ON CANONICAL FORMS, NON-NEGATIVE COVARIANCE MATRICES AND BEST AND SIMPLE LEAST SQUARES LINEAR ESTIMATORS IN LINEAR MODELS[1]

By George Zyskind

Iowa State University

0. Summary. Aspects of best linear estimation are explored for the model $y = X\beta + e$ with arbitrary non-negative (possibly singular) covariance matrix $\sigma^2 V$. Alternative necessary and sufficient conditions for all simple least squares estimators to be also best linear unbiased estimators (blue's) are presented. Further, it is shown that a linear function $w'y$ is blue for its expectation if and only if $Vw \varepsilon \mathcal{C}(X)$, the column space of X. Conditions on the equality of subsets of blue's and simple least squares estimators are explored. Applications are made to the standard linear model with covariance matrix $\sigma^2 I$ and with additional known and consistent equality constraints on the parameters. Formulae for blue's and their variances are presented in terms of adjustments to the corresponding expressions for the case of the unrestricted standard linear model with covariance matrix $\sigma^2 I$.

1. Introduction. Consider the general linear model

$$y = X\beta + e \tag{1}$$

where y is an $n \times 1$ vector of observable or known values, X is a known $n \times p$ matrix of rank r, β is a $p \times 1$ vector of parameters and e is an $n \times 1$ vector of errors with expectation $E(e) = 0$ and with variance $E(ee') = \sigma^2 V$, where σ^2 is positive and known or unknown and V is an $n \times n$ non-negative symmetric matrix which is either totally or partially known. One example of a situation which may be viewed as having a singular covariance matrix arises when the parameters are subject to specified linear constraints. Further, an important class of situations in which the model covariance matrix is naturally singular is that arising in randomized experiments when all errors are induced by the random assignment of treatments to subsets of the experimental material. Such situations are introduced and treated in the classic book by Fisher [4], and are mathematically explicitly exhibited, for example, in the books by Kempthorne [6], by Scheffé [16], and in Kempthorne, et al. [7]. For instance, the derived linear model for the observation on the kth treatment in the ith block in the case of a randomized block under additivity is given in Kempthorne [6] as

$$y_{ik} = \mu + b_i + t_k + \textstyle\sum_j \delta_{ij}^k e_{ij},$$

Received 22 April 1966; revised 20 March 1967.

[1] This research was supported by the United States Air Force under Contract No. AF 33(615)1737, monitored by the Aerospace Research Laboratories, Wright Air Development Division.

where the error $\sum_j \delta_{ij}^k e_{ij}$ associated with the observation y_{ik} is such that when it is summed over all the treatments in the given block it gives the value $\sum_k (\sum_j \delta_{ij}^k e_{ij}) = 0$. As a consequence singularities become introduced into the covariance matrix of the observations y_{ik}.

When the covariance matrix $\sigma^2 V$ is non-singular with V known, and the $n \times p$ matrix X is of full rank, i.e., of rank p, and when further the columns of the matrix X are all orthogonal eigenvectors of the matrix V then it is an elementary and easily verifiable fact that the best linear unbiased estimator of the vector β is identical with its usual simple least squares estimator. This fact was first pointed out perhaps by Anderson [1], and notice of it was taken soon after by Durbin and Watson [3]. An attempt at exploiting the fact in the case of a covariance structure of form like that of a randomized block, but with non-singular covariance matrix, was made by Box and Muller [2], and was further amplified by Muller and Watson in [9].

While blue estimators for contrasts in various randomized designs were obtained by specialized methods in the report ARL 149 by Kempthorne, et al., [7], mention was there also made of the alternative possibility of deriving the desired results by use of the fact that when a subset of r eigenvectors of the covariance matrix $\sigma^2 V$ may be chosen to form a basis of the design matrix X then the usual least squares estimators yield corresponding best linear estimators. A further statement on various conditions for the equality of best and simple least squares linear estimators under a non-singular covariance matrix, was made by Zyskind [17] and a short discussion of such conditions is presented by Zyskind, et al., [18]. Another statement and partial proof of a necessary and sufficient condition is also presented in the recent paper by C. R. Rao [15]. Also, a proof that the eigenvector condition is both necessary and sufficient for the corresponding best and simple least squares linear estimators to have the same covariance matrix is presented, with X of full rank and V non-singular, by Magness and McGuire [8].

In the present paper we discuss best linear estimation in linear models with arbitrary non-negative covariance structure. Because of the mode of arguments involved and of the above background we consider first the conditions under which all the simple least squares linear estimators are also blue's. Thus, in Section 2 we first verify for an arbitrary covariance matrix the validity of the generalization of the eigenvector condition, and we proceed to a proof and statement of a number of other equivalent conditions, most of which are direct generalizations of statements previously made by Zyskind [17] and Zyskind, et al. [18]. In Section 3 we provide a basic characterization of best linear estimators and then give conditions for the equality of subsets of best and simple least squares estimators. We finally consider the linear model involving known parametric restrictions, which may be regarded as known observations with zero variance, as an example of a model with a special singular covariance matrix. In dealing with the above problems we have found it both clarifying and fruitful to employ a canonical form of the general linear model.

2. On canonical forms and conditions under which all simple linear least squares estimators are also best linear unbiased estimators. A standard special form of the linear model is

$$y = X\beta + e, \qquad E(e) = 0, \qquad E(ee') = \sigma^2 I, \tag{2}$$

where the meaning of the symbols is as in (1) and where I represents the identity matrix. A canonical transformation of the above model may be obtained as follows. Let O_1' be any orthonormal basis for the r dimensional vector subspace of all linear combinations of the columns of X, hereafter called the column space of X, and abbreviated as $\mathcal{C}(X)$, and let O_0' be any orthonormal basis for the $(n - r)$ dimensional othogonal complement subspace to the column space of X. Then $O_0X = \emptyset$, a null matrix, and the chosen particular orthogonal matrix

$$O = \begin{pmatrix} O_1 \\ O_0 \end{pmatrix}$$

applied to the observational model (2) yields

$$z = \begin{pmatrix} z_1 \\ z_0 \end{pmatrix} = \begin{pmatrix} O_1 \\ O_0 \end{pmatrix} y = \begin{pmatrix} O_1 X \\ \emptyset \end{pmatrix} \beta + Oe = \begin{pmatrix} \delta \\ \emptyset \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_0 \end{pmatrix}$$

where $\delta = O_1X\beta$ is a set of r independent estimable functions and where the covariance matrix of the errors η is given by $\text{Cov}(\eta) = \sigma^2 I$. (We recall that a linear parametric function is said to be estimable if there exists a linear function of the observations whose expectation equals the parametric function for all permissible values of the parameters). We shall refer to the form

$$z = \begin{pmatrix} z_1 \\ z_0 \end{pmatrix} = \begin{pmatrix} \delta \\ \emptyset \end{pmatrix} + \eta, \tag{3}$$

obtained from (2) as described, as a canonical form of the model (2).

An estimation procedure, suggesting itself at once from inspection of the form (3) and without appealing to a "deeper" principle such as that of least squares, is to estimate the r independent estimable parametric functions $\delta = O_1X\beta$ by the corresponding vector z_1. It is a simple matter to establish rigorously the fact that when $\text{Cov}(z_1, z_0) = \emptyset$, then the best linear unbiased estimator of any linear parametric function $\nu'\delta = \nu'O_1X\beta$ is $\nu'z_1 = \nu'O_1y$, for then the adding of a linear function of components of z_0 cannot decrease the already achieved variance. We note that $\nu'z_1$ is in fact the unique unbiased estimator of the parametric function $\nu'\delta$ based alone on the transformed observational part z_1.

We now consider the model (1), in which the covariance matrix may be singular, and enquire into the conditions under which the simple least squares linear estimators of all the linearly estimable functions are also best linear unbiased estimators.

THEOREM 1. *A necessary and sufficient condition for all simple least squares* *linear estimators to be also best linear unbiased estimators of the corresponding estimable parametric functions* $\lambda'\beta$ *in the linear model*

$$y = X\beta + e, \qquad E(e) = 0, \qquad E(ee') = \sigma^2 V,$$

where V *is a symmetric non-negative matrix, is that there exist a subset of* r *orthogonal eigenvectors of* V *which forms a basis for the column space of the matrix* X.

PROOF. (a) Sufficiency. Suppose that V has r orthogonal eigenvectors which form a basis for the column space of X. Then V also has $(n - r)$ orthogonal eigenvectors which form a basis for the orthogonal complement of the column space of X. Let O_1' and O_0' be two matrices whose columns are made up respectively by each of the two sets of basis vectors. Then the orthogonal matrix $O' = (O_1', O_0')$ has for its columns a complete set of eigenvectors of V so that $OVO' = D$, a diagonal matrix. Also, $O_0 X = \varnothing$. Consequently, the transformation O applied to the original observational model yields

$$z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} O_1 y \\ O_0 y \end{pmatrix} = \begin{pmatrix} O_1 X\beta \\ \varnothing \end{pmatrix} + \begin{pmatrix} O_1 e \\ O_0 e \end{pmatrix} = \begin{pmatrix} \delta \\ \varnothing \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_0 \end{pmatrix}$$

where

$$\text{Cov} \begin{pmatrix} \eta_1 \\ \eta_0 \end{pmatrix} = \sigma^2 OVO' = \sigma^2 D = \sigma^2 \begin{pmatrix} D_1 & \varnothing \\ \varnothing & D_0 \end{pmatrix},$$

a diagonal matrix. It follows then easily that there exists no vector of linear functions of the observations with expectation $\delta = O_1 X\beta$ whose components have variances smaller than those of the components of $z_1 = O_1 y$. Hence, since the rows of the matrix O_1 provide a basis for the row space of X' it follows that the usual simple least squares linear estimators of all estimable functions $\lambda'\beta$ in the original model $y = X\beta + e$, $E(e) = 0$, $E(ee') = \sigma^2 V$, are also best linear unbiased estimators. Q.E.D.

(b) Necessity. Suppose that for every estimable function the simple least squares estimator is also a blue. Then every orthogonal matrix $O = (O_1', O_0')'$, with the $(n - r) \times n$ matrix O_0 such that $O_0 X = \varnothing$, leads to the corresponding transformed form

$$z = \begin{pmatrix} z_1 \\ z_0 \end{pmatrix} = Oy = \begin{pmatrix} O_1 X\beta \\ \varnothing \end{pmatrix} + Oe = \begin{pmatrix} \delta \\ \varnothing \end{pmatrix} + \eta$$

where z_1 is blue of δ, which can only be the case if Cov $(z_1, z_0) = \varnothing$, i.e., if the covariance of z is of the form

$$\text{Cov} \begin{pmatrix} z_1 \\ z_0 \end{pmatrix} = \begin{pmatrix} W_1 & \varnothing \\ \varnothing & W_0 \end{pmatrix},$$

where W_1 is of order $r \times r$. Let now R_1 be an orthogonal matrix such that $R_1 W_1 R'_1 = D_1$, and let R_0 be an orthogonal matrix such that $R_0 W_0 R_0' = D_0$, where both D_1 and D_0 are diagonal. Then the matrix

$$S = \begin{pmatrix} R_1 & \varnothing \\ \varnothing & R_0 \end{pmatrix} \begin{pmatrix} O_1 \\ O_0 \end{pmatrix} = \begin{pmatrix} R_1 O_1 \\ R_0 O_0 \end{pmatrix}$$

is orthogonal and

$$SVS' = D = \begin{pmatrix} D_1 & \varnothing \\ \varnothing & D_0 \end{pmatrix},$$

so that the columns of $S' = (S_1', S_0')$ form a set of eigenvectors of V. Further, the rows of R_1O_1 are all linear combinations of the rows of O_1 and the rows of R_0O_0 are all linear combinations of the rows of O_0, from which it follows that the columns of $S_1' = O_1'R_1'$ and of $S_0' = O_0'R_0'$ form respectively orthonormal bases of the column spaces of X and of its orthogonal complement. Q.E.D.

We remark that because in the preceding context certain elements of the matrix D_1 are permitted to have the value zero, it follows that some of the estimable functions may have blue's with variance zero, i.e., that certain estimable parametric functions may be subject to definite constraints. Further, when the value of a linear function of constraints is zero the corresponding linear function of components of z_1 may clearly be freely used in forming the algebraic expression for a linear estimator of any function without affecting its numerical value. The same also holds for any linear function of components of z_0 whose numerical value is always zero.

It seems appropriate to note at this point that in order for all the usual simple least squares estimators to be also blue's it is only needed that V has a subset of r eigenvectors as specified by Theorem 1. These eigenvectors need not be determined specifically. However, if we can in any way determine that V does have a subset of r eigenvectors forming a basis for the column space of X then we do know that all the usual simple least squares linear estimators are also blue's.

The eigenvector condition of Theorem 1 is not necessarily a simple one to apply in practice. We therefore now develop a number of other equivalent alternative conditions. It is well-known that the simple least squares estimator of $E(y) = X\beta$ is $X\tilde{\beta} = Py$, where $\tilde{\beta}$ is any vector such that $X'X\tilde{\beta} = X'y$ and P is the matrix orthogonal projection operator on the column space of X. One expression for P is $P = X(X'X)^*X'$, where $(X'X)^*$ is any matrix satisfying $(X'X)(X'X)^*X'X = X'X$. Any such matrix $(X'X)^*$ has been called a conditional inverse by R. C. Bose, since the middle 1950's, and a pseudo-inverse in the case of a singular matrix and a g-inverse in general by C. R. Rao, [12], [13], and [14]. It can be easily shown that the matrix P is symmetric, idempotent and that its column space is $\mathcal{C}(X)$, and thus that corresponding to the eigenvalue unity with multiplicity r, P has r eigenvectors in $\mathcal{C}(X)$ and corresponding to the eigenvalue zero with multiplicity $(n - r)$, P has $(n - r)$ eigenvectors in the orthogonal complement of $\mathcal{C}(X)$. Thus, under the conditions of Theorem 1 the complete set of n eigenvectors of V forms also a set of n eigenvectors of P, and hence also the matrices V and P are diagonalizable by the same orthogonal matrix. Equivalently then, by a known matrix theorem, the matrices V and P commute, or more conveniently, the product VP is symmetric. Since for standard situations, such as those of the common experimental designs, the vector Py is known, the matrix P can be obtained quickly, as can therefore also be a check for the symmetry of the matrix VP.

We also note that it follows immediately from the eigenvector condition that once the column space of X is specified the stated conditions on equality of estimators hold if and only if the class of the admissible V matrices is defined by the form

$$V = (O_1', O_0')D \begin{pmatrix} O_1 \\ O_0 \end{pmatrix},$$

where the matrix (O_1', O_0') is any orthogonal matrix with O_1' and O_0' as previously specified and D is any diagonal matrix with non-negative elements.

It is easy to verify that if a set of r eigenvectors of V belongs to $\mathcal{C}(X)$, and hence forms a basis for it, then for every vector a belonging to $\mathcal{C}(X)$ Va is also a vector in $\mathcal{C}(X)$, i.e., $\mathcal{C}(X)$ is an invariant subspace of V. Conversely, suppose that $\mathcal{C}(X)$ is an invariant subspace of V. Then, since for every vector a in $\mathcal{C}(X)$ the vector Va is also in $\mathcal{C}(X)$, it follows that for every set of matrices O_1' and O_0' defined as before $O_1 V O_0' = \emptyset$. Thus, by the argument of the necessity part of Theorem 1, the matrix V has a subset of r eigenvectors forming a basis for $\mathcal{C}(X)$. Further, it can be easily checked that $\mathcal{C}(X)$ is an invariant subspace of the matrix V if and only if $VX = XQ$, for some matrix Q. Also, since when V is non-singular, the eigenvectors of V are also eigenvectors of V^{-1}, it follows that then $VX = XQ$ if and only if $V^{-1}X = XR$, for some matrix R.

We note that when V is of the previously stated form

$$(O_1', O_0')D \begin{pmatrix} O_1 \\ O_0 \end{pmatrix}$$

then because the columns of the matrices O_1' and O_0' provide respectively bases for $\mathcal{C}(X)$ and $\mathcal{C}(\perp X)$, the orthogonal complement of $\mathcal{C}(X)$, it follows that there exist matrices P and Q such that $O_1' = XP$ and $O_0' = ZQ$, where $\mathcal{C}(Z) = \mathcal{C}(\perp X)$ so that $Z'X = 0$. Hence with D expressed in the form $L + k^2 I$, where L is a diagonal matrix, $L = \text{diag}(L_1, L_2)$ of the proper dimensions, the matrix V may be written as

$$O_1' D_1 O_1 + O_0' D_2 O_0 = XPL_1P'X' + ZQL_2Q'Z + k^2 I = X\Gamma X' + Z\Theta Z' + k^2 I$$

where $\Gamma = PL_1P'$ and $\Theta = QL_2Q'$.

With our σ^2 of equation (1) equal to unity this form for the covariance matrix is the one given by C. R. Rao in Lemma 5a of [15]. His condition for all the simple least squares estimators to be also blue's is there proved in the context of a known and non-singular covariance matrix Σ and a known full rank matrix X.

The condition is obviously related to those referred to in the introduction to the present paper and in particular is equivalent to the several conditions stated by Zyskind in [17], where incidentally each matrix on the right side of the equation in condition (5) should have an additional prime. Rao's actual statement of the condition and the proof of necessity (the proof of sufficiency, though perhaps implied, is not explicitly presented) are such as to allow the quantities Γ, Θ and k^2 to be unknown. His proof, using as it does the set of linear functions with expectations zero, does have a relationship with the known canonical form method which we use in this paper. Professor Rao adds that it is not essential in order for the condition to hold that the matrix X be of full rank. Also, he notes that further, the matrix Σ may be singular. It is interesting to realize that when Σ is singular the set of linear functions given by $Z'y$ does not in general provide the complete set of linear functions with zero expectation, as appears to be required in Professor Rao's proof. This incidental point demonstrates some of the differences which occur when one passes from a non-singular to a singular covariance matrix. Another point to note is that when the required conditions hold and Σ is singular, then all simple least squares estimators are also blue's but blue's may not be simple least squares estimators, since a particular estimable function may admit distinct blue's as functions of the observations, though it is a fact that all these functions have the same numerical value.

To see that C. R. Rao's condition in its most generalized version is not only implied by but also implies those which we have here derived we only need to note that for every vector w belonging to $\mathcal{C}(X)$ and for Σ any covariance matrix of the form $\Sigma = X\Gamma X' + Z\Theta Z' + k^2 I$ the vector $\Sigma w \,\varepsilon\, \mathcal{C}(X)$, since $Z'X = 0$, so that $\mathcal{C}(X)$ is an invariant subspace of the matrix Σ, which implies that a subset of r eigenvectors of Σ forms a basis for $\mathcal{C}(X)$, which is the starting point of the present development. In this reasoning the rank and possible singularity of Σ are irrelevant. Hence, Rao's condition, fully stated and proved in its most generalized form, like any of the conditions which we have stated in this section, might have been used as a basis for deriving interesting equivalent alternative statements.

With regard to the previous condition $VX = XQ$ it should be noted that Goldman and Zelen [5] have also demonstrated its sufficiency in the case where V is symmetric and non-negative and the matrix Q is non-singular. In effect, their restricted condition in addition to ensuring that a set of r eigenvectors of V forms a basis for $\mathcal{C}(X)$ stipulates also unnecessarily that the corresponding r eigenvalues be all different from zero, which implies in turn that the possible singularity of V induces no exact constraints on any estimable parametric functions (a condition which is not fulfilled, for instance, in the case of the complete randomized block design). For the case of Q non-singular these authors have also shown that $VX = XQ$ if and only if $V^+X = XQ^{-1}$, where V^+ denotes the Moore-Penrose [10], [11] generalized inverse of V. For any matrix A this inverse is defined as the (unique) matrix A^+ satisfying the joint conditions

(i) $AA^+A = A$,

(ii) $A^+AA^+ = A^+$,

(iii) $(AA^+)' = AA^+$,

(iv) $(A^+A)' = A^+A$.

(Note that the statement (i) alone restricts A^+ to be any conditional inverse of the matrix A.)

It is easy to show that for any real symmetric matrix A the Moore-Penrose generalized inverse is also symmetric and has identical sets of eigenvectors with A. Hence, $\mathcal{C}(X)$ is an invariant subspace of V if and only if it is an invariant subspace of V^+, and thus $VX = XQ$ for some matrix Q if and only if $V^+X = XR$ for some matrix R.

Because of the preceding arguments, Theorem 1 may be subsumed by the following statement on equivalent alternative conditions.

THEOREM 2. *Consider the model $y = X\beta + e$ where X is a known $n \times p$ matrix of rank r and the $n \times 1$ vector of errors, e, satisfies $E(e) = 0$, $E(ee') = \sigma^2 V$ where V is a symmetric non-negative matrix. Any one of the following conditions is both necessary and sufficient that the simple linear least squares estimator for every linearly estimable parametric function shall also be a corresponding best linear unbiased estimator.*

1. *A subset of r eigenvectors of V exists forming a basis for the column space of X.*

2. *A full rank reparametrization exists so that $E(y) = X\beta = W\theta$, where the columns of the $n \times r$ matrix W are mutually orthogonal eigenvectors of V.*

3. *The matrix V is expressible in the form*

$$V = (O_1', O_0')D\begin{pmatrix} O_1 \\ O_0 \end{pmatrix},$$

where the matrix $O' = (O_1', O_0')$ is orthogonal and O_1' is any orthonormal basis of the column space of X, O_0' is any orthonormal basis of the orthogonal complement of the column space of X, and D is any diagonal matrix with non-negative elements.

4. *The matrix V can be diagonalized by an orthogonal matrix specified as in* 3.

5. *If P denotes the orthogonal matrix projection operator on the column space of X then $VP = PV$, a relation which holds if and only if VP is symmetric.*

6. *A matrix Q exists satisfying the relation $VX = XQ$, and further, for V non-singular, a matrix R exists satisfying $V^{-1}X = XR$.*

7. *A matrix R exists such that $V^+X = XR$. (Note that when V is non-singular then $V^+ = V^{-1} = V^*$).*

8. *The column space of the matrix X is an invariant subspace of the matrix V,* i.e., *for every vector a belonging to $\mathcal{C}(X)$ the vector Va belongs also to $\mathcal{C}(X)$.*

It should be noted that the representation $y = X\beta + e$ expresses the model in merely one particular parametrization. The essential specification is that $E(y) \in \mathcal{C}(X)$, and clearly, under any of the conditions of Theorem 2 the simple least squares linear estimator of $E(y)$ is also its blue, regardless of the parametric representation employed.

An interesting illustration of the sufficiency part of condition 6, for example, occurs when the matrix V is expressible as $V = a_0 I + \sum_{i=1}^{k} a_i X_i Q_i$, where the matrices X_i, consist of disjoint sets of columns of X, for then clearly VX can be expressed as $VX = XQ$. The above form of V clearly subsumes the standard situation of $V = I$. Moreover, the matrix V is easily expressible in such a form in a number of concrete experimental design situations. For example, in the case of the randomized block design, under no measurement errors, mentioned earlier in the present paper, the covariance matrix of the observations, when $i = 1, 2, \cdots, b$; $k = 1, 2, \cdots, t$, as shown in Kempthorne, et al. [7], has the structure $\text{Cov}(y_{ik}, y_{i'k'}) = \delta_{ii'}\rho\sigma^2 + (1-\rho)\delta_{ii'}\delta_{kk'}\sigma^2$, where $\rho = -(t-1)^{-1}$ is the correlation of any two distinct observations in a block and δ_{ij} is the quantity defined by $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. It is easy to check in this case that if the observations are arranged lexicographically by blocks then the matrix V is given by

$$V = (1-\rho)I + \rho[\text{block diag } (J)],$$

where [block diag (J)] stands for a block diagonal matrix of order $bt \times bt$ whose $t \times t$ diagonal blocks consist of the $t \times t$ matrix J of elements unity everywhere. It is also easy to check that the matrix X, arising from $E(y_{ik}) = \mu + b_i + t_k$, has for its block part contribution an $n \times b$ matrix X_b such that $X_bX_b' =$ [block diag J]. Thus, by the stated consequence to condition 6, it follows at once that in the case of a randomized block design all estimable linear parametric functions are best estimated by their usual simple least squares estimators. Moreover, it is clear that when further additive measurement and technical errors with covariance structure k^2I are incorporated into the model, then the previous usual estimators still remain best.

3. On best linear estimators in the general case and on their subsets coinciding with simple least squares estimators. A convenient first stage canonical form for the general linear model (1) can be obtained as follows. Let the columns of O_0' form an orthonormal basis for the $(n-r)$ dimensional vector space orthogonal to the column space of X. Then as before $O_0X = \emptyset$, though in the present argument neither the orthogonality nor the normalization of the columns of O_0' are essential. Further, it is a known algebraic fact that, with $\mathcal{C}(\perp V)$ denoting the orthogonal complement of $\mathcal{C}(V)$ and with $k =$ dimension $[\mathcal{C}(O_0') \cap \mathcal{C}(\perp V)]$ the rank of O_0V is given by $n - r - k$. Choose now an $n \times r$ matrix U' of independent columns in such a way that the matrix $S' = (U', O_0')$ is nonsingular and satisfies the relation $UVO_0' = \emptyset$. The columns for the matrix U' can be obtained by choosing any set of r basis vectors which when augmented by a basis for that part of the column space of O_0' which is orthogonal to V forms a basis for the null space of the matrix O_0V, i.e., a basis for the space of vectors w such that $O_0Vw = \emptyset$. Consider now the operation of applying the matrix S to both sides of the model (1) to obtain

$$z = \begin{pmatrix} z_1 \\ z_0 \end{pmatrix} = \begin{pmatrix} U \\ O_0 \end{pmatrix} y = \begin{pmatrix} UX\beta \\ \emptyset \end{pmatrix} + \begin{pmatrix} U \\ O_0 \end{pmatrix} e = \begin{pmatrix} \delta \\ \emptyset \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_0 \end{pmatrix}. \tag{4}$$

We note that based on z_1 alone z_1 is a blue of the r independent estimable functions $\delta = UX\beta$. Further because the transformation used is non-singular and $\text{Cov}(z_1, z_0) = \sigma^2 UVO_0' = \emptyset$, it follows in fact that based on all the data z_1 is a blue of $\delta = UX\beta$, and thus that a blue of any estimable function $\lambda'\beta = \nu'UX\beta = \nu'\delta$ is $\nu'z_1 = \nu'Uy$, with variance $\sigma^2\nu'UVU'\nu$. It may be of interest to note that equivalent answers are also obtained by employing the simple normal equations pertaining to the model (4), and that since

$$(X'U', \emptyset)\begin{pmatrix} UX \\ \emptyset \end{pmatrix} = X'U'UX,$$

these normal equations expressed directly in terms of the vector parameter β, are $X'U'UX\beta = X'U'Uy$. In fact, a parametric function $\lambda'\beta$ is estimable if and only if it is expressible as a linear function of the left hand sides of these normal equations, say $\rho'X'U'UX\beta$, and when so expressible its blue is given by the same linear function, $\rho'X'U'Uy$, of the right hand sides of these equations. We further note that the possiblity of expressing the model in the form (4) implies that the original model form can always be transformed in such a way that a set of r eigenvectors of the new covariance matrix

$$\sigma^2\begin{pmatrix} U \\ O_0 \end{pmatrix} V(U', O_0') = \sigma^2\begin{pmatrix} UVU' & \emptyset \\ \emptyset & O_0VO_0' \end{pmatrix}$$

forms a basis for the column space of the $n \times p$ new matrix $\begin{pmatrix} UX \\ \emptyset \end{pmatrix}$ transformed from the original matrix X.

For certain purposes, such as for example the summarization of the data in an analysis of variance tabulation, it may be desirable to transform the model form further by an orthogonal matrix

$$\begin{pmatrix} R_1 & \emptyset \\ \emptyset & R_0 \end{pmatrix},$$

such that the matrices $R_1UVU'R_1'$ and $R_0O_0VO_0'R_0'$ are each diagonal, so as to achieve a diagonal matrix for the covariance of the vector

$$\begin{pmatrix} g_1 \\ g_0 \end{pmatrix} = \begin{pmatrix} R_1 & \emptyset \\ \emptyset & R_0 \end{pmatrix}\begin{pmatrix} U \\ O_0 \end{pmatrix} y$$

of the terminal canonical form. We note further that a reasonable assessment of the significance of an estimable parametric component $E(g_{1i})$ of $E(g_1)$ can be made if the variance of g_{1i} is known to be zero or if certain of the variances of the components of g_0 are known to be equal to the variance of g_{1i}.

We have seen that blue's of all estimable parametric functions $\lambda'\beta$ can be obtained by forming $\nu'z_1 = \nu'Uy = (U'\nu)'y$ where $\lambda' = \nu'UX$. Now the column space of U' is of dimension r, the column spaces of U' and O_0' are disjoint, and

$UVO_0' = \emptyset$. However, if VO_0' has dimension $(n - r - k)$ then there exists in fact a matrix W of order $(r + k) \times n$ and rank $(r + k)$ such that $WVO_0' = \emptyset$. Hence the row space of U is a subspace of the row space of W. Denote the row space of any matrix A by $\mathcal{R}(A)$. Let $\lambda'\beta$ be an arbitrary estimable parametric function and let $u'y$, with $u' \varepsilon \mathcal{R}(U)$ be some unbiased estimator of it. Then every unbiased estimator $s'y$ of $\lambda'\beta$, with $s' \varepsilon \mathcal{R}(U)$, is such that $(s' - u')y$ is zero with variance zero. We note that in the case where V is non-singular the only permissible choice of s' is $s' = u'$. The complete set of linear unbiased estimators of $\lambda'\beta$ is given by the set of functions of the form $w'y = u'y + (s' - u')y + e'y$, where the permissible values of s' are as before and the vectors e' range over the complete set of vectors of $\mathcal{R}(O_0)$. So, since $u'Ve = \emptyset$, it follows that the variance of $w'y$ is $\text{Var}(w'y) = \text{Var}(u'y) + \text{Var}(e'y)$. Hence $w'y$ is a blue if and only if $\text{Var}(e'y) = \sigma^2 e'Ve = 0$ which, with $\sigma^2 > 0$ and V non-negative and symmetric, is the case if and only if $Ve = 0$, and then $e'VO'_0 = 0$ so that $e' \varepsilon \mathcal{R}(W)$. Under this condition the vector $w'(= u' + (s' - u') + e')$ belongs to $\mathcal{R}(W)$. Thus, a linear function $w'y$ is a blue if and only if $w' \varepsilon \mathcal{R}(W)$.

It follows as an immediate consequence that linear functions of blue's are also blue's of their expectations. Further, an immediately equivalent necessary and sufficient condition for $w'y$ to be a blue of its expectation $E(w'y)$ is that the vector $Vw \varepsilon \mathcal{C}(X)$.

We now enquire into the conditions for equality of subsets of simple least squares estimators of estimable functions, under the original model $y = X\beta + e$, with their corresponding best linear unbiased estimators. It is a well-known fact that all the simple least squares estimators are of the form $w'y$, where w' is some vector belonging to the row space of X'. Hence, the set of all simple least squares estimators is given by the set of functions of observations $w'y$, where the vectors w are such that $O_0w = \emptyset$. On the other hand we have just stated that the complete set of all the blue's consists of vectors w such that $O_0Vw = \emptyset$. Thus, the set of vectors w leading to both the blue's and simple least squares estimators consists of those vectors w satisfying jointly the sets of equations

$$(5) \qquad O_0w = \emptyset; \qquad O_0Vw = \emptyset.$$

Thus, the vectors we seek are vectors belonging to the null space of the matrix $\begin{pmatrix} O_0 \\ O_0V \end{pmatrix}$. The rank of this null space is clearly between zero and r. Clearly also, if $2r > n$ the system (5) has non-trivial solutions.

In practice, the usual coefficient vectors w of simple least squares estimators, i.e., the vectors w satisfying $O_0w = \emptyset$, will often be well-known in advance for standard situations. It then only remains, in order to ascertain the subsets of these vectors w giving rise to blue's, to isolate those of them which satisfy also $O_0Vw = \emptyset$. Further, if we cannot easily obtain the linearly independent rows of O_0 but have the ordinary least squares solutions to the normal equations, so that we have in fact $X\hat{\beta} = X(X'X)^*X'y = Py$ (where $(X'X)^*$ is any matrix such that $X'X = (X'X)(X'X)^*(X'X)$ and P is the orthogonal matrix projection operator on the column space of X), then the matrix O_0 in the equations $O_0w = \emptyset$ and $O_0Vw = \emptyset$ may be replaced by the conveniently available matrix $(I - P)$ with resulting equations $(I - P)w = \emptyset$ and $(I - P)Vw = \emptyset$.

A sufficient condition on eigenvectors may also be easily derived. Thus if w is a vector belonging to the column space of X and is also an eigenvector of V then $Vw = kw \varepsilon \mathcal{C}(X)$, so that $w'y$ is a blue. Similarly if w is a vector belonging to the column space of X and is a linear combination of eigenvectors of V belonging to that subspace then also $Vw \varepsilon \mathcal{C}(X)$ and $w'y$ is therefore a blue.

We can easily see, in fact, that the essential condition, i.e., a necessary and sufficient condition, on any unbiased estimator $w'y$ to be both a blue and a simple least squares estimator is that both w and the product Vw be vectors belonging to the column space of X.

The conclusions of the latter part of the section may be summarized and extended somewhat as follows, where the model under consideration is the one given by equation (1) and as before O_0' is any matrix whose columns form an orthonormal basis for the orthogonal complement of $\mathcal{C}(X)$ and P denotes the orthogonal matrix projection operator on $\mathcal{C}(X)$.

THEOREM 3. *A linear function $w'y$ is a best linear unbiased estimator of its expectation $E(w'y)$ if and only if the vector w belongs to the null space of the matrix O_0V*, i.e., *if and only if w is such that $O_0Vw = \emptyset$, which holds if and only if $w'y$ is uncorrelated with every linear function of the form $\rho'O_0y$. Equivalently, $w'y$ is a* blue *if and only if the vector Vw belongs to $\mathcal{C}(X)$, which occurs if and only if $(I - P)V w = \emptyset$.*

We note at this point that the condition $Vw \varepsilon \mathcal{C}(X)$ is a direct generalization of the basic and well-known fact (see, for example the book by Scheffé [16]) that in the case where $V = I$ the function $w'y$ is the blue of its expectation if and only if $w \varepsilon \mathcal{C}(X)$.

In the case where the matrix V is non-singular it is easy to obtain from Theorem 3 the standard result that the blue's $w'y$ are all of the form $\gamma'X'V^{-1}y$, where γ ranges over the space of $p \times 1$ vectors, since then $Vw = VV^{-1}X\gamma = X\gamma$, a vector belonging to $\mathcal{C}(X)$, and since the dimension of the space spanned by the vectors $V^{-1}X\gamma$ is r.

THEOREM 4. *A linear function $w'y$ is both the simple least squares estimator and a best linear unbiased estimator of its expectation $E(w'y)$ if and only if w is such that $O_0w = \emptyset$ and $O_0Vw = \emptyset$, or equivalently such that $(I - P)w = \emptyset$ and $(I - P)Vw = \emptyset$, or equivalently if and only if both vectors w and Vw belong to $\mathcal{C}(X)$.*

COROLLARY 4.1. *If r, the rank of X, is such that $2r > n$ then some non-zero simple least squares estimators $w'y$ of $E(w'y)$ are also* blue's.

COROLLARY 4.2. *A simple least squares estimator $w'y$ of $E(w'y)$ is also a* blue *if and only if Vw belongs to $\mathcal{C}(X)$, which is the case if and only if $O_0Vw = \emptyset$, or equivalently if and only if $(I - P)Vw = \emptyset$. Further, the complete set of simple* *least squares estimators forms a set of* blue's *of all estimable linear parametric functions if and only if* $\mathcal{C}(X)$ *is an invariant subspace of* V.

Thus, as shown by Corollary 4.2, the independent general approach of the present section leads us once more into statements such as those of Theorem 2, concerning conditions under which all simple least squares linear estimators are also blue's.

COROLLARY 4.3. *If* w *is a vector belonging to* $\mathcal{C}(X)$ *and is a linear combination of eigenvectors of* V *belonging to* $\mathcal{C}(X)$ *then* $w'y$ *is both simple least squares estimator and* blue *of its expectation* $E(w'y)$.

4. Parametric constraints as an instance of a singular covariance matrix and of canonical form techniques. Consider the case of a standard linear model $y = X\beta + e$, $E(e) = \varnothing$, $E(ee') = \sigma^2 I$, for which in addition certain parametric constraints are known to hold. For example, the first three components of β, say β_1, β_2, β_3, might be known to be the measures in radians of the three angles of a particular triangle, so that a known condition would be $\beta_1 + \beta_2 + \beta_3 = \pi$. In general, suppose the known conditions are of the form $\Lambda\beta = c$, a known vector of constants, where Λ is a known matrix of order $k \times p$ and of rank m.

The present problem may be approached by several direct methods of which the treatments given by Zyskind, et al. [18], and the one in the recent book by C. R. Rao [15], are two useful examples. In the present paper we attack the problem simply, however, by using canonical forms and by viewing the overall covariance matrix of observations and of the conditions as an instance of a special singular covariance matrix.

The complete model for the given situation may be viewed as

$$g = \begin{pmatrix} y \\ c \end{pmatrix} = \begin{pmatrix} X \\ \Lambda \end{pmatrix} \beta + \begin{pmatrix} e \\ \eta \end{pmatrix} \tag{6}$$

where

$$\mathrm{Cov}(g) = \mathrm{Cov}\begin{pmatrix} y \\ c \end{pmatrix} = \mathrm{Cov}\begin{pmatrix} e \\ \eta \end{pmatrix} = \sigma^2 \begin{pmatrix} I & \varnothing \\ \varnothing & \varnothing \end{pmatrix},$$

a particular singular matrix.

We first treat the case where the row spaces of the matrices X and Λ are disjoint, i.e., where the components of $\Lambda\beta$ are non-estimable, as is also every linear function of these components, with the model $y = X\beta + e$ alone. Further, without loss of generality for the present situation, we may take the rows of Λ to be linearly independent. Then

$$\mathrm{rank}\begin{pmatrix} X \\ \Lambda \end{pmatrix} = \mathrm{rank}\, X + \mathrm{rank}\, \Lambda = r + k = \mathrm{rank}\begin{pmatrix} X \\ \varnothing \end{pmatrix} + \mathrm{rank}\begin{pmatrix} \varnothing \\ \Lambda \end{pmatrix},$$

so the column spaces of $\begin{pmatrix} X \\ \Lambda \end{pmatrix}$ and $\begin{pmatrix} \varnothing \\ \Lambda \end{pmatrix}$ are also disjoint. It can be seen at once that if the vectors $\nu_1, \nu_2, \cdots, \nu_r$ form any orthonormal basis for the column space of

X then those vectors are eigenvectors of the identity matrix I_n, and further that the vectors $(\nu_1', \emptyset')', (\nu_2', \emptyset')', \cdots, (\nu_r', \emptyset')'$ of order $(n + k) \times 1$ are eigenvectors of the matrix $\sigma^2\begin{pmatrix} I & \emptyset \\ \emptyset & \emptyset \end{pmatrix}$. Moreover, if a set of k orthogonal vectors of order $(n + k) \times 1$ is of the form $(\emptyset', u_1')', (\emptyset', u_2')', \cdots, (\emptyset', u_k')'$, where the vectors $u_1, u_2, \cdots, u_k$ form an orthogonal basis for the column space of the matrix Λ, then the totality of the vectors

$$(v_1', \emptyset')', (\nu_2', \emptyset')', \cdots, (\nu_r', \emptyset')', (\emptyset', u_1')', (\emptyset', u_2')', \cdots, (\emptyset', u_k')'$$

forms an orthogonal basis for the column space of $\begin{pmatrix} X \\ \Lambda \end{pmatrix}$ and forms also a set of $(r + k)$ orthogonal eigenvectors of the matrix $\sigma^2 \begin{pmatrix} I & \emptyset \\ \emptyset & \emptyset \end{pmatrix}$. Thus, by Theorem 1 of Section 2, all of the simple least squares estimators of estimable linear parametric functions under the model 6, with the restrictions specialized as noted, are also corresponding blue's. Hence blue's can be obtained by direct use of the simple normal equations which in the present case are

$$(7) \qquad (X'X + \Lambda'\Lambda)\beta = X'y + \Lambda'c.$$

We now briefly justify the fact, which is part of the folklore of the subject, that when all functions of the form $\rho'\Lambda\beta$ are non-estimable under the model $y = X\beta + e$ alone, then the blue's of all functions estimable under that model alone are also blue's of these estimable functions under the model (6), incorporating non-estimable constraints of the type stated. To see this we note that if $\hat{\beta}$ is any vector satisfying the normal equations (7) then because the column spaces of X' and Λ' are disjoint it must be true that

$$\Lambda'\Lambda\hat{\beta} - \Lambda'c = X'y - X'X\hat{\beta} = \emptyset, \qquad \text{so that} \qquad X'X\hat{\beta} = X'y.$$

Thus, any solution to the system (7) is also a solution to the simple normal equations $X'X\beta = X'y$, obtainable under the model $y = X\beta + e$ with no restrictions. Further, since the blue's of estimable functions under this model are invariant for all vectors $\tilde{\beta}$ such that $X'X\tilde{\beta} = X'y$, it follows that, when the model consists in addition of known restrictions on non-estimable functions, all linear functions of which are also non-estimable, then any solution to the equations $X'X\beta = X'y$ alone leads directly to blue's of functions of the form $\lambda'\beta = a'X\beta$ under the stated augmented model.

The argument just offered for the case where the $n \times 1$ vector y has covariance matrix $\sigma^2 I$ applies mutatis mutandis for the situation where the covariance of y is $\sigma^2 V$ and a set of r eigenvectors of V forms a basis for the column space of X. We may thus state the following:

THEOREM 5. *If in the model* (1) $y = X\beta + e$, $E(e) = \emptyset$, $E(ee') = \sigma^2 V$, *the relation between* V *and* X *is such that all simple linear least squares estimators of estimable functions are also* blue's *then under addition to the model of known con-*

sistent restrictions on non-estimable functions, all linear functions of which are also non-estimable, the previous simple least squares estimators remain blue's *in the joint augmented model.*

We note that, as largely demonstrated in [7], the derived mathematical models under the additivity assumption for the standard randomized experimental designs, such as the completely randomized, randomized complete blocks. Latin square, and split-plot designs, exhibit all of the conditions stated in Theorem 5. Thus, for these designs the usual linear simple least squares estimators based on the standard simplified infinite model are also best.

We next employ a canonical reduction to deal with the case where the parametric constraints $\Lambda\beta = c$ are all estimable. In this case let $A'y$ be the blue of $\Lambda\beta$ when there are no restrictions in the model and the covariance matrix is $\sigma^2 I_n$. Then $E(A'y) = A'X\beta = \Lambda\beta$ for all β, so that $A'X = \Lambda$. It is well-known that the columns of the $n \times k$ matrix A of rank m generate an m-dimensional subspace of the column space of X. Let the orthonormal columns of an $n \times m$ matrix O_{11}' form a basis for that subspace, and let O_{12}' be a matrix such that the columns of the matrix $O_1' = (O_{11}', O_{12}')$ form an orthonormal basis of the column space of X. Then, since O_{11}' and A generate the same column spaces, there exists a matrix B such that $O_{11} = BA'$ and hence $O_{11}X\beta$ satisfies $O_{11}X\beta = BA'X\beta = B\Lambda\beta = Bc$. Consider now the orthogonal matrix $O' = (O_{11}', O_{12}', O_0')$, where $O_0X = \emptyset$, and the canonical form

$$z = \begin{pmatrix} z_{11} \\ z_{12} \\ z_0 \end{pmatrix} = \begin{pmatrix} O_{11} \\ O_{12} \\ O_0 \end{pmatrix} y = \begin{pmatrix} O_{11}X \\ O_{12}X \\ \emptyset \end{pmatrix} \beta + Oe = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \emptyset \end{pmatrix} + \eta.$$

Then $\delta_1 = O_{11}X\beta = Bc$ is a known vector, so that there would be no point in using z to estimate it. The parametric components to be estimated are those of the vector $\delta_2 = O_{12}X\beta$ and, since $\mathrm{Cov}(z) = \sigma^2 I$, obviously the blue of δ_2 is z_{12}. The blue of any estimable function $\nu'\delta[= \nu_1'\delta_1 + \nu_2'\delta_2 = (\nu_1'O_{11}X + \nu_2'O_{12}X)\beta = \lambda'\beta]$ is clearly $\nu_1'\delta_1 + \nu_2'z_{12}$ with variance $\sigma^2\nu_2'\nu_2$. It is easy to verify that the blue $\nu_1'\delta_1 + \nu_2'z_{12}$ is the estimator that would also be obtained by minimizing the ordinary sum of squares for deviations from the expectations $E(y_i)$ subject to the stated constraints, and that the addition of further constraints on non-estimable functions, all linear functions of which are also non-estimable, does not affect the solutions presented.

It now only remains to express the preceding results in terms of the originally specified model. Let $b'y$ be the blue of $\lambda'\beta$ in the model with no restrictions. Then $b'y = \nu_1'z_{11} + \nu_2'z_{12} = (\nu_1'O_{11} + \nu_2'O_{12})y$ so that $b' = \nu_1'O_{11} + \nu_2'O_{12}$, and since the column spaces of O_{11}' and O_{12}' are mutually orthogonal it follows that $O_{11}'\nu_1$ is the orthogonal projection of b on the column space of O_{11}'. Thus, $O_{11}'\nu_1$ is expressible in the form $O_{11}'\nu_1 = P_Ab$, where P_A is the symmetric and idempotent matrix projection operator on the column space of A, where A is the matrix of coefficients of blue's of $\Lambda\beta$ under no restrictions. We therefore have blue of $\lambda'\beta =$

$$\begin{aligned}\text{blue of } \nu'\delta = (\widehat{\nu'\delta}) &= \nu_1'\delta_1 + \nu_2'z_{12} = \nu_1'z_{11} + \nu_2'z_{12} \\ &\quad -\nu_1'(z_{11} - \delta_1) = b'y - \nu_1'O_{11}(y - X\beta_0), \\ \text{(8)} \qquad &\qquad\qquad \text{where } \beta_0 \text{ is any vector such that } \Lambda\beta_0 = c, \\ &= b'y - b'P_A(y - X\beta_0).\end{aligned}$$

Further, the variance of the blue may be conveniently expressed as

$$\text{(9)} \qquad \begin{aligned}\text{Var}(\widehat{\nu'\delta}) &= \sigma^2\nu_2'\nu_2 = \sigma^2(\nu'\nu - \nu_1'\nu_1) \\ &= \sigma^2 b'b - \sigma^2 b'P_A b.\end{aligned}$$

The form of equations (8) and (9) points up the possibility of viewing the problem of the linear model with restrictions as one which decomposes into two simpler stages. In one stage one obtains results for the linear model without restrictions, and in the other, one computes the additional matrix orthogonal projection operator $P_A (= A(A'A)^*A')$ which as shown in [18], may be obtained by use of normal-type equations with coefficient matrix $A'A$. If the order of $A'A$ is small the additional matrix P_A may be fairly quickly obtained.

In certain cases of interest the results (8) and (9) quickly lead to the following special conclusions. When $b'y$ is the blue in the model $y = X\beta + e$ with no restrictions on the parameters, of a linear function of the components of $\Lambda\beta$ then b is a vector belonging to the column space of A, so that $P_A b = b$, and b is expressible as $b = A\rho$ for some ρ, and hence the blue of $\lambda'\beta$ is given by the formula (8) as

$$b'y - b'P_A(y - X\beta_0) = b'X\beta_0 = \rho'A'X\beta_0 = \rho'\Lambda\beta_0 = \rho'c,$$

a constant with variance zero.

Again if $b'y$ is a blue of a parametric function $\lambda'\beta$ under the model with no restrictions, and if $b'y$ is uncorrelated with every linear function of the form $\gamma'A'y$ then the vectors b and $A\gamma$ are orthogonal for every $k \times 1$ vector γ, and hence $P_A b = \emptyset$. Under the model subject to the stated restrictions the formulae derived yield at once $b'y$ with variance $\sigma^2 b'b$ for the blue and its variance respectively of the parametric function $\lambda'\beta$.

It may be of interest to add that the present example furnishes a simple illustration of the fact that when the covariance matrix is singular one may not validly construct a set of normal equations by direct use of the Moore-Penrose generalized inverse, [10], [11], in direct analogy with the well-known equations

$$\begin{pmatrix} X \\ \Lambda \end{pmatrix}' V^{-1} \begin{pmatrix} X \\ \Lambda \end{pmatrix} \beta = \begin{pmatrix} X \\ \Lambda \end{pmatrix}' V^{-1} \begin{pmatrix} y \\ c \end{pmatrix}$$

which would apply if the matrix V were non-singular. It is easy to verify that for

$$V = \begin{pmatrix} I_n & \emptyset \\ \emptyset & \emptyset \end{pmatrix}, \qquad V^+ = \begin{pmatrix} I_n & \emptyset \\ \emptyset & \emptyset \end{pmatrix},$$

and hence that the equations

$$\begin{pmatrix} X \\ \Lambda \end{pmatrix}' V^+ \begin{pmatrix} X \\ \Lambda \end{pmatrix} \beta = \begin{pmatrix} X \\ \Lambda \end{pmatrix}' V^+ \begin{pmatrix} y \\ c \end{pmatrix}$$

become $X'X\beta = X'y$, which are not the correct equations for obtaining blue's of estimable functions under the model in which certain estimable functions are known to satisfy specified constraints. The matter of how general normal equations may be validly constructed by using subclasses of the class of conditional inverses has been recently investigated by Zyskind and Martin [19].

We finally point out, as was also noted by Goldman and Zelen, [5], that the treatment of the general problem of a linear model with essentially known singular covariance structure may be reduced to that of the standard linear model with specified restrictions on certain estimable functions. For, if the covariance matrix is $\sigma^2 V$, where V is known and is of rank s, then we can find a non-singular matrix

$$T = \begin{pmatrix} T_1 \\ T_0 \end{pmatrix}$$

such that

$$TVT' = \left(\begin{array}{c|c} I_s & \varnothing \\ \hline \varnothing & \varnothing \end{array}\right),$$

with $T_1VT_1' = I_s$. Thus, when the transformation T is applied to the observations y then

$$z = \begin{pmatrix} z_1 \\ z_0 \end{pmatrix} = Ty = \begin{pmatrix} T_1 y \\ T_0 y \end{pmatrix}$$

where $z_1 = T_1X\beta + T_1e$, with $\text{Var}(z_1) = \sigma^2 I_s$, and $T_0X\beta = z_0$, a known constant vector once the data have been obtained, specifying restrictions on estimable parametric functions.

5. Acknowledgment. I wish to thank Professor C. R. Rao for permitting me to see a copy of the final version of his manuscript [15] before its publication.

REFERENCES

[1] Anderson, T. W. (1948). On the theory of testing serial correlation. *Skand. Aktuarietidsk.* **31** 88–116.

[2] Box, G. E. P. and Muller, M. E. (1958). Randomization and least squares estimates. Abstract published in the summaries of papers presented at the 118th Annual Meeting of the A.S.A. by the American Statistical Association. pg. 7.

[3] Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least squares regression. I. *Biometrika* **37** 409–428.

[4] Fisher, R. A. (1935). *The Design of Experiments.* Oliver and Boyd, Edinburgh.

[5] Goldman, A. J. and Zelen, M. (1964). Weak generalized inverses and minimum variance linear unbiased estimation. *J. Res. Nat. Bur. Stand.* **68B** 151–172.

[6] Kempthorne, Oscar. (1952). *The Design and Analysis of Experiments.* Wiley, New York.

[7] KEMPTHORNE, O., ZYSKIND, G., ADDELMAN, S., THROCKMORTON, T. N., and WHITE, R. F. (1961). Analysis of variance procedures. Aeronautical Research Laboratory Technical Report 149, Wright-Patterson Air Force Base, Ohio.
[8] MAGNESS, T. A. and MCGUIRE, J. B. (1962). Comparison of least squares and minimum variance estimates of regression parameters. *Ann. Math. Statist.* **33** 462–470.
[9] MULLER, M. E. and WATSON, G. S. (1959). Randomization and linear least squares estimation. Statistical Techniques Research Group Technical Report No. 32, Princeton University.
[10] MOORE, E. H. (1935). General Analysis I. *Mem. Am. Phil. Soc.* **1** 197.
[11] PENROSE, R. (1955). A generalized inverse for matrices. *Proc. Cambridge Phil. Soc.* **51** 406–413.
[12] RAO, C. RADHAKRISHNA. (1955). Analysis of dispersion for multiply classified data with unequal numbers in cells. *Sankhyā* **15** 253–280.
[13] RAO, C. RADHAKRISHNA. (1962). A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. *J. Roy. Stat. Soc. Ser. B* **24** 152–158.
[14] RAO, C. RADHAKRISHNA. (1965). *Linear Statistical Inference and Its Implications.* Wiley, New York.
[15] RAO, C. RADHAKRISHNA. (196?). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. To appear in *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* Univ. of California Press.
[16] SCHEFFÉ, H. (1959). *The Analysis of Variance.* Wiley, New York.
[17] ZYSKIND, G. (1962). On conditions for equality of best and simple linear least squares estimators. (Abstract). *Ann. Math. Statist.* **33** 1502–1503.
[18] ZYSKIND, G., KEMPTHORNE, O., WHITE, R. F., DAYHOFF, E. E., DOERFLER, T. E. (1964). Research on Analysis of Variance and Related Topics. Aerospace Research Laboratories, Technical Report 64-193, Wright-Patterson Air Force Base, Ohio.
[19] ZYSKIND, G. and MARTIN, FRANK B. (1966). A general Gauss-Markoff theorem for linear models with arbitrary non-negative covariance structure. (Submitted to *Ann. Math. Statist.*).