MULTIPLICATIVE EFFECTS IN MIXED MODEL ANALYSIS OF VARIANCE

by

Samuel D. Oman

Technical Report No. 247
April 1988

Prepared under the Auspices
of
National Science Foundation
DMS 87-08083
Ingram Olkin, Project Director

Department of Statistics
Stanford University
Stanford, California
MULTIPLICATIVE EFFECTS IN MIXED
MODEL ANALYSIS OF VARIANCE

by

Samuel D. Oman
Department of Statistics
Hebrew University of Jerusalem

and

Department of Mathematics
University of California, San Diego

March 1988
ABSTRACT

A class of two-way mixed analysis of variance models is proposed, in which the fixed and random effects enter multiplicatively. Equations are developed for iterative computation of maximum likelihood estimates via a scoring algorithm. Parameter estimation and hypothesis testing are illustrated on a set of plant genetics data.

Key words: Analysis of variance, multiplicative interaction, scoring algorithm.
I. INTRODUCTION

The mixed analysis of variance model, which contains both fixed and random effects, is widely used in the biological sciences. This paper extends the usual model in two respects: by allowing for unequal error variances and, more importantly, by including multiplicative effects.

Specifically, consider the usual two-way mixed analysis of variance model

\[ Y_{ijk} = m_{ij} + \varepsilon_{ijk} \]  \hspace{1cm} (1.1)

where

\[ m_{ij} = \mu + \alpha_i + b_j + c_{ij}. \] \hspace{1cm} (1.2)

Here \( i \) refers to the fixed factor, \( j \) to the random, \( k \) indexes replications and the independent \( N(0, \sigma^2) \) error terms \( \{ \varepsilon_{ijk} \} \) are independent of the \( \{ m_{ij} \} \). For example, in the machine-worker problem discussed by Scheffé (1959), \( m_{ij} \) is the average time it takes worker \( j \) to perform a task using machine \( i \).

We extend this model by allowing the variance of the error term \( \varepsilon_{ijk} \) to depend on the level \( i \) of the fixed factor, and by replacing (1.2) by

\[ m_{ij} = \mu + \alpha_i + b_j + \theta_i b_j + c_{ij}. \] \hspace{1cm} (1.3)

The \( \theta_i \) are fixed unknown parameters whose sum \( \theta_1 \) is zero (as are the \( \alpha_i \)), while the \( b_j \) and \( c_{ij} \) are mutually independent zero-mean normal variables with variances \( \sigma_b^2 \) and \( \sigma_c^2 \).
If the $\theta_i$ are not all zero then the random factor is modulated by, or interacts with, the fixed factor in a multiplicative manner. For example, suppose that $\alpha_i \gg \alpha_j$ in (1.3) for the machine-worker problem and that $\sigma_c^2 = 0$. Thus machine $i$ is, on the whole, better (faster). If $\theta_i \neq \theta_j$, then the difference between worker $j$'s average times on the two machines,

$$m_{ij} - m_{kj} = \alpha_i - \alpha_j + (\theta_i - \theta_j) b_j,$$

depends not only on the machines but on the worker's overall ability as measured by $b_j$. If $\theta_i > \theta_j$ then the difference is smaller for faster workers (with smaller $b_j$), possibly indicating that they manage to overcome the difficulties in using machine $i$. If we had $\theta_i < \theta_j$ then the difference would be greater for faster workers, perhaps because they could exploit some desirable property of machine $j$.

There is an extensive literature on multiplicative effects in fixed effects analysis of variance. In particular, if in the two-way model with one observation per cell,

$$Y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ij},$$

$\gamma_{ij} = \lambda_i \cdot \phi_j$ is assumed, then enough degrees of freedom remain to estimate the error variance and test the hypothesis of no interaction. Perhaps the best-known examples are Tukey's (1949) "one degree of freedom test" (with $\gamma_{ij} = \lambda_i \beta_j$) and Mandel's (1961) "bundle of straight lines" (with
Various generalizations have been studied by, among others, Tukey (1955), Ghosh and Sharma (1963), Milliken and Graybill (1970), Mandel (1971), Johnson and Graybill (1972), Hegemann and Johnson (1976a, b), Yochmowitz and Cornell (1978), and Marasinghe and Johnson (1982).

Krishnaiah and Yochmowitz (1980) give a comprehensive survey of the literature. Also, Denis (1983) uses multiplicative terms to model the error variances (as opposed to the response means) in the two-way fixed model. Gollob (1968, Section 5) treats multiplicative effects in the mixed analysis of variance model, but not in the same sense we do. He considers the case of two fixed factors measured on random subjects with an equal number of replications. A multiplicative structure is assumed for the interactions between the fixed factors, and heuristic sums of squares are presented for examining whether this structure varies among the subjects.

Our models are similar in spirit to those of Mandel (1961), but the randomness of the $b_j$ and $c_{ij}$ requires a different approach to estimation and hypothesis testing. In the next section, this paper's general model is precisely defined in terms of structured covariance matrices. It is a special case of the general linear model

$$y = X\alpha + Zb + e$$

(1.4)

treated by Harville (1977), in which $b$ is the vector of unobservable random effects and the "design matrix" $Z$ is unknown. Although Harville mentions the possibility of unknown $Z$, his explicit results largely deal with the case
of known $Z$. In Section 3 we discuss maximum likelihood estimation via a scoring algorithm for the unbalanced case of unequal numbers of replications in the different $(i, j)$ cells. Section 4 applies our models to a problem in plant genetics, while Section 5 discusses generalizations. The appendix contains the derivation of the equations for the scoring algorithm discussed in Section 3.
II. DEFINITION OF THE MODEL

Let $\mu_i = \mu + \alpha_i$ denote the fixed part of $m_{ij}$ in (1.3) and $\eta_{ij}$ the random part, so that (1.1) becomes

$$Y_{ijk} = \mu_i + \eta_{ij} + \epsilon_{ijk} \quad i = 1, \ldots, p$$

$$j = 1, \ldots, n$$

$$k = 1, \ldots, K_{ij}.$$  \hspace{1cm} (2.1)

We assume that all $K_{ij} \geq 1$ and $\sum_{j=1}^{n} K_{ij} > n$ for each $i$. The random terms $\{\eta_{ij}\}$ and $\{\epsilon_{ijk}\}$ are independent of each other, and the $\epsilon_{ijk}$ satisfy

$$\epsilon_{ijk} \sim N(0, \sigma_i^2)$$ \hspace{1cm} (2.2)

for unknown $\sigma_i^2$. It is convenient to formulate the assumptions on $\{\eta_{ij}\}$ in terms of the vectors

$$\eta(j) = (\eta_{1j}, \ldots, \eta_{pj})^t,$$ \hspace{1cm} (2.3)

which are assumed independent for different levels $j$ of the random factor.

Each $\eta(j)$ has the same distribution

$$\eta(j) \sim N(0, \Gamma),$$ \hspace{1cm} (2.4)

and different structures for $\Gamma$ correspond to different models for $\{\eta_{ij}\}$.

The following six structures will be considered:
\[1: \quad \Gamma = \sigma_c^2 \cdot \mathbf{I}\]
\[2: \quad \Gamma = \sigma_b^2 \cdot \mathbf{w} \mathbf{w}^t\]
\[3: \quad \Gamma = \sigma_b^2 \cdot \mathbf{w} \mathbf{w}^t + \sigma_c^2 \cdot \mathbf{I}\]
\[4: \quad \Gamma = \mathbf{v} \mathbf{v}^t\]
\[5: \quad \Gamma = \mathbf{v} \mathbf{v}^t + \sigma_c^2 \cdot \mathbf{I}\]
\[6: \quad \Gamma \text{ arbitrary.}\]

(2.5)

Here, \(\mathbf{w}\) denotes the \(p\)-dimensional vector \((1, \ldots, 1)^t\) of one's, while the non-negative scalars \(\sigma_b^2, \sigma_c^2\) and the vector \(\mathbf{v}\) are assumed unknown. It is also assumed in models 4 and 5 that \(\mathbf{v}^t \mathbf{w} > 0\). The different forms for \(\Gamma\) correspond to the following six models:

\[1: \quad \eta_{ij} = c_{ij}\]
\[2: \quad \eta_{ij} = b_j\]
\[3: \quad \eta_{ij} = b_j + c_{ij}\]
\[4: \quad \eta_{ij} = v_i \tilde{b}_j, \quad v_i > 0\]
\[5: \quad \eta_{ij} = v_i \tilde{b}_j + c_{ij}, \quad v_i > 0\]
\[6: \quad \eta_{ij} = c_{ij}, \quad \text{cov}[(c_{ij}, \ldots, c_{pj})^t] \text{ arbitrary.}\]

(2.6)

Model 6 is not of interest per se, but rather serves as a "supermodel" to allow testing the remaining models. In the remaining models the \(v_i\) denote fixed, unknown parameters to be estimated, \(b_j\) and \(c_{ij}\) denote independent zero mean normal variables with unknown variances \(\sigma_b^2\) and \(\sigma_c^2\), and the \(\tilde{b}_j\) are independent \(N(0,1)\) variables.
These models may be interpreted as follows. Model 3 corresponds to the usual mixed analysis of variance model (without the constraint \( \sum_i c_{ij} = 0 \)), generalized via (2.2) to the heteroscedastic case; while models 1 and 2 correspond to hypotheses on the variance components \( \sigma_b^2 \) and \( \sigma_c^2 \). Models 4 and 5 in (2.6) are obtained from (1.3) by taking \( v_i = \sigma_b(1+\theta_i) \) and \( \bar{b}_j = b_j/\sigma_b \); the constraints \( \sigma_b^2 > 0 \) and \( \theta_i = 0 \) translate into the requirement that \( v_i \) be positive. If in fact all \( \theta_i = 0 \) in (1.3), then one of models 2 or 3 results, depending on whether \( \sigma_c^2 = 0 \).

Aside from being convenient for estimation purposes, the formulation in terms of the \( v_i \) helps to clarify the difference between the \( \theta_i b_j \) and \( c_{ij} \) terms in (1.3). That is, consider the following three special cases of (1.3):

(A) \[ m_{ij} = \mu + \alpha_i + b_j \]

(B) \[ m_{ij} = \mu + \alpha_i + b_j + \theta_i b_j \]

(C) \[ m_{ij} = \mu + \alpha_i + b_j + c_{ij} \]

At first glance it might appear, analogous to the fixed effects case, that these models are nested within one another. However, the covariance matrices in (2.5) show this is not so: (A) corresponds to model 2, and the \( \Gamma \) for (B) (model 4) is obtained by changing the \( \sigma_b \cdot w \) in model 2's \( \Gamma \) to \( v \), while the \( \Gamma \) for (C) (model 3) is obtained by adding \( \sigma_c^2 \cdot I \) to the \( \Gamma \) in model 2.
To further clarify the difference between $v_i \tilde{b}_j$ in (2.6) and the interaction terms $c_{ij}$, consider model 5. Denoting means by $\bar{Y}_{ij}$, $\bar{Y}_{.j}$, etc., we have (assuming the sample means of the random components are $\approx 0$)

$$\bar{Y}_{ij} - \bar{Y}_{.} \approx \alpha_i + v_i \tilde{b}_j + c_{ij}$$

and

$$\bar{Y}_{.j} - \bar{Y}_{.} \approx \nu \tilde{b}_j$$

If as in Mandel (1961) we now graph $\bar{Y}_{ij} - \bar{Y}_{.}$ against $\bar{Y}_{.j} - \bar{Y}_{.}$ for each $i$, we will obtain (approximately) $p$ straight lines with different intercepts (if the $\mu_i$ are different). The $v_i$ determine the slopes of the lines, which are in particular parallel if all $v_i$ are equal, while the $c_{ij}$ determine the dispersion of the points about the lines. Alternatively, one may view model 5 in (2.5) as a factor analytic decomposition of $\Gamma$, with the communality $v_i^2$ giving the part of the variation in $\eta_{ij}$ due to the random factor and the specificity $\sigma_c^2$ giving the part due to $i$ alone. The $v_i$ may then be interpreted as factor loadings.
III. MAXIMUM LIKELIHOOD ESTIMATION

Maximum likelihood estimates will now be developed for the models specified by (2.1) - (2.4), when $\Gamma$ has one of the forms given in (2.5).

The vectors $Y_{(1)}, \ldots, Y_{(n)}$ and the variables $w_1, \ldots, w_p$ defined by

$$Y_{(j)} = (\overline{Y}_{1j}, \ldots, \overline{Y}_{pj})^t$$

and

$$w_i = \sum_{j=1}^n \sum_{k=1}^{K_{ij}} (Y_{ijk} - \overline{Y}_{ij})^2$$

are sufficient statistics; here, $\overline{Y}_{ij}$ denotes the mean of the $K_{ij}$ observations in cell $(i, j)$. The statistics are independently distributed as

$$Y_{(j)} \sim N(\mu, \Gamma + M_j)$$

and

$$w_i / \sigma_i^2 \sim \chi^2(m_i)$$

where

$$\mu = (\mu_1, \ldots, \mu_p)^t,$$

$$M_j = \text{diag}(\sigma_1^2 / K_{1j}, \ldots, \sigma_p^2 / K_{pj})$$

and

$$m_i = \sum_{j=1}^n K_{ij} - n.$$

Define

$$e_{(j)} = Y_{(j)} - \mu.$$
Except for an additive constant, the log-likelihood is given by

\[ \lambda = -\frac{1}{2} \sum_{j=1}^{n} \left[ \log |M_j + \Gamma| + e_{(j)}^t (M_j + \Gamma)^{-1} e_{(j)} \right] \]

\[ -\frac{1}{2} \sum_{i=1}^{p} \left[ w_i / \sigma_i^2 + m_i \log(\sigma_i^2) \right] . \]  

(3.3)

Before developing estimates, we comment on the similarities and differences between the models (3.1), (3.2), (2.5) and two other models for which explicit maximum likelihood estimation procedures have been derived and implemented.

Writing (3.1) for the fourth model in (2.5) as

\[ Y_{(j)} = \mu + v \tilde{b}_j + \delta_{(j)} , \]

where \( v = (v_1, \ldots, v_p)^t , \tilde{b}_j \sim N(0, 1) , \delta_{(j)} \sim N(0, M_j) \) and all variables are independent, shows the similarity to a factor analysis model with one factor (\( \tilde{b} \)) and loading matrix \( v \). However, in the standard factor analysis model the \( \delta_{(j)} \) all have the same unknown diagonal covariance matrix, as opposed to the present case of differing \( M_j \) which depend on known constants and unknown \( \sigma_i^2 \) (with accompanying statistics \( w_i \)).

The models in Jennrich and Schluchter (1986) are more similar to ours, including independent

\[ Y_{(j)} \sim N(\mu, \Sigma_{(j)}) \]

where the \( \Sigma_{(j)} \) can be quite general functions of known and unknown
parameters. Their log-likelihood consists of the first sum in (3.3), with
\[ M_j + \Gamma = \sum_{(j)} \]. In our case, however, the log-likelihood includes the
second sum in (3.3), with unknown \( \sigma_i^2 \) which also appear in the \( M_j \) in the
first sum. Moreover, in our case the specific structures in (2.5) allow for
fairly simple and explicit expressions for the relevant derivatives of \( \lambda \).

Formulas for the first order partial derivatives of \( \lambda \), together with
expected values of the mixed second order partials, are derived in the
appendix. These formulas can be used to compute maximum likelihood esti-
mates iteratively, via a scoring algorithm, and to obtain their asymptotic
covariances. The following notation is used in the formulas.

\[
A_j = (M_j + \Gamma)^{-1}
\]

\[
B_j = A_j e_{(j)} e_{(j)}^t A_j
\]

\[
u_{(i)} = (0, \ldots, 1, \ldots, 0)^t, \text{ the } i\text{-th unit vector in } \mathbb{R}^P.
\]

Also, if \( A_{p \times p} = (a_{ij}) \) is a matrix whose elements are functions of a scalar
parameter \( \xi \), then

\[
\frac{\partial}{\partial \xi} A = \left( \frac{\partial a_{ij}}{\partial \xi} \right)_{p \times p}.
\]

If \( \xi \) and \( \psi \) are vectors of length \( r \geq 1 \) and \( s \geq 1 \), then denote

\[
\nabla_{\xi, \lambda} = \left( \frac{\partial \lambda}{\partial \xi_i} \right)_{r \times 1}
\]

and

\[
H(\xi, \psi) = E \left( \frac{\partial^2 \lambda}{\partial \xi_i \partial \psi_j} \right)_{r \times s}.
\]
If the formulas in the appendix are used to compute maximum likelihood estimates iteratively via a scoring algorithm, in which the vector of estimates is updated at the k-th step by

\[ \xi_{(k+1)} = \xi_{(k)} - H \left[ \xi_{(k)}, \xi_{(k)} \right]^{-1} \cdot v \xi_{(k)} \],

(3.5)

then the structure of models 1-5 can be used to shorten computation time as follows. The formulas all depend on the matrices \( A_j \) in (3.4). Thus, at the k-th step the \( n \) matrices \((M_1 + \Gamma), \ldots, (M_n + \Gamma)\) (formed using the current values of \( \Gamma \) and \( \sigma_i^2 \)) need to be inverted, prior to forming the components of \( H \) and \( \nabla \lambda \) in (3.5). However, in models 1-5 the matrices \( M_j + \Gamma \) are either diagonal or of the form

\[ D_j + xx^t \]

where \( D_j \) is diagonal and \( x \) is a vector. Hence, the inverse is quickly computed by the well-known formula

\[ (D_j + xx^t)^{-1} = D_j^{-1} - (1 + x^t D_j^{-1} x)^{-1} D_j^{-1} xx^t D_j^{-1} \].

(3.6)

The scoring procedure does not take into account the restrictions that \( \sigma_b^2, \sigma_c^2 \) and \( \nu \) be positive when they occur in the model being fitted. For example, if in fact \( \sigma_c^2 = 0 \) and model 3 in (2.5) is fitted, there is a nonzero probability of obtaining a negative estimate of \( \sigma_c^2 \). If this happens, one can take this as evidence that \( \sigma_c^2 = 0 \) and "drop down" to model 2 in which \( \sigma_c^2 = 0 \). Similar remarks apply to \( \sigma_b^2 \). Regarding \( \nu \), \( \nu = 0 \) would correspond in (1.3) to

\[ m_{ij} = \mu + \alpha_i + \theta_i b_j + c_{ij} \]
with \( \theta_i = 0 \) and \( b_j \sim N(0, \sigma_b^2) \). Since \( \theta_i b_j = (A^{\theta_i}) \cdot (b_j/A) \) for any non-zero constant \( A \), \( \sigma_b^2 \) and \( \theta_i \) would no longer be identifiable. Moreover, since \( A \) could be -1, setting \( \sigma_b^2 = 1 \) would not help. Thus we have not considered models with \( \nu_i = 0 \).

If model 4 or 5 is fitted, maximum likelihood estimates of the \( \theta_i \) and \( \sigma_b^2 \) in (1.3) may be obtained from the \( \hat{\nu}_i \) by the transformation

\[
\theta_i = \frac{\nu_i}{\bar{\nu}} - 1 \quad 1 \leq i \leq p
\]

(3.7)

\[
\sigma_b^2 = (\bar{\nu})^2
\]

One can then estimate the asymptotic covariance matrix of \( (\hat{\theta}_1, \ldots, \hat{\theta}_p, \hat{\sigma}_b^2) \) by

\[
J(\hat{\nu}) \cdot \text{cov}(\hat{\nu}) \cdot [J(\hat{\nu})]^t
\]

(3.8)

where \( \text{cov}(\hat{\nu}) \) is obtained from the information matrix and

\[
J(\nu) = \begin{bmatrix}
\frac{p}{\nu} & I & - \frac{1}{\nu^2} & \vdots & \vdots & \vdots \\
\frac{1}{\nu} & v_1 & v_1 & \cdots & v_1 & \vdots \\
\frac{v_2}{v} & v_2 & v_2 & \cdots & v_2 & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\frac{v_p}{v} & v_p & v_p & \cdots & v_p & \vdots \\
\frac{2v}{p} & \frac{2v}{p} & \cdots & \frac{2v}{p} & \frac{2v}{p} & \frac{2v}{p} \\
\frac{v^2}{p} & \frac{v^2}{p} & \cdots & \frac{v^2}{p} & \frac{v^2}{p} & \frac{v^2}{p}
\end{bmatrix}
\]

is the \((p+1) \times p\) matrix of partial derivatives of the transformation (3.7).
4. AN EXAMPLE

To illustrate we consider some plant genetics data, kindly supplied by Yonatan Elkind of the Hebrew University Faculty of Agriculture. The data were collected to study the combined effects of a single gene (NOR) and remaining polygenes on the texture of tomato fruit. The NOR gene is known to have a major effect on tomato firmness, but also has adverse effects on tomato color and taste. Since these other traits are also controlled by polygenes, it may be possible to modify these undesirable NOR effects by manipulating the genetic background. For this to be successful in breeding firm tomatoes, however, it is first essential to understand the nature of the interaction (if any) between NOR and the polygenes in affecting the firmness of the fruit. More details can be found in Elkind (1987) and Elkind and Cahaner (1986).

In this example, \( n = 21 \) mother plants were randomly sampled from a genetically segregating population of tomato plants, heterozygous for the NOR gene. (A plant's NOR genotype was determined by observing the color of fruit picked from it.) From each mother plant, a family of 30 offspring plants was obtained via self-pollination. Fruit were then picked from the offspring plants and their softness measured. This was done by compressing the fruit between two metal plates a distance of approximately 0.5 mm and measuring the force required. The softness was then defined as the ratio of the distance travelled by the plates to the force required. Measurements were made on four fruit per offspring plant (and the measurements then averaged) at intervals of 1, 9 and 18 days after being picked.
The data matrix has the following structure: Let \( j = 1, 2, \ldots, 21 \) denote the family and \( i = 1, 2, 3 \) (corresponding to nor/nor, nor/+ and +/+ respectively) denote the NOR genotype of an offspring plant. Physical constraints precluded obtaining softness measurements on each of the 630 offspring, so within each family a number of nor/+ plants were randomly excluded. The number of offspring \( K_{ij} \) measured for each \( \text{NOR} \times \text{family} \) combination \((i, j)\) are given in the last three columns of Table 1. The data are then

\[
X_{ijkl} : \quad i = 1, 2, 3 \\
j = 1, \ldots, 21 \\
k = 1, \ldots, K_{ij} \\
\ell = 1, 9, 18
\]

where \( X \) (cm/kg) is the average of the softness measurements on the four fruit and \( \ell \) indexes the day after picking on which the measurements were made.

Clearly the \( X_{ijkl} \) should tend to increase with \( \ell \); moreover, a systematic type of increase (for example, linear) could make \( X \) potentially useful in defining an index of shelf life. To examine this, the differences

\[
u_{ijk} = X_{ijk9} - X_{ijkl}
\]

and

\[
v_{ijk} = X_{ijk18} - X_{ijk9}
\]

were computed. Figure 1 contains a scattergram of \{\((u_{ijk}, v_{ijk})\)\}, together with the line \( v = (9/8)u \). For the vast majority of observations, both \( u_{ijk} \)
and \( v_{ijk} \) are positive, indicating that \( X \) increases with time. Also, the dispersion about the line (on which the points would lie if the increases were exactly linear) does suggest a tendency to increase linearly. This tendency may not be strong enough to warrant treating the \( \{ X_{ijk} \} \) via a growth curve model, but it does indicate the usefulness of the given definition of softness for this data set. In the remainder of our discussion we shall concentrate on the softness measurements on the first day; that is,

\[
Y_{ijk} = X_{ijk1}
\]

in what follows.

Table 2 contains the results of fitting the model (1.2) to \( \{ Y_{ijk} \} \) and testing the hypothesis that \( \sigma^2_c = 0 \). The estimates were obtained via BMDP3V using maximum likelihood. Note that neither the estimated \( \sigma^2_c \) and its standard error under the full model nor the chi-squared statistic indicate a NOR-family interaction.

Normal plots (not shown) of the three sets of residuals \( \{ Y_{1jk} - \hat{\mu}_1 \} \), \( \{ Y_{2jk} - \hat{\mu}_2 \} \) and \( \{ Y_{3jk} - \hat{\mu}_3 \} \) were satisfactorily straight but had widely differing slopes, indicating unequal error variances. Figure 2 contains a graph of the \( \overline{Y}_{ij} - \overline{Y} \) versus \( \overline{Y}_{.j} - \overline{Y} \), as discussed at the end of Section 2. The graph suggests that a model with multiplicative effects might be appropriate.

A natural approach is to try transforming the data and then performing a standard analysis using the model (1.2). A logarithmic transform did not stabilize the variances (Barlett's test for equal \( \sigma^2_i \) gave a \( \chi^2 \) statistic of 33.14 based on 2 degrees of freedom). Transforming to \( 10/Y \) did in fact stabilize the variances, but the resulting response variable (which now
measures firmness) did not behave satisfactorily over time. This can be seen from Figure 3, which is analogous to Figure 1 except that the differences are computed using \( \tilde{X}_{ijkl} = 10/X_{ijkl} \). The correlation of the points in Figure 3 is -0.14, as opposed to 0.52 for Figure 1. The data were therefore analyzed in their original units, using the model (1.3). Regarding the question of transformations in general, note that transforming the data and then analyzing them via model (1.2) would have no particular computational or distributional advantages, since the resulting model would in any event be estimated by iterative maximum likelihood techniques with resulting asymptotic inferences. This is in contrast to regression, where transforming the dependent variable to obtain a linear model enables noniterative estimation with an exact distributional theory.

In performing the analysis, computations were done by a scoring algorithm (written in Fortran) with step halving to insure that the likelihood increased at each iteration. The models were fitted in reverse order (starting with model 6 in (2.5)), using the estimated \( \mu, \Gamma \) and \( \sigma_i^2 \) from each model as starting values for the next. Starting values for model 6 were obtained from the \( w_i \) and the sample mean and covariance matrix of the \( Y_{(j)} \); the \( w_i \) and \( Y_{(j)} \) are given in Table 1. Iteration stopped when the maximum difference in estimated parameters was 0.0001; the number of iterations required for each model is given in Table 3. Observe that the algorithms converged fairly rapidly.

Table 3 summarizes the results of fitting the six models to the data. In Table 3, the "\( \chi^2 \)" entry for a given submodel is twice the difference between the log-likelihood under model 6 and the log-likelihood under the
submodel, and "df" gives the corresponding asymptotic degrees of freedom. Thus a small \( \chi^2 \) (corresponding to a large p-value) indicates a satisfactory fit. Observe that model 5 alone provides an adequate fit. In particular, model 3 (the usual mixed analysis of variance model, modified to accommodate unequal error variances) does not fit the data.

Table 4 contains the parameter estimates and standard errors for model 5 in terms of \( \sigma^2_b \) and the \( \theta_i \) in (1.3). The estimated standard errors of \( \theta_i \) and \( \sigma^2_b \) were obtained using (3.8).

Comparing Tables 2 and 4, observe that the estimated mean levels for the NOR genotype, \( \hat{\mu}_i \), are virtually identical under both models. The main difference is that the estimated standard errors in Table 4 reflect the unequal error variances. The estimated variance component due to the family additive effects, \( \hat{\sigma}^2_b \), is also similar in both tables.

The results for the NOR - family interaction, however, are very different under the two models. Whereas Table 2 does not indicate a significant interaction, Tables 3 and 4 do. Moreover, the NOR - family factor can be broken down into two components:

- A multiplicative effect, with coefficients \( \hat{\theta}_1 \) significantly different from zero. This can be seen by comparing models 3 and 5 in Table 3 and obtaining a \( \chi^2 \) of 10.37 based on 2 degrees of freedom. Comparing the \( \hat{\theta}_i \) with their standard errors indicates that the multiplicative effect is primarily due to the nor/nor genotype (i = 1). In fact, we have the estimated contrast \( \hat{\psi} = \hat{\theta}_1 - \frac{1}{2}(\hat{\theta}_2 + \hat{\theta}_3) = 0.707 \), with an estimated standard error of 0.188. The standard error is obtained using the entire asymptotic covariance matrix.
\[
cov(\hat{\theta}) = \begin{bmatrix}
0.015723 & -0.002778 & -0.012946 \\
0.027824 & 0.025046 \\
0.037991 
\end{bmatrix},
\]

obtained as in (3.8).

- A random interaction, with a significant estimated variance component \( \hat{\sigma}^2 \). This can be seen by comparing models 4 and 5 in Table 3, giving \( \chi^2 = 6.00 \) based on one degree of freedom.
V. GENERALIZATIONS

In applications of mixed analysis of variance models often there are several fixed and random factors, possibly with nesting. When entering multiplicative terms into such a model three points are important:

(i) interpretability,
(ii) mathematical tractability
and
(iii) computational feasibility.

If a model has several fixed and random factors there are three types of multiplicative terms which might be considered, depending on the problem: fixed $\times$ fixed, fixed $\times$ random, and random $\times$ random. Fixed $\times$ fixed terms can be interpreted as in the various references cited in the introduction. Their estimation in the presence of random effects should present no particular difficulties, since the corresponding $H(\mu, \xi)$ terms (where $\xi$ is a parameter in $\Gamma$) will still be zero as indicated in the appendix.

Random $\times$ random terms appear intractable, and might be somewhat difficult to interpret if both factors have a large number of levels.

Additional fixed $\times$ random terms may make sense, and in principle can be accommodated by writing the model in the form (1.4) where some components of $Z$ are unknown. In practice, one would take advantage of the structure of the problem both in computing the relevant derivatives and programming an algorithm.

For example, in the machine-worker problem suppose that instead of comparing $p$ different machines, one is comparing $p = rs$ different
two-stage methods of manufacture, where the first stage can be done in one of \( r \) different ways and the second in one of \( s \). One model it might make sense to fit is

\[
m_{ij} = \mu + \alpha_i + \beta_f + \gamma_{if} + b_j + \theta_i b_j + \psi_f b_j + c_{ifj}
\]

where \( i = 1, \ldots, r \) indexes the first-stage method, \( f = 1, \ldots, s \) indexes the second, \( m_{ij} \) is the average time required by worker \( j \) when using method \( i \) followed by method \( f \), both \( \theta_i \) and \( \psi_f \) sum to zero as in (1.3), and \( \{b_j\} \) and \( \{c_{ifj}\} \) are independent random effects with variances \( \sigma_b^2 \) and \( \sigma_c^2 \).

If

\[
(m_{11j}', \ldots, m_{1sj}', \ldots, m_{r1j}', \ldots, m_{rsj})^t
\]

denotes worker \( j \)'s \( p \)-vector of expected times using the \( rs \) possible combinations and the corresponding vector

\[
\eta_{(j)} = (\eta_{11j}', \ldots, \eta_{pj})^t
\]

of random components is defined as in (2.3) (where now the first index in \( \eta_{hj} \) refers to a particular \((i,f)\) pair), then the covariance matrix of \( \eta_{(j)} \) may be written (analogous to model 5 in (2.5)) as

\[
\Gamma = \sigma_c^2 \cdot I + (Fz)(Fz)^t.
\]

Here,

\[
z = (z_1, z_2, \ldots, z_{r+s-1})^t
\]

is a full-rank reparameterization of the parameters.
\[ \{ \sigma_b, \sigma_{b1}, \ldots, \sigma_{br}, \sigma_b \psi_1, \ldots, \sigma_b \psi_s \} \]

subject to the constraints \( \theta = \psi = 0 \) and \( F \) is the corresponding fixed, known design matrix.

Model 5's corresponding derivatives with respect to \( v_i \), given in the appendix, are then replaced by (denoting the i-th column of \( F \) by \( f_{(i)} \))

\[
\frac{\partial}{\partial z_i} (M_j + \Gamma) = f_{(i)} (Fz)^t + (Fz) f_{(i)}^t ,
\]

\[
\nabla z \lambda = F^t \sum_{j=1}^{n} (B_j - A_j) Fz
\]

and

\[
H(z_i, z_{(k)}) = - \sum_{j=1}^{n} \left[ f_{(i)}^t A_j Fzz^t F^t A_j f_{(k)} + f_{(i)}^t A_j f_{(k)} \cdot z^t F^t A_j Fz \right]
\]

whence

\[
H(z, z) = - \sum_{j=1}^{n} \left[ T_j zz^t T_j + z^t T_j z \cdot T_j \right]
\]

for

\[
T_j = F^t A_j F .
\]

The computation of \( A_j \) can again be expedited by applying (3.6) with \( x \) replaced by the vector \( Fz \).

Similar simplifications would hold in other generalizations, but a systematic investigation of this question is beyond the scope of this paper.
ACKNOWLEDGMENTS

I would like to thank the following individuals for many useful conversations during the course of this research: Yonatan Elkind, Ian Abramson, Robert Jennrich, Ingram Olkin, Danny Pfefferman and John Rice.
APPENDIX. COMPUTATION OF DERIVATIVES

Derivatives for the log-likelihood \( \lambda \) are obtained as follows.

Regarding \( \mu \), it is immediate from (3.3) that

\[
\nabla_\mu \lambda = \sum_{j=1}^{n} A_j(Y_{(j)} - \mu)
\]

and

\[
H(\mu, \mu) = -\sum_{j=1}^{n} A_j.
\]

For \( \Gamma \), if \( \xi \) and \( \psi \) are scalar parameters on which \( \Gamma \) depends, then standard formulas for matrix derivatives give (see, e.g., Jennrich and Schluchter, 1986)

\[
\frac{\partial \lambda}{\partial \xi} = \frac{1}{2} \sum_{j=1}^{n} \text{tr} \left[ (B_j - A_j) \frac{\partial}{\partial \xi} (M_j + \Gamma) \right]
\]  \hspace{1cm} (A.1)

and

\[
H(\psi, \xi) = -\frac{1}{2} \sum_{j=1}^{n} \text{tr} \left\{ A_j \left[ \frac{\partial}{\partial \psi} (M_j + \Gamma) \right] A_j \left[ \frac{\partial}{\partial \xi} (M_j + \Gamma) \right] \right\} , \hspace{1cm} (A.2)
\]

where \( \text{tr} \) denotes the trace. Possible values of \( (\partial/\partial \xi)(M_j + \Gamma) \) for the various models in (2.5) are

\[
\frac{\partial}{\partial \sigma_j^2} (M_j + \Gamma) = I ,
\]

\[
\frac{\partial}{\partial \sigma^2_{cb}} (M_j + \Gamma) = ww^t ,
\]

24
\[ \frac{\partial}{\partial v_i} (M_j + \Gamma) = v_i u^t (i) + u^t (i) v_i \quad i = 1, \ldots, p \]

and

\[ \frac{\partial}{\partial \Gamma_{kl}} (M_j + \Gamma) = \begin{cases} u^t (k) u^t (k) & k = \ell = 1, 2, \ldots, p \\ u^t (k) u^t (\ell) + u^t (\ell) u^t (k) & k, \ell = 1, \ldots, p \\ & k \neq \ell \end{cases} \]

Substituting into (A.1) gives

\[ \frac{\partial \lambda}{\partial \sigma_c^2} = \frac{1}{2} \sum_{j=1}^{n} \text{tr}(B_j - A_j) , \]

\[ \frac{\partial \lambda}{\partial \sigma_b^2} = \frac{1}{2} \sum_{j=1}^{n} \omega^t (B_j - A_j) \omega \]

and

\[ \nabla \lambda = \sum_{j=1}^{n} (B_j - A_j) v \]

in the appropriate cases for models 1-5. For model 6,

\[ \frac{\delta \lambda}{\partial \Gamma_{kl}} = \begin{cases} \frac{1}{2} \sum_{j=1}^{n} (B_j - A_j)_{kk} & k = \ell = 1, \ldots, p \\ \sum_{j=1}^{n} (B_j - A_j)_{kl} & k, \ell = 1, \ldots, p \\ \sum_{j=1}^{n} (B_j - A_j)_{k\ell} & k \neq \ell \end{cases} \]

From (A.2) we have
\[ H(\sigma^2_c, \sigma^2_c) = -\frac{1}{2} \sum_{j=1}^{n} \text{tr}(A_j^2), \]

\[ H(\sigma^2_b, \sigma^2_b) = -\frac{1}{2} \sum_{j=1}^{n} (w^t A_j w)^2 \]

and

\[ H(v, v) = -\sum_{j=1}^{n} [A_j vv^t A_j + v^t A_j v \cdot A_j] \]

for models 1-5, and

\[ H(\Gamma_{kl}, \Gamma_{rs}) = -\frac{1}{2} \sum_{j=1}^{n} [(A_j)_{rk} (A_j)_{ls} (1 + 1) (r \neq s \text{ and } k \neq l)] \]

\[ + (A_j)_{rk} (A_j)_{ls} (1 + 1) (r \neq s) + (k \neq l)] \]

for model 6. The required mixed terms are as follows:

**Model 3:** \[ H(\sigma^2_b, \sigma^2_c) = -\frac{1}{2} \sum_{j=1}^{n} w^t A_j^2 w \]

**Model 5:** \[ H(v, \sigma^2_c) = -\sum_{j=1}^{n} A_j^2 v \]

Next consider the \( \sigma^2_i \). Since

\[ \frac{\partial}{\partial \sigma^2_i} (M_j + \Gamma) = K_{ij}^{-1} u_i u_i^t, \]

it is immediate from (3.3), (A.1), (A.2) and (3.2) that for \( 1 \leq i, k \leq p \).
\[
\frac{\partial \lambda}{\partial \sigma_i^2} = \frac{1}{2} \left[ \sum_{j=1}^{n} \left( B_j - A_j \right)_{ij} / K_{ij} + \frac{w_j}{2 \sigma_i^4} - \frac{m_i}{\sigma_i^2} \right]
\]

and

\[
H(\sigma_i^2, \sigma_k^2) = -\frac{1}{2} \left\{ \sum_{j=1}^{n} \left( \frac{(A_j)_{ij}}{K_{ij}} \right)^2 + \delta_{ij} \frac{m_i}{\sigma_i^4} \right\}
\]

here, \( \delta_{ik} \) is the Kronecker delta.

For the remaining mixed terms, if \( \xi \) is a parameter in \( \Gamma \) then

\[
H(\mu, \xi) = -\mathbb{E} [(\eta_{\mu, \lambda})(\partial \lambda / \partial \xi)] = 0,
\]

as it depends on odd moments of the \( \varepsilon_{(j)} \); similarly for \( H(\mu, \sigma_i^2) \). Regarding \( \sigma_i^2 \) and parameters of \( \Gamma \), substituting the appropriate partials of \( (M_j + \Gamma) \) into (3.6) gives for \( 1 \leq i, k, \ell \leq p \)

\[
H(\sigma_i^2, \sigma_c^2) = -\frac{1}{2} \sum_{j=1}^{n} \left( \frac{A_j^2}{K_{ij}} \right)_{ii} ,
\]

\[
H(\sigma_i^2, \sigma_b^2) = -\frac{1}{2} \sum_{j=1}^{n} \frac{(A_j w_{(j)})_{i}^2}{K_{ij}} ,
\]

\[
H(\sigma_i^2, v_k) = -\sum_{j=1}^{n} (A_j)_{ik} \cdot (A_j v)_{i} / K_{ij}
\]

and

\[
H(\sigma_i^2, \Gamma_{k\ell}) = \begin{cases}
-\frac{1}{2} \sum_{j=1}^{n} \frac{(A_j)_{ik} (A_j)_{i\ell}}{K_{ij}} & \text{k = } \ell \\
-\sum_{j=1}^{n} \frac{(A_j)_{ik} (A_j)_{i\ell}}{K_{ij}} & \text{k \neq } \ell
\end{cases}
\]
REFERENCES


TABLE 1: Tomato softness data.

<table>
<thead>
<tr>
<th>i (NOR genotype)</th>
<th>$\bar{Y}_{ij}$</th>
<th>K_{ij}</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>11.609</td>
<td>14.005</td>
<td>15.987</td>
</tr>
<tr>
<td>2</td>
<td>10.120</td>
<td>15.615</td>
<td>18.955</td>
</tr>
<tr>
<td>3</td>
<td>10.120</td>
<td>16.839</td>
<td>19.508</td>
</tr>
<tr>
<td>4</td>
<td>11.238</td>
<td>16.940</td>
<td>17.790</td>
</tr>
<tr>
<td>5</td>
<td>10.834</td>
<td>12.963</td>
<td>17.198</td>
</tr>
<tr>
<td>6</td>
<td>12.442</td>
<td>15.663</td>
<td>23.988</td>
</tr>
<tr>
<td>7</td>
<td>11.637</td>
<td>17.210</td>
<td>18.262</td>
</tr>
<tr>
<td>8</td>
<td>10.292</td>
<td>14.777</td>
<td>18.295</td>
</tr>
<tr>
<td>9</td>
<td>10.140</td>
<td>12.453</td>
<td>17.824</td>
</tr>
<tr>
<td>10</td>
<td>11.583</td>
<td>19.256</td>
<td>21.967</td>
</tr>
<tr>
<td>11</td>
<td>11.077</td>
<td>15.380</td>
<td>19.479</td>
</tr>
<tr>
<td>12</td>
<td>10.810</td>
<td>17.277</td>
<td>18.986</td>
</tr>
<tr>
<td>13</td>
<td>10.939</td>
<td>15.083</td>
<td>19.205</td>
</tr>
<tr>
<td>14</td>
<td>12.415</td>
<td>17.117</td>
<td>22.124</td>
</tr>
<tr>
<td>15</td>
<td>10.816</td>
<td>16.620</td>
<td>19.230</td>
</tr>
<tr>
<td>16</td>
<td>9.549</td>
<td>14.198</td>
<td>16.804</td>
</tr>
<tr>
<td>17</td>
<td>13.028</td>
<td>18.432</td>
<td>23.012</td>
</tr>
<tr>
<td>19</td>
<td>12.194</td>
<td>17.304</td>
<td>20.745</td>
</tr>
<tr>
<td>20</td>
<td>10.874</td>
<td>15.307</td>
<td>16.305</td>
</tr>
<tr>
<td>21</td>
<td>11.705</td>
<td>14.661</td>
<td>21.055</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>i</th>
<th>$w_i$</th>
<th>$m_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>112.420</td>
<td>102</td>
</tr>
<tr>
<td>2</td>
<td>952.036</td>
<td>178</td>
</tr>
<tr>
<td>3</td>
<td>1596.159</td>
<td>118</td>
</tr>
</tbody>
</table>
### TABLE 2: Results of additive fit.

A. **Full Model** \( Y_{ijk} = \mu_i + b_j + c_{ij} + \epsilon_{ijk} \)

<table>
<thead>
<tr>
<th>Source</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOR genotype</td>
<td>( \mu_1 )</td>
<td>11.154</td>
<td>0.378</td>
</tr>
<tr>
<td></td>
<td>( \mu_2 )</td>
<td>15.788</td>
<td>0.350</td>
</tr>
<tr>
<td></td>
<td>( \mu_3 )</td>
<td>19.115</td>
<td>0.370</td>
</tr>
<tr>
<td>Family</td>
<td>( \sigma_b^2 )</td>
<td>1.505</td>
<td>0.604</td>
</tr>
<tr>
<td>NOR – family interaction</td>
<td>( \sigma_c^2 )</td>
<td>0.337</td>
<td>0.287</td>
</tr>
<tr>
<td>Error</td>
<td>( \sigma^2 )</td>
<td>6.693</td>
<td></td>
</tr>
</tbody>
</table>

B. **Constrained Model** \( Y_{ijk} = \mu_i + b_j + \epsilon_{ijk} \)

<table>
<thead>
<tr>
<th>Source</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family</td>
<td>( \sigma_b^2 )</td>
<td>1.629</td>
<td>0.601</td>
</tr>
<tr>
<td>Error</td>
<td>( \sigma^2 )</td>
<td>6.914</td>
<td></td>
</tr>
</tbody>
</table>

C. **Likelihood Ratio Test for** \( H_0: \sigma_c^2 = 0 \)

\[ \chi^2 = 1.85 \]

**Degrees of freedom = 1**

**p-value = 0.17**
TABLE 3: Results for the different models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Form</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p-value</th>
<th>No. It.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\Gamma = \sigma^2_c \cdot I$</td>
<td>30.10</td>
<td>5</td>
<td>.00</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>$\Gamma = \sigma^2_b \cdot w w^t$</td>
<td>21.20</td>
<td>5</td>
<td>.00</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>$\Gamma = \sigma^2_b \cdot w w^t + \sigma^2_c \cdot I$</td>
<td>13.79</td>
<td>4</td>
<td>.01</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>$\Gamma = v v^t$</td>
<td>9.42</td>
<td>3</td>
<td>.02</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>$\Gamma = v v^t + \sigma^2_c \cdot I$</td>
<td>3.42</td>
<td>2</td>
<td>.18</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>$\Gamma$ arbitrary</td>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
</tbody>
</table>

$\chi^2_i = 2 \log(\text{max. lik. under model } i / \text{max. lik. under model } 6)$

df$_i = (\text{no. parameters in model } i) - (\text{no. parameters in model } 6)$

p-value$_i = P[\chi^2(\text{df}_i) > \chi^2_i]$

no. it. = number of iterations required
<table>
<thead>
<tr>
<th>Source</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOR genotype</td>
<td>$\mu_1$</td>
<td>11.128</td>
<td>0.201</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>15.791</td>
<td>0.358</td>
</tr>
<tr>
<td></td>
<td>$\mu_3$</td>
<td>19.113</td>
<td>0.487</td>
</tr>
<tr>
<td>Family</td>
<td>$\sigma^2_b$</td>
<td>1.487</td>
<td>0.581</td>
</tr>
<tr>
<td>NOR × family (multiplicative)</td>
<td>$\theta_1$</td>
<td>-0.471</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td>$\theta_2$</td>
<td>0.119</td>
<td>0.167</td>
</tr>
<tr>
<td></td>
<td>$\theta_3$</td>
<td>0.352</td>
<td>0.342</td>
</tr>
<tr>
<td>NOR - family interaction (random)</td>
<td>$\sigma^2_c$</td>
<td>0.237</td>
<td>0.145</td>
</tr>
<tr>
<td>Error</td>
<td>$\sigma^2_1$</td>
<td>1.094</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_2$</td>
<td>5.412</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_3$</td>
<td>13.156</td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 1: Paired differences for untransformed data.
FIGURE 2: Average softness vs family effect.
FIGURE 3: Paired differences after inverse transformation.