little confusing, but, it follows a very logical pattern when you think of the covariance parameters as they appear in the actual G matrix. Below is an example of a generic unstructured covariance matrix.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

As you can see, there are four estimated variances (i.e., the four elements on the diagonal of the matrix) and six covariances (i.e., the elements on the off-diagonal). Thinking of the SAS output above in Figure 10 from matrix perspective, one can better understand that the parameter labeled “UN (1,1)” is the variance for the element in the unstructured matrix that is located in the first row and first column; in our example, this is the variance of the intercept as it is the first variable listed on the RANDOM statement. Likewise, the parameters labeled “UN(2,2)” and “UN(3,3)” are the variances for the elements in the second row, second column of the matrix and the third row and third column of the matrix, respectively. In our example, these represent the random slopes for the variables LEP and EC, respectively. Following this logic, the parameter labeled “UN(2,1)” is the element in the second row, first column of the matrix – otherwise interpreted as the covariance between the first and second elements in the diagonal. In our example, UN(2,1) represents the covariance between the intercept and the LEP random slope. In addition, UN (3,1) is the covariance between the intercept and EC random slope, and UN(3,2) is the covariance between the LEP and EC random slopes.

Another important aspect of the unstructured covariance matrix is that during the estimation process the variance components are constrained to be nonnegative but the covariances are not. This is important to understand because when this more complex variance structure is used, it is common that one or more of the variance or covariance parameters will be zero or for the covariances a negative value. When this happens, SAS provides users a note in the log stating that the estimated g-matrix is not positive definite. When this happens, users should look at the

covariance output to determine which element of the matrix could not be estimated and then remove that parameter and re-estimate the model.

TYPE = AR(1) is a third variance structure option that is commonly used with repeated measures models (i.e., time nested within subjects). This option specifies a first-order autoregressive structure with homogenous variances. Within this structure, correlations in the matrix decline exponentially with distance. Specifically, within this structure option, the variability in the outcome being measured is the same, regardless of what time it is measured, yet, measurements that are adjacent to each other in time are going to have a strong correlation but as measurements get farther apart, they are less correlated (i.e., the correlation between time 1 and time 2 will be stronger than the correlation between time 1 and time 4).

The SAS syntax and corresponding covariance parameter output for our two-level growth model example using TYPE = AR(1) to estimate the covariance structures is below.

```
proc mixed COVTEST noclprint data = lib.eclsDDFM1 method=ml;
class childid;
model reading = sex timemos/solution ddfm = BETWITHIN;
random intercept timemos/sub=childid type=AR(1);
RUN;
```

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
Variance	CHILDDID	0.4150	0.06265	6.62	<.0001
AR(1)	CHILDDID	1.0000	0	.	.
Residual		47.4119	3.5542	13.34	<.0001

Figure 11. Covariance output using TYPE = AR(1)

Based on the results from this example, the reading data used in our growth curve example does not appear to fit an autoregressive covariance structure. As shown in Figure 11, autoregressive variance component could not be estimated. This is an example of the obtaining a non-positive definite G matrix. If this were to occur in a real research scenario, we recommend users specify their growth model using a simpler covariance structure, such as TYPE = VC.

REMEMBERING TO EXAMINE DISTRIBUTIONAL ASSUMPTIONS

As with any other statistical procedures, it is important for researchers to examine and report the appropriate assumptions when using multilevel modeling techniques. Under multilevel analyses, like with most parametric statistical procedures, there are certain distributional assumptions underlying the validity of the Type I error control. Therefore, the validity of inferences based on models estimated through these techniques depends on the degree to which assumptions are upheld about the structural and random parts of the model (Raudenbush & Bryk, 2002).

The assumptions associated with two-level linear models are similar to OLS model assumptions and include residual normality, independence, and homoscedasticity. With two-level linear models, these assumptions need to be examined for both level-1 and level-2 variables. In particular, the assumption of normally distributed errors must be made for both level-1 and level-2 variables, with violations adversely affecting level-2 estimated standard errors and inferential statistics (Raudenbush & Bryk, 2002). Also, the effects of the violation at level-1 may include distorted random effect coefficients and variance-covariance components. To examine the distribution of the errors, data analysts should consider the following options. First, one might plot the standardized residuals against their normal scores to observe how closely the plot follows a diagonal line (Hox, 2010). Second, to assess the distribution of the overall residuals at levels one and two, one might also use histograms or box-and-whisker plots. Similarly, box-and-whisker plots of the level-1 residuals for each level-2 unit and of the level-2 residuals for each level-2 effect will help assess deviations from a normal distribution and identify extreme values. Third, to assess assumptions of normality, linearity, and homogeneity of variance simultaneously, analysts should examine the plot of predicted values against the residuals. This should be done for level-1 residuals, as well as the residuals for each level-2 effect. Scatterplots that contain roughly equivalent frequencies of points above and below their mean value, with no particular structure, provide evidence that the assumptions have not been violated (Hox, 2010).

To supplement the visual level-1 assumption diagnostics, summary statistics (i.e., skewness, kurtosis, variance and standard deviation, and a statistical test for normality such as the Shapiro-Wilk or Kolmogorov-Smirnov test) for level-1 residuals within each level-2 unit should also be reviewed for information regarding level-1 normality and variance homogeneity. In addition, Levene's test can be used to assess homogeneity of variance of the level-1 residuals across each level-2 unit. Next, when evaluating the tenability of assumptions for the level-2 residuals, multivariate summary statistics such as skewness, kurtosis, and Mahalanobis distances should be generated and reviewed. Multivariate normality can also be evaluated visually through a histogram of Mahalanobis distances. Also, as with the level-1 residuals, Levene's test for assessing homogeneity of variance of the level-2 residuals can be conducted for each level-2 effect.

In addition to examining these assumptions, detection of influential observations is also important when conducting multilevel analyses. Littell, Milliken, Stroup, and Wolfinger (2006) recommend a "drill-down" approach to mixed model influence diagnostics, beginning with a global assessment of the influence on the overall model, followed by a more detailed exploration of the case-sets should they be warranted. The likelihood distance (Cook & Weisberg, 1982) provides an assessment of a unit's influence on the overall model. A group or unit's influence on parameter estimates can be determined using Cook's *D* or the multivariate DFFITS statistics. The larger the value for these statistics, the greater the influence a unit has on parameter estimates (i.e., the change in the parameter estimate is large relative to the variability of the estimate; Schabenberger, 2004). Researchers can also examine a unit's effect of the precision of an estimate through the covariance trace (COVTRACE) and covariance ratio (COVRATIO) statistics. For these statistics, benchmark values used to determine a unit's level of influence are zero and one, respectively (Schabenberger, 2004). Each of these influence diagnostics can be generated for both fixed effects and covariance parameters, however calculation for the latter requires an iterative process due to the potential impact of observations on covariance matrices. The INFLUENCE option within PROC MIXED allows the researcher to utilize non-iterative or iterative diagnostics, and provides the option to control the number of iterations when re-calculating estimates and covariance matrices (Littell et al., 2006).

In addition to examining a unit's influence on the change in parameter estimates and the change in the precision of estimates, influence on fitted and predicted values can also be examined through the Predicted Residual Sum of Squares (PRESS) statistic. The PRESS provides a comparison of the predicted marginal mean and the observed mean when the predicted value is calculated without the deleted observation in question (Schabenberger, 2004). Again, larger values suggest more influential units. Other measures of overall influence provided through the INFLUENCE option in PROC MIXED include RMSE and Restricted Likelihood Distance (RLD). RMSE values represent the model RMSE with a specific level-2 unit deleted. On the other hand, RLD functions more as an 'indicator' statistic in that it does not convey what part of the model is likely to change given the removal of a specific level-2 unit. Instead, it suggests that the overall influence of a particular level-2 unit stands out comparatively to other level-2 units (Littell et al., 2006). To determine what model components are influenced by a given level-2 unit, analysts need to examine the individual influence statistics such as MDFFITS or COVRATIO.

In an effort to make the process of checking the assumptions for multilevel models easier for the applied researcher, the MIXED_DX macro (Bell et al., 2010) provides analysts with a comprehensive approach for conducting two-level linear model diagnostics, including examinations of residual normality, linearity, homogeneity of variance, and influential outliers. By utilizing data from PROC MIXED ODS tables, the macro produces both visual output (e.g., box-and-whisker plots, histograms, scatter plots) and summary tables of statistical output for both level-1 and level-2 residuals. A unique feature of this macro is the calculation of level-2 predicted values for each level-2 effect which allows analysts to create and examine residual by predicted scatterplots for level-2 effects (something that cannot be done using the default SAS output alone). Furthermore, the MIXED_DX macro generates output for all level-2 units, with an option for a user-defined minimum of cases per level-2 unit to be included in these analyses. In addition, MIXED_DX facilitates detection of influential observations in a two-level linear model by creating a ranked summary table of the influence statistics automatically created in the SAS ODS influence table (Influence). We encourage users to read more about assumptions and how to test them in our MIXED_DX paper, Bell et al. (2010). A copy of the macro as well as a link to the macro itself is also provided in Bell et al. (2010).

CONCLUSIONS

This paper provides researchers with an extensive introduction to many of the most commonly used options and statements when estimating a variety of multilevel linear models via PROC MIXED. In addition, we introduced users to two macros that were developed to help researchers navigate the more complicated aspects of estimating multilevel models in PROC MIXED, including testing the distributional assumptions, which unfortunately is often overlooked in the published research. In writing this paper, our goal was to educate users that simply relying on the default options when using PROC MIXED is not always the best decision, especially when it comes to model estimation methods (ML vs. REML) as well as when estimating the model degrees of freedom. It is our hope that after reading this paper, users will realize the importance of becoming familiar with options available in PROC

MIXED; doing so will help ensure that the models being estimated are aligned with the research questions being examined and are correctly specified given characteristics of the data being examined.

REFERENCES

- Bell, B.A., Ene, M., Smiley, W., & Shonenberger, J.A. (2013). A multilevel primer using SAS® PROC MIXED. SAS Global Forum 2013 Proceedings. <http://support.sas.com/resources/papers/proceedings13/433-2013.pdf>
- Bell, B.A., Morgan, G.B., Schoeneberger, J. A., Kromrey, J.D., & Ferron, J.M. (2012; Available via Advanced Articles). How low can you go? An investigation of the influence of sample size and model complexity on point and interval estimates in two-level linear models. *Methodology*.
- Bell, B.A., Schoeneberger, J. A., Morgan, G.B., Kromrey, J.D., & Ferron, J.M. (2010). Fundamental diagnostics for two-level mixed models: The SAS® macro MIXED_DX. SAS Global Forum 2010 Proceedings. <http://support.sas.com/resources/papers/proceedings10/201-2010.pdf>
- Cook, R.D. & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall.
- Early Childhood Longitudinal Study - Kindergarten [ECLS-K]*. (n.d.). Retrieved from the National Center for Educational Statistics website, <http://nces.ed.gov/ecls/>
- Ene, M., Smiley, W., & Bell, B.A. (2013). MIXED_FIT: A SAS® macro to assess model fit and adequacy for two-level linear models. SAS Global Forum 2013 Proceedings. <http://support.sas.com/resources/papers/proceedings13/255-2013.pdf>
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R., D. (2006). *SAS for mixed models* (2nd ed.). Cary, NC: SAS Institute Inc.
- Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Schabenberger, O. (2004). *Mixed model influence diagnostics*. Proceedings of the 29th Annual SAS Users Group International Conference.
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 23 (4), 323-355.
- Snijders, T. A. B., Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advances multilevel modeling*. 2nd edn. Thousand Oaks, CA: Sage.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Bethany A. Bell
University of South Carolina
College of Education, Wardlaw 133
820 Main Street
Columbia, SC 29208
Work Phone: 803-777-2387
E-mail: babel@sc.edu
Web: <http://www.ed.sc.edu/bell/>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.