Paper 177-2007
Growing Up Fast: SAS® 9.2 Enhancements to the GLIMMIX Procedure

Oliver Schabenberger, SAS Institute Inc., Cary, NC

ABSTRACT

The GLIMMIX procedure was first released as a Web download add-on procedure to SAS/STAT® in SAS 9.1 in the 32-bit Windows operating environment. It was subsequently released as a Web download for SAS 9.1.3 in UNIX and Linux environments. This paper describes the key changes and enhancements to PROC GLIMMIX between the SAS 9.1 and SAS 9.2 releases.

Enhancements fall into several categories: statistical estimation, model building, post-processing, and miscellaneous other features. For example, new estimation methods and covariance structures enable parameter estimation with reduced bias in more flexible generalized linear mixed models. A new statement provides a comprehensive facility to test hypotheses about covariance parameters. New multiplicity corrections offer more ways to adjust inferences.

The goal of this paper is to highlight the new estimation and inference techniques and to demonstrate their usage with examples.

INTRODUCTION

The GLIMMIX procedure is a new procedure in SAS/STAT software. It estimates parameters in generalized linear mixed models and provides a full ensemble of post-processing features for this class of models. The GLIMMIX procedure was first released as a Web download add-on procedure to SAS/STAT in SAS 9.1 in the 32-bit Windows operating environment and later for SAS 9.1.3 in UNIX and Linux environments. Starting with SAS 9.2, the procedure is available as part of SAS/STAT in all operating environments.

SAS 9.2 brings a number of enhancements to the GLIMMIX procedure. (The ODDSRATIO facility was added to SAS 9.1.3 with the November 2005 production version.) The full potential of these enhancements comes to bear when they are applied in combination. For example, maximum likelihood estimation based on adaptive quadrature and the new COVTEST statement for testing hypotheses about covariance parameters are important enhancements by themselves. Together, they enable multiple likelihood ratio tests based on the marginal log likelihood in generalized linear mixed models for nonnormal data in a single run of the procedure—a novel and powerful facility in SAS mixed modeling software. The experimental EFFECT statement constructs special model effects, such as B-splines. New syntax to specify coefficients for linear hypotheses makes it simpler to work with interaction terms. In combination, these two features make comparisons of spline trends across levels of classification variables quite simple.

Not all of the new syntax is shown here. See the "Syntax" section of the PROC GLIMMIX documentation in the *SAS/STAT User's Guide* for all syntax elements, options, and suboptions. The "Details" section of that chapter contains numerous new subsections with in-depth mathematical coverage. New examples in the procedure documentation highlight various new features and enhancements.

STATISTICAL ESTIMATION

Maximum Likelihood by Laplace or Quadrature Approximation

Parameter estimation for generalized linear mixed models (GLMMs) with the GLIMMIX procedure in SAS 9.1 relies on pseudo-likelihood (PL) methods. These are based on linearizing a GLMM to transform it into a linear mixed (pseudo-)model. The techniques involve repeated fitting of the pseudo-models with interim updates of the pseudo-data. The broad applicability of pseudo-likelihood models makes them appealing as a default estimation method that can handle any type of model. However, the PL methods can produce biased estimates in situations where the number of observations per subject is small, particularly, for binary data

(see, for example, Breslow and Lin 1995, Lin and Breslow 1996, and Pinheiro and Chao 2006). A further disadvantage of this method is the absence of a true log likelihood, which complicates model comparisons and model selection based on information criteria.

Integral approximation methods that numerically evaluate the marginal log likelihood of the data can provide estimates with better statistical properties; they also provide the value of the log likelihood at the solution for testing and model comparisons. Integral approximation in mixed models with nonlinearity has been the purview of the NLMIXED procedure in SAS/STAT. The NLMIXED procedure assumes a general nonlinear mean function and cannot take advantage of the syntactic and computational simplifications that arise when it is known that the mean function contains a function that is linear in the parameters. The GLIMMIX procedure implements two integral approximation techniques in SAS 9.2 through the METHOD= option in the PROC GLIMMIX statement—Laplace and quadrature approximation.

PROC GLIMMIX METHOD=LAPLACE

If you choose METHOD=LAPLACE with a generalized linear mixed model, the GLIMMIX procedure approximates the marginal likelihood by using Laplace's method. Twice the negative of the resulting log likelihood approximation is the objective function that the procedure minimizes to determine parameter estimates. However, the class of models for which a Laplace approximation of the marginal log likelihood is possible is smaller compared to the class of models to which pseudo-likelihood estimation can be applied.

In order to construct a Laplace approximation, PROC GLIMMIX requires a conditional log likelihood, $\log\{p(\mathbf{y}|\boldsymbol{\gamma})\}$, and the log likelihood for the distribution of the G-side random effects, $\log\{p(\boldsymbol{\gamma})\}$. The random-effects distribution is assumed to be normal with zero mean and covariance structure determined by the RANDOM statement. The conditional distribution of the data given the random effects, $\log\{p(\mathbf{y}|\boldsymbol{\gamma})\}$, is determined by the DIST= option in the MODEL statement or the default associated with a particular response type. Because a valid conditional distribution is required, R-side random effects are not permitted with METHOD=LAPLACE.

The Laplace approximation implemented in the GLIMMIX procedure differs from that in Wolfinger (1993) and Pinheiro and Bates (1995). Their modified Laplacian approximation replaces the second derivative $f''(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta}; \widehat{\boldsymbol{\gamma}})$ with an (approximate) expected value, akin to scoring. The GLIMMIX procedure computes the full second derivative and does not include the fixed effects in the Laplace suboptimization. As a consequence, the optimization is carried out for the covariance parameters *and* the fixed effects. Referring to likelihood analysis in the MIXED procedure when data are missing at random, Verbeke and Molenberghs (1997, p. 243) call it a "pity" that standard errors based on inverting the entire Hessian (involving fixed effects and covariance parameters) are not available in that procedure. These standard errors are now available through maximum likelihood estimation based on Laplace and quadrature approximation in the GLIMMIX procedure.

PROC GLIMMIX METHOD=QUAD<(*quad-options*)>

If you choose METHOD=QUAD with a generalized linear mixed model, the GLIMMIX procedure approximates the marginal log likelihood by using an adaptive Gauss-Hermite quadrature. The *quad-options* enable you to control aspects of the quadrature (e.g., the number of nodes, the node search sequence, the method of obtaining starting values) and to request accuracy checks at convergence. For a discussion of the *quad-options* and the mathematical details about Gauss-Hermite quadrature in general and the adaptive quadrature implemented in PROC GLIMMIX, see the "Details" section of the PROC GLIMMIX documentation and the references therein. We note here only the following details:

- The quadrature rule in the GLIMMIX procedure is adaptive in the following sense:
  - Unless METHOD=QUAD(QPOINTS=n) is specified, the number of quadrature points is determined by evaluating the log likelihood at the starting values at an increasing number of nodes until successive log-likelihood evaluations change less than a tolerance value.
  - The GLIMMIX procedure centers and scales the quadrature points by using the empirical Bayes estimates (EBEs) of the random effects and the Hessian (second derivative) matrix from the EBE suboptimization to improve the likelihood approximation.

- The objective function for adaptive quadrature estimation in PROC GLIMMIX is constructed as follows. Suppose that $\boldsymbol{\beta}$ denotes the vector of fixed-effects parameters and that $\boldsymbol{\theta} = [\boldsymbol{\theta}^* \, \phi]$ denotes the vector of covariance parameters. The only R-side parameter is the scale parameter ϕ, provided that the conditional distribution has a scale. The marginal distribution of the data in a mixed model for subject i with random effects $\boldsymbol{\gamma}_i$ can be written as

$$p(\mathbf{y}_i) = \int \cdots \int p\left(\mathbf{y}_i | \boldsymbol{\gamma}_i, \boldsymbol{\beta}, \phi\right) p\left(\boldsymbol{\gamma}_i | \boldsymbol{\theta}^*\right) \, d\boldsymbol{\gamma}_i$$

For each subject, obtain the empirical Bayes estimates of $\boldsymbol{\gamma}_i$ as the vector $\widehat{\boldsymbol{\gamma}}_i$ that minimizes $-\log\left\{p(\mathbf{y}_i | \boldsymbol{\gamma}_i, \boldsymbol{\beta}, \phi) \, p(\boldsymbol{\gamma}_i | \boldsymbol{\theta}^*)\right\} = f\left(\mathbf{y}_i, \boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{\gamma}_i\right)$. This step comprises the quadrature (and Laplace) suboptimization. Suppose N_q denotes the number of quadrature points in each dimension, r denotes the number of random effects, $\mathbf{a}_j^*$ are the centered and scaled abscissas, and $\mathbf{w} = [w_1, \cdots, w_{N_q}]$ is the vector of Gauss-Hermite quadrature weights. The r-dimensional integral is then constructed by a sequence of one-dimensional rules

$$
\begin{aligned}
p(\mathbf{y}_i) &= \int \cdots \int p(\mathbf{y}_i | \boldsymbol{\gamma}_i, \boldsymbol{\beta}, \phi) \, p(\boldsymbol{\gamma}_i | \boldsymbol{\theta}^*) \, d\boldsymbol{\gamma}_i \\
&\approx 2^{r/2} |f''(\mathbf{y}_i, \boldsymbol{\beta}, \boldsymbol{\theta}; \widehat{\boldsymbol{\gamma}}_i)|^{-1/2} \sum_{j_1=1}^{N_q} \cdots \sum_{j_r=1}^{N_q} \left[p(\mathbf{y}_i | \mathbf{a}_j^*, \boldsymbol{\beta}, \phi) p(\mathbf{a}_j^* | \boldsymbol{\theta}^*) \prod_{k=1}^{r} w_{jk} \exp z_{j_k}^2 \right]
\end{aligned}
$$

where f'' denotes the second derivative matrix with respect to the random effects. The right-hand side of this expression, accumulated across subjects, is the objective function for adaptive quadrature estimation in the GLIMMIX procedure.

If you select the quadrature rule with a single node, namely METHOD=QUAD(QPOINTS=1), the results are identical to those of METHOD=LAPLACE. Computationally, however, the two methods are not identical. METHOD=LAPLACE can be applied to a larger class of models. For example, crossed random effects, models without subjects, or models with non-nested subjects can be handled with the Laplace approximation but not with quadrature. Furthermore, METHOD=LAPLACE draws on a number of computational simplifications that can increase its efficiency compared to a quadrature algorithm with a single node. For example, the Laplacian approximation is possible with unbounded covariance parameter estimates (NOBOUND option in the PROC GLIMMIX statement).

Compared to METHOD=LAPLACE, the models for which parameters can be estimated by quadrature are further restricted. In addition to the conditional independence assumption and the absence of R-side covariance parameters, it is required that models suitable for METHOD=QUAD can be processed by subjects.

Example

The following example demonstrates estimation by using quadrature with built-in and user-specified distributions. The DATA step in the example creates a SAS data set from a count process (the data are simulated). For each of $i = 1, \cdots, 18$ randomly select subjects, a randomly varying number n_i of observations were drawn from a count regression model with a single covariate.

```
data counts;
   input ni @@;
   sub = _n_;
   do i=1 to ni;
      input x y @@; output;
   end;
   datalines;
 6  2 0 82 5 33 0 15 2 35 0 79 1
 9 18 0 64 0 80 2  0 0 58 0  7 0 81 0 22 3 50 0
 2 34 0 95 0
```

```
13 28 1 31 0 63 2 14 0 74 0 44 0 75 3 65 0 74 1 84 5 57 0 29 0 41 0
 9 42 0  8 0 91 0 20 0 23 0 22 0 96 4 83 3 56 0
 3 64 0 64 1 15 0
 4  5 0 73 2 50 2 13 1
 2  0 0 41 0
 5 83 7 98 1 11 1 28 0 18 0
 7 91 0 25 1 51 4 20 0 61 1 34 0 33 2
14 60 0 87 0 94 0 29 0 41 0 78 0 50 0 37 0 15 0 39 0 22 0 82 0 93 0 3 0
12 68 0 26 1 19 0 60 1 93 3 65 0 16 0 79 0 14 0  3 1 90 0 28 3
 8 34 1 44 0 62 3 21 0  7 0 17 1  0 2 49 0
13 11 0 27 2 16 1 12 3 52 1 55 0  2 6 89 5 31 5 28 3 51 5 54 13 64 0
 9  3 0 36 0 57 0 77 0 41 0 39 0 55 0 57 0 88 1
 7  2 0 80 0 41 1 20 0  2 0 27 0 40 0
13 59 0 96 2 47 1 64 0 18 0 30 0 37 0 36 1 69 0 78 1 47 1 86 0 88 0
12 84 6 60 1 33 1 92 0 38 4  6 0 43 3 13 2 18 0 51 0 50 4 68 0
;
```

The next statements fit a model with conditional Poisson distribution and subject-specific random effects by maximum likelihood, using quadrature. Note that it is not necessary to specify starting values for the fixed effects or the covariance parameters in PROC GLIMMIX.

```
proc glimmix data=counts method=quad;
   class sub;
   model y = x / dist=poisson;
   random int / subject=sub;
run;
```

The quadrature setting is reflected in the "Model Information" table (Figure 1).

```
                    The GLIMMIX Procedure

                    Model Information

        Data Set                   WORK.COUNTS
        Response Variable          y
        Response Distribution      Poisson
        Link Function              Log
        Variance Function          Default
        Variance Matrix Blocked By  sub
        Estimation Technique       Maximum Likelihood
        Likelihood Approximation   Gauss-Hermite Quadrature
        Degrees of Freedom Method  Containment


                    Optimization Information

          Optimization Technique      Dual Quasi-Newton
          Parameters in Optimization  3
          Lower Boundaries            1
          Upper Boundaries            0
          Fixed Effects               Not Profiled
          Starting From               GLM estimates
          Quadrature Points           5
```

Figure 1. Model and Optimization Information

The "Optimization Information" table in Figure 1 shows two important aspects of quadrature and Laplace techniques in PROC GLIMMIX:

- The covariance parameters as well as the fixed effects are part of the optimization (same as PROC NLMIXED). The three parameters are the fixed intercept, the fixed slope, and the variance of the random intercept.

- The procedure determined by using the starting values that five quadrature nodes are adequate to approximate the marginal log likelihood with sufficient accuracy.

With METHOD=LAPLACE or METHOD=QUAD, the GLIMMIX procedure computes fit statistics for the marginal as well as the conditional distribution (Figure 2). Minus twice the marginal log likelihood at the converged estimates is 412.80. Because it is not a pseudo-likelihood, you can use this value for comparisons of nested models, and you can use the information criteria for comparisons of non-nested models. The scaled Pearson statistic for the conditional distribution (1.46) suggests that there is mild overdispersion in the conditional distribution.

```
                        Fit Statistics

            -2 Log Likelihood              412.80
            AIC  (smaller is better)       418.80
            AICC (smaller is better)       418.97
            BIC  (smaller is better)       421.48
            CAIC (smaller is better)       424.48
            HQIC (smaller is better)       419.17


               Fit Statistics for Conditional
                         Distribution

            -2 log L(y | r. effects)       368.56
            Pearson Chi-Square             216.34
            Pearson Chi-Square / DF          1.46
```

Figure 2.   Fit Statistics for Gauss-Hermite Quadrature

Suppose that you want to adjust the model by changing the conditional distribution from a Poisson to a negative binomial distribution. The variance function of a negative binomial random variable is $\text{Var}[Y] = \mu + k\mu^2$. The variance function of the Poisson distribution is a special case, $k = 0$.

You can estimate the parameters of a mixed model in which the conditional distribution is negative binomial by maximum likelihood with the following statements:

```
proc glimmix data=counts method=quad;
   class sub;
   model y = x / dist=negbin s;
   random int / subject=sub;
run;
```

Results from these statements are shown in Figures 3 and 4. The negative binomial mixed model has a lower -2 log likelihood than the Poisson mixed model. However, although for $k = 0$ the two distributions have identical variance functions, the log-likelihood functions are not nested. A likelihood ratio test comparing the -2 log likelihoods in Figures 2 and 3 is not possible. The information criteria do allow a comparison of non-nested models, however. All criteria in Figure 3 indicate that the negative binomial distribution provides a better fit.

```
                    The GLIMMIX Procedure

                        Fit Statistics

            -2 Log Likelihood              368.74
            AIC  (smaller is better)       376.74
            AICC (smaller is better)       377.02
            BIC  (smaller is better)       380.30
            CAIC (smaller is better)       384.30
            HQIC (smaller is better)       377.23


                 Fit Statistics for Conditional
                          Distribution

            -2 log L(y | r. effects)       337.95
            Pearson Chi-Square              99.25
            Pearson Chi-Square / DF          0.67
```

Figure 3. Fit Statistics for Gauss-Hermite Quadrature in Negative Binomial Model

The estimate for the parameter k of the negative binomial distribution is shown as "Scale" in Figure 4. A formal likelihood ratio test to address the question of whether the variance of the subject-specific random intercepts is significantly larger than zero is performed later in this paper by using the new COVTEST statement.

```
                    Covariance Parameter Estimates

                                            Standard
            Cov Parm      Subject    Estimate      Error

            Intercept     sub          0.8257      0.4796
            Scale                      1.2535      0.3973


                    Solutions for Fixed Effects

                        Standard
        Effect        Estimate       Error       DF     t Value    Pr > |t|

        Intercept      -1.2577      0.4022        17      -3.13      0.0061
        x              0.01683     0.005292       129      3.18      0.0018
```

Figure 4. Estimates of Covariance Parameters and Fixed Effects in Negative Binomial Model

Note that in order for the GLIMMIX procedure to compute an adaptive quadrature approximation to the marginal log likelihood, the data must be processed by subject. The following syntax specifies an equivalent model to the Poisson mixed model fit initially. But this syntax would not work for METHOD=QUAD.

```
proc glimmix data=counts method=laplace;
   class sub;
   model y = x / dist=poisson;
   random sub;
run;
```

Comparisons to the NLMIXED Procedure

The NLMIXED procedure also fits mixed models by using adaptive quadrature. If you choose QPOINTS=1 in the PROC NLMIXED statement, a Laplace approximation results. For models that can be fit with either procedure, the results for Laplace and quadrature approximations generally agree. However, there are a number of differences between the two procedures with respect to Laplace and quadrature approximations:

- PROC NLMIXED treats the Laplace approximation as a special case of the quadrature (with a single node). PROC GLIMMIX has a dedicated algorithm for METHOD=LAPLACE that takes advantage of additional computational simplifications. This allows more complex models and additional processing options, such as NOBOUND.

- PROC GLIMMIX has a full-featured mechanism to generate starting values. With METHOD=QUAD or METHOD=LAPLACE, starting values for the fixed effects are obtained from several iterations of a GLM model. Starting values for the covariance parameters are then determined by a MIVQUE0 step (Goodnight 1978). In addition, you can request that pseudo-likelihood steps precede the maximum likelihood step. With the NLMIXED procedure, starting values for all model parameters must be supplied with the PARMS statement; otherwise the default value of 1.0 is assigned.

- PROC GLIMMIX provides the METHOD=QUAD(QCHECK) option to recompute the log likelihood at the solution by using an increasing number of quadrature points. This enables you to study the stability of the integral computation at the solution.

- PROC GLIMMIX produces fit statistics for the marginal distribution (as PROC NLMIXED) and the conditional distribution (given the random effects).

When you request estimation by quadrature in nonlinear or generalized linear mixed models, you need to be aware of the computational limitations of the approach; these limitations apply to both procedures. As the number of random effects grows, quadrature quickly becomes computationally infeasible due to the high dimension of the integral. Consider the following statements and suppose that the A effect has four levels:

```
proc glimmix method=quad(qpoints=5);
   class A B id;
   model y = / dist=negbin;
   random A / subject=id;
run;
```

Computing the marginal log likelihood requires the numerical evaluation of a four-dimensional integral. As part of this evaluation, $5^4 = 625$ conditional log likelihoods need to be computed per observation on each pass through the data. Now suppose that an additional random effect with $b = 2$ levels is added and the previous RANDOM statement becomes the following:

```
   random A A*B / subject=id;
```

For each observation, this now requires evaluation of $5^{(4+8)} = 244140625$ conditional log likelihoods on each pass through the data. In contrast, the computational burden of the Laplace approximation does not depend as strongly on the dimension of the integral.

Empirical ("Sandwich") Estimators

In SAS 9.1, the GLIMMIX procedure offers the classical empirical ("sandwich") estimators for fixed effects and four bias-corrected estimators through the EMPIRICAL option in the PROC GLIMMIX statement. Bias-corrected forms of the empirical estimators are indicated particularly for small samples because the cross-product matrix of fitted residuals is not an unbiased estimator of the marginal variance matrix. There are two major enhancements in SAS 9.2 with respect to empirical estimators:

- If you fit a model by maximum likelihood with Laplace or adaptive quadrature approximation, and if you choose the classical empirical estimator in the PROC GLIMMIX statement, the procedure computes a likelihood-based "sandwich" estimator. This estimator is constructed from the Hessian matrix of the optimization and the outer product of the subject-specific gradients (White 1982). Since models that are fit with Laplace or quadrature approximation include the fixed effects and the covariance parameters in the optimization, the likelihood-based empirical estimator also adjusts the approximate covariance matrix of the covariance parameters.

- The GLIMMIX procedure now computes the residual-based design-adjusted empirical estimator by Morel (1989) and Morel, Bokossa, and Neerchal (2003) with the EMPIRICAL=MBN option in the PROC GLIMMIX statement. The estimator rests on an additive correction of the residual cross-products and a sample size correction. Besides having good statistical properties in terms of Type I error rates in situations where the number of independent sampling units (subjects, m) is small, the MBN estimator also has the desirable property to recover rank. The rank of the classical estimator is governed by the number of subjects that contribute to the sum in the "meat" part of the sandwich. When m is small, the classical estimator can suffer from a loss of rank that can impair inference for effects involving classification variables. The correction in the MBN estimator offsets the rank degradation. See the PROC GLIMMIX documentation for mathematical expressions and details.

MODEL BUILDING

Experimental EFFECT Statement

The experimental EFFECT statement enables you to construct special sets of columns for **X** or **Z** matrices in your model. These collections are referred to as *constructed effects* to distinguish them from the usual (old-style) model effects formed from continuous or classification variables and specified via the MODEL statement. You define a constructed effect through the EFFECT statement. For example, the EFFECT statement in the following program defines a constructed effect named spl. Its columns are formed from the data set variable x as a cubic B-spline basis with three equally spaced interior knots.

```
proc glimmix;
   class A B sub;
   effect spl = spline(x);
   model y = A B A*spl;
   random A*B / subject=sub;
run;
```

The general syntax for the EFFECT statement is as follows:

   EFFECT　*effect-name* = **effect-type** *(var-list < / options >)***;**

You can specify multiple EFFECT statements, and they must precede the MODEL statement. In SAS 9.2, the GLIMMIX procedure supports the following effect types in the EFFECT statement:

MULTIMEMBER　specifies a multimember classification effect whose levels are determined by one or more variables that appear in a CLASS statement. It is formed in such a way that each observation can be associated with one or more levels of the union of the levels of the classification variables. Multimember effects are useful, for example, in modeling the following effects:

- nurses' effects on patient recovery in hospitals
- teachers' effects on student scores
- lineage effects in genetic studies (see following example)

COLLECTION　specifies an effect that consists of one or more variables that are considered for estimation and inference as a unit with multiple degrees of freedom.

SPLINE   specifies a regression spline effect whose columns are univariate spline transformations of one or more variables. A spline transformation is a variable expansion where the original variable is expanded into a larger set of new variables. You can choose between a B-spline basis (default) and a truncated power function basis.

The experimental EFFECT statement is also available in the GLMSELECT and QUANTREG procedures in SAS 9.2.

Example

Cockerham and Weir (1977) apply variance component models in the analysis of reciprocal crosses. In these experiments it is of interest to separate genetically determined variation from variation determined by parentage. The basic model for a cross is

$$Y_{ijk} = \beta + \alpha_{ij} + \epsilon_{ijk}$$

where Y_{ijk} is the observation for offspring k from maternal parent i and paternal parent j. Their model (c) decomposes α_{ij}, the effect attributable to the parents, into

$$\alpha_{ij} = \eta_i + \eta_j + \mu_i + \phi_j + (\eta\eta)_{ij} + \kappa_{ij}$$

where η_i and η_j are contributions of the female and male parents, respectively, $(\eta\eta)_{ij}$ is an interaction term, and μ_i, ϕ_j are extranuclear maternal and paternal effects. The remaining interactions are captured by the κ_{ij} term.

The following DATA step creates a SAS data set for the diallel example in Cockerham and Weir (1977). A diallel is a mating design that consists of all possible crosses of a set of parental lines. The observations represent mean flowering times of *Nicotiana rustica* (Aztec tobacco) from crosses of inbred varieties grown in two blocks. The variables p and m identify the eight paternal and maternal lines, respectively. The variable sym is used to model the interaction between the parents, subject to $(\eta\eta)_{ij} = (\eta\eta)_{ji}$ (Figure 5).

```
data diallel;
   label time = 'Flowering time in days';
   do p = 1 to 8;
      do m = 1 to 8;
         if (m ne p) then do;
            sym = trim(left(min(m,p))) || ',' || trim(left(max(m,p)));
            do block = 1 to 2; input time @@; output; end;
         end;
      end;
   end;
   datalines;
14.4 16.2 27.2 30.8 17.2 27.0 18.3 20.2 16.2 16.8 18.6 14.4 16.4 16.0
15.4 16.5 14.8 14.6 18.6 18.6 15.2 15.3 17.0 15.2 14.4 14.8 10.8 13.2
31.8 30.4 21.0 23.0 24.6 25.4 19.2 20.0 29.8 28.4 12.8 14.2 13.0 14.4
16.2 17.8 11.4 13.0 16.8 16.3 12.4 14.2 16.8 14.8 12.6 12.2  9.6 11.2
14.6 18.8 12.2 13.6 15.2 15.4 15.2 13.8 18.0 16.0 10.4 12.2 13.4 20.0
20.2 23.4 14.2 14.0 18.6 14.8 22.2 17.0 14.3 17.3  9.0 10.2 11.8 12.8
14.0 16.6 12.2  9.2 13.6 16.2 13.8 14.4 15.6 15.6 15.6 11.0 13.0  9.8
15.2 17.2 10.0 11.6 17.0 18.2 20.8 20.8 20.0 17.4 17.0 12.6 13.0  9.8
;
```

The following PROC GLIMMIX statements fit model (c) in Cockerham and Weir (1977). The EFFECT statement defines the nuclear parental contributions as a multimember effect based on the CLASS variables p and m. Each observation has two nonzero entries in the design matrix for the effect that identifies the

paternal and maternal lines. The terms in the RANDOM statement model the variance components as follows: line $\to \sigma_\eta^2$, sym $\to \sigma_{(\eta\eta)}^2$, p $\to \sigma_\phi^2$, m $\to \sigma_\mu^2$, p*m $\to \sigma_\kappa^2$. The new OUTDESIGN= option in the PROC GLIMMIX statement writes the **Z** matrix to the SAS data set zmat. The EFFECT statement alleviates the need for complex coding, as in Section 2.3 of Saxton et al. (2004).

```
proc glimmix data=diallel outdesign(z)=zmat;
    class block sym p m;
    effect line = mm(p m);
    model  time = block;
    random line sym p m p*m;
run;
```

```
                   The GLIMMIX Procedure

                  Class Level Information

     Class    Levels   Values

     block         2   1 2
     sym          28   1,2 1,3 1,4 1,5 1,6 1,7 1,8 2,3 2,4 2,5 2,6
                       2,7 2,8 3,4 3,5 3,6 3,7 3,8 4,5 4,6 4,7 4,8
                       5,6 5,7 5,8 6,7 6,8 7,8
     p             8   1 2 3 4 5 6 7 8
     m             8   1 2 3 4 5 6 7 8
```

Figure 5. Class Levels in Diallel Example

The estimates of the variance components and the residual variance in Figure 6 agree with the results in Table 7 of Cockerham and Weir (1977).

```
              Covariance Parameter Estimates

                                 Standard
          Cov Parm     Estimate     Error

          line           5.1047    4.0021
          sym            2.3856    1.9025
          p              3.3080    3.4053
          m              1.9134    2.9891
          p*m            4.0196    1.8323
          Residual       3.6225    0.6908
```

Figure 6. Covariance Parameter Estimates

The following statements print the **Z** matrix columns that correspond to the multimember line effect for the first 10 observations in block 1 (Figure 7). For each observation there are two nonzero entries, and their column index corresponds to the index of the paternal and maternal parent.

```
proc print data=zmat(where=(block=1) obs=10);
  var p m time _z1-_z8;
run;
```

```
Obs    p    m    time    _Z1    _Z2    _Z3    _Z4    _Z5    _Z6    _Z7    _Z8

  1    1    2    14.4     1      1      0      0      0      0      0      0
  3    1    3    27.2     1      0      1      0      0      0      0      0
  5    1    4    17.2     1      0      0      1      0      0      0      0
  7    1    5    18.3     1      0      0      0      1      0      0      0
  9    1    6    16.2     1      0      0      0      0      1      0      0
 11    1    7    18.6     1      0      0      0      0      0      1      0
 13    1    8    16.4     1      0      0      0      0      0      0      1
 15    2    1    15.4     1      1      0      0      0      0      0      0
 17    2    3    14.8     0      1      1      0      0      0      0      0
 19    2    4    18.6     0      1      0      1      0      0      0      0
```

Figure 7. Z Matrix for Line Effect of the First 10 Observations in Block 1

Covariance Structures

A number of covariance structures have been added to the GLIMMIX procedure in SAS 9.2. With the addition of these structures you can draw on nearly the same covariance models as in the MIXED procedure. The following exceptions should be noted:

- PROC MIXED does not support the TYPE=CHOL and TYPE=PSPLINE covariance structures. The PSPLINE type is new in the GLIMMIX procedure in SAS 9.2. It is used to fit penalized B-splines (P-spline; Eilers and Marx 1996) with mixed model technology.
- The GLIMMIX procedure supports the Handcock and Stein (1993) and Handcock and Wallis (1994) parameterization of the Matérn family of covariance functions (Matérn 1986). The TYPE=SP(MAT) structure in PROC GLIMMIX corresponds to TYPE=SP(MATHSW) in the MIXED procedure.
- The GLIMMIX procedure does not support the GDATA= option and the Kronecker-type covariance structures. The GLIMMIX procedure does not support the SP(EXPA) and the geometric anisotropic structures (SP(EXPGA), SP(GAUGA), SP(SPHGA)).

AFTER THE MODEL FIT: POST-PROCESSING

COVTEST Statement

The new COVTEST statement provides a mechanism to obtain statistical inferences for the covariance parameters in a mixed model by likelihood-based tests comparing full and reduced models with respect to the covariance parameters. The comparisons of the models are based on the log likelihood or restricted log likelihood in models that are fit by maximum likelihood (ML) or restricted maximum likelihood (REML). With pseudo-likelihood methods, the calculations are based on the final pseudo-data of the converged optimization. Confidence limits and bounds are also available through the COVTEST statement. For details about the computation of likelihood-ratio tests and confidence bounds for covariance parameters, see the documentation for the GLIMMIX procedure in the *SAS/STAT User's Guide*.

You can specify multiple COVTEST statements. The general syntax of a COVTEST statement is as follows:

> **COVTEST** $<$*'label'*$>$ $<$*test-specification*$>$ $<$ */ options* $>$**;**

The *test-specification* is used to formulate the constraints on the covariance parameters—in other words, to formulate the hypothesis. Test specifications can be made by using keywords that represent a model condition, by specifying the values for covariance parameters, or by formulating linear contrasts.

The simplest approach is to use the predefined keywords. For example, the following statements fit a normal linear mixed model by maximum likelihood with two random effects. The covariance matrix of the random effects, the **G** matrix, is an unstructured (2×2) matrix.

```
proc glimmix method=mspl;
   class treat id;
   model y = treat*time;
   random intercept time / sub=id type=un;
run;
```

To compare the fitted model to a linear model without random effects, the **G** matrix must be eliminated from the model. This can be done, for example, by adding the following COVTEST statements:

```
covtest 'Ho: No ran eff.'   GLM;
covtest 'Ho: No ran eff.' ZEROG;
```

The keyword ZEROG requests that the reduced model be formed from the fitted model by imposing restrictions that reduce the **G** matrix to zero. The keyword GLM specifies that the mixed model be reduced to a generalized linear model. The keywords lead to identical model reductions in this case. Other predefined keywords enable you to test for a diagonal **G** matrix (DIAGG), to test for conditional independence (DIAGR), or to test for homogeneity of covariance parameters across groups (HOMOGENEITY). You can also specify values for the covariance parameters explicitly. For example, the following statement asks that, in the reduced model, the second covariance parameter be fixed at zero:

```
covtest 'Ho: Indep. ran eff.' . 0 . ;
```

Not specifying a parameter, or supplying a missing value, indicates that the parameter is not restricted beyond constraints imposed by the full model. Since the covariance structure is a (2×2) unstructured matrix, the previous COVTEST statement eliminates the covariance between the random intercept and the random slope.

Examples

In an earlier section, a mixed model for count data was fit by quadrature with conditional Poisson and conditional negative binomial distribution. To test whether the random intercepts vary significantly across subjects in the negative binomial model, you can add the following statements to the negative binomial fit:

```
covtest 'H: Var[random int]=0' 0;
covtest 'H: Var[random int]=0' GLM;
covtest 'H: Var[random int]=0' ZeroG;
```

The three statements are alternative ways of testing the same hypothesis; they produce Figure 8. The -2 log likelihood reported in Figure 8 is that for a negative binomial generalized linear model with regressor x. The single degree of freedom represents a single constraint that is imposed on the full model.

When parameters under the null hypothesis fall on the boundary of the parameter space, the distribution of the likelihood ratio statistic can be a complicated mixture of distributions. In certain situations it is known to be a straightforward mixture of central chi-square distributions. When the GLIMMIX procedure recognizes the model and hypothesis as a case for which the mixture is readily available, the *p*-value of the likelihood ratio test is determined accordingly as a linear combination of central chi-square probabilities. This is the case in this example. Because the null hypothesis specifies a value on the boundary of the parameter space for the radnom effects variance, the *p*-value in Figure 8 was adjusted based on Self and Liang (1987). This is indicated in the "Note" column and in the table footnote.

```
                        The GLIMMIX Procedure

                     Tests of Covariance Parameters
                         Based on the Likelihood

    Label                    DF    -2 Log Like      ChiSq    Pr > ChiSq    Note

    H: Var[random int]=0      1         385.31      16.57       <.0001     MI
    H: Var[random int]=0      1         385.31      16.57       <.0001     MI
    H: Var[random int]=0      1         385.31      16.57       <.0001     MI

  MI: P-value based on a mixture of chi-squares.
```

Figure 8.   Likelihood Ratio Test in Negative Binomial Model

A third method of specifying model reductions with respect to the covariance parameters is to use general contrasts. Consider the following example of a repeated measures study with four subjects in each of four treatment groups and four repeated measurements:

```
data covtex;
   array yvec{4} y0-y3;
   input tx $ y0 y1 y2 y3 @@;
   id = _n_;
   do time=0 to 3;
     y = log(yvec{time+1}); output;
   end;
   keep y time id tx;
   datalines;
A  .04   .20   .10   .08      A  .02   .06   .02   .02
A  .07 1.40   .48   .24      A  .17   .57   .35   .24
B  .10   .09   .13   .14      B  .12   .11   .10   .11
B  .07   .07   .07   .07      B  .05   .07   .06   .07
C  .03   .62   .31   .22      C  .03 1.05   .73   .60
C  .07   .83 1.07   .80      C  .09 3.13 2.06 1.23
D  .10   .09   .09   .08      D  .08   .09   .09   .10
D  .13   .10   .12   .12      D  .06   .05   .05   .05
;
```

The following GLIMMIX statements fit the model with fixed treatment and time effects, their interaction, and an unstructured residual (R-side) covariance matrix:

```
ods select FitStatistics CovParms CovTests;
proc glimmix data=covtex;
   class id tx time;
   model y = tx time tx*time;
   random _residual_ / sub=id type=un;
   covtest 'Comp. Symm.' contrast 1 0 -1                     ,
                                  0 0  1  0  0 -1            ,
                                  0 0  0  0  0  1  0  0  0 -1,
                                  0 1  0 -1                   ,
                                  0 1  0  0 -1               ,
                                  0 1  0  0  0  0 -1         ,
                                  0 1  0  0  0  0  0 -1      ,
                                  0 1  0  0  0  0  0  0 -1    / est;
   run;
```

The full model has a -2 residual log likelihood of 31.93 and 10 covariance parameters (Figure 9).

```
                        The GLIMMIX Procedure

                          Fit Statistics

             -2 Res Log Likelihood            31.93
             AIC  (smaller is better)         51.93
             AICC (smaller is better)         57.88
             BIC  (smaller is better)         59.66
             CAIC (smaller is better)         69.66
             HQIC (smaller is better)         52.33
             Generalized Chi-Square          48.00
             Gener. Chi-Square / DF           1.00


                  Covariance Parameter Estimates

         Cov                                  Standard
         Parm        Subject    Estimate       Error

         UN(1,1)     id            0.3509      0.1433
         UN(2,1)     id            0.3498      0.1685
         UN(2,2)     id            0.6228      0.2542
         UN(3,1)     id            0.4407      0.1946
         UN(3,2)     id            0.6433      0.2702
         UN(3,3)     id            0.7417      0.3028
         UN(4,1)     id            0.3814      0.1677
         UN(4,2)     id            0.5296      0.2275
         UN(4,3)     id            0.6299      0.2586
         UN(4,4)     id            0.5468      0.2232
```

Figure 9.   Fit Statistics and Covariance Parameter Estimates in Full Model

The COVTEST statement restricts the covariance parameters in such a way that the resulting covariance matrix has a compound-symmetric structure: common variance and common covariance. For example, the first three rows of the contrast impose equality of the four variances through three pairwise equality constraints:

$$\left.\begin{array}{l} \sigma_1^2 = \sigma_2^2 \\ \sigma_2^2 = \sigma_3^2 \\ \sigma_3^2 = \sigma_4^2 \end{array}\right\} \Leftrightarrow \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$$

The next five rows similarly set the covariances equal. The EST option in the COVTEST statement requests a listing of the estimated covariance parameters under the hypothesis.

The results of the COVTEST statement are shown in Figure 10. The -2 residual log likelihood under the restrictions is 71.189, and the difference from the full model is highly significant ($p < 0.0001$). The reduced model does not fit the data as well as the model with an unstructured covariance matrix. Under the restriction there are only two free parameters, whose estimates are 0.566 and 0.496.

```
                   Tests of Covariance Parameters
                   Based on the Restricted Likelihood

                                              -----Estimates H0-----
Label          DF  -2 Res Log Like    ChiSq  Pr > ChiSq  Est1    Est2    Est3

Comp. Symm.     8          71.1890    39.26      <.0001  0.566   0.496   0.566

                   Tests of Covariance Parameters
                   Based on the Restricted Likelihood

              --------------------Estimates H0--------------------
    Label       Est4    Est5    Est6    Est7    Est8    Est9    Est10   Note

    Comp. Symm.  0.496   0.496   0.566   0.496   0.496   0.496   0.566  DF

DF: P-value based on a chi-square with DF degrees of freedom.
```

Figure 10.   REML Likelihood Ratio Test of Compound Symmetry Hypothesis

Suppose you replace the previous RANDOM statement with the following:

```
random _residual_ / sub=id type=cs;
```

Then the fitted model will have -2 Res Log Likelihood = 71.189 and covariance parameters "CS" = 0.4958 and "Residual" = 0.0698; their sum is 0.566.

Another application of the general contrast formulation in the COVTEST statement is a test of sphericity. If the unstructured covariance matrix satisfies the Huynh-Feldt condition (Huynh and Feldt 1970), then

$$\text{Cov}[Y_i, Y_j] = \left\{ \begin{array}{ll} \sigma_i^2 & i = j \\ 0.5(\sigma_i^2 + \sigma_j^2) - \lambda & i \neq j \end{array} \right.$$

A test for sphericity of an unstructured (4×4) covariance matrix hence leads to the following contrast:

```
random _residual_ / sub=id type=un;
covtest 'Sphericity' contrast 0  -2  1  2  0 -1              ,
                              0  -2  1  0  0  0  2  0  0 -1,
                              1  -2  0  0  2 -1  0  0  0  0,
                              1  -2  0  0  0  0  0  2  0 -1,
                              0   0  1  0  0 -1  0 -2  2  0;
```

The full model fits the data significantly better than a model in which the covariance matrix satisfies the Huynh-Feldt condition ($p = 0.0012$; see Figure 11).

```
                        The GLIMMIX Procedure

                    Tests of Covariance Parameters
                    Based on the Restricted Likelihood

        Label           DF    -2 Res Log Like     ChiSq    Pr > ChiSq    Note

        Sphericity       5            52.0401      20.11        0.0012    DF

DF: P-value based on a chi-square with DF degrees of freedom.
```

Figure 11. Test of Sphericity

Alternative Syntax for Coefficients In Linear Hypotheses

In SAS 9.2, the GLIMMIX procedure supports a new syntax for specifying coefficients in ESTIMATE, CONTRAST, and LSMESTIMATE statements. This syntax is termed "nonpositional," to distinguish it from the current form of entering coefficients where the position in the user's value list corresponds to a column in the **L** matrix of a linear hypothesis. The need for this new syntax arises with effects constructed through the experimental EFFECT statement, but it can also be useful in models without constructed effects.

The general syntax for creating **L** matrix coefficients by using the nonpositional form is as follows:

effect-name [*multiplier* $<,>$ *level-values*] ... $<$[*multiplier* $<,>$ *level-values*]$>$

The first entry in square brackets is the multiplier that is applied to the elements of **L** for the effect after the *level-values* have been resolved and any necessary action forming **L** has been taken. Note that you enter the **levels** of the classification variables in the square brackets, not their formatted values (see the "Class Level Information" table). The *level-values* are organized in a specific form:

1. The number of entries equals the number of terms needed to construct the effect.
2. Values of continuous variables precede the level indicators of CLASS variables.
3. If the effect involves constructed effects, then you need to provide as many continuous and classification variables as are needed for the effect formation.
4. If an effect contains both regular (old-style) and constructed effects, then the order of the coefficients is as follows: continuous values for old-style effects, levels for CLASS variables in old-style effects, continuous values for constructed effects, and, finally, class levels needed for constructed effects.

You can alternate between positional and nonpositional syntax by effect, choosing for each the needed or most convenient formulation. In the following example the model contains a classification effect and a B-spline. The ESTIMATE statements produce predicted values for level one of C when the continuous variable x takes on the values 20 and 10, respectively.

```
proc glimmix;
   class c;
   effect spl = spline(x);
   model y = c spl;
   estimate 'C = 1 @ x=20' intercept 1 c 1 spl [1,20],
            'C = 1 @ x=10' intercept 1 c 1 spl [1,10];
run;
```

The GLIMMIX procedure computes the spline coefficients for the first ESTIMATE statement based on $x = 20$, and similarly in the second statement for $x = 10$. The difference between the predicted values can be estimated with the following statement:

```
estimate 'Difference'   spl [1,20] [-1,10];
```

Because the spline effect does not interact with the classification variable, this difference does not depend on the level of C. Suppose such an interaction is present, as in the following statements. To estimate the difference in predicted values between levels 1 and 2 of C, you specify coefficients for the C*spl effect with nonpositional syntax. By rule 4 in the preceding list, values for the old-style terms are specified in the square brackets before values for constructed effects:

```
proc glimmix;
   class c;
   effect spl = spline(x);
   model y = spl*c;
   estimate 'C2 = 1, x=20' intercept 1 c*spl [1,1 20] / e;
   estimate 'C2 = 2, x=20' intercept 1 c*spl [1,2 20] / e;
   estimate 'C diff at x=20' c*spl [1,1 20] [-1,2 20] / e;
run;
```

You should add the E option to the CONTRAST, ESTIMATE, or LSMESTIMATE statement to verify that the **L** matrix is formed according to your expectations.

The nonpositional syntax can also be useful in models without constructed effects, such as in assigning values to interaction effects. Consider, for example, these statements:

```
class a b;
model y = a b a*b;
contrast 'B1 vs B2 @ A2' b 1 -1 a*b [1, 2 1] [-1, 2 2];
```

The nonpositional syntax for the a*b effect assigns the value 1 to the interaction where A is at level 2 and B is at level 1 and assigns the value -1 to the interaction where both class variables are at level 2.

Stepdown Multiplicity and Multiple Comparison P-values

In SAS 9.1, the experimental STEPDOWN option in the ESTIMATE, LSMESTIMATE, and LSMEANS statements provides step-down multiplicity and multiple comparison adjustments of *p*-values when the STEPDOWN option is combined with ADJUST=SIMULATE or ADJUST=BON. In SAS 9.2, step-down adjusted *p*-values are a production feature and are available for more adjustment methods. You can use step-down adjustments in SAS 9.2 as follows:

- For the ESTIMATE and LSMESTIMATE statements, step-down adjusted *p*-values are supported for all ADJUST= methods.
- For the LSMEANS statement, step-down adjusted *p*-values are supported for pairwise differences of least-squares means and all adjustment methods except ADJUST=NELSON. For certain options and data, step-down adjustments are exact under a homoscedastic, *iid* normal model for the dependent variable, particularly for the following:
  - for ADJUST=DUNNETT when the means are uncorrelated
  - for ADJUST=TUKEY with STEPDOWN(TYPE=LOGICAL) when the means are balanced and uncorrelated

The first case is a consequence of the nature of the successive step-down hypotheses for comparisons with a control; the second case employs an extension of the maximum studentized range distribution appropriate for partition hypotheses (Royen, 1989). The STEPDOWN(TYPE=FREE) ADJUST=TUKEY combination employs the Royen (1989) extension in such a way that the resulting *p*-values are conservative.

SOME OTHER ENHANCEMENTS AND MISCELLANEOUS NEW FEATURES

ODS Statistical Graphics

The following functional changes were made to existing ODS Statistical Graphics:

Diffogram   Two new suboptions change the behavior of the diffogram. The CENTER suboption marks the center point for each comparison. This point corresponds to the intersection of two least-squares means. The NOLINES suboption suppresses the display of the line segments that represent the confidence bounds for the differences of the least-squares means.

Mean plots   When least-squares means are plotted by another effect with the PLOTBY= suboption, the view is maintained across plots to more clearly show the relationship between least-squares means across levels of another factor.

Two new plot types were added in SAS 9.2 and can be requested with the PLOTS= option in the PROC GLIMMIX statement. The BOXPLOT option requests box plots for model effects that consist of classification effects. The ODDSRATIO option requests a display of odds ratios and their confidence limits when the link function permits the computation of odds ratios. See the "Syntax" section in the PROC GLIMMIX documentation for a complete list of the PLOTS= syntax and the plot-specific options for box plots and graphs of odds ratios.

FIRSTORDER Suboption for DDFM=KENWARDROGER

As part of the inference adjustments associated with the DDFM=KENWARDROGER option in the MODEL statements of the MIXED and GLIMMIX procedures, the covariance matrix of the fixed effects (and the prediction variance matrix of the random effects) is corrected to account for the fact that covariance parameters have been estimated (Kenward and Roger 1997). The adjustment to the covariance matrix of the fixed effects in Kenward and Roger (1997) is based on Kackar and Harville (1984) and Harville and Jeske (1992), but differs in that Kenward and Roger (1997) incorporate a second-order term, $\mathbf{R}_{ij}$, that does not appear in the approach by Harville and co-workers. In models where the second derivatives of the covariance structure do not vanish, the second-order term in the Kenward-Roger adjustment can have the following consequences:

- It can lead to shrinkage of the standard errors compared to the unadjusted results.

- The resulting covariance matrix is not guaranteed to be positive definite.

- The inferential results might not be invariant under reparameterization.

The new FIRSTORDER suboption of the DDFM=KENWARDROGER option in the MODEL statement of the GLIMMIX procedure drops the second derivative term from the calculations. The resulting covariance adjustment is denoted as $\widetilde{m}^{@}$ in Harville and Jeske (1992). Consider, for example, a model with TYPE=FA0(q) and TYPE=UN(q) parameterization of the covariance structures. Provided that the parameter estimates for the unstructured covariance matrix lead to a positive (semi-)definite matrix, the two models are equivalent. Although the log likelihoods and the solutions for fixed and random effects are identical, the Kenward-Roger adjustments are not the same, and this affects test statistics and *p*-values. Specifying the following statement in the model with FA0(q) covariance structure reestablishes the invariance, avoids possible shrinkage, and guarantees that the adjusted covariance matrix of the fixed effects is at least positive semi-definite:

```
model <  > / ddfm=kr(firstorder);
```

Covariance structures where the FIRSTORDER suboption makes a difference, are, for example, the ANTE(1), AR(1), ARH(1), ARMA(1,1), CHOL, CSH, FA0(q), TOEPH, and UNR structures and all spatial structures.

Comparing and Diagnosing Covariance Matrices

If you use an empirical estimator with the EMPIRICAL= option in the PROC GLIMMIX statement, the procedure replaces the model-based estimator of the covariance of the fixed effects with the sandwich estimator. Similarly, if you choose the DDFM=KENWARDROGER degrees-of-freedom method in the MODEL statement, PROC GLIMMIX adjusts the model-based covariance matrix of the fixed effects according to Kenward and Roger (1997) or according to Kackar and Harville (1984) and Harville and Jeske (1992). In this situation, the new COVB(DETAILS) suboption in the MODEL statement has two effects:

- The procedure displays the adjusted and the unadjusted covariance matrix of the fixed effects.
- The procedure displays a table of statistics for the unadjusted and adjusted covariance matrix and for their comparison. The table contains scalar matrix summaries, matrix norms, an analysis of the eigenvalues, and measures of matrix closeness and concordance (see, for example, Vonesh, Chinchilli, and Pu 1996).

OUTDESIGN Option in PROC GLIMMIX Statement

The OUTDESIGN option in the PROC GLIMMIX statement creates a SAS data set that contains the contents of the **X** and/or the **Z** matrix. If the data are processed by subjects, then the OUTDESIGN data set contains a stacking of the $\mathbf{Z}_i$ matrices. This option is useful, for example, if you need to access the design matrices from a mixed model in a subsequent PROC IML step. You can use the OUTDESIGN option in combination with the NOFIT option if you want to produce **X** and/or **Z** matrices without fitting the model. For example, in the following statements a B-spline basis is generated from the variable E and saved to the data set xmat:

```
proc glimmix outdesign=xmat nofit;
  effect spl = spline(E);
  model NOx = spl / noint;
run;
```

REFERENCES

Breslow, N. E. and Lin, X. (1995), "Bias Correction in Generalised Linear Mixed Models with a Single Component of Dispersion," *Biometrika*, 81, 81–91.

Cockerham, C. C., and Weir, B. S. (1977), "Quadratic Analyses of Reciprocal Crosses," *Biometrics*, 33, 187–203.

Eilers, P. H. C. and Marx, B. D. (1996), "Flexible Smoothing with *B*-splines and Penalties," *Statistical Science*, 11, 89–121.

Goodnight, J. H. (1978), SAS Technical Report R-105, *Computing MIVQUE0 Estimates of Variance Components*, Cary, NC: SAS Institute Inc.

Handcock, M. S. and Stein, M. L. (1993), "A Bayesian Analysis of Kriging," *Technometrics*, 35(4), 403–410.

Handcock, M. S. and Wallis, J. R. (1994), "An Approach to Statistical Spatial-Temporal Modeling of Meteorological Fields (with Discussion)," *Journal of the American Statistical Association*, 89, 368–390.

Harville, D. A. and Jeske, D. R. (1992), "Mean Squared Error of Estimation or Prediction under a General Linear Model," *Journal of the American Statistical Association*, 87, 724–731.

Huynh, H. and Feldt, L. S. (1970), "Conditions Under Which Mean Square Ratios in Repeated Measurements Designs Have Exact *F*-Distributions," *Journal of the American Statistical Association*, 65, 1582–1589.

Kackar, R. N. and Harville, D. A. (1984), "Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models," *Journal of the American Statistical Association*, 79, 853–862.

Kenward, M. G. and Roger, J. H. (1997), "Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood," *Biometrics*, 53, 983–997.

Lin, X. and Breslow, N. W. (1996), "Bias Correction in Generalized Linear Mixed Models with Multiple Components of Dispersion," *Journal of the American Statistical Association*, 91, 1007–1116.

Matérn, B. (1986), *Spatial Variation*, Second Edition, Lecture Notes in Statistics, New York: Springer-Verlag.

Morel, J. G. (1989), "Logistic Regression under Complex Survey Designs," *Survey Methodology*, 15, 203–223.

Morel, J. G., Bokossa, M. C. and Neerchal, N. K. (2003), "Small Sample Correction for the Variance of GEE Estimators," *Biometrical Journal*, 4, 395–409.

Pinheiro, J. C. and Bates, D. M. (1995), "Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model," *Journal of Computational and Graphical Statistics,* 4, 12–35.

Pinheiro, J. C. and Chao, E. C. (2006), "Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models," *Journal of Computational and Graphical Statistics*, 15, 58–81.

Royen, T. (1989), "Generalized Maximum Range Tests for Pairwise Comparisons of Several Populations," *Biometrical Journal,* 31, 905–929.

Saxton, A., ed. (2004), *Genetic Analysis of Complex Traits Using SAS*, Cary, NC: SAS Institute Inc.

Self, S. G., and Liang, K.-Y. (1987), "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions," *Journal of the American Statistical Association*, 398, 605–610.

Verbeke, G. and Molenberghs, G. (1997), "Linear Mixed Models in Practice: A SAS-Oriented Approach," New York: Springer-Verlag.

Vonesh, E. F., Chinchilli, V. M., and Pu, K. (1996), "Goodness-of-Fit in Generalized Nonlinear Mixed-Effects Models," *Biometrics*, 52, 572–587.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica,* 50, 1–25.

Wolfinger, R. (1993), "Laplace's Approximation for Nonlinear Mixed Models," *Biometrika*, 80, 791–795.

CONTACT INFORMATION   Oliver Schabenberger, SAS Institute Inc., SAS Campus Drive, Cary, NC 27513
E-mail: oliver.schabenberger@sas.com