A GENERAL STRUCTURE FOR THE CLASS OF MIXED LINEAR MODELS

F. G. Giesbrecht, Dept. of Statistics, N. C. State University

1.1 Introduction

The object of this paper is to present a general structure for mixed models which can be used to unify a large number of linear models often encountered in agricultural research work. Researchers often encounter special cases of these models under such diverse conditions that they fail to recognize the links to a common structure. Consequently, in practice it is not unusual to find very special ad-hock analyses being developed when approaches based on well established general theory exist. The gains from this unification are two-fold. A general structure provides a basis that can be used in developing models that are appropriate for specific applications and it provides a framework within which computer programmers can develop software that can be used to perform relatively sophisticated analyses without unreasonable difficulty. The classical example here is the wide variety of least-squares analyses that are now routinely performed using commonly available linear regression packages. These calculations are almost invariably based on the fixed effects or Model I (Eisenhart, 1947) assumptions with independent identically distributed random errors. There is a need for similar tools for Model II, the random effects model and the large class of mixed models.

The first part of this paper gives a brief introduction and general discussion of the linear mixed model. This is followed by a series of examples drawn from a wide range of applications to agricultural research. The final sections are devoted to a series of computer coding segments that illustrate how MIXMOD, a user written SAS© procedure, unifies the techniques employed in using these models to analyse data. Copies of the source code for MIXMOD (IBM©version of PL1) are available from the author.[1]

The initial goal in the development of MIXMOD was to create a relatively general tool that researchers could use to estimate variance components in unbalanced data sets. The emphasis has always been to preserve generality while achieving reasonable computing efficiency. The coding was done in a well-known language in order to encourage critical examination and review of the methodology by users and particularly other Statisticians. An unfortunate corollary of this emphasis on maintaining openness and generality is that control codes tend to become complex for some models. At present it seems that the price for simpler controls will be a major reduction in flexibility. The author views MIXMOD as a developing

[1] SAS© is a registered trademark of SAS Institute Inc.
IBM© is a registered trademark of International Business Machines Corporation.

cooperative effort, rather than as a final complete package. The many helpful comments, criticisms and suggestions from colleagues, students and friends who have used the procedure (and occasionally had difficulties) are gratefully acknowledged. The hope is that this paper will encourage further comments and suggestions.

1.2 The General Mixed Model

The general linear mixed model which can be written in the form

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}\boldsymbol{e} \tag{1.1}$$

where Y denotes the column of n observations, X is an n×p matrix of known constants, β a column of p unknown constants or parameters, U a matrix of known constants and $\boldsymbol{e}$ a column of unknown random errors with mean zero and finite variances is deceptively simple. We find complete books devoted to methods for dealing with special cases where U is the n×n identity matrix and the elements of $\boldsymbol{e}$ are independent and have a common distribution. The more general $\mathbf{U} \neq \mathbf{I}$ is typically dismissed with little more than a paragraph or two. When we look at packaged computer programs the situation is even worse. The typical computer program is based on algorithms appropriate for Eisenhart's Model I. These computations are often completely inappropriate for Model II and the mixed model. Yet model (1.1) is often realistic and sufficiently general to model a very large number of phenomena frequently encountered in agricultural research. The object of this chapter is to explore some of these possibilities, some of the variations of this model and relate them to common problems encountered in agricultural research and subsequently illustrate a computational approach that can be used to fit these models to real data. A final section will describe an existing computer program that implements many of these ideas. Copies of this program are available from the author.

1.3 Analysis of Variance Models

The first and probably most common large class of models assumed under (1.1) consists of what may by called the analysis of variance models. For these models it is common to assume that all elements of $\boldsymbol{e}$ are mutually independent and to partition this vector into a number of segments, $\boldsymbol{e}' = (\boldsymbol{e}_1', \boldsymbol{e}_2', \ldots, \boldsymbol{e}_k')$ such that the m_i elements of $\boldsymbol{e}_i$ have common variance. With this partitioning the model becomes

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}_1\boldsymbol{e}_1 + \mathbf{U}_2\boldsymbol{e}_2 + \cdots + \mathbf{U}_k\boldsymbol{e}_k \tag{1.2}$$

where $\{\mathbf{U}_i\}$ are n×m_i known matrices. It is also possible to break β into several parts such that the elements within sub-vectors have something in common. However this does not appear to be particularly useful at this time. It is necessary however to point out that many authors will insist that $\mathbf{U}_k$ be the n×n identity matrix. For purposes of the discussion in this

chapter it is convenient to allow the greater generality. Models in this class include most if not all of the analysis of variance models one encounters in such standard Statistical Methods books as Snedecor and Cochran (1980), Steel and Torrie (1980), Federer (1955), Searle (1971) and Scheffé (1959).

If in addition to the assumptions made in connection with (1.1), one assumes that the elements of e_i are normal independent with mean zero and variance σ_i^2 for i = 1, 2, ... , k then it follows that the covariance matrix Y can be written as

$$\begin{aligned} V[Y] &= U_1U_1'\sigma_1^2 + U_2U_2'\sigma_2^2 + \cdots + U_kU_k'\sigma_k^2 \\ &= V_1\sigma_1^2 + V_2\sigma_2^2 + \cdots + V_k\sigma_k^2 \\ &= V. \end{aligned} \tag{1.3}$$

It is common to assume that there are sufficient restrictive assumptions on $\{\sigma_i^2\}$ to assure that V is positive definite. If the $\{\sigma_i^2\}$ were known, then either Gauss Markoff or maximum likelihood would lead one to search for estimable functions of the elements of β by examining the generalized least squares or normal equations

$$X'V^{-1}X\hat{\beta} = X'V^{-1}Y. \tag{1.4}$$

Fortunately, in many practical situations one can dispense with the V^{-1} matrix and simply work with the ordinary least squares equations

$$X'X\hat{\beta} = X'Y \tag{1.5}$$

and arrive at the same solutions. In particular this happens when the data happens to fall into the balanced analysis of variance situation. Note that for this to happen, both X and the $\{U_i\}$ need to have the proper form. In practice, if this happens one would write the model, (1.4) in the more common analysis of variance form. We note in passing that equations (1.5) are appropriate for Eisenhart's Model I and form the basis for most least squares or multiple regression programs.

Our aim here is not to examine conditions where solutions for (1.3) are identical with solutions for (1.5), but to assume that the solutions are not identical and to proceed with the more general analysis. In practice there are really two types of problems. Typically both β and the $\{\sigma_i^2\}$ are unknown. In some cases the elements of β or linear functions of the form $\lambda'\beta$ are of prime interest and the $\{\sigma_i^2\}$ are nuisance parameters while in others the interest is focused on the $\{\sigma_i^2\}$ and β represents a vector of nuisance parameters. Fortunately the basic computations are sufficiently similar that common software can be used for both cases.

1.4 A General Computational Solution

The underlying approach will be to compute C.R. Rao's (1972) MInimum Norm

Quadratic Unbiased Estimates (MINQUE) of the unknown variance components, and then use these estimates to construct the estimated generalized least squares equations if estimates of $\lambda'\beta$ are needed. Before proceeding, two comments are in order. First of all, computing the MINQUE requires prior values for the variance components. Experience indicates that all prior values equal to one leads to estimators with good properties. Alternatively one is tempted to iterate, to use one set of estimates as input to compute a new set and so on. If one assumes normality and if the scheme converges (it usually does so very quickly) to valid estimates, then one has the modified maximum likelihood estimates. (There is some confusion in the terminology used in the literature. Some authors use the term restricted maximum likelihood while others prefer modified maximum likelihood and still others use the term marginal maximum likelihood.)

Secondly, the MINQUE approach is <u>not</u> based on the notion of calculating sums of squares as in the analysis of variance and/or least squares. If the data set is balanced in the analysis of variance sense or happens to be of very special form, MINQUE and analysis of variance will lead to the same estimates. Generally this is not true.

The key implication of this difference from the analysis of variance and/or least squares approach is that attention must be focused on the $\{V_i\}$ in (1.3) rather than the $\{U_i\}$ in (1.2). If we define $\tilde{V} = V_1\tilde{\sigma}_1^2 + V_2\tilde{\sigma}_2^2 + \cdots + V_k\tilde{\sigma}_k^2$ where the $\{\tilde{\sigma}_i^2\}$ denote prior values for the $\{\sigma_i^2\}$ (these may all be equal to one), then the system of equations that must be solved to estimate the variance components can be written as:

$$\begin{bmatrix} \mathrm{tr}(Q_v V_1 Q_v V_1) & \cdots & \mathrm{tr}(Q_v V_1 Q_v V_k) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(Q_v V_k Q_v V_1) & \cdots & \mathrm{tr}(Q_v V_k Q_v V_k) \end{bmatrix} \begin{bmatrix} \hat{\sigma}_1^2 \\ \vdots \\ \hat{\sigma}_k^2 \end{bmatrix} = \begin{bmatrix} Y'Q_v V_1 Q_v Y \\ \vdots \\ Y'Q_v V_k Q_v Y \end{bmatrix} \qquad (1.6)$$

where $Q_v = \tilde{V}^{-1} - \tilde{V}^{-1}(X'\tilde{V}^{-1}X)^{-}\tilde{V}^{-1}$, $(X'\tilde{V}^{-1}X)^{-}$ is a generalized inverse of $X'\tilde{V}^{-1}X$ and $\mathrm{tr}(Q_v V_i Q_v V_j)$ is the trace, the sum of the elements along the main diagonal of $(Q_v V_i Q_v V_j)$. It needs to be pointed out that the $\{\tilde{\sigma}_i^2\}$ must be such that $\tilde{V}$ is always positive definite. This condition must be watched closely if one uses an iterative scheme since the solutions, $\{\hat{\sigma}_i^2\}$ obtained from (1.6) do not necessarily yield $\{\tilde{\sigma}_i^2\}$ that satisfy this condition. Note that this system is obtained from the $\{V_i\}$ rather than the $\{U_i\}$. This is important in those models where it is much easier to visualize the $\{V_i\}$ since they relate directly to the correlation structure in the data. This is particularly true in many of the Quantitative Genetics models and models with correlated error structures. As a companion to this system one can set up the estimated generalized least squares equations

$$X'\tilde{V}^{-1}X\hat{\beta} = X'\tilde{V}^{-1}Y \qquad (1.7)$$

to obtain estimates of functions of the fixed effects. The variance of $\lambda'\hat{\beta}$ can then be estimated by $\lambda'(X'\tilde{V}^{-1}X)^{-}\lambda$.

If the $\{\tilde{\sigma}_i^2\}$ are all replaced by ones then one obtains Rao's MINQUE (1972). Alternatively if one could replace the $\{\tilde{\sigma}_i^2\}$ by the true variance components then one would obtain the minimum variance quadratic unbiased estimates of the components. Also in that case, twice the inverse of the matrix of coefficients on the left-hand side of (1.6) is the variance-covariance matrix of the estimated coefficients. In practice, inserting any values that do not depend on the data at hand leads to unbiased quadratic but not necessarily minimum variance estimators. Putting $\tilde{\sigma}_k^2 = 1$ and $\tilde{\sigma}_i^2 = 0$ for $i = 1, 2, \ldots, k-1$ is appealing in the sense that the computational burden is reduced dramatically but at the cost of an almost equally dramatic increase in the variances of the estimators. Putting all prior values *i.e.* all $\{\tilde{\sigma}_i^2\}$ equal to 1 does not materially simplify (1.6), but leads to estimators with good statistical properties. Iterating is attractive, even though the unbiasedness property is lost. If the iteration converges to valid estimates (non-negative estimates) then one has the Patterson and Thompson (1971) marginal (modified or restricted) maximum likelihood estimates. Replacing Q_v by $\tilde{V}^{-1}$ only in the matrix of coefficients on the left of (1.6) and iterating leads to the usual maximum likelihood estimates provided the process converges and the final solution is in the parameter space.

We note that (1.6) will generally be a small system of equations that can be easily solved using any matrix package (such as, for example, PROC IML or PROC MATRIX in SAS). The difficult computational part is obtaining the $n \times n$ $\tilde{V}^{-1}$ where n is the number of observations. In fact, size (of n) presents problems throughout since Q and all of the $\{V_i\}$ are $n \times n$ matrices. With modern computing facilities one can proceed directly via the definitions using marix notation for small data sets. For example, data from most unbalanced split-plot and incomplete block designs can be handled in this way. Larger variance component estimation problems and cases where one wishes to combine data from a series of experiments require more sophisticated programming. Procedure MIXMOD is a procedure written specifically to make it easy to perform the computations implied by (1.6) and at the same time make it possible to handle relatively large problems.

Unfortunately, there is no nice coding scheme available that can be used to make a computer program construct the general types and varieties of $\{V_i\}$ matrices encountered in practice. One is forced to work via the $\{U_i\}$ matrices to achieve even moderate flexibility. In some specific problems this is quite natural, while in others it leads to cumbersome control

sequences.

1.5 Applications

Possibly the simplest example of the application of the methodology presented is the among and within classification. The model is commonly written as

$$y_{ij} = \mu + a_i + e_{ij} \qquad \text{for } j=1, 2, \ldots n_i \text{ and } i=1, 2, \ldots, m$$

where $\{e_{ij}\}$ are normal independent random variables with mean zero and variance σ_e^2, the $\{a_i\}$ are normal independent random variables with mean zero and variance σ_a^2 and the two sets of random variables are mutually independent. In matrix notation this model becomes

$$Y = 1\mu + U_1 a + Ie$$

and $V[Y] = U_1 U_1' \sigma_a^2 + I\sigma_e^2$. If the observations are in dictionary order then $U_1 U_1'$ will be block diagonal with blocks of ones of size $n_1 \times n_1$, $n_2 \times n_2$, ... , $n_m \times n_m$. The connection with (1.3) and subsequently (1.6) comes via the following match-ups:

$$\begin{aligned} U_1 U_1' &\leftrightarrow V_1 \\ \sigma_a^2 &\leftrightarrow \sigma_1^2 \\ I &\leftrightarrow V_2 \\ \sigma_e^2 &\leftrightarrow \sigma_2^2 \\ 1 &\leftrightarrow X \\ \mu &\leftrightarrow \beta. \end{aligned}$$

For purposes of statistical analysis, the split-plot experiment can be thought of as a slight generalization of this. The 1μ part of the model is replaced by a more complex $X\beta$ structure, in fact a structure induced by the factorial treatment arrangement in the split-plot. It is part of the design aspect of the split-plot that forces the concordance between the factorial treatment structure and the nested whole-plot, split-plot error structure. Equation (1.6) and subsequently (1.7) can be used to analyse any (possibly unbalanced) split-plot experiment.

Incomplete block experiments with the recovery of inter-block information fall into the same category. In this case there is typically no factorial treatment structure, while blocks and plots within blocks are the two random factors in the nested error structure. The α-designs developed by Patterson and Williams (1976) are special cases of the incomplete block designs that may have unequal block sizes.

Both the set of split-plot experiments and the set of incomplete block experiments can be extended. The split-plot to the split-split-plot and so on. The factorial structure, (X matrix), becomes more complex and the nested random portions of the model expands to allow

$$V[Y] = U_1 U_1' \sigma_a^2 + U_2 U_2' \sigma_{b(a)}^2 + I\sigma_e^2.$$

Both U_1U_1' and U_2U_2' are block diagonal with blocks of ones, with the first having larger blocks. The row-column designs (John,1987) with recovery of inter-row and inter-column information and the nested incomplete block designs (Dey,1986) can be thought of as direct generalizations of the incomplete block designs. In the case of the row-column design we have

$$V[Y] = V_r\sigma_r^2 + V_c\sigma_c^2 + I\sigma_e^2.$$

If the observations are put in order by the row classification then V_r will be block diagonal with blocks of ones, while if the observations are put in order by the column classification then V_c will be block diagonal with blocks of ones. The X matrix may display a factorial structure or it may simply display the treatment assignments as for example in a variety trial. In all of these cases, the introduction of extra unbalance due to possible loss of some observations does not introduce any computational complexities, though it will of course lead to a loss of information and cause extra correlations among the estimates.

An alternative useful class of models can be obtained if one relaxes the assumption that all of the residual error terms have common variance. For example, one could consider the model

$$Y = X\bar{\beta} + e$$

with $V[Y] = I_1\sigma_1^2 + I_2\sigma_2^2 + \cdots + I_k\sigma_k^2$ where the $\{I_i\}$ are n×n matrices with all elements either zero or one and ΣI_i equal to the n×n identity matrix. Here one obviously thinks of the I_i as V_i in (1.3). The first example that comes to mind here is the Agronomy experiment that is repeated over several years and one does not wish to assume that the residual variances are constant over years. This may be due simply to the fact that years are different or to such factors as different locations (soil types) or possibly even different plot sizes. Another application may be large variety trials conducted in incomplete block designs with blocks at various locations and different plot-to-plot variances. Combining data from a series of experiments also leads to models of this sort. This class of models will be referred to as the nonhomogeneous errors models. Models of this sort have been considered by Graybill (1954), Han (1969), Shukla (1962) and Levy (1975).

The series of papers by Brindley and Bradley (1985), Grubbs (1948,1973) and the book by Jaech (1985) deal with another class of linear mixed models that are closely related to the nonhomogeneous errors models. Typically these models can be written as:

$$y_{ij} = \mu + u_i + b_j + e_{ij} \qquad \text{for i} = 1, 2, \ldots, \text{r and j} = 1, 2, \ldots, \text{c}$$

where the u_i representing row effects are normal independent zero-mean random variables with variance σ_r^2, the b_j representing column effects are normal independent zero-mean random variables with variance σ_b^2, and the e_{ij} are independent zero-mean random variables with variance σ_j^2. It is assumed that all of the random variables are mutually independent. As a

specific example of the covariance structure implied by this model, the variance-covariance matrix of $(y_{11}, y_{12}, y_{13}, y_{21}, y_{22}, y_{23}, y_{31}, y_{32}, y_{33})'$ is

$$\begin{bmatrix} 1&1&1&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0\\ 0&0&0&1&1&1&0&0&0\\ 0&0&0&1&1&1&0&0&0\\ 0&0&0&1&1&1&0&0&0\\ 0&0&0&0&0&0&1&1&1\\ 0&0&0&0&0&0&1&1&1\\ 0&0&0&0&0&0&1&1&1 \end{bmatrix} \sigma_r^2 + \begin{bmatrix} 1&0&0&1&0&0&1&0&0\\ 0&1&0&0&1&0&0&1&0\\ 0&0&1&0&0&1&0&0&1\\ 1&0&0&1&0&0&1&0&0\\ 0&1&0&0&1&0&0&1&0\\ 0&0&1&0&0&1&0&0&1\\ 1&0&0&1&0&0&1&0&0\\ 0&1&0&0&1&0&0&1&0\\ 0&0&1&0&0&1&0&0&1 \end{bmatrix} \sigma_b^2 + \begin{bmatrix} 1&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0 \end{bmatrix} \sigma_1^2$$

$$+ \begin{bmatrix} 0&0&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&0&0 \end{bmatrix} \sigma_2^2 + \begin{bmatrix} 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&1 \end{bmatrix} \sigma_3^2.$$

In practice one encounters this sort of model if a number of randomly selected items (rows) are measured independently by m different measuring techniques (columns) and there is a question about the homogeneity of measurement errors across methods. The fact that the set of matrices as illustrated in the example above are linearly independent implies that the variance components are estimable. A special case of this model with the $\{b_i\}$ interpreted as unknown fixed constants leads to what is commonly referred to as the Grubbs's estimator. In fact Grubbs's estimators can be obtained as MINQUE when all prior values for the components are set equal to one.

Another large class of useful linear mixed models, the random regression coefficient models (Rao and Kleffe, 1988, Swamy, 1971) is obtained by writing

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{b} + \boldsymbol{e}$$

where X and Z are n×p and n×m matrices of observed constants, $\boldsymbol{\beta}$ is a column of p unknown parameters, $\boldsymbol{b}$ is a column of m independent zero-mean random variables with variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2$ and $\boldsymbol{e}$ is the usual column of n independent zero-mean random errors with common variance σ^2. All random variables are mutually independent. It follows that

$$\mathrm{V}[\mathbf{Y}] = \mathbf{Z}\mathbb{\Sigma}\mathbf{Z}' + \mathbf{I}\sigma^2$$

where $\mathbb{\Sigma}$ is the m×m diagonal matrix constructed from the $\{\sigma_i^2\}$. If $\mathbf{Z}_i$ is the i'th column in the matrix Z then this variance-covariance matrix can be written as

$$\mathrm{V}[\mathbf{Y}] = \sum_i \mathbf{Z}_i\mathbf{Z}_i'\sigma_i^2 + \mathbf{I}\sigma^2 \qquad (1.8)$$

and the parallel to (1.3) is clear. Generalizations to more complex structures for $\boldsymbol{e}$ are obvious and will not be examined. Generalizations that involve allowing the elements of $\boldsymbol{b}$ to have unknown means simply amount to redefining $X\beta$. The introduction of covariances among the elements of $\boldsymbol{b}$ can be handled but requires rewriting (1.8). If σ_{ij} is the covariance between element b_i and b_j of $\boldsymbol{b}$ and σ_{ii} the variance of b_i then (1.8) becomes

$$V[Y] = \sum_i Z_i Z_i' \sigma_{ii} + \sum\sum_{i<j} [Z_i Z_j' + Z_j Z_i'] \sigma_{ij} + I\sigma^2$$

in complete the analogy with (1.3). For future reference this can be rewritten as

$$V[Y] = \sum_i Z_i Z_i' (2\sigma_{ii} - \sum_l \sigma_{il}) + \sum\sum_{i<j} [(Z_i + Z_j)(Z_i + Z_j)'] \sigma_{ij} + I\sigma^2$$

where one thinks of estimating the $\{\theta_h\}$, with individual θ_h defined as the elements of $\{(2\sigma_{ii} - \sum_l \sigma_{il})\}$, $\{\sigma_{ij}\}$ with i≠j and σ^2 in some order. We note in passing that there is no logical reason for $(2\sigma_{ii} - \sum_l \sigma_{il})$ to be non-negative. The only inherent constraint is that the matrix previously referred to as $\tilde{V}$ be positive definite. Recall that $\tilde{V}$ is an estimate of V using the estimated variance and covariance components and is now computed as $\sum_h V_h \tilde{\theta}_h$.

While models of this sort are mentioned most frequently in the Social Science literature they would appear to have a real place in the Agricultural research. For example Rawlings and Giesbrecht (1987) used this sort of model while studying the effect of ozone concentration on plant yield in a series of field experiments. The regression coefficient of yield on ozone concentration in the air was modeled as a random coefficient since the actual realized magnitude of the coefficient in a specific experiment will depend on a host of other uncontrolled and/or unknown factors. It appears that yield responses to fertilizer application could be modeled in a similar manner. The growth and dose response curves discussed by Grizzle and Allen (1969) can be analysed very elegently using models of this type.

While the models discussed to this point are all univariate, they can be extended to the multivariate case to estimate covariance components as well as variance components. Rather than develop the extension in complete generality, the technique will be illustrated only for the extension of the basic separable model (1.2) to the r-variate case. Specifically it will be assumed that for each variable (j=1, 2, ... , r) one can write the a model of the form

$$Y_j = X\beta_j + U_1 e_{1j} + \cdots + U_k e_{kj}$$

where the dimensions of the matrices are as given for (1.2). In addition there are the assumptions $E[e_{ij}]=0$ for all i,j, $E[e_{ij}e_{ij}]=\delta_{jj'} I_{m_i} \sigma^{(i)}_{jj'}$ where $\delta_{ii'}=1$ if $i=i'$ and zero otherwise and I_{m_i} is the $m_i \times m_i$ identity matrix. Also $\sigma^{(i)}_{jj'}=\sigma^{(i)}_{j'j}$ denotes the covariance component, the covariance between variable j and variable j' for factor i in the model. In essence, the matrix of variances and covariances

$$\mathbb{\Sigma}_i=\begin{bmatrix} \sigma_{11}^{(i)} & \cdots & \sigma_{1r}^{(i)} \\ \vdots & & \vdots \\ \sigma_{r1}^{(i)} & \cdots & \sigma_{rr}^{(i)} \end{bmatrix}$$

in the multivariate model replaces the σ_i^2 component in the univariate model. All other covariances not mentioned are assumed to be zero.

In order to translate the multivariate model into the corresponding univariate model some notation is required. Let *vec*[A] denote the operation of replacing the m×n matrix A by the column of length mn obtained by replacing columns of A one below the other *i.e.*, $vec[A]=(a_{11}, a_{21}, \cdots, a_{m1}, a_{12}, a_{22}, \cdots, a_{mn})'$. Also the Kronecker product $\underset{m\times n}{A}\otimes\underset{r\times t}{B}$ is defined as the mr×nt matrix with mn submatrices of size r×t $(a_{ij}B)$.

Now write the r-variate model as a univariate model

$$Y=(X\otimes I_r)\beta+(U_1\otimes I_r)e_1+(U_2\otimes I_r)e_2+\cdots+(U_k\otimes I_r)e_k$$

where $Y=vec[(Y_1|Y_2|\cdots|Y_r)']$, $\beta=vec[(\beta_1|\beta_2|\cdots|\beta_r)']$ and $e_i=vec[(e_{i1}|e_{i2}|\cdots|e_{ir})']$ for i=1, 2, ... , k. It follows that

$$\begin{aligned} V[Y]&=(U_1\otimes I_r)(I_{m_1}\otimes\mathbb{\Sigma}_1)(U_1\otimes I_r)'+\cdots+(U_k\otimes I_r)(I_{m_k}\otimes\mathbb{\Sigma}_k)(U_k\otimes I_r)' \\ &=U_1U_1'\otimes\mathbb{\Sigma}_1+\cdots+U_kU_k'\otimes\mathbb{\Sigma}_k. \end{aligned}$$

Now define J_t to be a column of length r with element t equal to one and all other elements equal to zero. Using the set of r of these columns one can write

$$\begin{aligned} \mathbb{\Sigma}_i&=\sum_j\sum_h J_jJ_h'\sigma_{jh}^{(i)} \\ &=\sum_j J_jJ_j'\sigma_{jj}^{(i)}+2\sum_{j<h}\sum\sigma_{jh}^{(i)} \end{aligned}$$

and

$$V[Y]=\sum_i\sum_j\sum_h(U_i\otimes J_j)(U_i\otimes J_h)'\sigma_{jh}^{(i)}.$$

Finally to fit this into the pattern developed previously, this expression is rewritten as

$$\begin{aligned} V[Y]&=\sum_i\sum_j(U_i\otimes J_j)(U_i\otimes J_j)'(2\sigma_{jj}^{(i)}-\sum_h\sigma_{jh}^{(i)}) \\ &\quad+\sum_i\sum_{j<h}\sum\Big((U_i\otimes J_j)+(U_i\otimes J_h)\Big)\Big((U_i\otimes J_j)+(U_i\otimes J_h)\Big)'\sigma_{jh}^{(i)} \\ &=\sum_i\sum_j(U_i\otimes J_j)(U_i\otimes J_j)'\theta_j^{(i)}+\sum_i\sum_{j<h}\sum\Big((U_i\otimes J_j)+(U_i\otimes J_h)\Big)\Big((U_i\otimes J_j)+(U_i\otimes J_h)\Big)'\sigma_{jh}^{(i)}. \end{aligned}$$

One proceeds to estimate $\theta_j^{(i)}=(2\sigma_{jj}^{(i)}-\sum_h\sigma_{jh}^{(i)})$ and $\sigma_{jh}^{(i)}$ for $j\neq h=1, 2, \ldots, k$, for i=1, 2, ... , k and then transforms back to the desired variance components. Note that missing observations do not upset the approach. Their effect is to delete rows and columns from V[Y].

An alternative set of models that can be obtained as a special case of the multivariate mixed model is referred to in the literature as the seemingly unrelated regression model (Zellner, 1963). This is really a special case of the multivariate variance-covariance component model with k=1 and U_k the n×n identity matrix and possibly a distinct X matrix for each dependent variable.

The final set of examples is from Quantitative Genetics. A number of distinct mating designs are included under the general heading of reciprocal crosses. Cockerham and Weir (1977) present analyses of two distinct types of these designs under two alternative models. The analysis of data from these two designs will be examined via the mixed model methodology using the same basic models. The advantage of the methods presented here is that they are robust to loss of observations in the sense that the computations can still be performed with relative ease.

The general nature of reciprocal crosses is illustrated by the basic model

$$y_{ijk}=\mu + G_{ij} + e_{ijk}$$

where y_{ijk} denotes an observation on the k'th offspring of maternal parent (line) i mated to paternal parent (line) j with i≠j. The total genetic contribution is represented by G_{ij}. An observation on the reciprocal cross would be denoted by y_{jik}. The first step in the refinement of the model is to write $G_{ij}=M_i + P_j + (MP)_{ij}$ where M_i, P_j, and $(MP)_{ij}$ denote the total maternal, total paternal and interaction effects respectively. A more detailed model, called the diallel model is given by the expression

$$G_{ij} = g_i + g_j + s_{ij} + d_i - d_j + r_{ij} \tag{1.9}$$

with $s_{ij} = s_{ji}$ and $r_{ij} = -r_{ji}$. These random effects are assumed independent and to have variances σ^2_g, σ^2_s, σ^2_d and σ^2_r and are interpreted as general combining ability, specific combining ability, maternal-paternal difference effect and reciprocal effect respectively.

An alternative model, called the bio model is given by the expression

$$G_{ij} = n_i + n_j + t_{ij} + m_i + p_j + k_{ij} \tag{1.10}$$

with $t_{ij} = t_{ji}$ and where n_i and n_j represent the nuclear contributions of the parents, t_{ij} the interaction effect, m_i the maternal extranuclear effect, p_j the paternal extranuclear effect, and k_{ij} the higher order interaction effects. These random effects are again assumed independent and to have variances σ^2_n, σ^2_t, σ^2_m, σ^2_p and σ^2_k respectively.

Cockerham and Weir (1977) present the analyses of two alternative mating designs in terms of these two models. The first is a factorial mating design, consisting of all matings between two distinct sets of parents (parental lines). The second is a diallel consisting of all possible crosses from a single set of parents (lines).

The the diallel model for both mating designs and will be considered first. Assume that there are n observations and m parental lines (female or male) numbered from one to m involved in the design. Let U_M be the n×m matrix with the element u_{ij} equal to one if individual i has maternal parent j and zero otherwise. Note that columns of U_M may consist entirely of zeros in a design if not all lines are used as maternal parents. Let U_P be the

corresponding nxm matrix with the element u_{ij} equal to one if individual i has paternal parent j and zero otherwise. Again for some designs some columns may consist of only zeros. Let t be the number of crosses (including reciprocals) in the design. Define U_{MP} and U_{PM} be nxt matrices with rows corresponding to individuals and columns corresponding maternal-paternal and paternal-maternal crosses. Individual elements are one or zero, depending on whether that individual is a product of that particular cross or not. In general from the correlation structure one can write

$$V[Y] = V_g\sigma_g^2 + V_s\sigma_s^2 + V_d\sigma_d^2 + V_r\sigma_r^2 + I\sigma_e^2.$$

From the definitions of U_M, U_P, U_{MP} and U_{PM} it follows that

$$V[Y] = (U_M + U_P)(U_M + U_P)'\sigma_g^2 + (U_{MP} + U_{PM})(U_{MP} + U_{PM})'\sigma_s^2 + (U_M - U_P)(U_M - U_P)'\sigma_g^2 + (U_{MP} - U_{PM})(U_{MP} - U_{PM})'\sigma_s^2 + I\sigma_e^2.$$

This form corresponds exactly to (1.3). Note that care must be taken in the definitions of U_M and U_P to ensure that the specific column for individual or line h used as a male is added to (subtracted from) the corresponding column for individual or line h used as a female when forming $(U_M + U_P)$ and $(U_M - U_P)$. Similarly in U_{MP} and U_{PM} the column for female h and male k must match with the column for male h and female k. A further point that should be emphasized is that the key concern is the correlation (covariance) structure. Further complications such as a fixed covariate or blocking factor in the model and/or the design are easily taken care of. In fact it is also a simple matter to allow a random blocking factor, such as an incomplete block structure to enter into the design provided that nothing interferes with the existing covariances.

The bio model is handled in a very similar manner. Initially one can write

$$V[Y] = V_n\sigma_n^2 + V_t\sigma_t^2 + V_m\sigma_m^2 + V_p\sigma_p^2 + V_k\sigma_k^2 + I\sigma_e^2.$$

Using the definitions of U_M, U_P, U_{MP} and U_{PM} and the definitions of the random components in the bio model it follows that

$$V[Y] = (U_M + U_P)(U_M + U_P)'\sigma_n^2 + (U_{MP} + U_{PM})(U_{MP} + U_{PM})'\sigma_t^2 + U_MU_M'\sigma_m^2 + U_PU_P'\sigma_p^2 + U_{MP}U_{MP}'\sigma_k^2 + I\sigma_e^2$$

corresponding exactly to (1.3). Again further complications such as a fixed covariate or blocking factor in the model and/or the design are easily taken care of provided they are not correlated with the genetic factors in the bio model.

1.6 The Computing Algorithm

The purpose of this section is to provide a quick overview of the computing algorithm used in PROC MIXMOD (Giesbrecht, 1983) in order to clarify the connection between the models developed in section 1.5 and the computing examples to be given in section 1.7. An

important point to realize when trying to estimate parameters using maximum likelihood methods, is that the focus should be on the parameters being estimated, in this case the variance and covariance components themselves, or at least functions of them. One should think in terms of the total variance-covariance matrix of the vector of observations. Unfortunately, for the purpose of the computing algorithm used in procedure MIXMOD we are forced to work with the $\{U_i\}$ matrices. The alternative is to use a general purpose matrix manipulation facility with the inherent computational inefficiencies.

The algorithm begins with a linear model, such as defined in (1.2). The algorithm requires a set of values denoted by $\{\tilde{\theta}_i\}$ which must satisfy the side condition that $\tilde{V} = \sum_i V_i \tilde{\theta}_i$ is non-singular, the column vector Y, the matrix X and the set of matrices $\{U_i\}$ that define the $\{V_i\}$. The algorithm then computes the three systems of linear equations:

(1) The generalized normal equations $X'\tilde{V}^{-1}X\hat{\beta} = X'\tilde{V}^{-1}Y$.

These are referred to as the NOREQ in PROC MIXMOD.

(2) The "linearized marginal likelihood" equations.

$$\begin{bmatrix} \mathrm{tr}(Q_v V_1 Q_v V_1) & \cdots & \mathrm{tr}(Q_v V_1 Q_v V_k) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(Q_v V_k Q_v V_1) & \cdots & \mathrm{tr}(Q_v V_k Q_v V_k) \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_k \end{bmatrix} = \begin{bmatrix} Y'Q_v V_1 Q_v Y \\ \vdots \\ Y'Q_v V_k Q_v Y \end{bmatrix}. \quad (1.11)$$

These are referred to as the MMLEQ in PROC MIXMOD.

(3) The "linearized likelihood" equations

$$\begin{bmatrix} \mathrm{tr}(\tilde{V}^{-1}V_1\tilde{V}^{-1}V_1) & \cdots & \mathrm{tr}(\tilde{V}^{-1}V_1\tilde{V}^{-1}V_k) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(\tilde{V}^{-1}V_k\tilde{V}^{-1}V_1) & \cdots & \mathrm{tr}(\tilde{V}^{-1}V_k\tilde{V}^{-1}V_k) \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_k \end{bmatrix} = \begin{bmatrix} Y'Q_v V_1 Q_v Y \\ \vdots \\ Y'Q_v V_k Q_v Y \end{bmatrix}. \quad (1.12)$$

These are referred to as the MLEQ in PROC MIXMOD.

The generalized normal equations require little comment at this time. A point to note with regard to the two systems of equations, (1.11) and (1.12), when compared to (1.6) is that the $\{\sigma_i^2\}$ have been replaced by $\{\theta_i^2\}$. This is intentional in order to emphasize the fact one need not be estimating variance components. The algorithm is more general. If normality holds and one iterates system (1.11) using $\{\sigma_i^2\}$ and the process converges inside the parameter space then one has marginal (modified or restricted) maximum likelihood estimates. Similarly using (1.12) leads to the usual or conventional maximum likelihood estimates.

This also highlights one of the reasons for the absence of an automatic iteration feature

in MIXMOD. If one is estimating variance components, then the parameter space is obvious and one can easily design an iterative algorithm that keeps estimates inside. However, in the more complex cases being considered here, the relevant parameter space is not at all obvious. If the user wishes to iterate, the output from one cycle must be used as input to the next cycle with the user taking responsibility to insure that $\tilde{V}$ is positive definite. Violating this condition raises an error condition and causes the procedure to abort with an error message. Some users have used the SAS macro facility to tailor their own iterative version of MIXMOD. Also note that although PROC MIXMOD has options that permit the user to request solutions to (1.11) and (1.12) there are cases where it is not appropriate to solve the whole system of k equations in k unknowns. There is currently no option to request solutions to the NOREQ set of equations.

1.7 Examples

This section will present a series of examples to illustrate the use of PROC MIXMOD. It will be assumed that the reader is familiar with SAS and that conventions documented in the SAS User's Guide: Basics do not need to be explained. More complete documentation of PROC MIXMOD as well as source code are available from the author.

Example 1. The one-way nested model.

Consider the one-way classification with model

$$y_{ij} = \mu + a_i + e_{ij} \qquad \text{for j=1, 2, ... } n_i \text{ and i=1, 2, ... , m.}$$

The minimum required for MIXMOD to analyse this would be:

```
PROC MIXMOD DATA=SASdataset  SOLNMML;
MODEL Y=A;
LEVELS m;
```

where the variable A defines the grouping and m is the number of groups. One can contrast the LEVELS statement with the CLASSES statement in PROC GLM. The LEVELS statement tells the procedure that the first variable after the '=' in the MODELS statement will define m groups. Since no prior values for the variance components are given, the procedure uses ones. If for example the user wanted to use prior values 2.5 and 3.3 for σ_a^2 and σ_e^2 respectively than an extra statement, "PRIORS 2.5 3.3 ;" is needed. The SOLNMML option on the PROC statement requests that solutions to the MMLEQ and the natural logarithm of the marginal or modified likelihood evaluated at the priors be printed. The alternative option SOLNMML requests solutions to the MLEQ and the natural logarithm of the likelihood evaluated at the priors.

Example 2. The incomplete block design with t fixed treatments and b random blocks.

The basic control cards required are:

```
PROC MIXMOD DATA=SASdataset  SOLNMML;
MODEL Y=  T   B  /NF=1;
LEVELS   t    b;
```

The option NF=1 in the MODEL statement indicates that there is one fixed effect in the model. Fixed effects are always placed before the random effects. The intercept or mean is not included in the count of fixed effects. The intercept can be deleted by using the option NOINT on the MODEL statement. Again there are two variance components. Since no priors are given, they are assumed equal to one. T user can save the generalized normal equations by specifying NOREQ= SASdataset in the PROC statement. There is currently no option available to request solutions to the generalized normal equations. The LEVELS statement specifies that the first variable after the '=' has t distinct levels and the second has b levels. If in addition to the above, the user wished to include a covariate X in the model, then following set of statements would be required:

```
PROC MIXMOD DATA=SASdataset  SOLNMML;
MODEL Y= T  X   B  /NF=2;
LEVELS   t   1   b;
```

Now the LEVELS statement indicates that the second variable after the '=' is continuous *i.e.*, a covariate. Note that there are now two fixed effects in the model.

Interaction terms and nesting are specified in the model by using the '*' convention *i.e.*, A*B implies the interaction between A and B or nesting, depending on the context. A statement of the form "MODEL Y= A B A*B;" implies main effects A and B and an interaction, while "MODEL Y= A A*B;" implies A with B nested within A. The '|' and the '(.)' feature found in MODEL statements in PROC GLM are not supported. The total number of combinations of levels of A and levels of B that occur in the data set must be accounted for in the LEVELS statement if A*B is present.

Example 3. Models with non-homogeneous residual errors

Models with non-homogeneous residual errors are normally difficult, if not impossible to handle using ordinary analysis of variance techniques, but fit into the MIXMOD pattern very nicely. As an example, consider the previous incomplete block design with t treatments, a covariate X, b random blocks with data collected from three sources and the assumption that the experimental errors are different for the three. The appropriate control cards are:

```
PROC MIXMOD DATA = SASdataset NOREQ = SASdataset
                                  KE=3 GROUP=SOURCE  SOLNMML;
MODEL Y = TRT   X  BLOCK /NF=2;
LEVELS      t  1   b   n1  n2  n3;
```

The parameter KE in the PROC statement indicates that the residual errors are in three

homogenous groups. The GROUP parameter indicates that the variable SOURCE identifies the observations in the groups. The counts in each of the groups are given by n_1, n_2 and n_3. The order of the observed values for SOURCE defines the order of the groups. If a PRIORS statement is present, it must provide four values, $\tilde{\sigma}_b^2$, $\tilde{\sigma}_1^2$, $\tilde{\sigma}_2^2$ and $\tilde{\sigma}_3^2$. An asymptotic likelihood ratio test of H_0 that all residual varances are equal against the H_a that they are not all equal can be obtained by fitting the model first with the more complex model, the non-homogeneous errors and then fitting with the homogeneous errors model and looking at two times the difference between the log likelihoods.

Example 4. Measurement error models.

For this example consider a model of the type

$$y_{ij} = \mu + u_i + b_j + e_{ij} \qquad \text{for i = 1, 2, ... , r and j = 1, 2, ... , c}$$

where $u_i \approx N(0, \sigma_r^2)$, $b_j \approx N(0, \sigma_b^2)$ and $e_{ij} \approx N(0, \sigma_i^2)$ for i=1, 2, ... , r. This is often called the measurement errors model (Jaech, 1985) but can be thought of as just a more complex non-homogeneous errors model. The SAS statements for this model are:

```
PROC MIXMOD DATA = SASdataset SOLNMML   KE = r  GROUP = B;
MODEL Y = U   B   /NF=0;
LEVELS      r   c   n1  n2   ...   nr ;
PRIORS   σ̃r²  σ̃b²  σ̃1²  σ̃2²   ...   σ̃r² ;
```

Notice that the variable B in the data set is used in two places, to define the groups for non-homogeneous errors and also to define a random effect. The important $H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_r^2$ can be tested against the alternative hypothesis that not all are equal using the likelihood ratio statistic.

Example 5. Random regression models.

As an example for this type of model consider a hypothetical series of n experiments, each of which involves applying varying amounts of fertilizer to plots of grain. The basic assumption is that yield can be modeled as a linear and quadratic function of amount of fertilizer (at least in the ranges tested in the experiment). However the two regression coefficients are random in the sense that realized coefficients in a specific experiment depend on a host of random uncontrolled factors. Consider first the case with a linear and a quadratic term but with only the linear component random. The required SAS statements are:

```
PROC MIXMOD   NOREQ = SASdataset   ;
MODEL  YIELD =  AMT  AMT*AMT  EXPT  EXPT  / NF=2;
LEVELS            1     1      n     n      ;
RNDCOEF  AMT 4 ;
```

Several items need to be explained. The purpose of the second "EXPT" in the MODEL can be thought of as reserving space for the matrix of coefficients of σ_{11}, the variance of the linear regression coefficient. In response to the "EXPT" variables the MODEL sets up two

"U" type matrices consisting only of zeros and ones. The RNDCOEF statement multiplies the elements of the second matrix (fourth term in the MODEL) by the AMT values to produce the "Z" type matrix. One can think of the model being fitted as one with regression lines with both random intercepts and random slopes. It is possible and probably more realistic to also include a covariance term for the two. This will be illustrated in a subsequent example.

Consider first extending the model to allow the quadratic coefficient to be random and to allow a covariance between the linear and quadratic coefficients. The appropriate SAS statements to accomplish this are as follows:

```
DATA   ; SET SASdataset ;
AMTSQ = AMT*AMT; EXPTLL = EXPT ;
EXPTQQ = EXPT; EXPTLQ = EXPT;

PROC MIXMOD   NOREQ = SASdataset SOLNMML ;
MODEL  YIELD =  AMT  AMTSQ  EXPT  EXPTLL
                          EXPTQQ  EXPTLQ /NF=2;
LEVELS            1    1     n     n     n     n   ;
RNDCOEF  AMT 4  AMTSQ 5 ;
TRANS   EXPTLQ = 1.0*EXPTLL + 1.0*EXPTQQ ;
```

The purpose of the DATA step is simply to create four new variables. The MODEL statement now contains the four variables EXPT, EXPTLL, EXPTQQ and EXPTLQ to provide "U" type coefficient matrices for σ^2_{expt}, σ_{11}, σ_{22} and σ_{12} where the last three are the variances of the linear and quadratic regression coefficients and their covariance respectively. The RNDCOEF statement transforms the "U" type matrices to "Z" type matrices. The rule is that the variable named in the RNDCOEF statement provides the multiplier to be applied to the term identified by number in the MODEL *i.e.*, the EXPTLL coefficients multiplied by the AMT values and the EXPTQQ coefficients multiplied by the AMTSQ values. Finally the TRANS statement constructs a new "Z" type matrix (for the covariance) by adding together two existing "Z" type matrices and placing the product in the space reserved. An additional complication that must be stressed at this point is that this system does not estimate the variance and covariance components directly. Solving the MMLEQ gives estimates of σ^2_{expt}, $\sigma_{11} - \sigma_{12}$, $\sigma_{22} - \sigma_{12}$, σ_{12} and σ^2_e respectively. Recall that the residual error term in the model has variance σ^2_e. Values given in a PRIORS statement are also interpreted as priors for these linear combinations. Negative priors for $\sigma_{11} - \sigma_{12}$, $\sigma_{22} - \sigma_{12}$ and σ_{12} are acceptable provided they are consistent with $\tilde{V}$ being positive definite. Failure of this condition causes the procedure to abort. Once an iterative scheme has converged the user must solve for σ_{11} and σ_{22}.

Laird and Ware (1982) present an analysis for longitudinal data using a random regression model. They propose this model in preference to the commonly syggested

multivariate random effects model because of the robustness to missed and mistimed observations. While their model is similar their calculations are based on a combination of empirical Bayes and maximum likelihood using the EM algorithm. We note that all the calculations proposed in this paper are robust to unbalance due to missing and/or extra observations in the sense that the algorithm used in MIXMOD does not take advantage of balance, even if it exists. Obviously loss of observations affects precision and, if sufficiently severe, estimability.

Example 6. Growth curves.

To illustrate the use of MIXMOD to analyse growth curve data, the control cards for a simple analysis of the ramus height data given by Grizzle and Allen (1969) will be illustrated. The data consists of ramus heights of 20 boys at ages 8, $8\frac{1}{2}$, 9 and $9\frac{1}{2}$ years of age. Appropriate control cards are:

```
PROC MIXMOD SOLNMML;
MODEL HEIGHT = AGE  IND IND / NF=1 ;
LEVELS      1     20     20  ;
RNDCOEF    AGE    3;
```

This simple model allows for the growth rate of each individual to be modeled by a randomly selected line (intercept and slope). When all boys are considered the intercepts represent random variables with a nonzero mean and and the slopes represent random variables with a nonzero mean. There is an additional random observational error. As suggested earlier this model proved to be inadequate in the sense that it did not allow for a correlation between the intercept and the slope. The following code provided a much more reasonable model.

```
DATA  ; SET  ;
INDREG = IND ; INDCOV = IND;

PROC   MIXMOD   SOLNMML;
MODEL HEIGHT = AGE  IND   INDREG   INDCOV / NF=1 ;
LEVELS      1     20     20     20  ;
RNDCOEF    AGE    3;
TRANS   INDCOV = 1.0*IND  +  1.0*INDREG  ;
```

The preliminary DATA step is simply to define two new variables that can be referenced unambiguously by the TRANS statement. SOLNMML will provide modified or marginal likelihood estimates of $\sigma^2_{int} - \sigma_{int,slope}$, $\sigma^2_{slope} - \sigma_{int,slope}$, $\sigma_{int,slope}$ and σ^2_e.

Example 7. Seemingly unrelated regressions.

The appropriate coding for the seemingly unrelated regression type analysis is illustrated using the example from Zellner (1962). In this example the current gross investment for two firms (variables I1 and I2) are modeled as constants plus linear functions of the firms beginning-of-year capital stock (variables C1 and C2) and value of its outstanding shares at the beginning of the year (variables F1 and F2). The data set also contains a YEAR variable. There is data

for 20 years. A preliminary DATA step is required to reorganize the data. One can think of this step as shifting the data from the multivariate form given initially to a longer data set in the univariate mode.

```
DATA ; SET;
FIRM = 1 ; C = C1; F = F1 ; I=I1 ; IND1 = 1 ; IND2 = 0 ;OUTPUT;
FIRM = 2 ; C = C2; F = F2 ; I=I2 ; IND1 = 0 ; IND2 = 1 ;OUTPUT;

PROC MIXMOD NOREQ = SASdataset KE=2 GROUP = FIRM  SOLNMML ;
MODEL I = FIRM C*IND1 C*IND2  F*IND1  F*IND2  YEAR /NF=5  NOINT;
LEVELS      2     1   1   1    1    20     20      20  ;
```

The MODEL statement defines a model with six regression coefficients. These are the two intercepts, the two slopes of investment on capital stock and the two slopes of investment on value of outstanding shares. The NOREQ data set will contain the generalized normal equations (6 equations with 6 unknowns). The SOLNMML parameter leads to estimates for the covariance $\sigma_{f_2 f_2}$, $\sigma^2_{f_1}-\sigma_{f_1 f_2}$ and $\sigma^2_{f_2}-\sigma_{f_1 f_2}$. Prior value must be in terms of these functions.

It is informative at this point to consider several alternatives to the above analysis which corresponds to the analysis given by Zellner (1962). If the "KE=2 GROUP = FIRM" segment were deleted the result would be an analysis based on an among and within model, with independent firms nested within years and a bunch of fixed covariates. The assumed variance-covariance matrix would consist of 20 twoxtwo blocks $\begin{bmatrix} \sigma^2_y+\sigma^2_f & \sigma^2_y \\ \sigma^2_y & \sigma^2_y+\sigma^2_f \end{bmatrix}$ along the main diagonal. The "KE=2 GROUP = FIRM" segment redefines these twoxtwo blocks as $\begin{bmatrix} \sigma^2_y+\sigma^2_{f_1} & \sigma^2_y \\ \sigma^2_y & \sigma^2_y+\sigma^2_{f_2} \end{bmatrix}$. The procedure yields estimates of σ^2_y, $\sigma^2_{f_1}$ and $\sigma^2_{f_2}$. Zellner's analysis requires a matrix consisting of twoxtwo blocks like $\begin{bmatrix} \sigma^2_{f_1} & \sigma_{f_1 f_2} \\ \sigma_{f_1 f_2} & \sigma^2_{f_2} \end{bmatrix}$. Consequently the transformation at the end of the previous paragraph.

Example 8. Multivariate variance components.

Consider a 3-variate multivariate variance components estimation problem with an among and within classification for the random components and with $E[Y_i]=1\mu_i$ for i=1, 2 and 3. Assume that the data is in the form of variables Y1, Y2, Y3 and A, where A defines the groups. There are m groups and n observations in all. A preliminary DATA step is required to convert a multivariate data set to a univariate data set that is three times as long.

```
DATA LONG; SET MULTI ;
OBS12=_N_ ; OBS13=_N_ ; OBS23=_N_ ;
A11=A; A22=A; A33=A; A12=A; A13=A; A23=A;
VR = 1;Y = Y1;V1 = 1;V2 = 0;V3 = 0;
V12=V1+V2; V13=V1+V3; V23=V2+V3; OUTPUT;
VR = 2;Y = Y2;V1 = 0;V2 = 1;V3 = 0;
V12=V1+V2; V13=V1+V3; V23=V2+V3; OUTPUT;
VR = 3;Y = Y3;V1 = 0;V2 = 0;V3 = 1;
V12=V1+V2; V13=V1+V3; V23=V2+V3; OUTPUT;
```

```
PROC MIXMOD  DATA=LONG  KE=3  GROUP=VR  SOLNMML;
MODEL Y = VR A11 A22 A33 A12 A13 A23 OBS12 OBS13 OBS23 / NF=1;
LEVELS   3   m  m  m  m  m  m  n  n  n  n  n  n ;
PRIORS   θ̃1(1) θ̃2(1) θ̃3(1) σ̃12(1) σ̃13(1) σ̃23(1) σ̃12(2) σ̃13(2) σ̃23(2) θ̃1(2) θ̃2(2) θ̃3(2) ;
CORWTS V1  1  V2  2  V3  3  V12  7  V13  8  V23  9;
TRANS  A12 = 1.0*A11 + 1.0*A22 ;
TRANS  A13 = 1.0*A11 + 1.0*A33 ;
TRANS  A23 = 1.0*A22 + 1.0*A33 ;
```

The new 'LONG' data set contains 16 new variables in addition to Y. The three variables OBS12, OBS13 and OBS23 are needed in the model statement to generate coefficient matrices for the three within group covariances. The 'KE=3 GROUP=VR' segment generates space for the coefficients on the three within group variance components. The VR variable provides coefficients for μ_1, μ_2 and μ_3. More complex fixed effects models can be incorporated, but the coding tends to become messy. The six variables, A11, A22, A33, A12, A13 and A23 are in the MODEL statement to reserve space. Note that the six variables in the CORWTS statement are all zero-one variables. This statement functions like the RNDCOEF statement in that it is used to modify U_i matrices. The named variable is used as a multiplier for all entries in the row of the U matrix being generated in response to the identified term in the model. For example, if it were not for V1 the A11 term in the MODEL would lead to a coefficient for $\sigma_{11}^{(1)}$ in the variances for variates 2 and 3 as well. Finally the three TRANS statements define "U" matrices and store them in the locations reserved by A12, A13 and A23. The initial matrices produced for the A12, A13 and A23 by the combined action of the MODEL and the LEVELS statements are ignored. Notice that TRANS statements are used to generate the "U" matrices for the among covariances while the corresponding matrices for the within covariances are generated from variables V12, V13 and V23 defined in the preliminary step by deleting entries. It should be fairly obvious that one needs to use the TRANS statements for anything other than the within components and can not use them for the within components.

This model contains six variance and six covariance components. The six covariance components are estimated directly. The six varince components are obtained from the estimates of $\theta_j^{(i)} = (\sigma_{jj}^{(i)} - \sigma_{jh}^{(i)} - \sigma_{jk}^{(i)})$ j =1, 2, 3, h≠j, k≠j and h≠k for i=1, 2. There is no guarantee that the estimates obtained lead to nonnegative definite $\hat{\Sigma}_i$. Also it should be noted while it is common practice to estimate covariance components by analyzing new variables defined as the sum or difference of existing variables and then using the fact that the variance of the sum (difference) of two variables is the sum of the individual variances plus (minus) twice the covariance, the corresponding strategy does not yield either marginal (modified) or conventional maximum likelihood estimates.

Example 9. The reciprocal cross experiment in Genetics.

The analysis of the reciprocal crosses in Quantitative Genetics illustrates one of the more complex capabilities of the MIXMOD procedure. Analyses for two examples found in Cockerham and Weir (1977) will be given in some detail. Following the lead in the paper these will be referred to as the factorial example and the diallel example.

Consider the diallel first. The first coding segment that would be used to read the data given in the paper is the following:

```
DATA  DIALLEL;
DO    M = 1  TO  8;
    DO    P  =  1  TO  8;
          IF P=M  THEN GOTO SKIP;
          BLOCK = 1 ; INPUT  Y @@;  OUTPUT;
          BLOCK = 2 ; INPUT  Y @@;  OUTPUT;
         SKIP :   ;
         END ;
   END;
CARDS ;
```

data cards formatted as in Biometrics example

At this point the data set contains the 112 observations with maternal and paternal identification as well as the block indicator. The second segment of coding adds several new variables. Included here are the two variables D and G, both equal to the maternal identification variable in anticipation of their used in the diallel model. WP and WN are two indicator variables that can be used to generate the proper signs on the coefficients for the specific and reciprocal effects. The variables N, R, S and T are just dummies (place holders) at this time.

```
DATA TWO; SET DIALLEL;
D = M ; G = M ;
N = 1; R = 1;T = 1;S = 1;
WP = 0 ; IF M < P THEN WP = 1;
WN = 1 − WP ;
```

The third segment shows the coding for one iteration of MIXMOD:

```
PROC MIXMOD DATA=TWO   MMLEQ=THREE;
MODEL Y = BLOCK  G   D  M   P   S    R     M*P    P*M  / NF=1;
LEVELS       2     8  8  8   8   56  56    56      56 ;
PRIORS     σ²g   σ²d  0.0   0.0  σ²s   σ²r    0.0   0.0  σ²e ;
CORWTS      WN  7      WP     8  ;
TRANS   G = 1.0*M  +1.0*P  ;
TRANS   D = 1.0*M  −1.0*P ;
TRANS   S = 1.0*M*P + 1.0*P*M  ;
TRANS  R = 1.0 *M*P  - 1.0*P*M ;
```

Points to note are: (1) The MODEL and LEVELS statements together reserve space for random factors G, D, M, P, S, R, M*P and P*M even though as noted previously, variables R and S are just dummies. The CORWTS statement is executed before the TRANS statements

regardless of the order entered. The CORWTS statement multiplies the values found for the 7th random effect in the model by the value given as variable WN. (Actually the positive root of the value is used, but in this case we have only zeros and ones.) The effect is to eliminate the maternal by paternal interaction effect term if the maternal code is less than the paternal code. (They can never be equal since the design eliminates selfs.) Similarly, the WP variable is used to eliminate the paternal by maternal effect term if the maternal code exceeds the paternal code. (2) The TRANS statements implement the basic transformations required in the diallel model definitions. The first sets up the matrix of coefficients for the general combining ability, the contribution due to line contributed from either maternal or paternal parent and places the result in the space reserved for G. The program looks after details such as matching parental lines correctly. The second generates the matrix for the maternal-paternal difference effect and places it in the space reserved for D. The third generates the coefficient matrix for the specific combining ability (S) by combining the two parts of the M*P and P*M. Note the order of the elements in the interaction term is important because of the $s_{ij}=s_{ji}$ relationship. The fourth TRANS statement generates the coefficient matrix for the reciprocal effects. (3) Finally notice the PRIORS statement. One can not rely on the default values for the priors, since that would define $\tilde{V} = V_g + V_d + V_m + V_p + V_s + V_n + V_{M*P} + V_{P*M} + I$. Components V_m, V_p, V_{M*P} and V_{P*M} must be eliminated. This is accomplished by specifying zero prior values for the variances. (4) The MMLEQ stored in SASdataset THREE will contain coefficients for a system of nine equations with nine unknowns. The fourth and final segment of code uses PROC MATRIX to extract the proper equations and compute the solution. The SOLNMML and SOLNML options will yield incorrect answers.

```
PROC MATRIX;
FETCH EQ DATA=THREE;
SOL = INV(EQ(1 2 5 6 9,1 2 5 6 9))*EQ(1 2 5 6 9,10);
PRINT SOL;
```

The coefficients to estimate variance components corresponding to the first, second, fifth, sixth and ninth random effects in the model and the residual error term are extracted and the resulting system of simultaneous equations solved.

If the bio model is to be used then the following code is required:

```
PROC MIXMOD DATA=TWO  MMLEQ=FOUR;
MODEL Y = BLOCK N  M  P  T   M*P   P*M   M*P  / NF=1;
LEVELS      2    8    8  8  8   56     56     56 ;
PRIORS   σ̃²n  σ̃²m  σ̃²p  σ̃²t  0.0  0.0  σ̃²MP  σ̃²e ;
CORWTS    WN  5     WP    6  ;
TRANS  N = 1.0*M + 1.0*P  ;
TRANS  T = 1.0 *M*P  - 1.0*P*M ;
```

```
PROC MATRIX;
FETCH  EQ  DATA=FOUR;
SOL = INV(EQ(1  2  3  4  7  8,1  2  3  4  7  8 ))*EQ(1  2  3  4  7  8,9);
PRINT SOL;
```

In this case the model contains only two terms that are required at the TRANS stage, but must be eliminated from V via the zero priors and later from the likelihood equations via PROC MATRIX. PROC MIXMOD will sort control cards in the sense that it will use the LEVELS statement first to reserve space, will use the CORWTS statements next and finally the TRANS statements in the order that they are given. Order may be important and is preserved only within the set of TRANS statements.

The coding for the factorial example is very similar. The real differences are in coding the input. An example of suitable coding is :

```
DATA  FACTOR1 ;
DO  M= 1 TO 7 ;
    DO  P = 8 TO 14  ;
        BLOCK = 1 ; INPUT Y  @@;  OUTPUT;
        BLOCK = 2 ; INPUT Y  @@;  OUTPUT;
        END;
   END;
CARDS;
data cards formatted as in Biometrics example
DATA  FACTOR2 ;
DO  P= 1 TO 7 ;
    DO  M = 8 TO 14  ;
        BLOCK = 1 ; INPUT Y  @@;  OUTPUT;
        BLOCK = 2 ; INPUT Y  @@;  OUTPUT;
        END;
   END;
CARDS;
data cards formatted as in Biometrics example
DATA FACTOR ; SET FACTOR1  FACTOR2 ;
D = M ; G = M ;
N = 1; R = 1;T = 1;S = 1;
WP = 0 ; IF M < P THEN WP = 1;
WN = 1 - WP ;

PROC MIXMOD DATA=FACTOR    MMLEQ=FIVE;
MODEL Y = BLOCK  G  D  M  P  S  R  M*P  P*M  /NF=1;
LEVELS     2      14  14  14  14  98  98  98  98 ;
PRIORS    σ̃²g  σ̃²d  0.0  0.0  σ̃²s  σ̃²r  0.0  0.0  σ̃²e ;
CORWTS     WN  7     WP    8  ;
TRANS  G = 1.0*M  + 1.0*P  ;
TRANS  D = 1.0*M  - 1.0*P ;
TRANS  S = 1.0*M*P + 1.0*P*M  ;
TRANS  R = 1.0 *M*P  - 1.0*P*M ;
```

```
PROC MATRIX;
FETCH  EQ  DATA=FIVE;
SOL = INV(EQ(1  2  5  6  9,1  2  5  6  9))*EQ(1  2  5  6  9,10);
PRINT SOL;
```

To conclude one point should be stressed. While both data sets in the Cockerham and Weir paper are balanced in the sense that there is exactly one observation in each designated cell this fact is not used in the analysis. The usual SAS conventions can be used to enter missing values. MIXMOD recognizes them as such. The input segments need appropriate modification if some cells were to contain more than just one observation.

REFERENCES

Brindley, D.A. and R.A Bradley (1985) Some new results on Grubbs's estimators, *J. Amer. Statist. Assoc. 80, 711–719.*

Cockerham, C.C. and B.S. Weir (1977). Quadratic analyses of reciprocal crosses. *Biometrics 33,187–203.*

Dey, A. (1986). *Theory of block designs.* John Wiley & Sons, New York.

Eisenhart, C. (1947) The assumptions underlying the analysis of variance. *Biometrics 3,1–21.*

Federer, W.T. (1955). *Experimental Design.* Macmillan, New York.

Giesbrecht, F.G. (1983). An efficient procedure for computing MINQUE of variance components and generalized least squares estimates of fixed effects. *Commun. Statist.-Theor. Meth. 12(18),2169–2177.*

Graybill, F.A. (1954). Variance heterogeneity in a randomized block design, *Biometrics 10,516–520.*

Grizzle, J.E. and D.M. Allen (1969). Analysis of growth and dose response curves. *Biometrics 25,357–381.*

Grubbs, F.E. (1948). On estimating precision of measuring instruments and product variability, *J. Amer. Statist. Assoc. 43, 243–264.*

Grubbs, F.E. (1973). Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments, *Technometrics 15,53–66.*

Han, C. (1969). Testing the homogeneity of variances in a two-way classification, *Biometrics 25,153–158.*

Jaech, J.L. (1985). *Statistical analysis of measurement errors.* John Wiley & Sons, New York.

John, J.A. (1987). *Cyclic designs.* Chapman and Hall, London.

Laird, N.M. and J.H. Ware (1982). Random-effects models for longitudinal data, *Biometrics 38,963–974.*

Levy, K.J. (1975). A multiple range procedure for correlated variances in a two-way classification, *Biometrics 31,243–246.*

Patterson, H.D. and Thompson, R. (1971). Maximum likelihood estimation of variances *Biometrika, 58, 545–554.*

Patterson, H.D. and Williams, E.R. (1976). A new class of resolvable incomplete block designs. *Biometrika, 63, 83–92.*

Rao, C.R. (1971a). Estimation of variance covariance components—MINQUE theory. *J. Mult. Anal. 1,257–257.*

Rao, C.R. (1971b). Minimum variance quadratic unbiased estimation of variance components. *J. Mult. Anal. 1,257–257.*

Rao, C.R. and J. Kleffe (1988). *Estimation of variance components and applications.* North-Holland, Amstredam.

Rawlings, J.O. and F.G. Giesbrecht (1987). Modeling plant responses to environmental stresses using random coefficient linear models with implications for precision of regional inferences. *Environmetics 87 A conf. on statis. and math. meth. appl. to prob. in env. qual.* Wash., D.C.

Scheffé H. (1959). *The Analysis of Variance.* John Wiley & Sons, New York.

Shukla, G.K. (1973). An invariant test for the homogeneity of variances in a two-way classification, *Biometrics 28,1063–1072.*

Searle, S.R. (1971). *Linear Models.* John Wiley & Sons, New York.

Snedecor, G.W. and W.G. Cochran (1980). *Statistical Methods.* The Iowa State University Press, Ames Iowa.

Steel, R.G.D. and J.H.Torrie (1980). *Principles and Procedures of Statistics.* McGraw-Hill, New York.

Swamy, P.A.V.B. (1971). *Statistical inference in random coefficient regression models.* Springer Verlag, Berlin.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Am. Statist. Assoc. 57,348–368.*