ON USING PROC MIXED FOR LONGITUDINAL DATA

Walter W. Stroup

ON USING PROC MIXED FOR LONGITUDINAL DATA

Walter W. Stroup
Department of Biometry, University of Nebraska, Lincoln, NE 68583-0712

Abstract

PROC MIXED has become a standard tool for analyzing repeated measures data. Its popularity results from a wide choice of correlated error models compared to other software, e.g. PROC GLM. However, PROC MIXED's versatility comes at a price. Users must take care. Problems may result from MIXED defaults. These include: questionable criteria for selecting correlated error models; starting values that may impede REML estimation of covariance components; and biased standard errors and test statistics. Problems may be induced by inadequate design. This paper is a survey of current knowledge about mixed model methods for repeated measures. Examples are presented using PROC MIXED to demonstrate these problems and ways to address them.

Key Words: Repeated measures experiment, mixed model analysis, correlated error models

1. Introduction

Longitudinal data, also known as data from repeated measures designs, are common in research throughout most agricultural disciplines. Data analysts use several methods to analyze longitudinal data. Two of the most common are multivariate analysis of variance (MANOVA) and univariate linear models. MANOVA allows for correlated errors among repeated measures, but MANOVA is beyond the level of statistical training of most biological researchers, its assumed correlation model is inefficient, and its handling of missing data is even more so. Univariate linear models are far more accessible to biological researchers and much better at handling missing data. However, until recently, univariate linear models, as implemented by major statistical software packages (e.g. SAS PROC GLM) have been limited in their ability to handle correlated errors,. They have generally relied on assumptions that kept computing simple, but were not necessarily realistic. In the 1990's, comprehensive mixed model software has been introduced, notably SAS PROC MIXED. Because PROC MIXED makes accessible a comprehensive array of correlated error models, it has become a standard tool for analyzing longitudinal data.

However, PROC MIXED's versatility come at a price. The purpose of this paper is to review the underlying theory behind PROC MIXED's analysis of repeated measures data, and, more importantly, to review issues of which users of PROC MIXED should beware and how to cope with them. This paper is divided into three parts. Section 3 reviews what is meant by longitudinal data and repeated measures designs. Section 4 reviews relevant linear mixed model theory and problems associated with applying it to longitudinal data. Section 5 is an illustration using a PROC MIXED analysis of a hypothetical data set.

2. Repeated Measures Design and Model Background

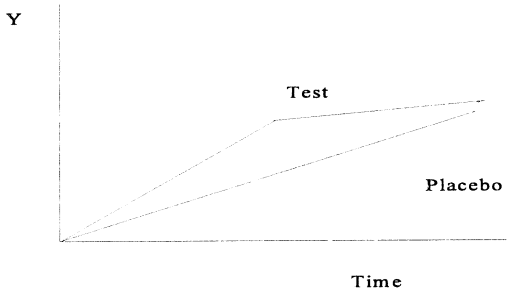
For the purposes of this paper a repeated measures design is understood to have the following features:

- There are 2 or more treatments. Let t (≥ 2) denote the number of treatments.

- Experimental units (subjects) are randomly assigned to each treatment. The number of subjects per treatment, denoted n_i, i=1, 2,..., t, need not be equal. Subjects may be assigned to treatments using any reasonable design, e.g. completely randomized, randomized complete or incomplete block, row-column designs such as Latin Squares, etc. "Reasonable" depends on the context of the particular experiment. To keep things simple, this paper presents examples using completely randomized designs.

- Each subject is observed at each of K times. The times are typically regularly spaced. They need not be equally spaced. Often, their timing reflects the biology of the subjects under study, e.g. growth stage.

The data of interest from repeated measures studies can usually be presented in graphical form. The following is a typical graph. In this example, 2 treatments,"test" and "placebo" are compared.



Inference typically focuses on the following effects:

- Treatment x time interaction
  Are changes over time the same (parallel on the graph) for all treatments?

- Time effects
  Assuming negligible treatment x time interaction, how does the mean response change over time? If non-negligible treatment x time interaction exists, there may be interest in how response changes over time for each treatment.

- Treatment effects
  How do the mean responses to treatments differ? Again, this may be averaged over all times

assuming negligible treatment x time interaction, or specific to each time, otherwise.

Assuming experimental units are assigned to treatments using a completely randomized design, the model equation is:

$$y_{ijk} = \mu_{ik} + s_{ij} + e_{ijk} \qquad (2.1)$$

where

y_{ijk} = observation on j^{th} subject, or experimental unit, on i^{th} treatment at k^{th} time

μ_{ik} = mean of i^{th} treatment at k^{th} time, often expanded in effects form as

$\mu + \alpha_i + \tau_k + \gamma_{ik}$,

where μ, α_i, τ_k, and γ_{ik} are the intercept, treatment main effect, time main effect, and treatment x time interaction, respectively

s_{ij} = ij^{th} subject effect ~ i.i.d. $N(0,\sigma_S^2)$

e_{ijk} = random error, i.e. random variation among repeated measurements on each subject

If a more complex design is used to assign subjects to treatments, it is reflected in the model, e.g. by adding a block effect if a randomized block design is used.

Superficially, equation (2.1) resembles the model for a split-plot experiment. However, a split-plot assumes random assignment of split-plot experimental units, and hence independent e_{ijk}'s, typically i.i.d. $N(0,\sigma^2)$. Because repeated measurements cannot be randomized, the e_{ijk}'s in a repeated measures experiment are at least potentially correlated. Denote $\mathbf{e_{ij}}$ as the vector of *errors* for the repeated measurements on the ij^{th} subject, that is, $\mathbf{e_{ij}}' = (e_{ij1}, e_{ij2,} \ldots e_{ijK})$.

The vector $\mathbf{e_{ij}}$ is assumed to be distributed MVN$(0,\Sigma)$, where Σ models the correlation among the e_{ijk}'s. Also, $\mathbf{e_{ij}}$ is assumed to be independent of $\mathbf{s_{ij}}$. Thus, the *observations*, $\mathbf{y_{ij}}' = (y_{ij1}, y_{ij2,} \ldots y_{ijK})$ ~ MVN$(\boldsymbol{\mu}_i , J\sigma_S^2 + \Sigma)$, where $\boldsymbol{\mu}_i'$ is the vector of means at the K times for the i^{th} treatment, i.e. $\boldsymbol{\mu}_i' = (\mu_{i1}, \mu_{i2}, \ldots , \mu_{iK})$. The y_{ijk}'s are thus distributed multivariate normal:

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{t1} \\ y_{12} \\ \vdots \\ y_{t2} \\ \vdots \end{pmatrix} \sim MVN \left(\begin{bmatrix} \mu_1 \\ \vdots \\ \mu_t \\ \mu_1 \\ \vdots \\ \mu_t \\ \vdots \end{bmatrix}, \begin{bmatrix} \sigma_s^2 J + \Sigma & \ldots & 0 & 0 & \ldots & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & \sigma_s^2 J + \Sigma & 0 & \ldots & 0 & \ldots \\ 0 & \ldots & 0 & \sigma_s^2 J + \Sigma & \ldots & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & \sigma_s^2 J + \Sigma & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \right)$$

The covariance matrix, Σ, can reflect various models of correlation among the e_{ijk}'s. Among the more common are:

- Independent Errors, $\Sigma = I\sigma^2$.

- Compound Symmetry (CS),

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & . & . & \rho \\ \rho & 1 & \rho & \rho & . & \rho \\ . & . & . & . & . & . \\ \rho & . & . & . & \rho & 1 \end{bmatrix}$$

where ρ= correlation between observations on the ij^{th} subject. Note that the compound symmetry model can be reexpressed as $\Sigma = J\sigma_S^2 + I\sigma^2$, and hence the correlation among repeated measurements is the same as the interclass correlation $\frac{\sigma_s^2}{\sigma^2 + \sigma_s^2}$ in the independent errors model. Thus, σ_S^2 in the independent errors model and ρ in the compound symmetry model are confounded; the two are actually equivalent expressions of the same model.

- First-order Autoregressive [AR(1)]

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & . & . & \rho^{k-1} \\ \rho & 1 & \rho & \rho^2 & . & \rho^{k-2} \\ . & . & . & . & . & . \\ \rho^{k-1} & . & . & . & \rho & 1 \end{bmatrix}$$

AR(1) assumes correlation between errors (e_{ijk}) *w* time periods apart is ρ^w.

- Toeplitz (TOEP)

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & . & . & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \rho_2 & . & \rho_{k-2} \\ . & . & . & . & . & . \\ \rho_{k-1} & . & . & . & \rho_1 & 1 \end{bmatrix}$$

- First-order Antedependence [ANTE(1)]

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 & . & . & \sigma_1\sigma_k\rho_1\rho_2\ldots\rho_{k-1} \\ . & \sigma_2^2 & \sigma_2\sigma_3\rho_2 & \sigma_2\sigma_4\rho_2\rho_3 & . & \sigma_2\sigma_k\rho_1\rho_2\ldots\rho_{k-2} \\ . & . & . & . & . & . \\ . & . & . & . & \sigma_{k-1}\sigma_k\rho_{k-1} & \sigma_k^2 \end{bmatrix}$$

- Unstructured (UN)
  $\Sigma = [\sigma_{ij}]$, where $\sigma_{ii} = \sigma_i^2$. Note that the unstructured covariance model is conceptually similar to the correlation structure assumed when one uses MANOVA to analyze longitudinal data, the distinction being that in MANOVA, the covariance matrix among the *observations*, y_{ijk}, is unstructured, whereas with $\Sigma=[\sigma_{ij}]$ the covariance among the *errors* is unstructured.

Note that ANTE(1) and UN allow for heterogeneous variances at each time of observation. There are modifications of CS, AR(1), and TOEP that also allow for heterogeneity at each time. Equally spaced times of observation are implicitly assumed for AR(1) and TOEP, whereas CS, ANTE(1), and UN allow for unequal spacing. See *SAS/STAT Software: Changes and Enhancements though Release 6.12* (SAS Institute, 1997) for a complete list of available covariance models.

With the flexibility PROC MIXED allows in choosing correlated error models, the following questions arise:

- Does the choice of correlated error, or covariance model matter? If so, how? Are type I error rates affected? Estimates of treatment/time effects? Standard errors of estimates?

- Assuming choice of covariance model *does* matter, how does one choose the "correct" model?

- Once an appropriate covariance model *is chosen*, how accurately does PROC MIXED compute test-statistics, degrees of freedom, standard errors, p-values, etc.

3. Linear Mixed Model Results

a. Model, estimation, and inference

The repeated measures model equation (2.1) is a special case of the mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \mathbf{e}$$

where $\begin{bmatrix} U \\ e \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}\right)$. Therefore, $\mathbf{y} \sim$ MVN(Xβ, V), where V=ZGZ'+R.

In the repeated measures model, X describes the treatment-time design, β is the vector of fixed treatment-time effects, Z describes the subjects design, **u** is the vector of random subject effects and **e** is the vector of random errors. Consistent with (2.1), $G=I\sigma_S^2$ and R is block diagonal, with each block equal to Σ, the within subject covariance matrix described above.

PROC MIXED obtains estimates of β and **U** by solving the mixed model equations:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix},$$

where **b** and **u** denote the solutions for β and **U**, respectively. Note that the solution for **b** is equal to the generalized least squares (GLS) solution, $\mathbf{b} = (X\mathbf{V}^{-1}X)^{-}X'\mathbf{V}^{-1}\mathbf{y}$.

Inference for the mixed model is based on **predictable functions**, i.e. functions of the form $K'\beta + M'\mathbf{U}$, where $K'\beta$ is an estimable function. The **best linear unbiased predictor** (BLUP) of $K'\beta + M'\mathbf{U}$ is $K'\mathbf{b} + M'\mathbf{u}$. Its **prediction error variance** is given by the formula

Var[K'(β-b)+M'(U-u)=L'CL, where L' = (K' M'), and $C = \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix}^{-}$, i.e. the generalized inverse of the left-hand side of the mixed model equations. For known G and R, inference proceeds as follows:

- **Confidence Interval**: use the formula $K'b + M'u \pm Z_\alpha \sqrt{L'CL}$, where Z_α is the standard normal table value for the 1-α level of confidence.

- **Test H_0: K'β +M'U = 0**: use the Wald statistic, (K'b + M'u)'(L'CL)$^{-1}$(K'b + M'u), which is approximately $\sim \chi^2_{\text{rank(L)}}$.

For unknown G and R, PROC MIXED substitutes **restricted maximum likelihood** (REML) estimates of the variance and covariance components, e.g. REML estimates of σ_S^2 and the components of Σ in repeated measures models, directly into the mixed model equations and the formula for prediction error variance. Let $\hat{C}$ denote the estimate of C obtained by substituting REML estimates of its variance and covariance components. Inference proceeds as follows.

- **Confidence Interval**: use the formula $K'b + M'u \pm t_{(\alpha,\ \nu)} \sqrt{L'\hat{C}L}$, where ν = d.f. to estimate L'CL. Note that ν may be "obvious" by inspection or it may require an approximation, e.g. PROC MIXED will compute Satterthwaite's procedure.

- **Test $H_0 : K'\beta + M'U = 0$**: use the Wald statistic divided by its degrees of freedom,

$$\frac{(K'b+M'u)'\ (L'\hat{C}L)^{-}\ (K'b+M'u)'}{rank(L)}$$, which is approximately ~ $F_{[rank(L),\ v]}$.

Kacker and Harville (1984) showed that, except for *balanced, variance components only* models, these procedures are biased. Specifically, standard errors obtained from $L'\hat{C}L$ underestimate the true standard errors based on known L'CL. Therefore, except for balanced, independent errors (and hence compound symmetry) models, PROC MIXED computes standard errors that are biased downward and test statistics that are biased upward for repeated measures experiments. For many commonly used covariance models, the bias can be quite severe, as shown in the examples below. The bias can be exacerbated by mispecifying Σ, particularly if the assumed Σ is quite different from the true Σ (e.g. assume CS when ANTE(1) more aptly describes the true structure).

So the questions are:

- How does one choose an appropriate covariance model, Σ?
- How does one correct for the downward bias of $L'\hat{C}L$?

b. Model selection

Suppose one wants to choose between V_1 and V_2 , the Var(**y**) resulting from two covariance models Σ_1 and Σ_2, respectively. PROC MIXED provides three methods.

- **Likelihood Ratio Test**: Compute REML log-Likelihood under V_1 and V_2. Assuming V_2 is a subset of V_1, the likelihood ratio statistic -2 log [REML L (model 1) / REML L (model 2)] is approximately distributed χ^2. The degrees of freedom for χ^2 equal the difference between the number of variance and covariance components in V_1 and V_2 .

The fact that the likelihood ratio test is possible only if V_2 is a subset of V_1 limits its usefulness. Alternatively, one can compare REML log-likelihoods, preferably with a penalty function to account for differences in the number of covariance parameters among models. PROC MIXED offers two such criteria:

- **Schwarz's Bayesian Information Criterion**, called *SBC* in PROC MIXED, given by.
  SBC = REML log-likelihood - ½q[(log(N-p)],
  where q = # parameters in covariance model,
  N = # observations*
  p = # independent fixed effects parameters

* **Important:** N should be the number of *subjects*. Versions of PROC MIXED through Release 6.12 use N equal to the *total number of observations*, i.e. the number of

subjects multiplied by the number of times each subject is observed. Thus, SBC over penalizes covariance models with more covariance parameters. This is corrected in version 7.0 and subsequent versions of SAS. The difference is illustrated below in Section 5.

- **Akaike Information Criterion** (AIC).
  AIC = REML log-likelihood - q.

There is no consensus on which procedure for comparing covariance models is "best." Keselman et al. (1998) conducted a simulation study comparing the three methods used by PROC MIXED for a number of different covariance models. They found that all of the procedures choose the "wrong" covariance model over 50% of the time. However, they did not distinguish between "slightly wrong" and "substantially wrong" choices. That is, choosing an AR(1) model when the true model is TOEP may have negligible consequences, whereas choosing UN when the true model is CS may be quite serious. Their study found that AIC choose the "right" model more often than SBC, but suggested that because the study was done using Release 6.12 of SAS, this result was probably an artifact of the error in the way SBC is computed. Wolfinger (1999) suggested that Bayesian weighting among candidate models, along the lines described by Carlin (1996), may be preferable to choosing any one covariance model. Further work needs to be done in this area.

d. Accounting for bias in standard errors and test-statistics

As noted in the introduction (section 3a), repeated measures experiments usually focus on treatment/time effects. In other words, inference depends on estimable functions $K'\beta$, i.e. on fixed effects only. For estimable functions, the prediction error variance reduces to

$$Var(K'b) = [K' \; 0' \;] \; C \begin{bmatrix} K \\ 0 \end{bmatrix} = K'(X'V^{-1}X)^{-1}K$$, where **0** denotes the matrix M=0 in the predictable function $K'\beta+M'\mathbf{U}$. Thus, the "model-based" REML estimate, $L'\hat{C}L$, reduces to $K'(X'\hat{V}^{-1}X)^{-1}K$, where $\hat{V}$ is V using the REML variance and covariance component estimates. As with $L'\hat{C}L$, except for balanced, variance components only models, e.g. independent errors longitudinal models, $K'(X'\hat{V}^{-1}X)^{-1}K$ is a downward biased estimate of $K'(X'V^{-1}X)^{-1}K$, resulting in underestimated standard errors and inflated test statistics. Also, the degrees of freedom for estimating $K'(X'V^{-1}X)^{-1}K$, used by PROC MIXED to obtain p-values for t- and F-tests and to obtain confidence intervals via the t-distribution, often must be approximated. The Satterthwaite procedure PROC MIXED uses is not always appropriate.

How should the bias and degree of freedom issues be handled? Possible approaches available using PROC MIXED are

- The **empirical** estimate of $K'(X'V^{-1}X)^{-}K$, also called the "robust" or "sandwich" estimate. This estimate is described by Liang & Zeger (1986) and Diggle et al. (1994). They propose replacing the model-based REML estimate $(X'\hat{V}^{-1}X)^{-}$ by

$$(X'\hat{V}^{-1}X)^{-}[\Sigma X_i'\hat{V}_i^{-1}e_ie_i'\hat{V}_i^{-1}X_i](X'\hat{V}^{-1}X)^{-} \qquad (3.1)$$

where e_i is the vector of observed residuals, obtained fitting the independent errors model, for the i^{th} subject. This approach assumes that the number of subjects per treatment is substantially greater than the number of times of observation. When the number of observation times is equal to or greater than the number of subjects per treatment, as often happens in agricultural experiments, the empirical estimate of Var(K'**b**) may actually be *less than* the model-based estimate and the resulting test-statistics may be wildly inflated. Hardly a solution to the problem of downward biased standard errors and upward biased test statistics!

- **Kenward and Roger** (1997) propose a **bias adjustment** for the model-based REML estimate, $K'(X'\hat{V}^{-1}X)^{-1}K$, and a **degree of freedom** approximation more general than the Satterthwaite approximation. This option is available in Version 7.0 and subsequent versions of SAS.

Following Kenward and Roger, let $\Phi = X'V^{-1}X$ and $\hat{\Phi} = (X'\hat{V}^{-1}X)^{-1}$. Kenward and Roger obtain $E(\hat{\Phi})$. They use this result to adjust Var(K'**b**) and F_{obs} and to obtain a generally applicable approximation for ν. The main results are

- $E(\hat{\Phi}) = \Phi + \frac{1}{2}\sum\sum W_{ij}\frac{\partial^2\Phi}{\partial\sigma_i\partial\sigma_j}$

where $\frac{\partial^2\Phi}{\partial\sigma_i\partial\sigma_j} = \Phi(P_i\Phi P_j + P_j\Phi P_i - Q_{ij} - Q_{ji} + R_{ij})\Phi,$

$$P_i = X'\frac{\partial V^{-1}}{\partial\sigma_i}X, \quad Q_{ij} = X'\frac{\partial V^{-1}}{\partial\sigma_i}V\frac{\partial V^{-1}}{\partial\sigma_j}X, \quad R_{ij} = X'V^{-1}\frac{\partial^2 V}{\partial\sigma_i\partial\sigma_j}V^{-1}X,$$

and W_{ij} is ij^{th} element of $\mathrm{Var}(\hat{\sigma})$, where σ is the vector of variance and covariance components.

- Hence $\hat{\Phi}_{adj} = \hat{\Phi} + 2\,\hat{\Phi}\,[\Sigma\Sigma W_{ij}(Q_{ij} - P_i\,\hat{\Phi}\,P_j - \tfrac{1}{4}R_{ij}]\,\hat{\Phi}$. W_{ij} can be approximated from the information matrix $I_E = E\left(\frac{\partial^2 L}{\partial\sigma_i\partial\sigma_j}\right)$, where L is the REML log-likelihood, or, as with PROC MIXED, version 7.0, the observed information matrix

$$I_{obs} = \left[\frac{\partial^2 L}{\partial\sigma_i\partial\sigma_j}\right] = H$$ where **H** is the Hessian matrix.

- Using $\hat{\Phi}_{adj}$, F_{obs} is adjusted as $F^* = \lambda F_{obs}$, where $\lambda = \frac{\nu}{E(F_{obs})(\nu-2)}$ and ν is equal to the approximate denominator degrees of freedom.
- The denominator degrees of freedom are approximated by the formula

$$\nu = 4 + \frac{rank(K) + 2}{rank(K)\times\rho - 1}, \text{ where } \rho = \frac{Var(F_{obs})}{2E(F_{obs})^2}.$$

The approximations for $E(F_{obs})$, $Var(F_{obs})$ are quite involved. See Kenward and Roger (1997) for details.

The Kenward-Roger and Satterthwaite degrees of freedom approximations yield equivalent results for variance-component-only mixed models (independent errors or CS for longitudinal data) when all the variance component estimates are positive. However, the Kenward-Roger approximation is generally applicable, whereas the Satterthwaite procedure often breaks down when variance component estimates are negative or when covariance models are used.

4. Examples

This section present an example illustrating the issues discussed in the previous sections. The data are given in Table 1. They are from a repeated measures experiment with 4 treatments, 6 subjects per treatment, and 8 times of observation. The data in Table 1 appear in the format appropriate for analyzing longitudinal data with the REPEATED option in SAS PROC GLM. For PROC MIXED the data need to be rearranged so that there is one data line per treatment x subject x time combination, e.g. using the following SAS statements:

```
DATA A;
 INPUT TRT SUBJ T1-T8;
DATA B;
 SET A;
  TIME=1; Y=T1; OUTPUT;
  TIME=2; Y=T2; OUTPUT;
  TIME=3; Y=T3; OUTPUT;
  TIME=4; Y=T4; OUTPUT;
  TIME=5; Y=T5; OUTPUT;
  TIME=6; Y=T6; OUTPUT;
  TIME=7; Y=T7; OUTPUT;
  TIME=8; Y=T8; OUTPUT;
 DROP T1-T8;
```

The model for this experiment is exactly as given in equation (2.1). The basic PROC MIXED statements for the analysis are:

```
PROC MIXED;
 CLASSES SUBJ TRT TIME;
 MODEL Y= TRT TIME TRT*TIME/DDFM=SATTERTH;
 RANDOM SUBJ(TRT);
 REPEATED TIME / TYPE=< type of covariance model > SUBJECT=SUBJ(TRT);
```

Readers are referred to Littell et al. (1996) for further details on programming MIXED for repeated measures and to SAS documentation (SAS Institute, 1997) for further information on types of covariance models available and other options in PROC MIXED. Here, note only that the RANDOM statement identifies the between subjects random model effect and REPEATED identifies the type of covariance model (independent, AR(1), etc.) and the subject on which the block diagonal structure of the covariance matrix R is based.

As noted above, some covariance models contain the between subjects random effects. In these model, one *must not* include both RANDOM and REPEATED statements. For example, compound symmetry covariance model and random subject + independent errors are equivalent models, and hence the following two programs

```
PROC MIXED;
 CLASSES SUBJ TRT TIME;
 MODEL Y= TRT TIME TRT*TIME/DDFM=SATTERTH;
 RANDOM SUBJ(TRT);
TITLE 'MIXED - INDEP ERRORS (SPLIT PLOT IN TIME)';
```

and

```
PROC MIXED;
 CLASSES SUBJ TRT TIME;
 MODEL Y= TRT TIME TRT*TIME/DDFM=SATTERTH;
 REPEATED TIME / TYPE=CS SUBJECT=SUBJ(TRT);
TITLE 'MIXED - COMPOUND SYMMETRY';
```

produce the same result. Key output:

```
                Covariance Parameter Estimates (REML)

                            INDEP ERRORS

                Cov Parm        Estimate

                SUBJ(TRT)     27.71786905
                Residual       4.79758929

                          COMPOUND SYMMETRY
```

Cov Parm	Subject	Estimate
CS	SUBJ(TRT)	27.71786905
Residual		4.79758929

Model Fitting Information for Y

INDEP ERRORS

Res Log Likelihood	-419.696
Akaike's Information Criterion	-421.696
Schwarz's Bayesian Criterion	-424.771
-2 Res Log Likelihood	839.3910

COMPOUND SYMMETRY

Res Log Likelihood	-419.696
Akaike's Information Criterion	-421.696
Schwarz's Bayesian Criterion	-424.771
-2 Res Log Likelihood	839.3910

Tests of Fixed Effects

INDEP ERRORS

Source	NDF	DDF	Type III F	Pr > F
TRT	3	20	0.74	0.5425
TIME	7	140	109.04	0.0001
TRT*TIME	21	140	1.98	0.0106

COMPOUND SYMMETRY

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
TRT	3	20	0.74	0.5425
TIME	7	140	109.04	0.0001
TRT*TIME	21	140	1.98	0.0106

The same equivalence exists for the **unstructured** covariance model, and hence the RANDOM SUBJ(TRT) statement *should not* be used along with a REPEATED / TYPE=UN

SUBJECT=SUBJ(TRT) statement. For other models, the random subject effect is not necessarily confounded with any covariance parameters, and so both statements may be used. For example, for AR(1):

```
PROC MIXED;
 CLASSES  SUBJ TRT TIME;
 MODEL Y= TRT TIME TRT*TIME/DDFM=SATTERTH;
 RANDOM  SUBJ(TRT);
 REPEATED / TYPE=AR(1) SUBJECT=SUBJ(TRT);
TITLE 'MIXED - AR(1) ERRORS';
```

produces distinct estimates of σ_S^2 and ρ. In some data sets σ_S^2 and ρ are too closely identified. The typical symptom of this problem is failure of the REML algorithm to converge. In theory, σ_S^2 and the ρ_i in the Toeplitz structure are identified and the random subject effect and repeated statements may both be used. In practice, many data sets do not behave well (i.e. REML does not converge or gives nonsense estimates) when one tries to fit both the random subject effect and the REPEATED TYPE=TOEP covariance model. The same is true for the first-order antedependence model. The random subject effect is identifiable *in theory*. However, fitting the random subject effect along with REPEATED TYPE=ANTE(1) rarely "works," because *in practice* the terms are usually not sufficiently identifiable to permit variance-covariance estimation algorithms to obtain solutions.

Fitting the AR(1) model to the data in Table 1, σ_S^2 and ρ are well identified and REML yields meaningful estimates:

Cov Parm	Subject	Estimate
SUBJ(TRT)		22.50472051
AR(1)	SUBJ(TRT)	0.75734258
Residual		9.61359585

However, recalling Kenward and Roger's work, the Satterthwaite approximation does not generalize to the covariance models in repeated measures, and hence one often sees peculiar results. The MIXED program for the AR(1) model yields:

Source	NDF	DDF	Type III F	Pr > F
TRT	3	2.48	0.75	0.6009
TIME	7	0	60.55	.
TRT*TIME	21	0	1.48	.

In Version 6 (Release 6.12), the alternatives are to either use the default degrees of freedom, using the **containment rule** (in essence, the degrees of freedom for error from the ANOVA table that results looking at the model as a split-plot) or supplying user-determined denominator

degrees of freedom. It is not clear what one would use for the latter. The default yields

Source	NDF	DDF	Type III F	Pr > F
TRT	3	20	0.75	0.5344
TIME	7	140	60.55	0.0001
TRT*TIME	21	140	1.48	0.0921

Comparing the p-values for the AR(1) vs. CS suggests the impact of covariance model choice: the test for the main effect of treatment not appreciably affected but for the time and time x treatment tests, the simpler the covariance model, the greater the F-values - and hence the lower the p-values -- tend to be. Model misspecification can result in **excessive type I error rate** when the **covariance model** used in the analysis is **too simple** relative to the true serial correlation (e.g. analyze the data with a compound symmetry model when there is actually substantial, patterned correlation - AR(1), TOEP, etc.) or it can result in excessive type II error and hence **inadequate power** when the **covariance model** used in the analysis is **more complicated than necessary** (e.g. using UN when AR(1) adequately describes serial correlation).

Table 2 summarizes the model selection results and the F-values for the time and treatment effects using a number of different covariance models. Note that the likelihood ratio test for AR(1) tests AR(1) versus the independent error (or, equivalently, compound symmetry) model. The tests for TOEP, ANTE(1), and UN test these models versus AR(1), the logic being that whereas AR(1) clearly shows an advantage relative to CS, none of the more complex models show an advantage over AR(1); hence AR(1) is used here as a benchmark because it is the most complex model showing any advantage over its less complex competing models. Table 2 also gives test results for the MANOVA and Huynh-Feldt corrected p-value options available in PROC GLM. The main points of Table 2 are:

- By all three model selection criteria, AR(1) is the covariance model of choice
- Under-modeling serial correlation (independent errors/CS) results in inflated F-values and hence overly significant p-values
- Over-modeling serial correlation (TOEP, ANTE(1), UN) results in deflated F-values and hence understated statistical significance

In other words, choosing an appropriate covariance model matters. Failing to choose a reasonable covariance model can result in serious misinterpretation of the data.

- The MANOVA option in GLM also over models serial correlation and hence its p-values severely understate statistical significance
- The Huynh-Feldt option in GLM does not adequately account for the AR(1) structure, and hence does not adequately correct for inflated F-values

In other words, the options provided by traditional fixed effects software are inadequate to

provide accurate analysis of longitudinal data.

- MANOVA in GLM and TYPE=UN in MIXED ought to produce similar results, yet the p-values are quite far apart.

This point requires additional discussion. It is a symptom of the bias that results when test statistics are computed using model-based estimates of $X'V^{-1}X$, that is, substituting REML estimates into V and using the resulting $\hat{V}$ as if it was known. As discussed in Section 4, Kacker and Harville (1984) and Kenward and Roger (1997) showed that test statistics are biased upward and standard errors are biased downward.

Recalling Section 4, one suggestion for solving the bias problem is the **empirical** (or **robust** or "Sandwich") estimator. This can be implemented by using the EMPIRICAL option in the PROC MIXED statement:

```
PROC MIXED EMPIRICAL;
 CLASSES  SUBJ TRT TIME;
 MODEL Y= TRT TIME TRT*TIME/DDFM=SATTERTH;
 RANDOM  SUBJ(TRT);
 REPEATED / TYPE=AR(1) SUBJECT=SUBJ(TRT);
```

This results in the following tests of the treatment and time effects:

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
TRT	1	20	3.49736E29	0.0001
TIME	1	140	2.29324E31	0.0001
TRT*TIME	1	140	1.78875E29	0.0001

Note that the F-values are wildly inflated. This is typical when the number of time periods exceeds the number of subjects per treatment. Deleting the RANDOM statement helps somewhat. That is, the MIXED program:

```
PROC MIXED EMPIRICAL;
 CLASSES  SUBJ TRT TIME;
 MODEL Y= TRT TIME TRT*TIME/DDFM=SATTERTH;
 REPEATED / TYPE=AR(1) SUBJECT=SUBJ(TRT);
```

yields

Source	NDF	DDF	Type III F	Pr > F
TRT	3	20	1.31	0.2981

TIME	7	140	121.57	0.0001
TRT*TIME	20	140	9.04	0.0001

Recall that using the model based default, $F_{TRT*TIME}$ with the AR(1) covariance model was 1.48. Using the **empirical** estimate, $F_{TRT*TIME} = 9.04$, hardly a solution to the problem of upward bias! While not shown here, the standard errors of treatment, time, and treatment-time combination means and various differences are similarly unreasonable.

Note also when the **empirical** option is used, the TYPE= specification in the REPEATED statement is not used. Instead, test statistics and standard errors are computed using the "sandwich" estimator of $(X' V^{-1}X)^{-}$ given above in formula (3.1). See *SAS/STAT Software: Changes and Enhancements though Release 6.12* (SAS Institute, 1997) for further details. With the repeated measures experiments typically used in agriculture, where the number of subjects per treatment is often small relative to the number of times of observation, the empirical estimator is not a viable approach. A promising alternative, however, **is** available in Release 7.0 and subsequent releases of SAS. In these releases, the adjustment to $X'\hat{V}^{-1}X$ and denominator degrees of freedom approximation proposed by Kenward and Roger (1997) are available. They are implemented via the DDFM options in the MODEL statement, DDFM=KR.

Recalling the AR(1) example, the PROC MIXED code to compute the Kenward-Roger adjustment is:

```
PROC MIXED;
 CLASSES  SUBJ TRT TIME;
 MODEL Y= TRT TIME TRT*TIME/DDFM=KR;
 RANDOM  SUBJ(TRT);
 REPEATED / TYPE=AR(1) SUBJECT=SUBJ(TRT);
```

This yields the following:

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
TRT	3	20.5	0.77	0.5219
TIME	7	109	50.90	<.0001
TRT*TIME	21	117	1.24	0.2330

Note the adjustment of the degrees of freedom and the F-values. For instance, without the adjustment, $F_{TRT*TIME} = 1.48$ (p=0.0921) whereas with the adjustment, the F-value is 1.24 (p=0.2330).

Version 7.0 also corrects the error in Release 6.12's computing the Schwarz Bayesian Criterion noted above. Table 3 gives an update of Table 2 with corrections in the SBC and Kenward-Roger adjusted test-results compared to default tests. The following main points result:

- The SBC does not penalize more complex covariance models as heavily as it appeared to in Release 6.12. Although the AR(1) is still clearly the preferred covariance model here, some conclusions could change. Note the reversal of independent errors/CS relative to ANTE(1).
- The Kenward-Roger adjustment does not affect the independent errors /CS models, because default statistics are exact for balanced variance-components-only mixed models.
- There is considerable adjustment of several of the F-values.
- The inadequacy of the Huynh-Feldt adjustment is even more apparent.
- The results for TYPE=UN in MIXED and MANOVA in GLM, while not identical, are only trivially different.

What about the affect on standard errors? For AR(1), adding the statement

```
LSMEANS TRT*TIME / DIFF SLICE=TIME SLICE=TRT:
```

yields estimates the mean response at every treatment-by-time combination, all possible differences among them, and tests of simple effects of treatment for each time and time for each treatment. Below are selected results: the treatment-time LSMEANS for treatments 1 and 2, differences between time 1 and each subsequent time for treatment 1, differences between treatment 1 and 2 for each time, and each SLICE.

For the LSMEANS, the results are:

Model-based "Naive" Standard Errors

Least Squares Means

Effect	TRT	TIME	LSMEAN	Std Error	DF	t	Pr > |t|
TRT*TIME	1	1	30.33333333	2.30980355	140	13.13	0.0001
TRT*TIME	1	2	24.95000000	2.30980355	140	10.80	0.0001
TRT*TIME	1	3	21.21666667	2.30980355	140	9.19	0.0001
TRT*TIME	1	4	20.65000000	2.30980355	140	8.94	0.0001
TRT*TIME	1	5	20.10000000	2.30980355	140	8.70	0.0001
TRT*TIME	1	6	19.98333333	2.30980355	140	8.65	0.0001
TRT*TIME	1	7	20.58333333	2.30980355	140	8.91	0.0001
TRT*TIME	1	8	21.06666667	2.30980355	140	9.12	0.0001
TRT*TIME	2	1	32.13333333	2.30980355	140	13.91	0.0001
TRT*TIME	2	2	24.56666667	2.30980355	140	10.64	0.0001
TRT*TIME	2	3	21.31666667	2.30980355	140	9.23	0.0001
TRT*TIME	2	4	18.21666667	2.30980355	140	7.89	0.0001
TRT*TIME	2	5	18.66666667	2.30980355	140	8.08	0.0001
TRT*TIME	2	6	17.60000000	2.30980355	140	7.62	0.0001
TRT*TIME	2	7	18.45000000	2.30980355	140	7.99	0.0001
TRT*TIME	2	8	18.40000000	2.30980355	140	7.97	0.0001

Kenward-Roger Adjustment

Least Squares Means

Standard

Effect	TRT	TIME	Estimate	Error	DF	t Value	Pr > |t|
TRT*TIME	1	1	30.3333	2.3098	30.8	13.13	<.0001
TRT*TIME	1	2	24.9500	2.3098	30.8	10.80	<.0001
TRT*TIME	1	3	21.2167	2.3098	30.8	9.19	<.0001
TRT*TIME	1	4	20.6500	2.3098	30.8	8.94	<.0001
TRT*TIME	1	5	20.1000	2.3098	30.8	8.70	<.0001
TRT*TIME	1	6	19.9833	2.3098	30.8	8.65	<.0001
TRT*TIME	1	7	20.5833	2.3098	30.8	8.91	<.0001
TRT*TIME	1	8	21.0667	2.3098	30.8	9.12	<.0001
TRT*TIME	2	1	32.1333	2.3098	30.8	13.91	<.0001
TRT*TIME	2	2	24.5667	2.3098	30.8	10.64	<.0001
TRT*TIME	2	3	21.3167	2.3098	30.8	9.23	<.0001
TRT*TIME	2	4	18.2167	2.3098	30.8	7.89	<.0001
TRT*TIME	2	5	18.6667	2.3098	30.8	8.08	<.0001
TRT*TIME	2	6	17.6000	2.3098	30.8	7.62	<.0001
TRT*TIME	2	7	18.4500	2.3098	30.8	7.99	<.0001
TRT*TIME	2	8	18.4000	2.3098	30.8	7.97	<.0001

Note that the estimates and standard errors are not affected for LSMEANS, but the denominator degrees of freedom, and hence the p-values, are. For differences:

Model-based "Naive" Standard Errors

Differences of Least Squares Means

Effect	TRT	TIME	_TRT	_TIME	Difference	Std Error	DF	t
TRT*TIME	1	1	1	2	5.38333333	0.88100697	140	6.11
TRT*TIME	1	1	1	3	9.11666667	1.22306461	140	7.45
TRT*TIME	1	1	1	4	9.68333333	1.47079848	140	6.58
TRT*TIME	1	1	1	5	10.23333333	1.66795230	140	6.14
TRT*TIME	1	1	1	6	10.35000000	1.83190291	140	5.65
TRT*TIME	1	1	1	7	9.75000000	1.97178415	140	4.94
TRT*TIME	1	1	1	8	9.26666667	2.09315416	140	4.43
TRT*TIME	1	1	2	1	-1.80000000	3.26655551	140	-0.55
TRT*TIME	1	2	2	2	0.38333333	3.26655551	140	0.12
TRT*TIME	1	3	2	3	-0.10000000	3.26655551	140	-0.03
TRT*TIME	1	4	2	4	2.43333333	3.26655551	140	0.74
TRT*TIME	1	5	2	5	1.43333333	3.26655551	140	0.44
TRT*TIME	1	6	2	6	2.38333333	3.26655551	140	0.73
TRT*TIME	1	7	2	7	2.13333333	3.26655551	140	0.65
TRT*TIME	1	8	2	8	2.66666667	3.26655551	140	0.82

Kenward-Roger Adjustment

Effect	TRT	TIME	_TRT	_TIME	Estimate	Standard Error	DF	t Value
TRT*TIME	1	1	1	2	5.3833	0.9081	139	5.93
TRT*TIME	1	1	1	3	9.1167	1.2609	147	7.23
TRT*TIME	1	1	1	4	9.6833	1.5165	154	6.39

TRT*TIME	1	1	1	5	10.2333	1.7199	158	5.95
TRT*TIME	1	1	1	6	10.3500	1.8890	160	5.48
TRT*TIME	1	1	1	7	9.7500	2.0331	160	4.80
TRT*TIME	1	1	1	8	9.2667	2.1580	157	4.29
TRT*TIME	1	1	2	1	-1.8000	3.2666	30.8	-0.55
TRT*TIME	1	2	2	2	0.3833	3.2666	30.8	0.12
TRT*TIME	1	3	2	3	-0.1000	3.2666	30.8	-0.03
TRT*TIME	1	4	2	4	2.4333	3.2666	30.8	0.74
TRT*TIME	1	5	2	5	1.4333	3.2666	30.8	0.44
TRT*TIME	1	6	2	6	2.3833	3.2666	30.8	0.73
TRT*TIME	1	7	2	7	2.1333	3.2666	30.8	0.65
TRT*TIME	1	8	2	8	2.6667	3.2666	30.8	0.82

Note that in addition to the denominator degrees of freedom, the standard errors of the differences between times within treatment, but *not between treatments within time,* are affected. For SLICES:

Model-based "Naive" Slices

Tests of Effect Slices

Effect	TIME	NDF	DDF	F	Pr > F
TRT*TIME	1	3	140	1.83	0.1454
TRT*TIME	2	3	140	1.85	0.1412
TRT*TIME	3	3	140	1.26	0.2911
TRT*TIME	4	3	140	0.68	0.5644
TRT*TIME	5	3	140	0.35	0.7887
TRT*TIME	6	3	140	0.73	0.5328
TRT*TIME	7	3	140	0.18	0.9075
TRT*TIME	8	3	140	0.40	0.7517

Effect	TRT	Num DF	Den DF	F Value	Pr > F
TRT*TIME	1	7	140	8.53	<.0001
TRT*TIME	2	7	140	15.57	<.0001
TRT*TIME	3	7	140	15.48	<.0001
TRT*TIME	4	7	140	12.30	<.0001

Kenward-Roger Adjustment

Effect	TIME	Num DF	Den DF	F Value	Pr > F
TRT*TIME	1	3	30.8	1.83	0.1633
TRT*TIME	2	3	30.8	1.85	0.1591
TRT*TIME	3	3	30.8	1.26	0.3058

TRT*TIME	4	3	30.8	0.68	0.5698
TRT*TIME	5	3	30.8	0.35	0.7889
TRT*TIME	6	3	30.8	0.73	0.5392
TRT*TIME	7	3	30.8	0.18	0.9069
TRT*TIME	8	3	30.8	0.40	0.7524

Effect	TRT	Num DF	Den DF	F Value	Pr > F
TRT*TIME	1	7	139	8.02	<.0001
TRT*TIME	2	7	139	14.65	<.0001
TRT*TIME	3	7	139	14.56	<.0001
TRT*TIME	4	7	139	11.57	<.0001

Again, the denominator degrees of freedom are affected. In addition, for the SLICES comparing TIME effects within each treatment, the F-values are corrected downward. The following are the "naive" (PROC MIXED default) and Kenward-Roger corrected SLICE statistics for the effect of time for each treatment for CS, TOEP, ANTE(1), and UN.

COMPOUND SYMMETRY = INDEP ERRORS - MIXED DEFAULT = KENWARD-ROGER CORRECTION

Effect	TRT	Num DF	Den DF	F Value	Pr > F
TRT*TIME	1	7	140	16.10	<.0001
TRT*TIME	2	7	140	31.19	<.0001
TRT*TIME	3	7	140	44.31	<.0001
TRT*TIME	4	7	140	23.37	<.0001

MIXED DEFAULT - TOEPLITZ COVARIANCE

Effect	TRT	Num DF	Den DF	F Value	Pr > F
TRT*TIME	1	7	140	11.09	<.0001
TRT*TIME	2	7	140	19.27	<.0001
TRT*TIME	3	7	140	21.23	<.0001
TRT*TIME	4	7	140	16.28	<.0001

KENWARD-ROGER CORRECTION - TOEPLITZ COVARIANCE

Effect	TRT	Num DF	Den DF	F Value	Pr > F
TRT*TIME	1	7	55.1	10.36	<.0001
TRT*TIME	2	7	55.1	18.00	<.0001
TRT*TIME	3	7	55.1	19.83	<.0001
TRT*TIME	4	7	55.1	15.20	<.0001

MIXED DEFAULT - ANTEDEPENDENCE COVARIANCE

Effect	TRT	Num DF	Den DF	F Value	Pr > F
TRT*TIME	1	7	140	8.58	<.0001
TRT*TIME	2	7	140	14.44	<.0001
TRT*TIME	3	7	140	15.65	<.0001
TRT*TIME	4	7	140	12.56	<.0001

KENWARD-ROGER CORRECTION - ANTEDEPENDENCE COVARIANCE

Effect	TRT	Num DF	Den DF	F Value	Pr > F
TRT*TIME	1	7	42.9	7.49	<.0001
TRT*TIME	2	7	42.9	12.63	<.0001
TRT*TIME	3	7	42.9	13.68	<.0001
TRT*TIME	4	7	42.9	10.96	<.0001

MIXED DEFAULT - UNSTRUCTURED ERROR COVARIANCE

Effect	TRT	Num DF	Den DF	F Value	Pr > F
TRT*TIME	1	7	20	19.16	<.0001
TRT*TIME	2	7	20	27.84	<.0001
TRT*TIME	3	7	20	31.29	<.0001
TRT*TIME	4	7	20	27.12	<.0001

KENWARD-ROGER CORRECTION - UNSTRUCTURED ERROR COVARIANCE

Effect	TRT	Num DF	Den DF	F Value	Pr > F
TRT*TIME	1	7	14	13.41	<.0001
TRT*TIME	2	7	14	19.49	<.0001
TRT*TIME	3	7	14	21.90	<.0001
TRT*TIME	4	7	14	18.98	<.0001

The main points arising from comparing SLICEs are

- The Kenward-Roger correction is noticeable for all covariance models , except CS where it is equivalent to default results.
- The Kenward-Roger correction is greater for TYPE=UN, the most complex of the models compared in this example.

This example is consistent with Kacker and Harville (1984) and Kenward and Roger (1997) papers: except for CS, the default methods for computing the SLICE F-values are biased upward and the Kenward-Roger correction will always decrease them. In their paper, Kenward and Roger included the results of simulation studies they did on their correction. There are other as-yet

unpublished simulations of which the author is aware. More small-sample studies are certainly needed, as the adjustment has only been tested for a few of the situations for which it will probably be used. However, the studies conducted to date suggest that using DDFM=KR probably should be considered "standard operating procedure" when using PROC MIXED to analyze longitudinal data.

One final note comparing the above SLICE output:

- The F-values are affected by choice of covariance model. The difference from the F-values obtained using model is greatest for CS and UN. However, discrepancies also exist among AR(1), TOEP, and ANTE(1).

There is an "oral tradition," for lack of a better expression, that the *exact* choice of covariance model is not overly critical, but choosing a model that is "in the ball park" *is important*. In other words, if the true serial correlation is approximately AR(1), using TOEP should yield similar results, but using a seriously misspecified model, e.g. CS in one direction or UN in the other, is more likely to adversely affect the accuracy of one's conclusions. In this example, discrepancies are not exactly consistent with conventional wisdom, but neither are they overly contradictory. More work needs to be done to validate or debunk the "oral tradition" just described. This would be especially useful in conjunction with studying the small sample behavior of the Kenward-Roger correction. Studies to address these question are in progress and we hope to be able to report on our findings in the future.

5. Summary

PROC MIXED represents a distinct step forward for the analysis of longitudinal data. Compared to MIXED, previously available MANOVA or Huynh-Feldt adjusted univariate analyses, such as have been available with PROC GLM, are not flexible enough to handle the type of correlated error structure typical of longitudinal data in agriculture. GLM's lack of flexibility is not trivial: severely misleading conclusions can result. Moreover, GLM's - indeed, MANOVA's - approach to missing data is unacceptably Draconian.

Several cautions are in order when using PROC MIXED. Among the more important issues are:

- Several covariance structures (e.g. CS and UN) contain the random subject effect.

Hence, one should not include the corresponding random and repeated statements in the same program. Other structures [e.g. AR(1), TOEP, and ANTE(1)] are technically not confounded, but in certain data sets their covariance parameters are too closely identified with the random model effect to permit both to be modeled. TOEP and ANTE(1) are especially susceptible.

- The Satterthwaite degree of freedom approximation is inappropriate for most covariance models.

However, the Kenward-Roger procedure, to be available beginning with Version 7.0 of SAS, appears to work quite well. In the special case of variance-components-only mixed models (and hence, CS) the Kenward-Roger and Satterthwaite approximations are equivalent.

- For non-trivial covariance models, PROC MIXED's default statistics are biased.

F-values for effects involving time are biased upward and the corresponding default standard errors are biased downward. The so-called "sandwich" or empirical estimator is not suitable for most agricultural applications. However, the Kenward-Roger adjustment appears to address this bias. More testing is needed to fully verify its small-sample behavior, but the indications to date are sufficiently favorable to recommend the Kenward-Roger option as standard operating procedure.

REFERENCES

Carlin, B.P. 1996. *Bayes and Empirical Bayes Methods and Applications*. New York: Chapman & Hall.

Diggle, P.J., K.-Y. Liang, and S.L. Zeger. 1994. *Analysis of Longitudinal Data*. New York: Oxford University Press.

Kacker, R.N. and D.A. Harville. 1984 "Approximations for standard errors of estimators of fixed and random effects in mixed linear models." *J. Amer. Statist. Assoc.* **79**: 853-862.

Kenward, M.G. and J.H. Roger/ 1997. "Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**: 983-997.

Keselman, H.J., J. Algina, R.K. Kowalchuk, R.D. Wolfinger. 1998. "A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements." *Commun. Statist. - Simula.* **27**(3): 591-604.

Liang, K.-Y. and S.L. Zeger. 1986. "Longitudinal data analysis using generalized linear models." *Biometrika* **73**: 13-22.

Littell, R.C., G.A. Milliken, W.W. Stroup, and R.D. Wolfinger. 1996. *Sas System for Mixed Models*. Cary, NC: SAS Institute.

SAS Institute, Inc. 1997. *SAS/STAT Software: Changes and Enhancements through Release 6.12*. Cary NC: SAS Institute.

Wolfinger, R.D. 1999. Personal communication.

Table 1. Longitudinal data example*.

OBS	TRT	SUBJ	T1	T2	T3	T4	T5	T6	T7	T8
1	1	1	27.7	19.8	18.4	15.8	16.5	16.1	15.5	17.1
2	1	2	26.7	20.1	13.5	12.2	11.5	12.4	14.9	18.3
3	1	3	31.6	29.6	25.1	24.4	23.0	24.0	20.8	19.1
4	1	4	25.8	22.3	17.6	19.8	19.1	16.0	17.9	18.0
5	1	5	37.3	30.2	29.3	29.1	29.0	30.1	33.9	33.8
6	1	6	32.9	27.7	23.4	22.6	21.5	21.3	20.5	20.1
7	2	1	30.6	25.4	22.7	18.7	20.8	20.4	19.1	19.8
8	2	2	31.9	22.5	19.9	15.0	15.8	14.9	19.9	18.5
9	2	3	27.8	21.8	14.5	11.4	14.8	13.2	11.1	7.0
10	2	4	37.4	29.8	25.6	25.4	27.1	26.1	26.2	26.5
11	2	5	35.8	30.7	29.0	26.5	24.5	22.1	24.9	27.9
12	2	6	29.3	17.2	16.2	12.3	9.0	8.9	9.5	10.7
13	3	1	35.7	27.9	23.9	19.9	20.9	18.4	19.0	20.4
14	3	2	40.8	32.6	31.9	28.7	26.7	28.4	28.4	25.6
15	3	3	30.8	26.1	19.5	19.6	19.0	17.8	16.2	17.3
16	3	4	43.5	37.4	34.4	27.8	24.8	22.3	20.3	19.8
17	3	5	35.5	31.4	26.4	20.1	15.7	19.0	20.3	22.0
18	3	6	36.7	29.2	23.0	21.2	22.1	24.1	18.3	20.7
19	4	1	34.9	29.0	24.0	26.3	22.9	24.8	26.4	27.7
20	4	2	28.2	21.0	20.6	21.2	20.3	20.0	19.9	20.8
21	4	3	27.7	17.8	15.7	9.7	13.1	10.3	12.5	11.7
22	4	4	30.5	25.3	20.8	18.7	15.1	13.3	14.4	13.7
23	4	5	25.2	19.3	17.1	18.8	14.6	12.2	15.9	10.7
24	4	6	38.7	32.0	30.7	29.8	26.1	25.1	27.4	26.8

* Variables in data set: TRT denotes treatment, SUBJ denotes subject (experimental unit), T1 denotes observation at time 1, T2 denotes observation at time 2, ... T8 denotes observation at time 8.

Table 2. Model selection criteria and tests of treatment and times effects for various covariance models using Release 6.12 of SAS

Model	-2 REML LL	LR χ^2	df	AIC	SBC
Indep errors	839.39			-422.0	-424.7
CS	839.39			-422.0	-424.7
AR(1)	**788.65**	**50.74**	**1**	**-397.3**	**-401.5**
Toeplitz	784.94	3.71	6	-400.5	-412.8
ANTE(1)	780.65	8.00	12	-405.3	-428.4
Unstructured	760.45	28.20	33	-416.2	-471.6

F-values under various models

Model	TRT	TIME	TRT*TIME
I & CS	0.74 (.5425)	109.04 (.0001)	1.98 (.0106)
AR(1)	0.75 (.5433)	60.55 (.0001)	1.48 (.0921)
TOEP	0.75 (.5335)	63.60 (.0001)	1.43 (.1698)
ANTE(1)	0.77 (.5223	47.30 (.0001)	1.31 (.2103)
UN	0.74 (.5425)	101.31 (.0001)	1.37 (.2450)
MANOVA		p=.0001	p=.5581
H-F		p=.0001	p=.0284

Table 3. Model selection & test results. Table 2 updated with corrected SBC and Kenward-Roger corrected F-tests.

Model	-2 REML LL	AIC	SBC-6.12	SBC-7.0
I & CS	839.39	-422.0	-424.7	-422.9
AR(1)	788.65	-397.3	-401.5	-399.1
Toeplitz	784.94	-400.5	-412.8	-405.2
ANTE(1)	780.65	-405.3	-428.4	-414.2
Unstructured	760.45	-416.2	-471.6	-437.4

F-values under various models

Model	TRT*TIME - default	TRT*TIME - KR adjustment
I & CS	1.98 (.0106)	1.98 (.0106)
AR(1)	1.48 (.0921)	1.24 (.2330)
TOEP	1.43 (.1698)	1.32 (.1893)
ANTE(1)	1.31 (.2103)	1.12 (.3516)
UN	1.37 (.2450)	0.90 (0.5982)
MANOVA		p=.5581
H-F		p=.0284