Abstract

This paper provides motivation for the use of mixed linear models (i.e. fixed and random effects models) as a generalization of the standard linear model. Several case scenarios from various fields of study are presented illustrating their power and unity. A brief introduction to mixed models theory is given, as well as a review of current software. A new SAS/STAT® procedure, MIXED, is introduced and applied to some common examples.

KEY WORDS: Mixed Model, Linear Model, Best Linear Unbiased Predictor, Repeated Measures, Covariance Structures, Maximum Likelihood, Restricted Maximum Likelihood, Newton-Raphson.

Introduction

Mixed model methodology provides a unified framework that comfortably encompasses most of the add-on features that burden traditional linear models methods. The notion of a random effect has been considerably broadened in recent literature, and software is under development to exploit this generality. To see how this software can be used, consider the following case scenarios.

Case 1

You are an agricultural scientist with a split-plot experiment. You must figure out the layers in the design, and for each effect, which other effect to use as an error term. You are careful to get the design balanced so that the ratios work out. You add a TEST statement to specify the error terms to get correct tests. You must take care not to trust the automatic tests, and not to trust the standard errors printed for the least-squares means. If the data are unbalanced, you must use the RANDOM statement in the GLM procedure with the TEST option in order to construct synthetic mean squares to use as denominators, with Satterthwaite approximations. You must also endure the less than optimal properties of the variance components estimators, which are implicitly used, but never printed.

Imagine how much simpler it would be to do the analysis with a mixed model by simply declaring the random effects properly. The software automatically takes care of tests, standard errors, and imbalance. It also produces the best possible estimates, and offers a choice of "inference spaces" to which to generalize.

Case 2

You are a medical scientist running a longitudinal study. You model the study as a multivariate repeated-measures model. If you run it as a pure univariate model, you worry about the price of ignoring the covariances between measurements within a subject. So you form your measurements as separate variables by declaring a separate design across these response variables. You choose whether to use the covariance matrix in a multivariate framework or go with adjusted univariate estimators.

If you use the adjusted univariate estimators, you must select the proper adjustment (G-G,H-F, or unadjusted). If you take multivariate approach, you face four different test statistics, all of which may have only approximate significance p-values printed. The full procedure will involve several separate multivariate functions across the response, including the sum and various contrasts that need to be orthonormalized in some cases. You also must worry about the price in power using the covariance across measurements, but not estimating it parsimoniously. If you have a missing value, you must throw away the whole observation, because the framework does not work for incomplete data. If the data are not lined up by time, the whole approach fails. If there is a covariate that varies across time, the whole approach fails.

Imagine how much simpler it would be to use a mixed model. You can model the structure of the covariance across measurements in a variety of different ways, and you can even test the different covariance structures.
You can model the measurements as a simple subject-induced constant correlation, or a time-series process where measurements closer in time are more correlated. Or you can use the fully parameterised covariance matrix equivalent to MANOVA. The mixed model approach uses all the non-missing measurements; you do not have to throw away a whole subject if one measurement is missing. You can have covariates that vary across the measurements. There is one set of familiar test statistics. You can even predict the growth curve for each subject.

Case 3

You are an econometrician running a regression. You use time-series data, and thus want to incorporate an autoregressive process in the error term. Usually you go to a specialised regression program for the analysis. But if you have missing data, the program may not handle it (the AUTOREG procedure will). If the data have an unequal spacing in time, the specialized procedure cannot handle it. If the data are not sorted properly, the procedure will not produce the right answers. Also, what if you have a model with cross sections, i.e. several time series from different regions, that might or might not have the same error structure, and you want to generalise to all regions? Perhaps the regions are spatially correlated, with regions that are closer to each other being more highly correlated.

All of these issues can be handled by a mixed model in a graceful way. You can even perform such methods as seemingly unrelated regression, errors in variables, and ridge regression in a mixed models framework.

Case 4

You are a sports statistician and must estimate/predict the baseball batting averages of the players. The problem is that it is early in the season and each player has only had a few dozen times at bat (Stroup, 1989). With such little data, any ordinary estimates would have an unacceptable degree of variability. So, you adopt an empirical Bayesian point of view, and shrink your estimates in a way suggested by the structure of variability of the data. Later you want to use the same method in a way that is appropriate for more data, with less shrinkage.

This, too, is a mixed model. The shrinkage is done through the effect of the random effects in the model. The predictors from the mixed models are called BLUPs, and they are equivalent to empirical-Bayes and Stein-type estimators. As a mixed model, all the estimation and shrinkage is an automatic part of the method.

Case 5

You have a very simple one-way analysis of variance. But you suspect that the variances are different. You want to test it and adjust for it. It isn't easy, unless you have a Welch ANOVA already canned in. But if you think of it as a mixed model, it is all quite easy. You end up with all your variances estimated and your tests adjusted for automatically. If you add a blocking factor, the interblock information is incorporated as well. If you extend to a two-way analysis, the confusion over parameterisations (e.g. Searle, 1971) is elucidated by the mixed model (see McLean, Sanders, and Stroup, 1991).

The Case for Mixed Models

The case for using mixed models is strong. There are only two impediments to their widespread use: education and software. We appeal to educators to stress mixed models, while SAS Institute addresses the software issue.

Be warned, however, that even with good software, you can expect mixed models to be computationally expensive to estimate. When developing statistical software, one of the most important design goals is to address issues in the most general way possible so that the software handles a broad class of problems, rather than a few particular ones. Mixed models provide an opportunity to address a broad class of issues. If it were not for the computational expense, you would do well to use a mixed models program for all your fitting work.

The Creed of a Mixed Modeler

- A mixed modeler searches for generalised conclusions. The medical scientist wants conclusions to apply to all people, not just the ones in his study. The agricultural scientist wants conclusions to apply to all similar animals, all similar plants, or plots, or whatever. Because the effect is considered random, this method will estimate diversity (variance components) and take that into account to produce results that generalise to whole populations. A mixed modeler thinks very carefully about the measurement process, and how to make inferences at various levels.
- A mixed modeler looks at the second moments as well as the first. The mixed model incorporates
the variances and covariances in a structure. People who do not use mixed models do not think about the variance until it is testing time, and therefore make unrealistic assumptions about the covariance structure of their data.

• A mixed modeler wants to use all the information available, and none that is not available. If there is information in the covariances, it should be used. If there is little information in parts of the data, the estimators should be shrunken as appropriate. If some measurements are missing, the nonmissing measurements should be used, and no imputations of missing measurements are needed. If the variances are unequal, then they should be modeled.

• A mixed modeler looks at relationships in terms of shared effects. Covariances between measurements reflect the sharing of a subject or a plot. Covariances across time reflect the partial sharing of temporal neighborhoods. Covariances across space reflect partial sharing of spatial neighborhoods.

• A mixed modeler wants to use unified general methods, rather than special purpose methods.

Mixed Models Theory

Mixed models have a well-established theoretical base; a few of the more prominent literature entries are

• Rao (1971a, 1971b) MINQUE and MIVQUE

• Searle (1971) Linear models and variance components

• Patterson and Thompson (1971) REML estimation

• Harville (1977, 1988, 1990) Maximum likelihood theory and algorithms, prediction

• Henderson (1984, 1990) Mixed model equations, genetics, BLUP


For an easy-to-read introduction, see McLean, Sanders, and Stroup (1991), and McLean (1989). The theory presented in these articles compels you to consider fitting mixed models to your data. The following is a brief introduction to that theory.

A mixed linear model is of the form

\[ y = XB + ZU + e \]

where \( y \) represents univariate data, \( B \) is a vector of fixed effects with design matrix \( X \), \( U \) is a vector of random effects with design matrix \( Z \), and \( e \) is an error vector. Assume

\[
\begin{bmatrix}
  U \\
  e
\end{bmatrix}
\] = \[
\begin{bmatrix}
  0 \\
  0
\end{bmatrix}
\]

\[
\text{Var}
\begin{bmatrix}
  U \\
  e
\end{bmatrix}
\] = \[
\begin{bmatrix}
  G & 0 \\
  0 & R
\end{bmatrix}
\]

where both \( G \) and \( R \) are nonsingular. The variance of \( y \) is thus

\[ V = ZGZ' + R. \]

We require that \( G \) has small enough dimensions to be computationally invertible, and that \( R \) is block-diagonal with blocks similarly small. The simplest such \( R \) has the form \( \sigma^2I \); and, if \( U \) is the null vector, the mixed model reduces to the standard linear model.

You can arbitrarily place random effects in either \( U \) or \( e \), although it is computationally convenient to put those effects that have a \( Z \) with few columns into \( U \), the rest going into \( e \). You can furthermore specify the structure of \( G \) and \( R \) by assigning a type to each of the random effects. Possible types include

• Simple (random effect)

• Unstructured (full covariance)

• Time Series (autoregressive)

• Distance (Kriging-type)

or a block diagonal combination of any of these. An example is the compound symmetry structure found in split-plot experiments. It results from a simple random effect defined over the whole plots, with the usual error term accommodating the subplots.

If \( G \) and \( R \) and hence \( V \) is known, the mixed models theory gives estimates of both \( B \) and the realized value of \( U \) as

\[
\begin{align*}
  b &= (X'V^{-1}X)^{-1}X'V^{-1}y \\
  u &= GZ'V^{-1}r
\end{align*}
\]

respectively, where \( r = y - Xb \). Under the further assumption of normality of \( U \) and \( e \) (which we make in constructing likelihoods), \( b \) is the best linear unbiased estimator (BLUE) of \( B \) and \( u \) is the best linear unbiased predictor (BLUP) of \( U \). These predictors can also be written as the solution to Henderson's (1984) mixed model equations:

\[
\begin{bmatrix}
  X'R^{-1}X & X'R^{-1}Z \\
  Z'R^{-1}X & Z'R^{-1}Z + G^{-1}
\end{bmatrix}
\]

\[
\begin{bmatrix}
  b \\
  u
\end{bmatrix}
\] = \[
\begin{bmatrix}
  X'R^{-1}y \\
  Z'R^{-1}y
\end{bmatrix}
\]

1382
You can subsequently define arbitrary estimable linear combinations of \( b \) and \( u \), which can then be used to test hypotheses and construct confidence intervals. To be more specific, define

\[
p' = (b' \ u')
\]

\[
P' = (B' \ U')
\]

\[
L' = (K' \ M')
\]

where \( K \) is assumed to be estimable (i.e., the rows of \( K' \) lie in the row space of \( X \)) and \( M \) is chosen to reflect the desired inference space (see Stroup, 1989). Assuming knowledge of \( V \) and normality, McLean, Sanders, and Stroup (1991) list the following results:

- \( K'b \) is the BLUE of \( K'B \), provided \( K'B \) is estimable
- \( L'p \) is the BLUP of \( L'P \), provided \( K'B \) is estimable
- \( \text{Var}[L'p - L'P] = L'CL \)

where \( C \) is a generalized inverse of the coefficient matrix in the mixed model equations.

The mixed models software estimates the unknown parameters of \( V \) using normal theory maximum likelihood (ML) or restricted maximum likelihood (REML). These estimated parameters are plugged back into \( V \) to form \( \tilde{V} \) and the resulting \( \tilde{p} \) and \( \tilde{C} \). This yields a prediction interval for \( L'P \) of the form

\[
L'\tilde{p} \pm t_{\nu} \sqrt{L'CL}
\]

where \( t_{\nu} \) represents an appropriate point from the \( t \) distribution with \( \nu \) degrees of freedom. In all but special balanced cases, \( \nu \) must be approximated, usually using Satterthwaite's method. The fact that we are using \( L'CL \) as the variance of \( L'p - L'P \) is naive because we are using \( V \) instead of \( \tilde{V} \). Kackar and Harville (1984) have proposed using

\[
L'\tilde{C}L + \text{tr}(AB)
\]

as the variance, where \( A \) is the variance matrix of \( \delta L'\tilde{p}/\theta \), and \( B \) is the inverse of the appropriate information matrix.

If \( L' \) has multiple rows, and you want to test

\[
H : L'P = 0
\]

use

\[
F = \tilde{p}'L(L'\tilde{C}L)^{-1}L'\tilde{p}'/\text{rank}(L).
\]

The numerator degrees of freedom is \( \text{rank}(L) \), and the denominator degrees of freedom is some approximation to \( \nu \) discussed above. McCarroll and Helms (1987) provide some alternatives to this \( F \)-statistic.

### Current Mixed Models Software

SAS Institute is currently developing general purpose mixed models software. Before discussing it, we review the state of the art.

### Random Effects in SAS Software Now

The GLM procedure does provide ways to calculate expected mean squares with respect to random effects, and, with some effort, yields estimates of variance components. But it does not offer the feature to use the resulting covariances in re-estimating the fixed effect parameters. PROC GLM also does not offer a ML or REML solution, which are preferred (Searle, 1988). While the VARCOMP procedure does do this, it lacks many of the features in PROC GLM such as continuous effects and least-squares means. Several authors, including Milliken and Johnson (1984), have gone to great lengths to circumvent the above restrictions. In addition to PROC GLM and PROC VARCOMP, special cases of mixed models are fit with the CALIS, NESTED, INBREED, and HARVEY procedures, as well as in SAS/ETS® software with the ARIMA and TSCSREG procedures. On the other hand, a very general MINQUE algorithm is available in Giesbrecht's (1983,1989) MIXMOD procedure.

### Additional Software

The following are a few of the current programs that fit mixed models.

- **BMDP-5V** (Schluchter, 1988) Unbalanced Repeated Measures
- **GLMM** (Blouin, 1989) Variance components
- **NONMEM** (Beal and Sheiner, 1980) Nonlinear mixed models
- **VARCL** (Longford, 1986) Variance components
- **GROWTH** (Gates and Dahm, 1989) Growth models
- **SUPER CARP** (Hidiroglou et al., 1980) Nested errors
- **CADA, BRAP, SEARCH** (see Goel, 1988) Bayesian programs

We are working on the following enhancements:

- Multiple effects labeled by type
• GLM-type grammar
• Subject and group variables that enable blocking and heteroscedasticity, respectively
• Computational speed and accuracy (W and Q transforms and the best optimization methods).

The software incorporating these features is scheduled to appear in JMP® software and in the MIXED procedure in SAS/STAT® software; the latter is discussed below.

The MIXED Procedure

PROC MIXED can be considered as a mixed models extension of PROC GLM. We now discuss its model specification and computing details.

Model Specification

Two GLM statements, RANDOM and REPEATED, are modified to specify the random effects contained in G and R, respectively. The model specification grammar for MIXED is as follows:

PROC MIXED option;
CLASS variables;
MODEL dependent = <fixed effect> / options;
RANDOM <vartype = random effect in G>;
REPEATED <vartype = random effect in R>;

where objects occurring within < > can occur multiple times. The current values for vartype are IND (default), UN, AR1, ARN, KR, SUBJECT, and GROUP. SUBJECT is a special type that indicates independent realizations of the effect, and GROUP is used to specify heteroscedasticity.

We now illustrate the model specification in PROC MIXED with some examples.

Example 1. Random Coefficients

This example comes from a pharmaceutical stability data simulation performed by Obenchain (1990). The observed response is the shelf-life of a pharmaceutical product, measured in “percent of label claim.” The desired mixed model involves batches of the product that differ randomly in intercept and slope, with error terms that follow a nested structure (Fuller and Battese, 1973). The PROC MIXED code to fit this model is the following:

```plaintext
proc mixed;
  class batch monthc;
  model y = month;
  random intercept month monthc
    subject=batch;
run;
```

The MODEL statement constructs MONTH and an intercept as fixed effects. The RANDOM statement also specifies these variables as random effects along with MONTHC which specifies the nested error term. SUBJECT=BATCH constructs a block-diagonal G with identical blocks corresponding to each batch. These blocks contain the variance components for the three given random effects, appropriately levelised. R by default becomes \( \sigma^2 I \), yielding a total of four unknown parameters.

Example 2. Repeated Measures

The data are from Pothoff and Roy (1964) and consists of growth measurements for 11 girls and 16 boys at ages 8, 10, 12, and 14. Jennrich and Schluchter (1986) analyze these data extensively using various covariance structures. The PROC MIXED code to fit an unstructured variance matrix is

```plaintext
proc mixed;
  class person sex agec;
  model y = sex age sex(age);
  repeated un=agec subject=person;
run;
```

You can replace UN= with AR1= or ARN= to obtain the first or nth order autoregressive covariance matrices, respectively. To fit a random coefficients model, use the line

```plaintext
random un={one age} subject=person;
```

This constructs a block-diagonal G with identical 2 x 2 unstructured blocks. R by default becomes \( \sigma^2 I \). Finally, to fit a compound symmetry structure, use the following line:

```plaintext
repeated intercept agec subject=person;
```

Here, INTERCEPT specifies the common covariance and AGEC the diagonal enhancement. This is the model Jennrich and Schluchter deemed best among the ones they fit. To go one step further, you can specify heteroscedasticity of this structure across girls and boys by using the following line:
repeated intercept agec subject=person
group=sex;

Some information criteria, as well as the likelihood ratio test, declare that this model fits the data significantly better than the homogeneous compound symmetry, as was noted by Lee (1988).

Example 3. Teachers and Students

Sanders (1989) presents an example in which you can use both RANDOM and REPEATED. The data are structured as follows:

data a;
  input student teacher topic y;
cards;
  1 Fred English .
  2 Mary English .
  2 Mary Math .
  3 Fred English .
  3 Mary Math .
  4 Mary English .
  4 Fred Math .
;

Suppose you want G to be a block-diagonal matrix with two identical unstructured blocks, which corresponds to Fred and Mary's covariance structures across topics. Additionally, you want R to be block-diagonal with unstructured blocks corresponding to the students across topics. The PROC MIXED code is the following:

proc mixed;
  class student teacher topic;
  model y = topic;
  random un=topic subject=teacher;
  repeated un=topic subject=student;
  run;

PROC MIXED constructs the desired matrices and outputs BLUPs that can be used to rank Fred and Mary.

Potential output from the above examples includes parameter estimates along with their iteration history, the observed inverse Fisher information matrix (which asymptotically is a variance matrix for the estimates), various information criteria, and t-, F-, and Wald tests for the fixed and random effects. You can furthermore use CONTRAST, ESTIMATE, and LSMEANS statements to carry out desired inferences just as in PROC GLM. Appropriate standard errors are provided, although the choice for degrees of freedom is still an open question. We are currently planning to do something simple, since the current df approximations (Giesbrecht and Burns, 1985, Jeske and Harville, 1988, McLean and Sanders, 1988) are computationally intensive. Suggestions are welcome.

Computing Details

PROC MIXED uses a Newton-Raphson (NR) algorithm, which is considered preferable to the EM algorithm (Lindstrom and Bates, 1988). The NR algorithm requires the computation of second derivatives of the objective function with respect to the variance parameters. The fixed effects parameters are concentrated out of the likelihood and are not part of the NR step.

Three estimation methods are currently available. The first is MIVQUE0 (Rao, 1971b), which is non-iterative. The second is REML, or iterated MIVQUE0, and uses a portion of the Gaussian likelihood not involving the fixed effects as its objective function. The third is ML, which uses the entire Gaussian likelihood as its objective function. A scoring option is also available with REML and ML; it turns off when close to convergence.

Two different methods are used to evaluate the likelihood and derivatives. The first is the W-transformation, discussed in Hemmerle and Hartley (1973) and in Goodnight and Hemmerle (1979). This is most effective for G-side calculations. The other method we call the Q-transformation, which essentially involves a direct inversion of V. This is obviated by using a subject effect that induces a block-diagonal structure in R, and is thus appropriate for R-side effects.

Both of the above transformations make heavy use of the sweep operator (Goodnight, 1979). They can also be combined when necessary, yielding a flexible and efficient computing algorithm.

Future enhancements to PROC MIXED may include

1. Additional covariance structures
   - General linear
   - Factor analytic
   - ARMA and GARCH (via a Kalman filter)
2. Quasi-Newton optimization methods (Scott, 1988)
3. Sparse matrix techniques (Fellner, 1987)
4. QR decomposition (Lindstrom and Bates, 1988a)
5. Linearised likelihoods (Callanan and Harville, 1989)
6. Maximum quasilikelihood (Longford, 1988)
7. Generalized linear models (Liang and Zeger, 1986)
8. Symbolic representation of G and R
9. Sophisticated df calculations (McLean and Sanders, 1988)
10. Bayesian methods (Strenio et al., 1983, Cook et al., 1990)
11. Approximate REML and pseudo Bayesian methods (Harville, 1977)
12. Different MINQUE norms (Kim and Lee, 1989)
13. Nonlinear mixed models (Lindstrom and Bates, 1988b)

Conclusion

We make the following recommendations.

Data Analysts: Include the mixed model as one of your standard analysis tools.

Educators: Teach the basics of mixed modelling at an introductory level using the "inference space" notion.

Students: Learn mixed models now; use them when you get out of school (they are not as difficult as some textbooks make them appear).

The time for the widespread use of mixed models has come. They unify and simplify a large variety of traditional statistical methods. Furthermore, they are developed both theoretically and computationally to the extent that a comprehensive computer program is now possible.

PROC MIXED promises to be that program, offering the ability to define multiple random effects labeled by types. This induces a covariance structure in the fitted model, the unknown parameters of which are estimated by the software. This enables you to exploit the generalisability of the mixed model and thus improve your inferences.

References


Blouin, David C. (1989) "Integrating Fixed Effects Hypotheses Testing with Variance Component Estimation: A Prototype of a Mixed Model Program,


Giesbrecht, Francis G. (1989) "A General Structure for the Class of Mixed Linear Models," in Applications of Mixed Models in Agriculture and Related Disciplines, Southern Cooperative Series Bulletin No. 343, Louisiana Agricultural Experiment Station, Ba-
ton Rouge, 183-201.


Agricultural Experiment Station, Baton Rouge, 9-38.


Patterson, H.D. and Thompson, Robin (1974) "Recovery of Inter-Block Information when Block Sizes are Unequal," *Biometrika*, 58, 545-554.


JMP, SAS, SAS/ETS, and SAS/STAT are registered trademarks of SAS Institute Inc., Cary, NC.