QUADRATIC UNBIASED ESTIMATION OF VARIANCE COMPONENTS IN LINEAR MODELS WITH AN EMPHASIS ON THE ONE-WAY CLASSIFICATION

by

James Howard Goodnight

Institute of Statistics
Mimeo Series No. 850
November 1972

TABLE OF CONTENTS

1. INTRODUCTION

The classical approach to the unbiased estimation of variance components for unbalanced data is one of choosing several different quadratic functions of the data, equating them to their expected values, and solving the resultant system of equations.

There are, of course, infinitely many quadratic functions available for equating observed to expected values, and solving to provide unbiased estimation. However, much of the previous work in this area centers around quadratic forms which bear analogy to those used with balanced data; in particular, much work has been done using the methods outlined by Henderson [1953]. Rao [1972] states that the classical methods lack a clear theoretical basis and that the classical procedures are: "*ad hoc* and much seems to depend on intuition". Numerous authors[1] have, in fact, compared two or more of the classical estimators to determine which, if any, has the smaller variance for a particular design. However, the decision as to which estimator among a given set is "best", always seems to depend upon the unknown values of the components being estimated. Hence, the idea of achieving a uniformly "best" unbiased variance component estimator in the "generally" unbalanced situation[2] appears to be improbable. In

---

[1] Review articles by Searle [1971] and Harville [1969b] describe much of the earlier work in variance component estimation and give some of the important references.

[2] Situations such as those that arise from the loss of data, or lack of data, etc. are implied here. Conceivably, specific designs with planned imbalanced may be found which yield uniformly "best" estimators. Rao [1971b] provides some necessary conditions for this to occur.

fact, Read [1961] has proven that there exists no quadratic estimator of the "between" component in the unbalanced one-way classification for which the variance (assuming normality) is uniformly smaller than that of every other quadratic estimator.

Recognizing, perhaps, the fundamental difficulties which arise from the classical approaches to variance component estimation, authors of recent papers have focused their attention on choosing quadratic forms with some sorts of optimal properties. Harville [1969a], for example, has used the results of Hultquist and Graybill [1965] on minimal sufficient statistics in conjunction with Koch's [1967] lemma on the variance of quadratic forms to establish the basic form that the matrix of a quadratic form should have when it is to be used for estimating the components in a one-way classification. Townsend [1968] gives locally best quadratic unbiased estimators for the variance components associated with the one-way classification with zero mean. Recently published papers by Rao [1970, 1971a, 1971b, 1972] outline new techniques referred to as MINQUE (Minimum Norm Quadratic Unbiased Estimation) and MIVQUE (Minimum Variance Quadratic Unbiased Estimation). Although Rao develops MINQUE and MIVQUE with fairly relaxed assumptions on the distributional properties of the random effects involved, he does consider the special case where the random effects are normally distributed. In this case MINQUE and MIVQUE estimators are the same and provide locally "best" quadratic unbiased estimators.

The present paper is restricted to the quadratic unbiased estimation of the variance components in linear models for which the random effects are taken to be normally distributed. The basic objectives of this paper are to extend MIVQUE theory to provide estimators whose variance is functionally dependent on as few of the unknown parameters as possible, to provide computational techniques for MIVQUE and its extensions, and to apply MIVQUE techniques to the unbalanced one-way classification in order to develop an estimator for which no *a priori* knowledge about the variance components is necessary, yet the efficiency of which is in some sense optimal.

2. STATEMENT OF THE PROBLEM

2.1 The Mathematical Model

Let the $N \times 1$ vector of random variables Y have the linear structure

$$Y = X_0\beta_0 + X_1\beta_1 + \ldots + X_m\beta_m , \tag{2.1}$$

where X_i $(i = 0, \ldots, m)$ is an $N \times n_i$ matrix of given values (with $X_m = I_N$) , β_0 is an $n_0 \times 1$ vector of unknown non-stochastic parameters, and each β_i $(i = 1, \ldots, m)$ is an $n_i \times 1$ vector of uncorrelated random variables assumed to be normally distributed with mean zero and variance $\sigma_i^2 I_{n_i}$. Furthermore, each β_i and β_j $(i \neq j)$ are assumed to be uncorrelated. From the above it follows that

$$Y \sim \text{Normal}\ (X_0\beta_0, V) ,$$

where

$$V = X_1X_1'\sigma_1^2 + X_2X_2'\sigma_2^2 + \ldots + X_mX_m'\sigma_m^2 . \tag{2.2}$$

2.2 Unbiasedness

The problem of quadratic unbiased estimation of the variance components $\sigma_1^2, \ldots, \sigma_m^2$ is one of choosing quadratic forms $Y'Q_iY$ such that the $E(Y'Q_iY) = \sigma_i^2$ $(i = 1, \ldots, m)$. The expectation of any quadratic form in Y , as defined by (2.1), is

$$E(Y'QY) = \beta_0'X_0'QX_0\beta_0 + \sum_{i=1}^{m} \text{tr}\ X_i'QX_i\sigma_i^2 .$$

Thus a quadratic unbiased estimator of σ_i^2 $(i = 1, \ldots, m)$ requires a matrix Q_i such that $X_0'Q_iX_0 = 0$ and

$$\operatorname{tr} X_j'Q_iX_j = \left\{\begin{matrix} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{matrix}\right\} \qquad \text{for} \quad j = 1, \ldots, m .$$

2.3 Invariance Concepts

By imposing additional restrictions on Q_i $(i = 1, \ldots, m)$ other desirable properties of the estimator $Y'Q_iY$ may be obtained. The concept of "invariance on the translation of the β_0 parameter" is considered by Rao, Harville, and others. A quadratic form $Y'QY$ is said to be "invariant on the translation of the β_0 parameter" if

$$Y'QY = (Y - X_0\beta_0^*)'Q(Y - X_0\beta_0^*) \tag{2.3}$$

for any choice of values for the $n_0 \times 1$ vector β_0^*. A sufficient condition for invariance with respect to β_0 is $X_0'Q = 0$. This condition is also necessary[1] unless $\beta_0^* = 0$. The appeal for this type of β_0-invariance can be seen in that recoding of the data by subtracting $X_0\beta_0^*$ does not alter the estimate for a particular effect. In addition, quadratic forms in Y, as defined by (2.1), do not contain any elements of $\beta_0\beta_0'$ in their variance if they are β_0-invariant.

The variance of a symmetric quadratic form $Y'QY$ is

$$\operatorname{Var}(Y'QY) = 2 \sum_{ij} \operatorname{tr} QVQV + 4\beta_0'X_0'QVQX_0\beta_0 . \tag{2.4}$$

---

[1] Expansion of the right hand side of (2.3) with the definition of Y from (2.1) involves the term $X_0'QX_m$ or $X_0'QI$ which must be zero for (2.3) to hold for any non-trivial β_0^*.

Applying the definition of V given in (2.2), the above variance may be expressed as

$$\mathrm{Var}(Y'QY) = 2 \sum_{ij} \mathrm{tr}\, X_i'QX_jX_j'QX_i\sigma_i^2\sigma_j^2 + 4 \sum_i \beta_0'X_0'QX_iX_i'QX_0\beta_0\sigma_i^2 \text{ .}$$

Furthermore, by defining, for any real matrix A , the matrix operator $\mathrm{ssq}(A)$ to be equal to the sum of the squared elements of A , we have at once that $\mathrm{tr}\,(AA') = \mathrm{ssq}(A)$. Thus (2.4) may be expressed more succinctly as

$$\mathrm{Var}(Y'QY) = 2 \sum_{ij} \mathrm{ssq}(X_i'QX_j)\sigma_i^2\sigma_j^2 + 4 \sum_i \mathrm{ssq}(\beta_0'X_0'QX_i)\sigma_i^2 \text{ .} \qquad (2.5)$$

Clearly, if $Y'QY$ is β_0-invariant, which implies $X_0'Q = 0$, (2.5) contains no elements of $\beta_0\beta_0'$. Thus the variance of any β_0-invariant quadratic form in Y is

$$\mathrm{Var}(Y'QY) = 2 \sum_{ij} \mathrm{ssq}(X_i'QX_j)\sigma_i^2\sigma_j^2 \text{ .} \qquad (2.6)$$

Harville also considers a more general type of invariance in regard to the quadratic forms arising from variance component estimation in the one-way classification. What he terms "α-invariance" (which implies that the variance of the estimator for the within component contains no terms involving the between component) is expanded here to include invariance with respect to any random or non-random effect when estimating the components associated with (2.1).

Definition 2.1: For any Y defined by (2.1), the quadratic form $Y'QY$ is said to be β_i-invariant (for a fixed value of i between 0 and m) provided the variance of $Y'QY$ does not contain any β_0 terms if $i = 0$ and does not contain any terms involving σ_i^2 and $i \neq 0$.

Theorem 2.1: A necessary and sufficient condition for the quadratic form $Y'QY$ to be β_i-invariant in any non-trivial situation is that $X_i'Q = 0$.

Proof: Sufficiency is immediately seen upon examination of (2.5). Recalling that $X_m = I_N$ establishes necessity also on inspection.

The following lemmas although obvious are presented to show to what extent additional invariance restrictions may be placed on quadratic forms and still maintain unbiasedness when estimating σ_i^2 $(i = 1, \ldots, m)$.

Lemma 2.1: If a quadratic form $Y'QY$ is β_i-invariant, then it is β_j-invariant if $X_j = X_i L$ for some $n_i \times n_j$ matrix L.

Lemma 2.2: No quadratic unbiased β_i-invariant estimator exists for σ_i^2 $(i = 1, \ldots, m)$.

Lemma 2.3: No quadratic unbiased β_i-invariant estimator exists for σ_j^2 if $X_j = X_i L$ for any $n_i \times n_j$ matrix L.

Definition 2.2: The quadratic form $Y'QY$ is said to be a maximally-invariant quadratic unbiased estimator of σ_i^2 $(i = 1, \ldots, m)$ provided that it is β_j-invariant for as many j's as possible among the set $j = 0, 1, \ldots, m$ and still maintains its unbiasedness property.

By using a maximally-invariant quadratic unbiased estimator, one insures that the "goodness" of the estimator depends on the value of the parameter being estimated and on as few of the other parameter values as possible. This property of maximally-invariant estimators is particularly applicable when little or no a priori knowledge is available concerning the relative magnitudes of the variance components being estimated.

2.4 Locally Best Quadratic Unbiased Estimators

When the normality of Y is assumed, Rao's [1971a] MINQUE and Rao's [1971b] MIVQUE estimators coincide and provide locally best β_0-invariant QUE's[1] for $\sum_{i=1}^{m} K_i\sigma_i^2$, where the K_i's are known constants. In the context of (2.1) Rao [1972] proves that for any positive definite matrix W, the minimum of $\operatorname{tr} QWQW$ subject to the conditions that $X_0'Q = 0$ and $\operatorname{tr} X_i'QX_i = K_i$ $(i = 1, \ldots, m)$ is obtained when

$$Q = \sum_{i=1}^{m} \delta_i \, RX_iX_i'R \, , \tag{2.7}$$

where

$$R = W^{-1} - W^{-1}X_0(X_0'W^{-1}X_0)^{-}X_0'W^{-1} \tag{2.8}$$

and $\delta' = (\delta_1, \delta_2, \ldots, \delta_m)$ is determined from the equation $S\delta = K$ where $K' = (K_1, K_2, \ldots, K_m)$, and the (i,j)th element of S is $\operatorname{ssq}(X_i'RX_j)$.

A quadratic form, $Y'QY$, where Q is defined by (2.7), satisfies

$$E(Y'QY) = \sum_{i=1}^{m} K_i\sigma_i^2 \, ,$$

---

[1] QUE is henceforth used to denote quadratic unbiased estimator.

$$\mathrm{Var}(Y'QY) = 2 \sum_{ij} \mathrm{ssq}(X_i'QX_j)\sigma_i^2\sigma_j^2 = 2 \ \mathrm{tr}\ QVQV ,$$

and tr QWQW is a minimum subject to the unbiasedness and invariance restrictions.

When $W = V$ the minimum variance β_0-invariant QUE of $\Sigma K_i\sigma_i^2$ is obtained. When reliable a priori estimates $(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_m^2)$ are available for the components being estimated, then the W matrix of (2.8) may be computed as

$$W = \sum_{i=1}^{m} X_iX_i'\hat{\sigma}_i^2 .$$

Furthermore, since the restrictive minimization of tr QVQV is equivalent to the restrictive minimization of $\sum_{ij} \rho_i\rho_j \ \mathrm{ssq}(X_i'QX_j)$ where $\rho_i = \sigma_i^2/\sigma_k^2$ $(i = 1, \ldots, m)$ for any $k = 1, 2, \ldots, m$, the following W matrix may be used in (2.8) to obtain a minimum variance β_0-invariant QUE of $\Sigma K_i\sigma_i^2$:

$$W = \sum_{i=1}^{m} X_iX_i'\rho_i .$$

Thus the minimum variance β_0-invariant QUE of $\Sigma K_i\sigma_i^2$ may be realized if the ratios of all variance components to a common variance component are known.

Rao [1971b] presents necessary conditions for obtaining a minimum variance β_0-invariant QUE irrespective of the values of the variance components being estimated. However, due to the restrictive conditions required of the X_i $(i = 1, \ldots, m)$ matrices, it would seem that few of the generally unbalanced designs met in practice could qualify. Further inspection of these conditions may lead to specially constructed unbalanced designs (perhaps nested) for which minimum variance β_0-invariant QUE's are possible.

Rao [1972] suggests that when no a priori knowledge is available for the components being estimated that the W matrix of (2.8) be computed as

$$W = \sum_{i=1}^{m} X_i X_i' \ . \tag{2.9}$$

Using this W matrix corresponds to assigning equal a priori weights to the unknown variance components, and results in the minimization of

$$\sum_{ij} \operatorname{ssq}(X_i' Q X_j) \ .$$

Due to the lack of additional invariance constraints, the variance of any β_0-invariant QUE in the unbalanced case will, in general, have all terms in the variance expression (2.6) greater than zero. Hence, the "goodness" of any estimator will depend on the actual values of all of the components being estimated. In the event that some components are large relative to others, use of (2.9) as Rao suggests could lead to estimators with possibly undesirable variances. Perhaps in situations where a priori knowledge is not available, maximally-invariant QUE's would be less risky. Rao's β_0-invariant QUE's are easily extended to include β_i-invariant QUE's $(i = 1, \ldots, m)$ and thus maximally-invariant QUE's .

2.5 Locally Best Quadratic Unbiased Estimators With Additional Invariance Restrictions

Locally best β_0-invariant QUE's which are in addition β_i-invariant QUE's for one or more β_i $(i = 1, \ldots, m)$ are considered here. Let β_α denote that collection of β_i's $(i = 0, \ldots, m)$ for which β_i-invariance is sought. In the context of this section β_α

is assumed to contain at least β_0. Obviously, based on Lemma 2.3, β_α can not contain β_m. Also denote by β_γ that collection of β_i's $(i = 1, \ldots, m)$ not contained in β_α. The term β_α-invariant is used to denote β_i-invariance with respect to all β_i's contained in β_α.

Furthermore, let α_n and γ_n represent the number of elements in β_α and β_γ respectively, $(\alpha_n + \gamma_n = m+1)$. Denote by $X_{\alpha_1}, \ldots, X_{\alpha_n}$ the incidence matrices associated with the elements in β_α, and denote by $X_{\gamma_1}, \ldots, X_{\gamma_n}$ the incidence matrices associated with the elements in β_γ. Let $\hat{\sigma}^2_{\gamma_1}, \ldots, \hat{\sigma}^2_{\gamma_n}$ be a <u>priori</u> estimates of the variance components $\sigma^2_{\gamma_1}, \ldots, \sigma^2_{\gamma_n}$ respectively, associated with the elements of β_γ.

The problem, then, is one of finding β_α-invariant QUE's for

$$\sum_{i=\gamma_1}^{\gamma_n} K_i\sigma_i^2 , \tag{2.10}$$

(where $K_{\gamma_1}, \ldots, K_{\gamma_n}$ are a set of predetermined values) for which

$$\sum_{\gamma_i\gamma_j} \operatorname{ssq}(X'_{\gamma_i} QX_{\gamma_j})\hat{\sigma}^2_{\gamma_i}\hat{\sigma}^2_{\gamma_j} \tag{2.11}$$

is a minimum.

In light of Lemma 2.3, (2.10) is not estimable if β_α-invariance implies invariance with respect to any of the elements of β_γ for which the corresponding K value in (2.10) is non-zero.

By constructing the matrix

$$W_\gamma = \sum_{i=\gamma_1}^{\gamma_n} X_iX_i'\hat{\sigma}_i^2 , \tag{2.12}$$

(2.11) may be rewritten as

$$\mathrm{tr}\ QW_\gamma QW_\gamma$$

and by letting

$$X_\alpha = [X_{\alpha_1} | X_{\alpha_2} | \cdots | X_{\alpha_n}]$$

the solution to β_α-invariant QUE's is apparent by applying Rao's basic MINQUE theorem from the previous section. To be specific, the β_α-invariant QUE for (2.10), if it exists, for which (2.11) is minimized is $Y'QY$ where Q is determined as follows:

$$Q = \sum_{i=\gamma_1}^{\gamma_n} \delta_i R X_i X_i' R \ , \tag{2.13}$$

where

$$R = W_\gamma^{-1} - W_\gamma^{-1} X_\alpha (X_\alpha' W_\gamma^{-1} X_\alpha)^- X_\alpha' W_\gamma^{-1}$$

and $\delta' = (\delta_{\gamma_1}, \ldots, \delta_{\gamma_n})$ is determined from the equation $S\delta = K$ where $K' = (K_{\gamma_1}, \ldots, K_{\gamma_n})$, and the (i,j)th element of S is $\mathrm{ssq}(X_{\gamma_i}' R X_{\gamma_j})$.

The above results imply that if β_i-invariant QUE's for one or more values of $i = 1, \ldots, m$ are sought in addition to β_0-invariant QUE's, then one has only to include those β_i's (for which β_i-invariance is sought) with the set β_0 of fixed effects and proceed with Rao's MINQUE estimation procedure as if all the random effects included with β_0 were fixed. The following theorem is an interesting result of the above.

Theorem 2.2: The unique minimum variance maximally-invariant QUE for σ_m^2 in model (2.1) if rank $(X) < N$ is

$$\hat{\sigma}_m^2 = Y'(I - X(X'X)^{-}X')Y/[N - \text{rank}(X)] ,$$

where $X = [X_0|X_1|\ldots|X_{m-1}]$.

Proof: The variance of a maximally-invariant QUE , $Y'QY$, for σ_m^2 is

$$2 \text{ ssq}(Q)\sigma_m^4 = 2 \text{ tr } QIQI\sigma_m^4 ,$$

which is minimized whenever tr QIQI is a minimum, regardless of the value of σ_m^4 . Thus, to find a maximally-invariant QUE for σ_m^2 , it is necessary to find Q such that:

$$\text{tr } Q = 1, \quad X'Q = 0, \quad \text{and} \quad \text{tr } QIQI \text{ is a minimum} .$$

From (2.13), $Q = \delta_1 RR$, where $R = I - X(X'X)^{-}X'$ and δ_1 is determined from the equation

$$\text{ssq}(R)\delta_1 = 1 .$$

Since the idempotent matrix R is unique, and ssq(R) = tr (R) = N - rank(X) the theorem is proven.

2.6 A Suggestion For Obtaining Invariant Quadratic Unbiased Estimators

β_0-invariant QUE's which are also β_i-invariant for some fixed value of i between 1 and m have the property that

$$\mathrm{ssq}(X_i'QX_j) = 0 \text{ for all } j = 0, 1, \ldots, m , \tag{2.14}$$

since β_i-invariance implies $X_i'Q = 0$. If (2.7) is computed with a matrix

$$W = \sum_{i=1}^{m} X_i X_i' \hat{\sigma}_i^2 , \tag{2.15}$$

where $\hat{\sigma}_i^2$ is an a priori estimate of σ_i^2 for $i = 1, \ldots, m$, then

$$\sum_{ij} \mathrm{ssq}(X_i'QX_j)\hat{\sigma}_i^2\hat{\sigma}_j^2 \tag{2.16}$$

is minimized subject to the necessary constraints. It would seem that "practical" β_i-invariance could be achieved by replacing the corresponding $\hat{\sigma}_i^2$ in (2.15) by a number whose magnitude is sufficiently larger than that of any of the $\hat{\sigma}_j^2$'s . Hence when (2.16) is minimized, the coefficients, $\mathrm{ssq}(X_i'QX_j)$, associated with the large values of $\hat{\sigma}_i^2$ will be forced to be relatively small.

3. COMPUTING LOCALLY BEST QUADRATIC UNBIASED ESTIMATORS

3.1 Forming The Matrices Needed

Although equation (2.7) provides the theoretical basis for β_0-invariant QUE's in models described by (2.1), a more computationally oriented procedure is given here.

Suppose that *a priori* estimates of $\rho_i = \sigma_i^2/\sigma_m^2$ $(i = 1, \ldots, m)$, denoted by $r_1, \ldots, r_m$ are available. Then upon defining

$$W = \sum_{i=1}^{m} r_i X_i X_i' \qquad (3.1)$$

and by applying equation (2.7) to this W for some set of K_i's, a matrix Q will be generated such that:

$$E(Y'QY) = \sum_i K_i \sigma_i^2 ,$$

$$\mathrm{Var}(Y'QY) = 2 \sum_{ij} \mathrm{ssq}(X_i' Q X_j)\sigma_i^2 \sigma_j^2 ,$$

and

$$\sum_{ij} \mathrm{ssq}(X_i' Q X_j) r_i r_j$$

will be minimized subject to the conditions $X_0'Q = 0$ and $\mathrm{tr}\, X_i' Q X_i = K_i$ $(i = 1, \ldots, m)$.

Thus, values of r_i "close" to ρ_i will yield an estimator whose variance is "close" to that of the minimum variance β_0-invariant QUE of $\sum_i K_i \sigma_i^2$.

Using a suitably chosen positive definite matrix W as is given in (3.1), define the following matrices:

$$A = \left[\begin{array}{c|c|c} X_0'W^{-1}X_0 & X_0'W^{-1}X & X_0'W^{-1}Y \\ \hline X'W^{-1}X_0 & X'W^{-1}X & X'W^{-1}Y \end{array}\right], \tag{3.2}$$

where

$$X = [X_1 | X_2 | \cdots | X_m] \; ;$$

$$B = \left[\begin{array}{c|c} (X_0'W^{-1}X_0)^{-} & 0 \\ \hline -(X'W^{-1}X_0)(X_0'W^{-1}X_0)^{-} & I \end{array}\right] \tag{3.3}$$

and the matrix product:

$$BA = \left[\begin{array}{c|c|c} (X_0'W^{-1}X_0)^{-}X_0'W^{-1}X_0 & (X_0'W^{-1}X_0)^{-}X_0'W^{-1}X & (X_0'W^{-1}X_0)^{-}X_0'W^{-1}Y \\ \hline X'RX_0 & X'RX & X'RY \end{array}\right], \tag{3.4}$$

where $R = W^{-1} - W^{-1}X_0(X_0'W^{-1}X_0)^{-}X_0'W^{-1}$.

Since the submatrix $X'RX$ of (3.4) can be partitioned as:

$$X'RX = \left[\begin{array}{c|c|c|c} X_1'RX_1 & X_1'RX_2 & \cdots & X_1'RX_m \\ \hline X_2'RX_1 & X_2'RX_2 & \cdots & X_2'RX_m \\ \vdots & \vdots & & \vdots \\ \hline X_m'RX_1 & X_m'RX_2 & \cdots & X_m'RX_m \end{array}\right],$$

the quantities $S_{ij} = \mathrm{ssq}(X_i'RX_j)$, $(i, j = 1, \ldots, m)$, needed in equation (2.7), can be computed directly from (3.4).

Also note that the Q matrix defined by (2.7) is:

$$Q = \sum_{i=1}^{m} \delta_i RX_iX_i'R ,$$

hence

$$Y'QY = \sum_{i=1}^{m} \delta_i Y'RX_iX_i'RY ,$$

$$= \sum_{i=1}^{m} \delta_i \ \mathrm{ssq}(X_i'RY) . \tag{3.5}$$

Thus the elements $\mathrm{ssq}(X_i'RY)$ are also available from (3.4) since the submatrix $X'RY$ may be partitioned as:

$$X'RY = \left[\begin{array}{c} X_1'RY \\ \hline X_2'RY \\ \hline \vdots \\ \hline X_m'RY \end{array}\right] .$$

The δ_i's of (3.5) are computed from the equation $S\delta = K$. Provided that S is non-singular $\delta = S^{-1}K$ and when estimating σ_i^2 $(i = 1, \ldots, m)$ the elements of K will be:

$$K_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} .$$

Thus, since S and S^{-1} (if it exists) are symmetric, the δ vector associated with the estimation of σ_i^2 will be the $j^{\underline{th}}$ row of S^{-1}. Hence by defining $\hat{\Sigma}' = [\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_m^2]$ and

$$T' = [\mathrm{ssq}(X_1'RY),\ \mathrm{ssq}(X_2'RY),\ \dots,\ \mathrm{ssq}(X_m'RY)] ,$$

$$\hat{\Sigma} = S^{-1}\,T . \tag{3.6}$$

However, (3.6) may also be derived by equating the T values to their expected values since:

$$E(T_i) = \sum_{j=1}^{m} \mathrm{tr}\,(X_j'RX_iX_i'RX_j)\sigma_j^2 ,$$

$$= \sum_{j=1}^{m} \mathrm{ssq}(X_i'RX_j)\sigma_j^2 .$$

Hence

$$E(T) = S\,\Sigma$$

where

$$\Sigma' = [\sigma_1^2,\ \sigma_2^2,\ \dots,\ \sigma_m^2] .$$

Since the solution to the system of equations

$$T = S\,\hat{\Sigma}$$

is invariant under linear row operations, provided S is of full rank, any set of linear combinations of the T_i's , say $L'T$, can be equated to its expected value and $\hat{\Sigma}$ be solved for, provided $L'S$ is of full rank. Thus by forming the quadratic forms:

$$Y'Q_iY,\ i = 1,\ 2,\ \dots,\ m ,$$

where

$$Q_i = RX_iX_i'R \ ,$$

then equating each quadratic form to its expected value and solving, β_0-invariant QUE's for each σ_i^2 are obtained which are minimum variance β_0-invariant QUE's provided $r_i = \rho_i$ $(i = 1, \ldots, m)$.

3.2 Computing The A Matrix

Although the matrix A of (3.2) can be computed directly, an alternate method will be presented which in general will require less computer storage.

Since the matrix W is assumed to be positive definite, there exists a non-singular matrix P such that $W = PP'$. One such P would be $P = C\Lambda^{\frac{1}{2}}$ where the C matrix represents the characteristic vectors of W stored columnwise, and $\Lambda^{\frac{1}{2}}$ represents a diagonal matrix, whose diagonal elements are the square roots of the characteristic roots of W . Unfortunately computer routines require both the upper triangular portion of the W matrix and the C matrix which is N × N to be resident in core.

An alternate method for deriving P which requires only the upper triangular portion of W to be in core can be used. Since the Forward Doolittle method as described by Rohde [1964] factors a matrix, say W , into two triangular matrices, the A_d matrix and B_d matrix, such that

$$W = B_dA_d \ ,$$

where A_d is an upper triangular matrix and $B_d = (DA_d)'$, where D is a diagonal matrix with diagonal elements $d_{ii} = \frac{1}{A_{ii}}$. Hence

$$W = (DA_d)'A_d \ ,$$

$$= A_d'DA_d \ ,$$

$$= A_d'D^{\frac{1}{2}}D^{\frac{1}{2}}A_d \ ,$$

and by letting

$$P = A_d'D^{\frac{1}{2}} \ , \tag{3.7}$$

the W matrix may be expressed as

$$W = PP' \ .$$

(A computational algorithm for forming P' is given in Appendix 7.1). Hence

$$W^{-1} = (P')^{-1}(P)^{-1} \ .$$

Since the P' described above is an upper triangular matrix, it may be inverted in place using a slightly modified version of the "sweep procedure" outlined by Schatzoff _et_ _al_. [1968] as described in Appendix 7.2.

Thus the A matrix may be computed efficiently via the following:

STEP 1: Form the upper triangle portion of the W matrix.

STEP 2: Use row operations to convert it in place to the A_d matrix of the Forward Doolittle.

STEP 3: Divide each row of the resultant matrix by the square root of the diagonal element of that row.

STEP 4: Invert the resultant matrix in place using the modified sweep procedure outlined in Appendix 7.2 and name the inverse matrix L' (hence $W^{-1} = L'L$).

STEP 5: Compute $Z_0 = LX_0$, $Z = LX$, and $G = LY$.

STEP 6: Compute the A matrix of (3.2) using the matrices of STEP 5 above as follows:

$$A = \left[\begin{array}{c|c|c} Z_0'Z_0 & Z_0'Z & Z_0'G \\ \hline Z'Z_0 & Z'Z & Z'G \end{array}\right], \tag{3.8}$$

however, only those elements of A for which $i \leq j$ need be stored.

3.3 Computing The BA Matrix

Once the A matrix has been computed using (3.2) or (3.8), by applying Rohde's modification of the Doolittle procedure (to produce a generalized inversion routine) to the sweep routine outlined by Schatzoff (see Appendix 7.3), the A matrix may be swept on the first n_0 columns to produce the BA matrix of (3.4) in place..

Once this is accomplished, all the elements needed to estimate all variance components are at hand and equation (3.6) may be employed.

3.4 Computing Other Invariant Estimators

From the results of section 2.5, locally best β_0-invariant QUE's which are in addition locally best β_i-invariant QUE's for one or more values of $i = 1, \ldots, m$, may be achieved by augmenting the X_0 matrix with the X_i matrices associated with those β_i's for which invariance is sought, and by redefining the estimated variance-covariance matrix as in (2.12). Hence, the preceding techniques may be used to compute estimators which are locally best β_i-invariant QUE's in addition to being locally best β_0-invariant QUE's.

Once locally best β_0-invariant QUE's for σ_i^2 $(i = 1, \ldots, m)$ are computed using (3.6), the BA matrix of (3.3) could be swept on the columns associated with one of the β_i's as if they were columns associated with β_0. The remaining submatrices in $X'RX$ and $X'RY$ could then be used to compute new S and T matrices, and β_i-invariant QUE's for the remaining variance components could be achieved.

4. QUADRATIC UNBIASED ESTIMATORS FOR THE ONE-WAY CLASSIFICATION

4.1 Foreword

The preceding sections outline the theory and the computational procedures for obtaining locally best β_0-invariant QUE's for the variance components associated with model (2.1). Section 2.5 extends the concept of β_0-invariance to include β_i-invariance for one or more values of $i = 1, \ldots, m$, which in turn leads to the concept of maximally-invariant estimators. Although maximally-invariant QUE's heavily restrict the form of Q, uniqueness can not be claimed in general (except for σ_m^2). Thus, in the "generally" unbalanced case *a priori* estimates of the variance components (for which invariance is not sought) must be supplied to completely specify the estimator for a particular component. When *a priori* estimates are not available, are there other relatively "safe" prior estimates, based on the design characteristics, which could be used to prevent the estimator from having a variance much larger than that of the "best" estimator were the true variance components known? In seeking a partial answer to this question, we take the approach suggested by Harville and work with a simplified model, that of the unbalanced one-way classification. As stated by Harville, "virtually every problem that arises in constructing a complete body of theory for the estimation of the variance components associated with a complex possibly-unbalanced random or mixed model is also encountered, though in simpler form, in carrying out that process for the possibly-unbalanced one-way random model". By extending our knowledge of this model, "we effectively place an 'upper-bound' on what we can hope to achieve in the way of theoretical results for more complicated models".

4.2 The Mathematical Model

The one-way classification model is represented here as:

$$Y = X_0\mu + X_1\alpha + I\epsilon , \qquad (4.1)$$

where

$Y' = (y_{11}, \ldots, y_{1n_1}, \ldots, y_{a1}, \ldots, y_{an_a})$,

$N = n_1 + \ldots + n_a$,

X_0 is an $N \times 1$ vector of 1's ,

μ is a fixed but unknown constant,

X_1 is an $N \times a$ matrix whose $j^{\underline{th}}$ column is zero except for 1's in elements R_1 thru R_2 where $R_2 = n_1 + \ldots + n_j$ and $R_1 = R_2 - n_j + 1$,

$\alpha' = (\alpha_1, \ldots, \alpha_a)$ where α is assumed to be distributed Normal $(0, I_a\sigma^2_\alpha)$,

$\epsilon' = (\epsilon_{11}, \ldots, \epsilon_{1n_1}, \ldots, \epsilon_{a1}, \ldots, \epsilon_{an_a})$ where ϵ is assumed to be distributed Normal $(0, I_N\sigma^2_e)$, and α_i and ϵ_{ij} for all i and j are assumed to be uncorrelated.

From the above,

$$Y \sim \text{Normal}\ (X_0\mu,\ V)\ ,$$

where

$$V = X_1X_1'\sigma_a^2 + I\sigma_e^2\ .$$

4.3 Invariant Quadratic Unbiased Estimators For The Variance Components

Although Harville provides locally best μ-invariant QUE's for σ_a^2 and σ_e^2 through use of Lagrangian functions, a simplier derivation is given here by applying equation (2.7).

The variance of any μ-invariant QUE , $Y'QY$, for $K_1\sigma_a^2 + K_2\sigma_e^2$ (K_1 and K_2 are predetermined constants) is:

$$\begin{aligned}\text{Var}(Y'QY) &= 2\ \text{ssq}(X_1'QX_1)\sigma_a^4 + 4\ \text{ssq}(X_1'Q)\sigma_a^2\sigma_e^2 + 2\ \text{ssq}(Q)\sigma_e^4\ ,\\ &= 2\sigma_e^4[\text{ssq}(X_1'QX_1)\rho^2 + 2\ \text{ssq}(X_1'Q)\rho + \text{ssq}(Q)]\ ,\end{aligned}$$

where

$$\rho = \sigma_a^2/\sigma_e^2\ .$$

Hence, infinitely many μ-invariant QUE's for $K_1\sigma_a^2 + K_2\sigma_e^2$ can be generated by choosing <u>a priori</u> estimates, r's, of ρ and determining the matrix Q_r such that: $X_0'Q_r = 0$, $\text{tr}\ X_1'Q_rX_1 = K_1$, $\text{tr}\ Q_r = K_2$, and minimizing

$$\text{ssq}(X_1'Q_rX_1)r^2 + 2\ \text{ssq}(X_1'Q_r)r + \text{ssq}(Q_r)\ . \qquad (4.2)$$

When $r = \rho$, the minimum variance μ-invariant QUE for $K_1\sigma_a^2 + K_2\sigma_e^2$ is obtained.

Since (4.2) is equivalent to $\operatorname{tr} Q_rW_rQ_rW_r$ where

$$W_r = I + rX_1X_1' ,$$

we have from (2.7) that the μ-invariant QUE of $K_1\sigma_a^2 + K_2\sigma_e^2$ which minimized (4.2) is $Y'Q_rY$, where:

$$Q_r = \delta_1RX_1X_1'R + \delta_2RR , \tag{4.3}$$

$$R = W_r^{-1} - W_r^{-1}X_0(X_0'W_r^{-1}X_0)^{-1}X_0'W_r^{-1} ,$$

and δ_1 and δ_2 are determined from the equation:

$$\begin{bmatrix} \operatorname{ssq}(X_1'RX_1) & \operatorname{ssq}(X_1'R) \\ \operatorname{ssq}(X_1'R) & \operatorname{ssq}(R) \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} . \tag{4.4}$$

Since W^{-1} can be expressed in closed form, (4.3) simplifies to:

$$Q_r = \begin{bmatrix} C_{11}J_{11} & C_{12}J_{12} & \cdots & C_{1a}J_{1a} \\ C_{12}J_{21} & C_{22}J_{22} & \cdots & C_{2a}J_{2a} \\ \vdots & \vdots & & \vdots \\ C_{1a}J_{a1} & C_{2a}J_{2a} & \cdots & C_{aa}J_{aa} \end{bmatrix} + I\delta_2 , \tag{4.5}$$

where each J_{ij} represents an $n_i \times n_j$ matrix of 1's, and each

$$C_{ij} = \delta_1[(n_i + n_j)t_{ij} + h_{ij} + 1_{(i=j)}] + 2\delta_2t_{ij} + \delta_2m_{ij} .$$

The quantity $1_{(i=j)}$ represents the indicator function, which has a value of one if $i = j$ and zero otherwise. The remaining quantities t_{ij}, h_{ij}, and m_{ij} are computed as follows:

Let

$$g_i = r/(1 + rn_i) \quad \text{for} \quad i = 1, \ldots, a , \qquad (4.6)$$

and

$$G = -1/(N - \sum_i n_i^2 g_i) .$$

Then

$$t_{ij} = G(n_i g_i - 1)(n_j g_j - 1) - g_i 1_{(i=j)} ,$$

$$h_{ij} = \sum_k n_k^2 t_{ik} t_{kj} ,$$

and

$$m_{ij} = \sum_k n_k t_{ik} t_{kj} .$$

The elements of (4.4) can be expressed as:

$$ssq(R) = N + \sum_{ij} n_i n_j t_{ij}^2 + 2 \sum_k n_k t_{kk} ,$$

$$ssq(X_1'R) = N + \sum_{ij} n_i^2 n_j t_{ij}^2 + 2 \sum_k n_k^2 t_{kk} , \quad \text{and}$$

$$ssq(X_1'RX_1) = \sum_k n_k^2 + \sum_{ij} n_i^2 n_j^2 t_{ij}^2 + 2 \sum_k n_k^3 t_{kk} .$$

Although the above equations could possibly be simplified, they do serve the purpose of reducing the amount of computer storage required to study μ-invariant QUE's of $K_1\sigma_a^2 + K_2\sigma_e^2$ which minimize (4.2).

Examination of (4.5) leads to the fact that the estimator $Y'Q_rY$ is a function of the sufficient statistics[1] $(Y_{1.}, Y_{2.}, \ldots, Y_{a.}$, i.e., the class sums) and $\sum_{i,j} y_{ij}^2$. Furthermore (4.5) meets the necessary conditions specified by Harville [1969a] for "quadmissibility" since $C_{ij} = C_{i'j'}$ if $n_i = n_i'$ and $n_j = n_j'$.

For any Q_r matrix computed using (4.5),

$$E(Y'Q_rY) = K_1\sigma_a^2 + K_2\sigma_e^2 ,$$

and

$$\begin{aligned} Var(Y'Q_rY) = {} & 2\ ssq(X_1'Q_rX_1)\sigma_a^4 + 4\ ssq(X_1'Q_r)\sigma_a^2\sigma_e^2 \\ & + 2\ ssq(Q_r)\sigma_e^4 , \end{aligned}$$

where

$$2\ ssq(X_1'Q_rX_1) = 2\sum_i n_i^2(n_iC_{ii} + \delta_2)^2 + 4\sum_{i,j>i} n_i^2n_j^2C_{ij}^2 ,$$

$$4\ ssq(X_1'Q_r) = 4\sum_i n_i(n_iC_{ii} + \delta_2) + 2\sum_{i,j>i} n_in_j(n_i + n_j)C_{ij}^2 ,$$

and

$$2\ ssq(Q_r) = 2\sum_i [n_i(C_{ii} + \delta_2)^2 + (n_i^2 - n_i)C_{ii}^2] + 4\sum_{i,j>i} n_in_jC_{ij}^2 .$$

---

[1] See Hultquist and Graybill [1965].

The only restriction that need be placed on the value of r is that it not be equal to $-1/n_i$ for any $i = 1, \ldots, a$; in this case (4.6) becomes infinite for some value of i .

4.4 Relative Efficiencies

Having stated the model and provided a technique for generating infinitely many μ-invariant QUE's for $K_1\sigma_a^2 + K_2\sigma_e^2$, we are now in a position to study the behavior of the variances of estimators generated by using (4.5) with a variety of values for $r \geq 0$, and to compare them with the variance of the minimum variance μ-invariant QUE.

Throughout this section Q_r will denote the matrix generated by (4.5) for a specified value of $r \geq 0$ and for some specified values of K_1 and K_2 . The matrix Q_ρ denotes the matrix generated by (4.5) with $r = \rho = \sigma_a^2/\sigma_e^2$.

For any values of K_1 and K_2 (not both zero), $Y'Q_\rho Y$ is the minimum variance μ-invariant QUE for $K_1\sigma_a^2 + K_2\sigma_e^2$. The $\mathrm{Var}(Y'Q_\rho Y) \leq \mathrm{Var}(Y'Q_r Y)$ for any $r \geq 0$. The efficiency of $Y'Q_r Y$ for any $r \geq 0$ relative to $Y'Q_\rho Y$ denoted by

$$\mathrm{Eff}(Q_r | Q_\rho) = \frac{\mathrm{Var}(Y'Q_\rho Y)}{\mathrm{Var}(Y'Q_r Y)} \leq 1 .$$

Since little can be shown theoretically about $\mathrm{Eff}(Q_r|Q_\rho)$ due to the algebraic complexity of the problem, the current author employed equation (4.5) in a computer program to examine $\mathrm{Eff}(Q_r|Q_\rho)$ over a range of values of r and ρ for a number of different one-way classification designs. Tables 4.1, 4.2, and 4.3 are a sample of the tables produces in studying the efficienty of the μ-invariant QUE for σ_a^2 $(K_1 = 1, K_2 = 0)$. Tables 4.4, 4.5, and 4.6 are a sample of the

Table 4.1 Eff($Q_r|Q_\rho$) when estimating σ_a^2 in a one-way classification model with cell frequencies: 3, 5, 59, 20, 50, 21, and 89

	ρ=0	ρ=.25	ρ=1	ρ=5	ρ=10	ρ=100	ρ=1000	ρ=10000
r=0	1.000000	0.545024	0.441949	0.399638	0.393582	0.387957	0.387385	0.387328
r=.25	0.150487	1.000000	0.933258	0.874939	0.865512	0.856543	0.855619	0.855527
r=1	0.053123	0.898088	1.000000	0.983854	0.979049	0.974069	0.973535	0.973481
r=5	0.034322	0.788522	0.981067	1.000000	0.999621	0.998596	0.998456	0.998441
r=10	0.032262	0.770279	0.974931	0.999613	1.000000	0.999668	0.999598	0.999590
r=100	0.030465	0.752912	0.968404	0.998537	0.999662	1.000000	0.999997	0.999996
r=1000	0.030288	0.751126	0.967695	0.998388	0.999589	0.999997	1.000000	1.00000$^-$
r=10000	0.030271	0.750947	0.967624	0.998373	0.999581	0.999996	1.00000$^-$	1.000000
ANOVA[1]	0.587925	0.807093	0.671921	0.611748	0.602986	0.594819	0.593987	0.593903

---

[1] The efficiency of the standard ANOVA estimator (described in section 4.6) is included here for comparative purposes

Table 4.2 Eff($Q_r|Q_\rho$) when estimating σ_a^2 in a one-way classification model with cell frequencies: 22, 52, 33, 88, 68, 48, and 25

	$\rho=0$	$\rho=.25$	$\rho=1$	$\rho=5$	$\rho=10$	$\rho=100$	$\rho=1000$	$\rho=10000$
r=0	1.000000	0.655821	0.620161	0.609706	0.608368	0.607159	0.607037	0.607025
r=.25	0.559071	1.000000	0.996790	0.994632	0.994316	0.994022	0.993992	0.993989
r=1	0.513653	0.996715	1.000000	0.999715	0.999637	0.999559	0.999551	0.999550
r=5	0.500592	0.994470	0.999713	1.000000	0.999995	0.999983	0.999981	0.999981
r=10	0.498931	0.994140	0.999635	0.999995	1.000000	0.999996	0.999995	0.999995
r=100	0.497430	0.993833	0.999555	0.999983	0.999996	1.000000	1.00000^-	1.00000^-
r=1000	0.497279	0.993802	0.999547	0.999981	0.999995	1.00000^-	1.000000	1.00000^-
r=10000	0.497264	0.993799	0.999546	0.999981	0.999995	1.00000^-	1.00000^-	1.000000
ANOVA[1]	0.844936	0.891645	0.859781	0.849871	0.848586	0.847420	0.847303	0.847291

[1] The efficiency of the standard ANOVA estimator (described in section 4.6) is included here for comparative purposes

Table 4.3 Eff$(Q_r|Q_\rho)$ when estimating σ_a^2 in a one-way classification model with cell frequencies: 1, 33, 94, 78, 1, 64, 91, 69, 72, 1, 24, and 42

	ρ=0	ρ=.25	ρ=1	ρ=5	ρ=10	ρ=100	ρ=1000	ρ=10000
r=0	1.000000	0.743460	0.662315	0.571334	0.550680	0.528928	0.526564	0.526326
r=.25	0.184756	1.000000	0.934801	0.815404	0.786971	0.756779	0.753486	0.753154
r=1	0.011197	0.766674	1.000000	0.925139	0.898059	0.867824	0.864453	0.864113
r=5	0.002280	0.338290	0.863808	1.000000	0.995820	0.983743	0.982032	0.981856
r=10	0.001775	0.280231	0.805017	0.995427	1.000000	0.995623	0.994662	0.994560
r=100	0.001401	0.231693	0.739401	0.980770	0.995258	1.000000	0.999948	0.999937
r=1000	0.001367	0.227077	0.732190	0.978587	0.994171	0.999947	1.000000	0.999999
r=10000	0.001364	0.226618	0.731462	0.978361	0.994055	0.999936	0.999999	1.000000
ANOVA[1]	0.646481	0.906109	0.816273	0.706138	0.680848	0.654158	0.651255	0.650962

---

[1] The efficiency of the standard ANOVA estimator (described in section 4.6) is included here for comparative purposes

Table 4.4 $\mathrm{Eff}(Q_r|Q_\rho)$ when estimating σ_e^2 in a one-way classification model with cell frequencies: 3, 5, 59, 20, 50, 21, and 89

	ρ=0	ρ=.25	ρ=1	ρ=5	ρ=10	ρ=100	ρ=1000	ρ=10000
r=0	1.000000	0.700228	0.140769	0.006756	0.001706	0.000017	0.000000	0.000000
r=.25	0.992288	1.000000	0.992585	0.838361	0.569698	0.013387	0.000136	0.000001
r=1	0.990311	0.999020	1.000000	0.996873	0.987058	0.434284	0.007642	0.000077
r=5	0.989969	0.998695	0.999833	1.000000	0.999979	0.996338	0.730918	0.026453
r=10	0.989955	0.998681	0.999819	0.999995	1.000000	0.999755	0.975695	0.286479
r=100	0.989949	0.998676	0.999814	0.999990	0.999997	1.000000	0.999997	0.999732
r=1000	0.989949	0.998676	0.999814	0.999990	0.999997	1.00000^-	1.000000	1.00000^-
r=10000	0.989949	0.998676	0.999814	0.999990	0.999997	1.00000^-	1.00000^-	1.000000
ANOVA[1]	0.989949	0.998676	0.999814	0.999990	0.999997	1.00000^-	1.00000^-	1.00000^-

[1] The efficiency of the standard ANOVA estimator (described in section 4.6) is included here for comparative purposes

Table 4.5 $\mathrm{Eff}(Q_r|Q_\rho)$ when estimating σ_e^2 in a one-way classification model with cell frequencies: 22, 52, 33, 88, 68, 48, and 25

	$\rho=0$	$\rho=.25$	$\rho=1$	$\rho=5$	$\rho=10$	$\rho=100$	$\rho=1000$	$\rho=10000$
r=0	1.000000	0.707101	0.142656	0.006806	0.001717	0.000017	0.000000	0.000000
r=.25	0.997278	1.000000	0.999675	0.991101	0.965369	0.218622	0.002792	0.000028
r=1	0.997227	0.999976	1.000000	0.999955	0.999809	0.980934	0.339785	0.005121
r=5	0.997223	0.999972	0.999998	1.000000	1.00000^-	0.999966	0.996596	0.745375
r=10	0.997223	0.999972	0.999998	1.00000^-	1.000000	0.999998	0.999784	0.978849
r=100	0.997223	0.999972	0.999998	1.00000^-	1.00000^-	1.000000	1.00000^-	0.999998
r=1000	0.997223	0.999972	0.999998	1.00000^-	1.00000^-	1.00000^-	1.000000	1.00000^-
r=10000	0.997223	0.999972	0.999998	1.00000^-	1.00000^-	1.00000^-	1.00000^-	1.000000
ANOVA[1]	0.997223	0.999972	0.999998	1.00000^-	1.00000^-	1.00000^-	1.00000^-	1.00000^-

[1] The efficiency of the standard ANOVA estimator (described in section 4.6) is included here for comparative purposes

Table 4.6 $\mathrm{Eff}(Q_r|Q_\rho)$ when estimating σ_e^2 in a one-way classification model with cell frequencies: 1, 33, 94, 78, 1, 64, 91, 69, 72, 1, 24, and 42

	ρ=0	ρ=.25	ρ=1	ρ=5	ρ=10	ρ=100	ρ=1000	ρ=10000
r=0	1.000000	0.785052	0.201057	0.010270	0.002597	0.000026	0.000000	0.000000
r=.25	0.997504	1.000000	0.995835	0.908862	0.728468	0.028563	0.000297	0.000003
r=1	0.995061	0.998645	1.000000	0.990550	0.962757	0.216443	0.002787	0.000028
r=5	0.993360	0.997080	0.999123	1.000000	0.999822	0.970352	0.247723	0.003286
r=10	0.993222	0.996942	0.998990	0.999949	1.000000	0.997570	0.802567	0.039101
r=100	0.993166	0.996887	0.998935	0.999902	0.999972	1.000000	0.999969	0.996867
r=1000	0.993166	0.996886	0.998935	0.999901	0.999972	1.00000^-	1.000000	1.00000^-
r=10000	0.993166	0.996886	0.998935	0.999901	0.999972	1.00000^-	1.00000^-	1.000000
ANOVA[1]	0.993166	0.996886	0.998935	0.999901	0.999972	1.00000^-	1.00000^-	1.00000^-

---

[1] The efficiency of the standard ANOVA estimator (described in section 4.6) is included here for comparative purposes

tables produced in studying the efficiency of the μ-invariant QUE for σ_e^2 $(K_1 = 0, K_2 = 1)$.

Close examination of Tables 4.1 thru 4.6 seems to indicate the following: when estimating σ_a^2 or σ_e^2 using equation (4.5), as r approaches ρ (fixed) from above or below, $\mathrm{Eff}(Q_r|Q_\rho)$ monotonically increases to a value of one. Furthermore, it appears that as ρ approaches r (fixed) from above or below, $\mathrm{Eff}(Q_r|Q_\rho)$ monotonically increases to a value of one. Also, when estimating σ_e^2 , $\mathrm{Eff}(Q_r|Q_\rho)$ seems to approach an asymptotic value rather quickly as r increases beyond ρ (fixed). In tables not shown, the $\mathrm{Eff}(Q_r|Q_\rho)$ was equal to one any time $n_1 = n_2 = \dots = n_a$.

4.5 Choosing Reliable Estimates With Limited Prior Knowledge

If a reliable estimate is available for ρ , then use of that estimate in (4.5) should yield fairly efficient estimators for both σ_a^2 and σ_e^2 due to the seemingly monotonic properties of $\mathrm{Eff}(Q_r|Q_\rho)$. In many situations it might be possible to bracket ρ with feasible upper and lower bounds. In other words, if we are confident that the true value of ρ lies somewhere between, say, ρ_0 and ρ_1 , then choosing a particular value of r between ρ_0 and ρ_1 to use in (4.5) would seem the most logical choice.

By observing Tables 4.1 thru 4.6, one can see that the line associated with any choice of r has its smallest efficiency either in the first or last position of the line. This smallest efficiency represents the "worst" one can do if he uses that particular r and the true value of ρ in fact lies somewhere within the upper and lower bounds of the table. If we are certain that the true value of ρ lies between ρ_0 and ρ_1, then, for any value of r we use in (4.5), the worst efficiency we could possibly have would be the smaller of either $\mathrm{Eff}(Q_r|Q_{\rho_0})$ or $\mathrm{Eff}(Q_r|Q_{\rho_1})$. Conceptually, we could build a list containing all possible r values along with the worst possible efficiencies associated with each of them assuming ρ lies between ρ_0 and ρ_1. To minimize the risk we are taking in choosing a value of r, we could pick the r value associated with the largest efficiency in this list (of worst efficiencies). By using this value of r in (4.5) we could provide not only the variance component estimates based on this r value, but also a "guaranteed" efficiency level for the estimators, which would merely be the worst possible efficiency associated with that value of r.

On closer examination of Tables 4.1 thru 4.6, one can see that the search area for the "best" r may be restricted to values between ρ_0 and ρ_1. If we denote by r^* the value chosen by this technique, then r^* is equal to the value of r for which

$$F(r) = \min[\mathrm{Eff}(Q_r|Q_{\rho_0}) ,\ \mathrm{Eff}(Q_r|Q_\rho)] \qquad (4.7)$$

is a maximum over the range $\rho_0 \leq r \leq \rho_1$.

The following graph (Figure 4.1) illustrates the above points.

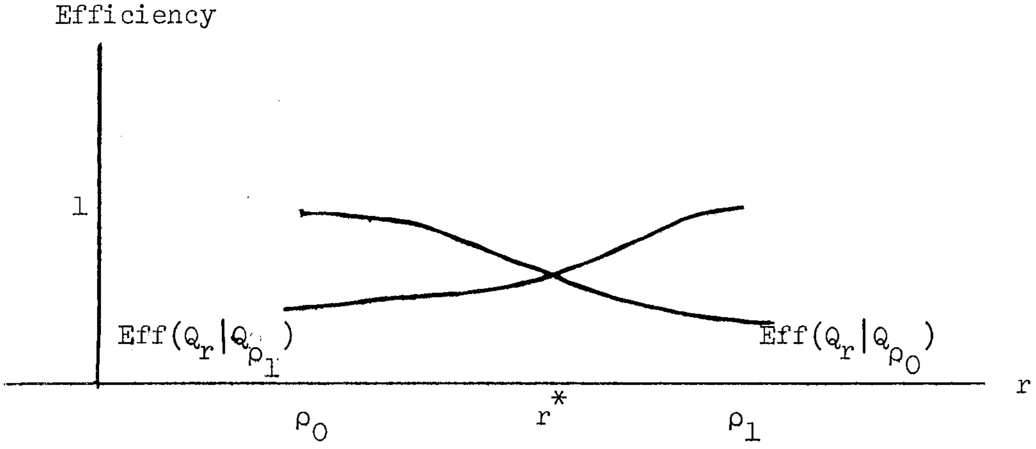

Figure 4.1 Typical efficiencies when estimating σ_a^2 or σ_e^2

Although the above curves are merely representations of the efficiencies, the following point can be made: Since $\mathrm{Eff}(Q_r|Q_{\rho_0})$ has a value of 1 when $r = \rho_0$ and monotonically decreases as r approaches ρ_1 from below and since $\mathrm{Eff}(Q_r|Q_{\rho_1})$ has a value of 1 when $r = \rho_1$ and monotonically decreases as r approaches ρ_0 from above, the function $F(r)$ of (4.7) is represented by the portion of the two curves closest to the r axis. As can be seen $F(r)$ achieves a maximum at the point of intersection of the two curves, and it is at this intersection that r^* is realized. Since the theoretical formulas for these curves are intractable, the following algorithm can be used for determining r^*.

Algorithm 4.1:

STEP 0: Assign the desired values for K_1, K_2, and set ϵ_0 = desired accuracy of $F(r^*)$ (i.e., the algorithm terminates when a value of r is found such that

$$|\mathrm{Eff}(Q_r|Q_{\rho_0}) - \mathrm{Eff}(Q_r|Q_{\rho_1})| \le \epsilon_0) .$$

```
STEP 1:  Set $r_L = \rho_0$ and $r_H = \rho_1$ .
STEP 2:  Set $r = r_L + (r_H - r_L)/2$ .
STEP 3:  Compute $Q_r$ using equation (4.5).
STEP 4:  Set $F = \mathrm{Eff}(Q_r|Q_{\rho_0}) - \mathrm{Eff}(Q_r|Q_{\rho_1})$ .
STEP 5:  If $|F| \le \epsilon_0$ , set $r^* = r$ and terminate, otherwise:
STEP 6:  If $F < 0$ , go to STEP 7; otherwise set $r_L = r$ , and go to
         STEP 2 .
STEP 7:  Set $r_H = r$ and go to STEP 2.
```

Since the search area for r^* is halved at each iteration in the above algorithm, convergence is quite rapid. Several examples are presented in section 4.6 which compare the Q_{r^*} estimators (r^* being generated by algorithm 4.1) with the standard analysis of variance estimators for σ_a^2 and σ_e^2 .

When no knowledge about ρ is available other than assuming $\rho \ge 0$, then algorithm 4.1 may be applied by setting $\rho_0 = 0$ and ρ_1 equal to a pseudo value for infinity. A choice of $\rho_1 = 1000$ would in most instances suffice. Of course the smaller ρ_1 can prudently be made the better, especially when estimating σ_a^2 . However, when estimating σ_e^2 , one need not be concerned about making ρ_1 as small as possible, since it appears that for any fixed value of ρ , the $\mathrm{Eff}(Q_r|Q_\rho)$ approaches an asymptotic value quite rapidly as r increases beyond the value of ρ. In other words, if it were known that ρ were less than say 100, then any choice of $\rho_1 \geq 100$, whether it be 100 or 10,000 would not make any appreciable change in the choice of r^*. Based on the above discussion, we formally define an estimator which uses the r^* of algorithm (4.1).

Definition 4.1: The Relatively Safe Quadratic Unbiased Estimator (RESQUE) for $K_1\sigma_a^2 + K_2\sigma_e^2$ in the one-way classification is $Y'Q_{r^*}Y$, where r^* is determined by algorithm 4.1 and Q_{r^*} is determined by equation (4.5) with $r = r^*$.

4.6 Comparison of Invariant Quadratic Unbiased Estimators To The Standard Analysis of Variance Estimators

The standard analysis of variance estimators for σ_e^2 and σ_a^2 are:

$$\hat{\sigma}_e^2 = Y'EY ,$$

and

$$\hat{\sigma}_a^2 = Y'AY ,$$

where

$$E = \frac{1}{N-a}[I - X_1X_1'\Lambda] ,$$

$$A = \frac{N}{N^2-\Sigma n_i^2}\left[\frac{N-1}{N-a} X_1X_1'\Lambda - \frac{1}{N} X_0X_0' - \frac{a-1}{N-a} I\right] ,$$

and where Λ is an $N \times N$ diagonal matrix whose first n_1 diagonal elements are $1/n_1$, whose next n_2 diagonal elements are $1/n_2$, etc. Both $Y'EY$ and $Y'AY$ are μ-invariant estimators, and the matrices E and A have the same structure as does the Q_r matrix of (4.5). In terms of (4.5), the E matrix has:

$$C_{ij} = \begin{cases} -\left(\frac{1}{N-a}\right)\left(\frac{1}{n_i}\right) & \text{if } i = j , \\ 0 & \text{otherwise} \end{cases} ,$$

and

$$\delta_2 = \frac{1}{N-a} .$$

In terms of (4.5), the A matrix has:

$$C_{ij} = \begin{cases} \frac{K_a}{n_i}\left(\frac{N-1}{N-a}\right) - \frac{K_a}{N} & \text{if } i = j , \\ -\frac{K_a}{N} & \text{otherwise} \end{cases} ,$$

and

$$\delta_2 = - K_a \left(\frac{a-1}{N-a}\right) ,$$

where

$$K_a = \frac{N}{N^2 - \Sigma n_i^2} .$$

Since both the A and E matrices have the same structure as Q_r, is there a value of r which could be used in (4.6) to generate the A and the E matrices? By Theorem (2.2), the estimator $\hat{\sigma}_e^2$ is a maximally-invariant estimator. Therefore the E matrix may be generated by setting $r = \infty$. Hence, if an upper bound on ρ is known, setting r equal to that upper bound will produce a more efficient estimator than the standard analysis of variance estimator

$\hat{\sigma}_e^2$. The corresponding value of r associated with the matrix A is not as easily determined. However, by making numerous computer runs, the current writer found that a value of $r = 1/n_h$ (where n_h is the harmonic mean of the n_i's) produced a matrix Q_r such that $Y'Q_rY$ approximately equaled $Y'AY$ for all cases tested. Thus, if prior knowledge is available which indicates that ρ has a value differing from $1/n_h$, then use of that knowledge in algorithm 4.1 should produce a more efficient estimator.

To compare further the standard analysis of variance estimators with μ-invariant QUE's, the following tables give the worst efficiency achievable provided ρ lies between ρ_0 and ρ_1 , for the analysis of variance estimators, the RESQUE estimators, and the estimators based on setting $r = 1$ in (4.5).

Table 4.7 Smallest obtainable efficiencies when $0 \leq \rho \leq 10000$ of RESQUE, ANOVA and $Q_{r=1}$ estimators for a one-way classification model with cell frequencies: 3, 5, 59, 20, 50, 21, and 89

Component to be Estimated	RESQUE r^* value	Smallest RESQUE Efficiency	Smallest ANOVA Efficiency	Smallest $Q_{r=1}$ Efficiency
σ_a^2	.0427	.62780	.58792	.05312
σ_e^2	40.2069	.98994	.98994	.00007

Table 4.8 Smallest obtainable efficiencies when $1 \leq \rho \leq 10$ of RESQUE, ANOVA and $Q_{r=1}$ estimators for a one-way classification model with cell frequencies: 3, 5, 59, 20, 50, 21, and 89

Component to be Estimated	RESQUE r^* value	Smallest RESQUE Efficiency	Smallest ANOVA Efficiency	Smallest $Q_{r=1}$ Efficiency
σ_a^2	1.9097	.99419	.60298	.97904
σ_e^2	3.2500	.99985	.99981	.98705

Table 4.9 Smallest obtainable efficiencies when $0 \leq \rho \leq 10000$ of RESQUE, ANOVA and $Q_{r=1}$ estimators for a one-way classification model with cell frequencies: 22, 52, 33, 88, 68, 48, and 25

Component to be Estimated	RESQUE r^* value	Smallest RESQUE Efficiency	Smallest ANOVA Efficiency	Smallest $Q_{r=1}$ Efficiency
σ_a^2	.0222	.84952	.84493	.51365
σ_e^2	16.7847	.99722	.99722	.00512

Table 4.10 Smallest obtainable efficiencies when $1 \leq \rho \leq 10$ of RESQUE, ANOVA and $Q_{r=1}$ estimators for a one-way classification model with cell frequencies: 22, 52, 33, 88, 68, 48, and 25

Component to be Estimated	RESQUE r^* value	Smallest RESQUE Efficiency	Smallest ANOVA Efficiency	Smallest $Q_{r=1}$ Efficiency
σ_a^2	1.8438	.99990	.84858	.99963
σ_e^2	5.5000	.99999	.99999	.99980

Table 4.11 Smallest obtainable efficiencies when $0 \leq \rho \leq 10000$ of RESQUE, ANOVA and $Q_{r=1}$ estimators for a one-way classification model with cell frequencies: 1, 33, 94, 78, 1, 64, 91, 69, 72, 1, 24, and 42

Component to be Estimated	RESQUE r^* value	Smallest RESQUE Efficiency	Smallest ANOVA Efficiency	Smallest $Q_{r=1}$ Efficiency
σ_a^2	.0574	.70109	.64648	.01119
σ_e^2	82.0923	.99316	.99316	.00002

Table 4.12 Smallest obtainable efficiencies when $1 \leq \rho \leq 10$ of RESQUE, ANOVA and $Q_{r=1}$ estimators for a one-way classification model with cell frequencies: 1, 33, 94, 78, 1, 64, 91, 69, 72, 1, 24, and 42

Component to be Estimated	RESQUE r^* value	Smallest RESQUE Efficiency	Smallest ANOVA Efficiency	Smallest $Q_{r=1}$ Efficiency
σ_a^2	2.0624	.96166	.68084	.89805
σ_e^2	3.6016	.99925	.99893	.96275

5. SUMMARY AND SUGGESTIONS FOR FUTURE RESEARCH

5.1 Summary

Rao's recent works on MINQUE and MIVQUE provide estimators of the variance components in linear models. If normality is assumed, then MINQUE and MIVQUE coincide, and provide locally best quadratic unbiased estimators which are invariant to translation of the fixed effects in the model. As pointed out in section 2, β_0-invariant estimators, which are based on a priori estimates of the components being estimated, have variances which are, in general, functionally dependent on all of the true values of the parameters in the model. The basic weakness of MIVQUE is then, that when a priori estimates are not "close" enough to their true values, or when no a priori estimates are available, the variance of MIVQUE estimators could in many instances be unacceptably large. To help reduce this risk, the concept of invariance with respect to one or more of the random components was introduced. These additional invariance concepts, remove certain terms from the variance of an estimator in order to protect against situations for which these terms might be large. The concept of maximally-invariant estimators was introduced to denote those estimators for which the variances are as free as possible of terms involving other components in the model.

In section 3, techniques are presented for computing MIVQUE estimators, and also for locally best quadratic unbiased estimators which are invariant to other random effects in the model. In that section, one can see that MIVQUE is actually equivalent to selecting a set of sums of squares from a weighted regression analysis, to equate them to their expected values, and to solve to get variance component estimates.

To provide some insight into the choice of the prior estimates to use in MIVQUE, when either limited or no a priori knowledge is available, section 4 examines what can be done in this respect when working with the one-way classification. A new estimator, referred to as the Relatively Safe Quadratic Unbiased Estimator (RESQUE), is developed which in some sense minimizes the risk one takes in choosing the prior estimates of components to use in MIVQUE.

5.2 Suggestions For Future Research

One of the problems which arises in MIVQUE is the size of the weighting matrix which must be inverted. Although in section 3 techniques which can be used are presented, these still require in general the inversion of the $N \times N$ weighting matrix (where N represents the number of observations). Unless techniques can be developed which circumvent this inversion process, MINQUE will be rendered an interesting but somewhat useless technique. Along this same line, using an identity matrix as the weighting matrix might prove quite efficient for those designs which are not very unbalanced, such as those arising from experiments which were initially balanced, but for which some observations have been lost.

One of the other computational problems, so far as the size is concerned, is the amount of storage required to compute the $X'RX$ portion of (3.4). Assuming the last effect in the model being analyzed is $I\sigma_e^2$, where I is $N \times N$ and σ_e^2 is the error variance, then the quantities $ssq(X_1'R)$, $ssq(X_2'R)$, ..., $ssq(X_{m-1}'R)$, and $ssq(R)$ are needed in the expectation of each sums of squares being computed. Since $X_1'R$, $X_2'R$, ..., $X_{m-1}'R$, and R represent the major portion of the $X'RX$ matrix, computing them in some other fashion would reduce considerably the amount of computer storage required.

In addition to the computational problems described above, there remains the problem of what can be done when no a priori estimates are available. Hopefully the work presented in section 4 describing what can be done in this situation for the one-way classification can be extended to cover the more general linear models discussed in section 2.

Although Rao [1971b] presents the basic theory for obtaining minimum mean square quadratic estimators when the true values of the variance components are known, Rao [1972] suggests that iterative estimation using MINQUE's may provide estimators with interesting properties. These "interesting properties" might include small mean square errors, obtained through the use of equation (4.5) and the more general techniques of section 3.

As aid to future research in this area, and to make available MIVQUE's for experiments of under 250 observations, the present writer has developed a computer procedure which is available in the Statistical Analysis System of North Carolina State University at Raleigh. A description of the system and the MIVQUE procedure is given by Service [1972].

6. LIST OF REFERENCES

Harville, D. A. [1969a]. Quadratic unbiased estimation of variance components for the one-way classification. Biometrika 56:313-326.

Harville, D. A. [1969b]. Variance component estimation for unbalanced one-way random classification - a critique. ARL Report 69-1080.

Henderson, C. R. [1953]. Estimation of variance and covariance components. Biometrics 9:226-252.

Hultquist, R. A. and F. A. Graybill [1965]. Minimal sufficient statistics for the two-way classification mixed model design. JASA 60:182-192.

Koch, G. G. [1967]. A general approach to the estimation of variance components. Technometrics 9:93-118.

Rao, C. R. [1970]. Estimation of heteroscedastic variances in linear models. JASA 65:161-172.

Rao, C. R. [1971a]. Estimation of variance and covariance components - MINQUE theory. Journal of Multivariate Analysis 1:257-275.

Rao, C. R. [1971b]. Minimum variance quadratic unbiased estimation of variance components. Journal of Multivariate Analysis 1:445-456.

Rao, C. R. [1972]. Estimation of variance and covariance components in linear models. JASA 67:112-115.

Read, R. R. [1961]. On quadratic estimates of the interclass variance for unbalanced designs. J.R.S.S. B 23:493-497.

Rohde, C. A. [1964]. Contributions to the theory, computation, and application of generalized inverses. Ph.D. Thesis, Department of Experimental Statistics, North Carolina State University at Raleigh, Raleigh, North Carolina. Univerisity Microfilms, Ann Arbor, Michigan.

Schatzoff, M., R. Tsao, and S. Fienberg [1968]. Efficient calculation of all possible regressions. Technometrics 10:769-779.

Searle, S. R. [1971]. Topics in variance components estimation. Biometrics 27:1-76.

Service, J. W. [1972]. A User's Guide To The Statistical Analysis System. North Carolina State University Students Supply Stores, Raleigh, North Carolina.

Townsend, E.C. [1968]. Unbiased estimators of variance components in simple unbalanced designs. Ph.D. Thesis, Biometrics Unit, Cornell University, Ithaca, New York. University Microfilms, Ann Arbor, Michigan.

7. APPENDIX

7.1 An Algorithm For Factoring A Symmetric Positive Definite Matrix in Place

Denoting by W the symmetric positive definite matrix which is to be factored, only the upper triangular portion need be computed and stored in core. The algorithm presented here computes the P' outlined in (3.7) and stores it in the same core positions previously occupied by the upper triangular portion of W without using any other storage locations for intermediate results.

```
STEP 0:  Compute the $N(N+1)/2$ elements of $W$ and denote them by
         $W_{ij}$ $(i = 1, \ldots, N), (j = i, \ldots, N)$ and set $k = 0$ .
STEP 1:  Set $k = k + 1$ and set $D = W_{kk}$ .
STEP 2:  Set $\ell = K + 1$ and if $\ell > N$ , then go to STEP 5.
STEP 3:  Set $W_{ij} = W_{ij} - W_{ki}W_{kj}/D$ for each $(i = \ell, \ldots, N)$ and
         $(j = i, \ldots, N)$ .
STEP 4:  If $k < N$ , then go to STEP 1.
STEP 5:  Set $i = 0$ .
STEP 6:  Set $i = i + 1$ and $D = \sqrt{W_{ii}}$ .
STEP 7:  Set $W_{ij} = W_{ij}/D$ for each $j = i, \ldots, n$ .
STEP 8:  If $i < N$ , then go to STEP 6, otherwise stop; the elements
         labeled $W_{ij}$ are now $P'_{ij}$ .
```

The above algorithm stores only the upper triangular portion of P' since the lower triangle is all zeros. In the event that the matrix W is not of full rank, a singularity check may be inserted following STEP 1 as follows.

```
STEP 1a: If $D < \epsilon_1$ (where $\epsilon_1 > 0$ represents a pseudo value for
         zero), then stop, $W$ is singular.
```

7.2 An Algorithm For The Inversion Of An Upper Triangular Matrix in Place

Assuming that the elements above and including the diagonal of an upper triangular matrix denoted by W_{ij} $(i = 1, \ldots, N)$, $(j = i, \ldots, N)$ have been stored, the following algorithm replaces the elements of W with the elements of W^{-1} without additional working storage needed, and is a modification to the sweep technique presented by Schatzoff, *et al*.

```
STEP 0: Set $k = 0$.
STEP 1: Set $k = k + 1$, set $D = W_{kk}$, and set $i = 0$.
STEP 2: Set $i = i + 1$ and if $i = k$, then go to STEP 5; otherwise
        set $j = k - 1$.
STEP 3: Set $j = j + 1$ and if $j = k$, then go to STEP 4; otherwise
        set $W_{ij} = W_{ij} - W_{ik}W_{kj}/D$.
STEP 4: If $j < N$, then go to STEP 3.
STEP 5: If $i < k$, then go to STEP 2.
STEP 6: Set $W_{ik} = -W_{ik}/D$ for each $i = 1, \ldots, k$.
STEP 7: Set $W_{kj} = W_{kj}/D$ for each $j = k, \ldots, N$.
STEP 8: Set $W_{kk} = 1/D$.
STEP 9: If $k < N$, then go to STEP 1; otherwise stop; the upper
        triangular portion of $W$ has been replaced by the upper
        triangular portion of $W^{-1}$.
```

The above algorithm takes advantage of the fact that the inverse of an upper triangular matrix is an upper triangular matrix and hence neither W nor W^{-1} needs the lower triangular portion of zeros to be stored. In the event that W might be singular, a singularity check as was given in Appendix 7.1 may be inserted following STEP 1.

7.3 An In Place Generalized Inverse Sweep Algorithm

As pointed out by Rohde the abbreviated Doolittle method may be modified to produce a generalized inverse by setting any row of the A_d matrix of the forward Doolittle to zero if its diagonal element goes to zero during the forward Doolittle procedure, and by setting the corresponding column of the B_d matrix to zero except for the diagonal element which is set of 1. Applying the backward Doolittle to the A_d and B_d matrices thus defined, then produces a generalized inverses. This same technique applied to the sweep routine outlined by Schatzoff et al. would imply that if any pivot element goes to zero then the matrix will be considered swept on that pivot element once the row and column containing that pivot element have been set to zero.

Using the modified sweep routine on pivot elements 1, 2, ..., n_0 of the matrix A of (3.2) would thus be equivalent to multiplying A by the matrix B of (3.3) yielding the resultant matrix product BA of (3.4).

Assuming only the elements A on and above the diagonal have been stored (i.e., $(i = 1, \ldots, n_r)$ and $(j = i, \ldots, n_c)$), where n_r and n_c represent the number of rows and columns respectively of A , the following algorithm performs a sweep on each of the pivot elements $1, \ldots, n_0$, with the result that the elements A_{ij} are replaced by the corresponding elements of the BA matrix of (3.4).

```
STEP 0:  Set $k = 0$ .
STEP 1:  Set $k = k + 1$ and $D = A_{kk}$ .
STEP 2:  If $D > \epsilon_1$ (where $\epsilon_1$ is a pseudo value for zero), then go
         to STEP 3; otherwise set $A_{ik} = 0$ for $i = 1, \ldots, k$ and
         set $A_{kj} = 0$ for $j = k, \ldots, n_r$ ; then go to STEP 10.
STEP 3:  Set $i = 0$ .
STEP 4:  Set $i = i + 1$ and if $i = k$ then go to STEP 8; otherwise
         set $j = i - 1$ .
STEP 5:  Set $j = j + 1$ and if $j = k$ , then go to STEP 7.
STEP 6:  Set $B = \begin{cases} A_{ik} & \text{if } i < k \\ A_{ki} & \text{otherwise} \end{cases}$ , and
         set $C = \begin{cases} A_{kj} & \text{if } k < j \\ -A_{kj} & \text{otherwise} \end{cases}$ , and
         set $A_{ij} = A_{ij} - BC/D$ .
STEP 7:  If $j < n_c$ , then go to STEP 5.
STEP 8:  If $i < n_r$ , then go to STEP 4.
STEP 9:  Set $A_{ik} = -A_{ik}/D$ for each $i = 1, \ldots, k$ .
         Set $A_{kj} = A_{kj}/D$ for each $j = k, \ldots, n_c$ .
         Set $A_{kk} = 1/D$ .
STEP 10: If $k < n_0$ , then go to STEP 1; otherwise stop; the BA
         matrix of (3.4) has replaced the A matrix of (3.2).
```