IMPROVED ESTIMATION OF VARIANCE COMPONENTS
IN MIXED MODELS

BY

ALAN E. GELFAND and DIPAK K. DEY

TECHNICAL REPORT NO. 434
AUGUST 30, 1990

PREPARED UNDER CONTRACT
N00014-89-J-1627 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
IMPROVED ESTIMATION OF VARIANCE COMPONENTS
IN MIXED MODELS

BY

ALAN E. GELFAND and DIPAK K. DEY

TECHNICAL REPORT NO. 434
AUGUST 30, 1990

Prepared Under Contract
N00014-89-J-1627 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ABSTRACT

Taking Albert's (1976) formulation of a mixed model ANOVA, we consider improved estimation of the variance components for balanced designs under squared error loss. Two approaches are presented. One extends the ideas of Stein (1964). The other is developed from the fact that variance components can be expressed as linear combinations of chi-square scale parameters. Encouraging simulation results are presented.

1. INTRODUCTION

Albert (1976) exhibits necessary and sufficient conditions for a sum of squares decomposition under a mixed model to be an ANOVA, i.e., for the terms of the decomposition to be independent and to be distributed as multiples of chi-square. We consider improved estimation of the variance components under squared error loss in such a set-up. We make no attempt to discuss the enormous literature on this problem. See Harville (1977) for such a review. Rather, we specialize to the "balanced" case considering designs consisting of crossed and nested classifications and combinations thereof. Rules of thumb for formalizing the associated ANOVA table are thus well known (see, e.g., Searle (1971, Chap. 9)). Customary estimators of the variance components are the unbiased ones obtained as described in Searle, pp. 405-6. Under normality these estimators are UMVU (Graybill (1954), Graybill and Wortham (1956)) and, in fact, restricted maximum likelihood (REML) (Thompson (1962)). However, positive part corrections are usually taken yielding improved mean square error but sacrificing these "optimalities." Bayesian approaches to variance component estimation in this setting are discussed in, e.g., Hill (1965) and Box and Tiao (1973).

Since the positive part estimators are not smooth and, thus, not admissible under squared error loss (SEL), it is natural to
seek dominating estimators. The earliest work of this type is due to Klotz, Milton and Zacks (1969) for the one-way layout. They show, for example, that the MLE of the "between" variance component (see Herbach (1959)), which is itself a positive part estimator, dominates the UMVU and, using ideas of Stein (1964), that it in turn can be dominated.

The objective of this paper is to describe two general approaches for creating improved estimators of the variance components under SEL. In Section 2 we develop a method by extending the aforementioned Stein idea. We have discussed a special case of this approach in linear regression models in Gelfand and Dey (1987b). Since the variance components are linear combinations of chi-square scale parameters, we can draw upon some literature for improved estimation of linear combinations of scale parameters. This second approach is offered in Section 3. Work of Dey and Gelfand (1987) for arbitrary scale parameter distributions and of Klonecki and Zontek (1985, 1987) for the Gamma family of distributions is pertinent here. Finally, in Section 4 we present some simulation results.

In the remainder of this section we develop notation for and features of the model we will be working with. Consider the general balanced mixed model of the form

\[ Y = \mu 1 + \sum_{r=1}^{p} \Theta_{r} + X \beta + \varepsilon \]

(1.1)

where \( Y \) is an \( n \times 1 \) vector of observations, \( \mu \) is an overall mean effect or intercept, \( \Theta_{r} \) are known \( n \times m \) incidence matrices where \( H(1 = 1 \quad \text{and} \quad H^{T}H = \nu I \) (i.e., \( \nu \) is the number of nonzero entries in a typical column of \( H \), \( \tau_{r} \) are independent distributed as \( N(0, \sigma_{r}^{2} I_{r}) \), \( X \) is a known \( n \times s \) design matrix involving
possibly fixed effects and covariates, $\beta$ is the associated $s \times 1$ vector of coefficients and $\varepsilon$ is an $n \times 1$ vector of errors distributed $N(0, \sigma^2 \mathbf{I})$ independent of the $r$. Thus, $Y \sim N(m, W)$ where

$$m = \mu_1 \mathbf{1}_{nx1} + X\beta$$

and $W$ is the patterned covariance matrix

$$W = \sigma^2 \mathbf{I} + \sum_{r=1}^{p} \sigma_r^2 \mathbf{H}_r \mathbf{H}_r^T$$

Let $(\sigma^2) = (\sigma^2_e, \sigma^2_1, \ldots, \sigma^2_p)^T$. Our primary interest is in estimating the $\sigma^2_r$ individually (as it has been done historically) although we shall say something in Section 3 about simultaneous estimation.

As in Albert (1976) we consider a complete set of orthogonal projections, $P_1, P_2, \ldots, P_p, P_e, P_e'P_e, P_e'P_\mu, \sum_{i=1}^{p} P_i + P_e + P_\mu + P_\beta = \mathbf{I}_n$.

In particular, $P_e$ is associated with the error, i.e., $Y^TP_eY$ is the full model error sum of squares. $P_\mu$ is associated with the intercept ($P_\mu = \mu^{-1}_{nxn}$), i.e., $Y^TP_\mu Y = \frac{1}{n} \bar{Y}^2$, where $\bar{Y}$ is the average of the $Y$'s. $SS_e$ is the model sum of squares for the reduced ANOVA model, i.e., $SS_e = Y^T(HH^T)H^TY = \sum_{i=1}^{p} Y_i^T P_i Y + \frac{1}{n} \sum_{i=1}^{p} Y_i$.

where $H = (\mathbf{H}_1 \mathbf{H}_2 \ldots \mathbf{H}_p)$. Note that we have a sum of squares for each random effect. Finally, let $SS_{\beta | H}$ be the sum of squares for the fixed effects and covariates adjusted for the ANOVA, i.e.,
Typically, $\nu_\beta$ is itself expressed as a sum of orthogonal pieces.

According to Albert (1976) (see also Brown (1984) and Harville (1984) in this regard), we have an ANOVA if and only if

$$ H H^T p = \lambda P, \quad k = 1, \ldots, r, \quad H H^T_r \mu = \lambda_r P_r $$

$$ H H^T_r \mu = \lambda_r P_r $$

and

$$ H H^T_r \beta = \lambda_r P_r $$

where $\lambda_r = 0$ or $\nu_r$ according to whether or not $H^T \beta = 0$. Then the $Q_\gamma = Y^T P Y$ are independent and distributed as $\gamma \chi^2_{(k-1,\delta_\gamma)}$ where

$$ \gamma = \frac{2}{p} \sum_{r=1}^{k} \frac{\sigma^2_r}{\nu_r} f = \text{rank}(P) $$

$$ \delta_\gamma = 0, \quad \delta_r = 0, \quad \delta_\mu = 0 $$

Thus, $\gamma = \frac{2}{p} \sum_{r=1}^{k} \frac{\sigma^2_r}{\nu_r}$.

Since $H^T P = 0$ and $H^T \beta = 0$, $\gamma_0 = 0$ and $\gamma_\mu = 0$ determine explicitly without specifying the design.

However, for two random effects, with respective sums of squares $Y^T P Y$ and $Y^T P Y$, if the latter is any nested or crossed effect involving all the factors in the former, then $\gamma_k \geq \gamma_k'$. This is, in fact, Rule 12 of Searle (1971, p. 393). Obviously, $\gamma_k \leq \gamma_k \leq \gamma_\mu$ and typically there is a partial ordering amongst the $\gamma_k$.

Finally, again as in Searle (1971, p. 405), if we define $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_p)$ with $\gamma_0 = \sigma^2_e$ we have $\gamma = A \sigma^2$ where

$$ A = \begin{pmatrix} 1 & \Omega \\ \Omega^T & \lambda \end{pmatrix} $$

with $[\Omega]_{kr} = \lambda_{kr}$ whence
\[ \sigma^2 = A^{-1} \gamma. \] (1.2)

Expression (1.2) reveals a key point. The variance components are expressible as linear combinations of chi-square scale parameters. In fact, this expression is usually employed to create the familiar unbiased estimators of the \( \sigma \) using \( f^{-1} Q_k' k \), the unbiased estimator of \( \gamma_k \).

2. IMPROVED ESTIMATORS USING STEIN'S METHOD

Consider estimating \( a \gamma_k + b \gamma_k' \). For appropriate choices of \( a, b, k, k' \) (in fact \( ab < 0 \)), this parameter will be a variance component. To proceed we utilize the following elementary lemma whose proof is immediate.

**Lemma 2.1.** Let \( S_1 \) be an estimator of \( \theta_1 \) and let \( T_1 \) dominate \( S_1 \) under SEL. Let \( S_2 \) be an estimator of \( \theta_2 \) where \( S_2 \) is independent of \( S_1 \) and \( T_1 \). Then in estimating \( a \theta_1 + b \theta_2 \), \( a T_1 + b S_2 \) dominates \( a S_1 + b S_2 \) under SEL if

\[ ab \mathbb{E}_{\theta_1} (S_1 - T_1) \mathbb{E}_{\theta_2} (S_2 - \theta_2) \geq 0 \] (2.1)

In our applications we will meet (2.1) by having \( ab < 0 \), \( T_1 \leq S_1 \), \( \mathbb{E}_{\theta_2} (S_2) \leq \theta_2 \).

We also require the following result which is a minor generalization of a theorem stated and proved in Gelfand and Dey (1987b).

**Theorem 2.1.** Let \( s_i \sim \gamma_i \xi_i^2 + s_i \sim (\gamma_0 + \phi_i) \lambda_i^2 \),

\( i = 1, \ldots, t \) all independent where \( \phi_i \geq 0, \lambda_i \geq 0 \). Define \( R_j = \sum_{i=1}^{j} c_j S_i \) where \( c_j = \sum_{i=1}^{j} n_i + 2 \) and let \( \delta = \min(R_0, R_1, \ldots, R_j) \).
Then in estimating $\gamma_0$ under SEL $\delta$ \ll $\delta$ \ll $\delta$ \ll $\delta$ \ll $\delta$ where
$\delta$ \ll $\delta$ means $\delta$ dominates $\delta$.

Lastly we need a lemma which appears, for example, in Klotz, Milton and Zacks (1969, p. 1394).

**Lemma 2.2.** If $T \leq S$ and, in estimating $0 > 0$, $S \ll T$ under SEL then $S \ll T$, under SEL where $+$ denotes positive part.

These results will be synthesized in the following way.

Assume $a\gamma_k + b\gamma_{k'} > 0$ and w.l.o.g. that $a > 0$, $b < 0$. This will be the case if $a\gamma_k + b\gamma_{k'}$ defines a variance component. Find the set of all $\gamma_r = \gamma_k$ (excluding $\gamma_{k'}$, regardless). This set is nonempty since at the very least $\gamma_{\mu} \geq \gamma_k$. $Q_r$ and the associated set of $Q_r$ form the $S_0$ and $S_1$, respectively, for Theorem 2.1 and enable the creation of a decreasing sequence of estimators which dominate $(f_k + 2)^{-1} Q_k$, the best invariant estimator of $\gamma_k$. The resultant $\delta$ suitably defined play the role of $S_0$ and $T_1$ in Lemma 2.1 and will be independent of $S_2 = (f_k + \delta)^{-1} Q_{k'}$, $\delta \geq 0$ whence Lemma 2.1 holds. Finally, using Lemma 2.2, $[aT_1 + bS_2] \ll [aT_1 + bS_2]$.

**Remark 2.1.** Theorem 2.1 allows for a variety of improved estimators for $\sigma^2_e$. Let $S = Q_r$ with $S_0$ being the $Q_r$, $r = 1, \ldots, p$ as well as $Q_\alpha$ and $Q_\beta$. Then $t = p + 2$ and we may readily create $\delta$. In fact, corresponding to any specified permutation, $\eta$, of the $S_1$ there will be a resultant $\delta^{\eta}$, i.e., there will be $t!$ such estimators. How might we combine them to produce a permutation invariant estimator? It can be argued that the minimum of these will be "too small" and that the average is a better practical choice. See Gelfand and Dey (1987b) for details.

We illustrate using the one-way ANOVA, $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$,

$i = 1, \ldots, I, j = 1, \ldots, J$, $\alpha_i \sim N(0, \sigma^2_\alpha)$, $\epsilon_{ij} \sim N(0, \sigma^2_e)$, all

independent. In this case $p = 1$ with $Q_1 \sim (\sigma^2_e + J\sigma^2_\alpha)^2 I^{-1}$,
\[ Q_u \sim (\sigma_e^2 + \sigma_\alpha^2) \chi_1^2, \quad \text{and} \quad Q_e \sim \frac{2}{2} \frac{I \!J}{2(\sigma_e^2 + \sigma_\alpha^2)} \chi_{(J-1)}^2. \]

In estimating \( \sigma_e^2 \) we may dominate the best invariant estimator \( R_0 = (I(J-1) + 2)^{-1} Q \) using \( \delta_{e,1} = \min(R_0, R_1) \) which in turn is dominated by

\[ \delta_{e,2} = \min(R_0, R_1, R_2) \quad (2.2) \]

or using \( \delta_{e,1} = \min(R_0, R_1') \) which in turn is dominated by

\[ \delta_{e,2} = \min(R_0, R_1', R_2) \quad (2.3) \]

Here \( R_1 = (I+1)^{-1}(Q_e + Q_1) \), \( R_1' = (I(J-1) + 3)^{-1}(Q_e + Q_{1l}) \) and \( R_2 = (I+2)(Q_e + Q_1 + Q_{1l}) \). Estimators \( \delta_{e,1} \) and \( \delta_{e,2} \) appear in Klotz, Milton and Zacks (1969). In practice, if we suspect \( \sigma_\alpha^2 \) small we would use \( \delta_{e,2} \); if we suspect \( \mu \) small we would use \( \delta_{e,2} \) and if we have no prior suspicions we would recommend

\[ (\delta_{e,2} + \delta_{e,2}')/2 \quad (2.4) \]

Turning to \( \sigma_\alpha^2 \) we may write \( \sigma_\alpha^2 = J^{-1}(\gamma_1 - \gamma_e) \) whence the usual unbiased estimator is given by \( J^{-1}[I(I-1)^{-1}Q_1 = (I(J-1))^{-1}Q_e]. \)

By Lemma 2.1 this is immediately dominated by \( J^{-1}[I(I+1)^{-1}Q_1 - (I(J-1))^{-1}Q_e] \) which in turn is dominated by \( \delta_{\alpha,1} = \]

\[ J^{-1}[\min((I+1)^{-1}Q_1, (I+2)^{-1}(Q_{1l} + Q_{1l}')) - (I(J-1))^{-1}Q_e]. \]

Using Lemma 2.2 we arrive at the positive part version \( \delta_{\alpha,1}^+ \).
Alternatively, again by Lemma 2.1, the usual unbiased estimator is dominated by \( J^{-1}[(I-1)^{-1}Q_1 - (I(\bar{J} - 1) + 2)^{-1}Q_e] \)

which is dominated by \( J^{-1}[(I + 1)^{-1}Q_1 - (I(\bar{J} - 1) + 2)^{-1}Q_e] \) which is dominated by \( J^{-1}[(I + 1)^{-1}Q_1, (I + 2)^{-1}(Q_1 + Q_\mu)] \)

- \( (I(\bar{J} - 1) + 2)^{-1}Q_e \). Again by Lemma 2.2, we arrive at \( \delta^+ \).

The estimator \( \delta_{a,1} \) appears in Klotz, Milton and Zacks. Note that while \( \delta_{a,1} \ll \delta_{a,2} \) since \( \delta_{a,2} > \delta_{a,1} \) we cannot conclude regarding \( \delta_{a,1}^+ \) and \( \delta_{a,2}^+ \). In fact, the simulation results in Section 4 show that neither dominates the other.

In concluding this section we remark that utilizing the ideas in Gelfand and Dey (1987a) and in Gelfand (1987), along with the aforementioned results, we can improve in the estimation of the ratio \( \gamma_k / \gamma_{k'} \). This allows for improved estimation of, e.g., \( \sigma^2 / \sigma^2_{r,e} \). See Loh (1986) in this regard. We omit the details.

Unfortunately, we cannot extend this to, e.g., the intraclass correlation coefficient since it is a non-linear function of such ratios.

3. IMPROVED ESTIMATES USING A GEOMETRIC MEANS APPROACH

In this section we develop a method for obtaining improved estimates which arises from expression (1.2), the fact that the variance components are expressible as linear combinations of chi-square scale parameters. Consider a single \( \sigma^2_r \) which we write as

\[
\sigma^2_r = \sum_{k=0}^{p} c_k \gamma_k \text{ and let } \sum_{k=0}^{p} \ell_k Q_k \text{ be a candidate estimator. Here we denote } Q_e \text{ by } Q_0. \text{ When can } \sum_{k=0}^{p} \ell_k Q_k \text{ be dominated and what is the }
\]
form of the dominating estimator? Dey and Gelfand (1987) discuss this problem when the $\gamma_k$ are scale parameters from arbitrary distributions. Klonecki and Zontek (1985, 1987), assuming the $\gamma_k$ are scale parameters from Gamma distributions, obtain conditions which enable assessment of linear admissibility for $\sum_k \gamma_k^2$, i.e., admissibility within the class of linear estimators of $\sigma_r^2$. They also offer a slightly broader class of dominating estimators than in Dey and Gelfand (1987).

More precisely, the following result appears in Dey and Gelfand (1987).

Theorem 3.1. Let $Y_{i1} \sim \theta_{i1}$, $i = 1, 2, \ldots, t$, $t \geq 2$, independent and such that $E Y_{i1}^2 < \infty$. Consider the estimator

$$
\sum_{i=1}^{t} \frac{Y_{i1}^{1/t}}{i} + b \frac{1}{t} \sum_{i=1}^{t} \frac{Y_{i1}}{i}
$$

(3.1)

Let $d_{i1} = c_i - \frac{a_i}{i}$, where $a_i = E(Y_{i1}^{1+t-1})/E(Y_{i1}^{t-1}) | \theta_i = 1)$ and

$$
d_{(1)} = \min d_{i1}, \quad d_{(t)} = \max d_{i1}.
$$

Then (3.1) dominates $\sum_i \gamma_i$ under SEL in estimating $\sum_i \gamma_i$ if either

(i) $d_{(1)} > 0$ and $0 < b < 2td_{(1)}$

(ii) $d_{(t)} < 0$ and $2td_{(t)} < b < 0$

where $\rho = \frac{1}{t} \frac{E(Y_{i1}^{t-1}) | \theta_i = 1)/E(Y_{i1}^{2t-1}) | \theta_i = 1)}{E(Y_{i1}^{1+t-1})/E(Y_{i1}^{t-1}) | \theta_i = 1)}$.

If we denote by $G(\alpha_{i1}, \theta_{i1})$ the gamma density $f_{i1}(y)$ =
\[ \alpha_i^{-1} \text{e}^{-y_i/\theta_i} \]
\[ \theta_i \sim \Gamma(\alpha_i) \]

then \( a_i = \alpha_i + t^{-1} \) and \( \rho = \prod_i (\alpha_i + t^{-1})/ (\alpha_i + 2t^{-1}) \).

Now let \( D \) be a diagonal matrix whose diagonal entries are the \( \alpha_i \) and let \( G \) be of the form \( E_d^{-1} \) where \( \Sigma \) is a nonnegative definite matrix such that \( (\Sigma)_{ij} \geq 0 \), \( (\Sigma)_{ii} > 0 \) and \( E_d \) is a diagonal matrix whose diagonal entries are \( (\Sigma)_{ii} \). Then Klonecki and Zontek (1985) show:

**Theorem 3.2.** If \( Y_i \sim G(\alpha_i, \theta_i) \), \( i = 1 \ldots t \) independent, then

\[ \Sigma^I Y_i \] is linearly admissible for \( \Sigma C_i \theta_i \) if and only if there exists a matrix \( G \) of the above form such that \( (I + GD) \Sigma = GC \) where

\[ \Sigma^T = (\Sigma_1, \ldots, \Sigma_t), \quad C^T = (c_1, \ldots, c_t). \]

Theorem 3.1 is often too restrictive. If instead we allow a more general product \( \Sigma Y_{ij} \), we can choose \( q_j \) to achieve suitably defined \( "d" \) all having the same sign. In fact, for a specified set of \( q_j \geq 0 \) such that \( \Sigma q_j = 1 \) Klonecki and Zontek (1987), again for gamma distributions, provide necessary and sufficient conditions for the existence of an estimator of this form which improves upon \( \Sigma^I Y_i \). We state a version of their Lemma 1 which is in a form parallel to Theorem 3.1.

**Theorem 3.3.** If \( Y_i \sim G(\alpha_i, \theta_i) \), \( i = 1 \ldots t \) independent there exists \( b \neq 0 \) such that the estimator

\[ \Sigma^I Y_{ij} + b \Sigma Y_{ij}, \quad (3.2) \]

where \( q_j \geq 0 \), \( \Sigma q_j = 1 \), dominates \( \Sigma^I Y_i \) under SEL, in estimating

\[ \Sigma C_i \theta_i \] if and only if either (i) or (ii) below holds. Define \( d_{i*} = \)
\( c_i = (\alpha_i + q_i) \), \( d^* = \min_{i=1}^{p} d_i^* \) and \( \max_{i=1}^{p} d_i^* \).

(i) \( d^*_i \geq 0, q_i = 0 \) if \( d_i^* = 0 \) and \( 0 < b < b^* \)

(ii) \( d^*_i \leq 0, q_i = 0 \) if \( d_i^* = 0 \) and \( -b^* < b < 0 \)

where \( b^* = \frac{2}{\sum \left( \frac{a_j + q_j}{\frac{a_j}{j} + 2q_j} \right) \frac{d_j}{q_j} \mid \{j : d_j \neq 0\}} \).

Remark 3.1. Theorem 3.3 holds more generally than for the gamma family. Its proof only requires specification of the increasing functions \( w_i(q) = E(Y_i^{1+q} \mid \theta_i = 1) / E(Y_i^{q} \mid \theta_i = 1) \).
(In the gamma case \( w_i(q) = \alpha_i + q \).) Given \( w_i \) we can characterize the sets of \( q_i \)'s which make the corresponding \( d_i^* \)'s all have the same sign, thus enabling domination by (3.2).

Remark 3.2. The bounding of the risk difference in Theorem 3.1 is not as sharp as is possible under the Gamma assumption in Theorem 3.3; hence, the resulting bounds on \( \eta \) in Theorem 3.3 when all \( q_j \) are equal are more liberal than those in Theorem 3.1.

Returning to the estimation of a variance component \( \sigma_r^2 = \sum_{k=0}^{p} \frac{c_k \gamma_k}{\sum_{k=0}^{p} \ell_k Q_k} \) where \( \ell_k = k \).

Consider the estimator \( \sum_{k=0}^{p} \frac{c_k \gamma_k}{f_k + 2\varepsilon_k} \), \( 0 < \varepsilon_k \leq 1 \). The terms \( \ell_k Q_k \) range from the unbiased to the invariant estimator of \( c_k \gamma_k \) as \( \varepsilon_k \) ranges from 0 to 1. Since \( \gamma_k^2 f_k \) is \( G(\frac{1}{2} f_k \gamma_k, 2f_k) \),

\[ d^*_k = \frac{c_k \gamma_k (\varepsilon_k - q_k)}{f_k + 2\varepsilon_k} \] (3.3)

If \( c_k = 0 \) we must set \( q_k = 0 \). Thus, if \( \gamma_k \) does not appear in \( \sigma_r^2 \), using Theorem 3.3, \( Q_k \) does not help in estimating \( \sigma_r^2 \). This clearly differs from the approach in Section 2 where, for example,
in estimating $\sigma^2_e$ all the $Q_k$ can be used to improve upon the best invariant estimator. Note that with $d_k^*$ as defined above, from (3.3), the sign of $d_k^*$ depends only upon $\text{sgn}(c_k(\varepsilon_k - q_k))$; for specified $\varepsilon_k$ and $q_k$ the magnitude of $c_k$ does not play a role with respect to whether an estimator of the form (3.2) can dominate.

From (3.3) if all $\varepsilon_k = 0$ or all $\varepsilon_k = 1$ this approach will provide a dominating estimator if and only if at least two $c_k$ differ from 0 and all nonzero $c_k$ have the same sign. For a variance component some pair of $c_k$ may have opposite signs. Therefore, a dominating estimator will not be obtained if for any such pair both $\varepsilon$'s are 0 or both $\varepsilon$'s are 1. If we can choose $q_k$ to make $d_k^* < 0$ then the dominating estimator in (3.2) will be a "shrinker." Hence, using Theorem 2.3, the positive part of

\[ (3.2) \text{ will dominate } \left\{ \sum_{k=0}^{p} \varepsilon_k Q_k \right\}^+ \]

In this spirit it is natural to ask whether the approach of this section can be combined with that of the previous section. Can we improve upon the estimators developed through Lemma 2.1 and Theorem 2.1 using a more general version of Theorem 3.3 as suggested in Remark 3.1? The answer appears to be no since in the notation of Theorem 2.1 $\gamma_0$ is not a scale parameter for the distribution of $\delta^*_j$. The reader might suggest that $\gamma_0$ could be viewed as a scale parameter for the distribution of $\delta^*_j$ under suitable conditioning. Following the argument leading to Theorem 3.3, while $b$ must be chosen unconditionally, it would have to provide improvement at each conditional level. We can readily show that even in the simplest case, $t = 2$, no $b$ unequal to 0 can achieve this.

As an example, we turn again to the one-way ANOVA using the notation in Section 2. Recalling $\sigma^2 = J^{-1}(\gamma_1 - \gamma_e)$ we consider dominating the estimator
\[ J^{-1} [(I - 1 + 2e_1)^{-1} q_1 - (I(J - 1) + 2e) e^{-1} q_e] \]  

(3.4)

Thus, \( d_1 \geq 0 \) as \( q_1 \leq t_1, d_e \geq 0 \) as \( q_e \leq e \). As noted above, this approach unfortunately does not provide a dominating estimator in the two important cases where \( \varepsilon_1 = \varepsilon_e = 0 \) and where \( \varepsilon_1 = \varepsilon_e = 1 \). Instead, we take

(i) \( \varepsilon_1 = 1, \varepsilon_e = 0 \) for which any \( q_1, q_2, > 0, q_1 + q_2 = 1 \) work with

\[
0 < b < \frac{2}{J(I - 1) + 4q_1} \frac{I(J - 1) + 2q_2}{I(J - 1) + 4q_2} x
\]

\[
\frac{1 - q_1}{q_1(I + 1)} q_1 \left( \frac{q_2}{I(J - 1)} \right) q_2
\]

and (3.2) becomes

\[ J^{-1} [(I + 1)^{-1} q_1 - (I - 1)^{-1} q_e] + b q_1 q_e \]  

(3.5)

(ii) \( \varepsilon_1 = 0, \varepsilon_e = 1 \) for which any \( q_1, q_2 > 0, q_1 + q_2 = 1 \) work with

\[
0 > b > -\frac{2}{J(I - 1) + 4q_1} \frac{I(J - 1) + 2q_2}{I(J - 1) + 4q_2} x
\]

\[
\left( \frac{1}{I - 1} \right) q_1 \left( \frac{1 - q_2}{q_2[I(J - 1) + 1]} \right) q_2
\]

and (3.2) becomes
\[ J^{-1} \left[ (I-1)^{-1} Q_1 - (I(J-1) + 2)^{-1} Q_e \right] + b Q_1 q_1 Q_e q_e \]  \hspace{1cm} (3.6)

We conclude this section with a remark.

Remark 3.3. Results applicable to the simultaneous estimation of variance components under unweighted SEL are given in Klonecki and Zontek (1987). In particular, extensions of Theorems 3.2 and 3.3 are given for the estimation of a vector \( C^T \theta \) using \( L^T Y \). The special case \( C = I \) (not of interest here) has been extensively discussed. See, e.g., Berger (1980), Das Gupta (1986), Dey and Gelfand (1987), and Das Gupta, Dey and Gelfand (1987).

4. SIMULATION RESULTS

In the one-way ANOVA we studied improved estimation of both \( \sigma_e^2 \) and \( \sigma_\alpha^2 \) by undertaking a substantial simulation study over various values of \( I, J, \mu, \sigma_\alpha^2 \) and \( \sigma_e^2 \). Each case received 10,000 replications. Even with so many replications, resimulation of particular cases suggests that the stated percent improvements (PI's) will only be accurate within 2%.

In estimating \( \sigma_e^2 \) some selected cases are presented in Table 1. In this table PI is relative to the best invariant estimator given above (2.2). Not surprisingly (2.2) outperforms (2.3) when \( \mu \) is small, and vice versa when \( \sigma_\alpha^2 \) is small. The estimator (2.4) seems like a good compromise. For fixed \( (\mu, \sigma_\alpha^2, \sigma_e^2) \), PI's increase in \( I \), decrease in \( J \). Although the PI's are small the fact that (2.2)-(2.4) are so simple to calculate encourages their use.

In the estimation of \( \sigma_\alpha^2 \), the reference estimator is the positive part of the unbiased estimator.

\[ \frac{Q_1}{I - 1} - \frac{Q_2}{I(J-1)} + \frac{Q_1 q_1 Q_e q_e}{J} \]  \hspace{1cm} (4.1)
TABLE 1
PERCENT IMPROVEMENTS IN ESTIMATING $\sigma^2_e$

<table>
<thead>
<tr>
<th>(\mu, \sigma_a^2, \sigma_e^2)</th>
<th>PI for (2.2)</th>
<th>(2.3)</th>
<th>(2.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I = 2, J = 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0, 1, 10)</td>
<td>2.60</td>
<td>2.39</td>
<td>2.65</td>
</tr>
<tr>
<td>(1, .1, 10)</td>
<td>1.95</td>
<td>2.27</td>
<td>2.28</td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>1.39</td>
<td>0.78</td>
<td>1.24</td>
</tr>
<tr>
<td>I = 5, J = 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0, 1, 10)</td>
<td>4.09</td>
<td>2.98</td>
<td>3.72</td>
</tr>
<tr>
<td>(1, .1, 10)</td>
<td>3.64</td>
<td>3.36</td>
<td>3.95</td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>0.72</td>
<td>0.15</td>
<td>0.50</td>
</tr>
<tr>
<td>I = 10, J = 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0, 1, 10)</td>
<td>4.97</td>
<td>3.86</td>
<td>4.53</td>
</tr>
<tr>
<td>(1, .1, 10)</td>
<td>6.09</td>
<td>3.92</td>
<td>5.69</td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>0.15</td>
<td>0.03</td>
<td>0.10</td>
</tr>
</tbody>
</table>

As shown in Section 2, (4.1) is dominated by

$$\frac{Q_1}{I + 1} - \frac{Q_e}{I(J - 1)} / J$$

and by

$$\frac{Q_1}{I + 1} - \frac{Q_e}{I(J - 1) + 2} / J$$

(4.3)

Neither of (4.2) and (4.3) dominates the other. However, from Section 2, $\xi_{c,1}^+$ dominates (4.2), $\xi_{a,2}^+$ dominates (4.3). For $b$ sufficiently small (3.5) dominates (4.2) ignoring the positive parts. With positive parts applied to both estimators this is no longer true. Since there is no obvious optimal choice we took $b$
at the middle of the allowable range.

In Table 2 we compare $\delta^+_{a,1}$ and $\delta^+_{a,2}$ with (4.1). We see enormous improvement for both, that the PI's are essentially indistinguishable and that neither of the $\delta$'s dominates the other. We would draw the same conclusions in the comparison of (4.2) and (4.3) with (4.1). Of course, if $\sigma^2_\alpha > \sigma^2_e$ then $\delta^+_{a,1}$ will tend to be nonnegative whence the domination result in Section 2 argues for $\delta^+_{a,2}$. Turning to a comparison of $\delta^+_{a,2}$ with (4.3) we see that if $\mu$ is small the gain may be substantial. A comparison of $\delta^+_{a,1}$ with (4.2) would yield essentially the same magnitudes of improvement. Again, since these estimators are so simple to calculate, their use is encouraged. Finally, the comparison of the positive part of (3.5) with (4.2) is discouraging when $\sigma^2_\alpha$ is smaller than $\sigma^2_e$. Modest improvement will usually occur when $\sigma^2 > \sigma^2_e$. This is reasonable since then the positive part modification is rarely applied and the dominance result comes into play.

We conclude by recommending (2.4) for $\sigma^2_e$ and $\delta^+_{a,2}$ for $\sigma^2_\alpha$.

ACKNOWLEDGMENT

The authors acknowledge Brad Carlin for performing the computations.
<table>
<thead>
<tr>
<th>( (\nu, \sigma^2_\alpha, \sigma^2_\varepsilon) )</th>
<th>( \delta_{\alpha^2_1}^{(4.1)} ) vs ( \delta_{\alpha^2_2}^{(4.1)} )</th>
<th>( \delta_{\alpha^2_1}^{(4.1)} ) vs ( \delta_{\alpha^2_2}^{(4.3)} )</th>
<th>( \delta_{\alpha^2_2}^{(4.3)} ) vs ( (3.5) ) vs ( (4.2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I = 2, J = 5 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0, 1, 1)</td>
<td>92.22</td>
<td>91.96</td>
<td>17.67</td>
</tr>
<tr>
<td>(0, 1, 1)</td>
<td>66.49</td>
<td>66.69</td>
<td>1.10</td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>70.61</td>
<td>71.31</td>
<td>2.10</td>
</tr>
<tr>
<td>(1, .1, 1)</td>
<td>90.80</td>
<td>90.32</td>
<td>1.87</td>
</tr>
<tr>
<td>(1, 1, .1)</td>
<td>66.44</td>
<td>66.64</td>
<td>.90</td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>71.51</td>
<td>72.14</td>
<td>1.74</td>
</tr>
<tr>
<td>( I = 5, J = 5 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0, .1, 1)</td>
<td>68.55</td>
<td>67.72</td>
<td>14.90</td>
</tr>
<tr>
<td>(0, 1, .1)</td>
<td>33.33</td>
<td>33.61</td>
<td>1.57</td>
</tr>
<tr>
<td>(0, 1, 1)</td>
<td>35.24</td>
<td>37.03</td>
<td>0.03</td>
</tr>
<tr>
<td>(1, 1, .1)</td>
<td>32.08</td>
<td>32.33</td>
<td>0.63</td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>33.42</td>
<td>35.08</td>
<td>0.71</td>
</tr>
<tr>
<td>( I = 10, J = 5 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0, .1, 1)</td>
<td>44.49</td>
<td>43.92</td>
<td>8.73</td>
</tr>
<tr>
<td>(0, 1, .1)</td>
<td>19.55</td>
<td>19.73</td>
<td>1.35</td>
</tr>
<tr>
<td>(0, 1, 1)</td>
<td>19.39</td>
<td>20.88</td>
<td>1.33</td>
</tr>
<tr>
<td>(1, .1, 1)</td>
<td>40.47</td>
<td>39.21</td>
<td>0.00</td>
</tr>
<tr>
<td>(1, 1, .1)</td>
<td>18.81</td>
<td>18.91</td>
<td>0.10</td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>18.39</td>
<td>19.68</td>
<td>0.20</td>
</tr>
</tbody>
</table>
REFERENCES


**Title:** Improved Estimation of Variance Components in Mixed Models

**Authors:** Alan E. Gelfand and Dipak K. Dey

**Performing Organization:**
Department of Statistics  
Stanford University  
Stanford, CA 94305

**Controlling Office:**
Office of Naval Research  
Statistics & Probability Program Code 1111

**Report Date:** August 30, 1990

**Number of Pages:** 22

**DISTRIBUTION STATEMENT:** Approved for public release: distribution unlimited

**KEY WORDS:** variance components; ANOVA decomposition; balanced designs; squared error loss.

**ABSTRACT:**
Taking Albert's (1976) formulation of a mixed model ANOVA, we consider improved estimation of the variance components for balanced designs under squared error loss. Two approaches are presented. One extends the ideas of Stein (1964). The other is developed from the fact that variance components can be expressed as linear combinations of chi-square scale parameters. Encouraging simulation results are presented.