SOME CONSIDERATIONS ABOUT THE EFFECT OF REDUNDANCIES AND RESTRICTIONS IN THE GENERAL LINEAR REGRESSION MODEL

by

Gary G. Koch
University of North Carolina

Institute of Statistics Mimeo Series No. 459

February 1966

This research was supported by the National Institutes of Health Institute of General Medical Sciences Grant No. GM-12868-02.

The following paper represents a further study of topics considered in Statistics 150 and Statistics 254 at the University of North Carolina at Chapel Hill and Statistics 691 at North Carolina State University at Raleigh. Its primary purpose is to investigate the effect of restrictions and redundancies in linear models. An additional topic considered is the effect of a variance-covariance matrix which is different from the identity matrix and may be singular.

In the paper, Section 2 is based upon the notes of R. C. Bose for Statistics 150; Section 6, upon the notes of R. C. Bose for Statistics 254; Section 4, upon the notes of H. L. Lucas for Statistics 691. Results in the other sections are discussed also in Rao [1965].

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
Chapel Hill, N. C.

ACKNOWLEDGMENTS

The author is indebted to R. C. Bose for providing him with the tools with which to deal with linear models and to H. L. Lucas for stimulating his interest in linear models subject to restrictions.

Some Considerations About the Effect of Redundancies and Restrictions in the General Linear Regression Model

1. Introduction. In this paper, we consider the ways in which the presence of redundancies in the design matrix and/or restrictions on the parameters of a general linear regression model influence the class of estimable functions of the parameters. Depending on the model, this class may be partitioned into disjoint subsets to which correspond disjoint uncorrelated sets of linear functions of the observations. The sets of these partitions play a fundamental role in the construction of best linear unbiased estimators (b.l.u.e.) of estimable linear functions of the parameters and in the construction of tests of linear hypotheses involving estimable functions. Finally, the theory developed provides a means of dealing with linear models in which the variance-covariance matrix of the observations may be singular.

It should be noted that most of the results given here are already well-known (see Lucas [1965], Rao [1945b, 1946, 1965], Plackett [1960]). However, the paper is of interest since its principal purpose is an interpretation of these results.

2. The non-restricted model. Let $y_1, y_2, \ldots, y_n$ be uncorrelated random variables with a common unknown variance σ^2 and with expected values which are linear functions with known coefficients of m unknown parameters $p_1, p_2, \ldots, p_m$ as follows

(1)
$$\begin{aligned}
\mathcal{E}(y_1) &= a_{11}p_1 + a_{21}p_2 + \cdots + a_{m1}p_m \\
\mathcal{E}(y_2) &= a_{12}p_1 + a_{22}p_2 + \cdots + a_{m2}p_m \\
\cdots & \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\
\mathcal{E}(y_n) &= a_{1n}p_1 + a_{2n}p_2 + \cdots + a_{mn}p_m
\end{aligned}$$

where the a_{ij}'s are known constants. Let $\underline{y} = \underline{y}(n \times 1)$ denote the random column vector

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}$$

and let $\underline{p} = \underline{p}(m \times 1)$ denote the column vector of unknown parameters

$$\underline{p} = \begin{bmatrix} p_1 \\ p_2 \\ \cdots \\ p_m \end{bmatrix} .$$

Also let $A = A(m \times n)$ denote the matrix

(2) $$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} .$$

The model can now be compactly written as

(3) $$\mathcal{E}(\underline{y}) = A'\underline{p},\ \mathrm{Var}(\underline{y}) = \mathcal{E}\{(\underline{y} - \mathcal{E}(\underline{y}))\ (\underline{y} - \mathcal{E}(\underline{y}))'\} = I_n\sigma^2;$$

the matrix A' is called the design matrix and is assumed to have rank $n_o \leq m < n$. The following approach to linear estimation in the model specified by (3) is due to Bose [1950].

Definition: A linear function of $p_1, p_2, \ldots, p_m$ will be said to be (linearly) estimable with respect to the model (3) if there exists a linear function of $y_1, y_2, \ldots, y_n$ which is an unbiased estimator of it. i.e., if $\underline{\ell}' = (\ell_1, \ell_2, \ldots, \ell_p)$, $\underline{\ell}'\underline{p}$ is estimable if there exists $\underline{c}' = (c_1, c_2, \ldots, c_n)$

such that $\underline{c}'\ \underline{y}$ is an unbiased estimator of it.

Hence, we see that $\underline{\ell}'\ \underline{p}$ will be estimable if and only if there exists $\underline{c}'$ such that $\mathcal{E}\{\underline{c}'\ \underline{y}\} = \underline{c}'\ A'\ \underline{p} = \underline{\ell}'\ \underline{p}$. As a result, we see that $\underline{\ell}'\ \underline{p}$ will be estimable if and only if the equations $A\underline{c} = \underline{\ell}$ have a solution which will be the case if and only if $\underline{\ell}'$ lies in the vector space spanned by the rows of the design matrix A'. As a result, we see that every estimable function of the parameters must be of the form $\underline{c}'A'\underline{p} = \underline{c}'\mathcal{E}(\underline{y})$.

Definition: A linear function of $y_1, y_2, \ldots, y_n$ is said to belong to error if its expectation vanishes independently of the parameters; i.e., if $\underline{e}' = (e_1, e_2, \ldots, e_n)$ and $\underline{e}'\underline{y}$ belongs to error, then $\mathcal{E}(\underline{e}'\underline{y}) = \underline{e}'A'\underline{p} = 0$ for all values of $\underline{p}$.

Thus, we see that $\underline{e}'\underline{y}$ belongs to error if and only if $A\underline{e} = \underline{0}$; i.e., if and only if $\underline{e}'$ is orthogonal to the rows of A. With this in mind, we call the set of (1 x n) vectors orthogonal to the rows of A the error space and the corresponding set of linear functions of $y_1, y_2, \ldots, y_n$ the error set. Since the rank of A is n_o, there exist $n_e = n - n_o$ independent (1 x n) vectors orthogonal to the rows of A; thus the rank of the error space is n_e.

Definition: A linear function of $y_1, y_2, \ldots, y_n$ is said to be an estimating function if it is uncorrelated with all linear functions belonging to error; i.e., if $\underline{c}' = \underline{c}'(1 \times n)$ and $\underline{c}'\ \underline{y}$ is an estimating function, then $\text{Cov}\ (\underline{c}'\underline{y}, \underline{e}'\underline{y}) = \underline{c}'\underline{e}\sigma^2 = 0$ for every $\underline{e}'\underline{y}$ belonging to error.

The set of (1 x n) vectors which are the coefficient vectors of estimating functions is called the estimation space and the corresponding set of linear functions of $y_1, y_2, \ldots, y_n$ is called the estimation set. For the model specified by (3), we see that $\underline{c}'\ \underline{y}$ is an estimating function if $\underline{c}'$ is orthogonal to all vectors in the error space which will be the case if and only if $\underline{c}'$ lies in the vector space spanned by the rows of A. Thus, in this case, the rows of A span the estimation space. Finally, we note that the estimation space and the error space are disjoint (i.e., the only vector they have in common is the zero vector) since vectors which are orthogonal to one another are necessarily linearly independent.

We now quote some useful lemmas, due to Bose [1950], about vectors, matrices, and linear equations:

Lemma 1: Given a matrix $A = A(m \times n)$, the matrix $A^g = A^g(n \times m)$ is defined to be a conditional (generalized) inverse of A if $AA^gA = A$. For every matrix A, there exists a conditional inverse A^g.

Lemma 2: If the equations $A\underline{u} = \underline{V}$ are consistent, then $\underline{u}_o = A^g\underline{v}$ is a solution if A^g is any conditional inverse of A. Conversely, if $\underline{u}_o = A^*\underline{v}$ is a solution of the equations $A\underline{u} = \underline{v}$ for every $\underline{v}$ for which they are consistent, then A^* is a conditional inverse of A.

Lemma 3: The equations $AA'\underline{u} = A\underline{v}$ for determining $\underline{u}$ have a solution and $A'\underline{u}$ is uniquely determined.

Lemma 4: If $A = A(m \times n)$, we can find $X = X(m \times n)$ such that $AA'X = A$; in fact X is given by $X = (AA')^gA$; moreover, $A'X = A'(AA')^gA$ is uniquely determined, symmetric, and idempotent.

From the above lemmas and the concepts of estimability, error space, estimation space, the following fundamental theorem of linear estimation in the model specified by (3) may be proven.

Theorem 1: If $\underline{\ell}'\underline{p}$ is any estimable function of the parameters $p_1, p_2, \ldots, p_m$, then

(i) there exists a unique linear function $\underline{c}'\underline{y}$ of the random variables $y_1, y_2, \ldots, y_n$ such that $\underline{c}'$ belongs to the estimation space and

$\mathcal{E}(\underline{c}'\underline{y}) = \underline{\ell}'\underline{p}$; also the variance of $\underline{c}'\underline{y}$ is smaller than the variance of any other linear unbiased estimator of $\underline{\ell}'\underline{p}$;

(ii) the best estimator of $\underline{\ell}'\underline{p}$ may be written as either

(a) $\underline{q}'A\underline{y}$ where $\underline{q}'AA' = \underline{\ell}'$ or

(b) $\underline{\ell}'\hat{\underline{p}}$ where $\hat{\underline{p}}$ is a value of $\underline{p}$ which minimizes the sum of squares $S^2 = (\underline{y} - A'\underline{p})'(\underline{y} - A'\underline{p})$; i.e., $\hat{\underline{p}}$ satisfies the equations $AA'\underline{p} = A\underline{y}$;

(iii) the variance of the best estimator of $\underline{\ell}'\underline{p}$ is uniquely determined and given by $\underline{\ell}'\,(AA')^{g}\underline{\ell}$.

For a proof of Theorem 1, the reader is referred to Bose [1950]. However, certain basic parts of the proof will be duplicated in the derivation of the results of Sections 3 and 4.

Let us now consider the problem of finding a solution to the equations

$$AA'\underline{p} = A\underline{y} \tag{4}$$

which are called the normal equations. From Lemmas 1, 2, 3, it follows that the equations (4) have a solution which is given by

$$\hat{\underline{p}} = (AA')^{g} A\underline{y} \quad . \tag{5}$$

Alternatively, we may obtain a solution of (4) as follows. Since A has rank n_o, there exist $(m - n_o)$ linearly independent $(m \times 1)$ column vectors which are linearly independent of the columns of A (if $n_o = m$, what follows is formally correct but has a vacuous meaning). Let these column vectors form the columns of a matrix $R = R(m \times (m - n_o))$. We may then say that the matrix $B = B\ (m \times m)$ defined by

$$B = [A \quad R] \begin{bmatrix} A' \\ R' \end{bmatrix} = [AA' + RR'] \tag{6}$$

has rank m and hence is non-singular. If we let $\underline{\psi} = \underline{\psi}\ ((m - n_o) \times 1)$ be a vector of arbitrary constants, then we may consider the equations

$$(7) \qquad \begin{aligned} AA'\underline{p} &= A\underline{y} \\ R'\underline{p} &= \underline{\psi} \end{aligned}$$

which may be solved by noting that $RR'\underline{p} = R\underline{\psi}$ and thus that

$$(8) \qquad A\underline{y} + R\underline{\psi} = AA'\underline{p} + RR'\underline{p} = B\underline{p} \qquad ;$$

Hence, a solution of (7) and also a solution of (4) is given by

$$(9) \qquad \hat{\underline{p}} = B^{-1}\ (A\underline{y} + R\underline{\psi})$$

Let $Q = Q\ (n_o \times m)$ be defined to be a matrix of rank n_o whose rows are orthogonal to the columns of R; i.e., $QR = 0_{n_o,\ m - n_o}$. If $\underline{\ell}'$ is a linear function of the rows of QB, then there exists $\underline{b}' = \underline{b}'\ (1 \times n_o)$ such that $\underline{\ell}' = \underline{b}'\ QB$ and

$$(10) \qquad \underline{\ell}'\hat{\underline{p}} = \underline{b}'\ QBB^{-1}\ (A\underline{y} + R\underline{\psi}) = \underline{b}'QA\underline{y} = \underline{\ell}'B^{-1}A\underline{y}.$$

Since $QB = Q\ [AA' + RR'] = QAA'$, it follows that linear functions of the rows of QB are linear functions of the rows of A'. Conversely, since B is non-singular, the rank of QB is n_o which implies that the rank of QA is n_o. Hence, the rows of QB represent n_o linearly independent linear functions of the rows of A' and thus form a basis of the vector space generated by the rows of A'. As a result, it follows that $\underline{\ell}'\underline{p}$ is estimable if and only if it can be written as a linear combination of the rows of QB in which case its best estimate is given by (10). Also, one can see that from the standpoint of estimation of estimable functions, we may take $\underline{\psi} = \underline{0}_{m - n_o}$ without any loss of generality; in this case (9) becomes

(11) $$\underline{\hat{\beta}} = B^{-1} A\underline{y}$$

which implies $B^{-1} = (AA')^g$ and that $A'B^{-1}A$ is uniquely determined (in the sense that it does not depend on R). Finally, we note that a particular choice of R has a meaningful interpretation when we take $\underline{\Psi} = \underline{0}$.

<u>Definition</u>: The vector $\underline{r}' = \underline{r}'$ (1 x m) is said to correspond to a redundancy in the model (3) if it is orthogonal to the columns of A; i.e., $\underline{r}' A = \underline{0}_n'$.

Hence redundancies are in a 1 - 1 correspondence with linear relationships among the rows of A. Also, we see that all (1 x m) vectors corresponding to redundancies are independent of the rows of A' and thus are not estimable. As a result, we may take the rows of R' to be $(m - n_o)$ independent (1 x m) vectors corresponding to redundancies. If this is done, the equations (7) may be written as

(12) $$AA'\underline{p} = A\underline{y}$$
$$R'\underline{p} = \underline{0}$$

where the rows of AA' and the rows of R' are orthogonal; equations of the form (12) arise in many applied problems (eg., those involving analysis of variance models).

3. <u>Models subject to non-estimable restrictions</u>. Suppose now that we know that the parameters of the model specified by (3) satisfy $s_1 \leq (m - n_o)$ linear restrictions of the form

(13) $$R_1' \underline{p} = \underline{\Psi}_1$$

Where $R_1' = R_1'$ (s_1 x m) and $\underline{\Psi}_1 = \underline{\Psi}_1$ (s_1 x 1) are known. Without loss of generality, we may assume that the rows of R_1' are linearly independent.

We further assume that the restrictions (13) are non-estimable.

Definition: The restrictions specified by (13) are said to be non-estimable if the rows of R_1' are linearly independent of the rows of A'.

We now compactly write the model (3) subject to (13) as

$$(14) \qquad \mathcal{E}(\underline{y}) = A'\underline{p}, \; \text{Var}\,(\underline{y}) = I_n \sigma^2$$
$$\underline{\Psi}_1 = R_1' \underline{p}$$

where the matrix $[A \quad R_1]$ has rank $(n_o + s_1)$.

Let $\phi = \underline{\ell}'\underline{p}$ be a linear function of the parameters. We will say that ϕ is estimable with respect to the model (14) if there exists $\underline{c}' = \underline{c}'(1 \times n)$ and $\underline{d}' = \underline{d}'(1 \times s_1)$ such that $\mathcal{E}(\underline{c}'\underline{y} + \underline{d}'\underline{\Psi}_1) = \underline{\ell}'\underline{p}$. If ϕ is estimable, $\hat{\phi} = \underline{c}'\underline{y} + \underline{d}'\underline{\Psi}_1$ is its best estimate provided $\mathcal{E}(\hat{\phi}) = \underline{c}'A'\underline{p} + \underline{d}'R_1'\underline{p} = \underline{\ell}'\underline{p}$ and $\text{Var}\,(\hat{\phi}) = \underline{c}'\underline{c}\sigma^2$ is a minimum; we can find $\hat{\phi}$ by the method of Lagrangian multipliers to minimize $\underline{c}'\underline{c}$ subject to $\underline{c}'A' + \underline{d}'R_1' = \underline{\ell}'$.

Let

$$(15) \qquad v(\underline{c}, \underline{d}, \underline{\lambda}) = \underline{c}'\underline{c} - 2\underline{\lambda}'(A\underline{c} + R_1\underline{d} - \underline{\ell}).$$

where $\underline{\lambda}' = \underline{\lambda}'(1 \times m)$ is the vector of Lagrangian multipliers. Differentiating with respect to the elements of $\underline{c}, \underline{d}, \underline{\lambda}$, we obtain the following system of equations for the stationary values of v.

$$(16) \qquad \begin{aligned} \underline{c} &= A'\underline{\lambda} \\ \underline{0} &= R_1'\underline{\lambda} \\ A\underline{c} + R_1\underline{d} &= \underline{\ell} \end{aligned}$$

We may obtain a solution of the equations (16) as follows. First

$$\underline{Ac} = AA'\underline{\lambda} \text{ implies } \underline{\lambda} = (AA')^g \, \underline{Ac}$$

(17) $$\underline{0} = R_1' \, \underline{\lambda} = R_1' \, (AA')^g \, \underline{Ac}$$

$$R_1' \, (AA')^g \underline{Ac} + R_1' \, (AA')^g R_1 \underline{d} = R_1' \, (AA')^g \, R_1 \underline{d} = R_1' \, (AA')^g \, \underline{\ell}$$

One possible choice for $(AA')^g$ is the matrix B^{-1} where $B = B(m \times m)$ is defined by

$$B = [A \quad R_1 \quad R_2] \begin{vmatrix} A' \\ R_1' \\ R_2' \end{vmatrix} = [AA' + R_1 \, R_1' + R_2 \, R_2']$$

where $R_2 = R_2(m \times (m - n_o - s_1))$ is a matrix of rank $(m - n_o - s_1)$ whose columns are independent of the columns of A and the columns of R_1. It follows that B has rank m and is non-singular. Hence a solution of (16) is given by

$$\underline{d} = [R_1' \, B^{-1} \, R_1]^{-1} \, R_1' \, B^{-1} \, \underline{\ell}$$

(19) $$\underline{c} = A'\underline{\lambda} = A'B^{-1}A\underline{c} = A'B^{-1} \, (\underline{\ell} - R_1 \, \underline{d})$$
$$= A'B^{-1}\underline{\ell} - A'B^{-1}R_1 \, [R_1'B^{-1}R_1]^{-1}R_1'B^{-1}\underline{\ell}.$$

Hence, we have that

(20) $$\hat{\phi} = \underline{\ell}'B^{-1}A\underline{y} - \underline{\ell}' \, B^{-1}R_1(R_1'B^{-1}R_1)^{-1} \, (R_1'B^{-1}A\underline{y} - \underline{\psi}_1) \quad .$$

If $\underline{\ell}' = \underline{d}^{*\prime}R_1'$ where $\underline{d}^{*\prime}(1 \times s_1)$, then $\hat{\phi} = \underline{d}^{*\prime} \, \underline{\psi}_1$ and Var $(\hat{\phi}) = 0$. On the other hand, suppose there exists $\underline{b}' = \underline{b}'(1 \times n_o)$ such that

(21) $$\underline{\ell}' = \underline{b}' \, QB$$

where $Q = Q(n_o \times m)$ is defined to be a matrix of rank n_o whose rows are orthogonal to the columns of $R = R(m \times (m - n_o))$ defined by $R = [R_1 \quad R_2]$;

i.e., $QR = [QR_1 \quad QR_2] = 0_{n_o,\, m - n_o}$. Then, we would have by substituting (21) into (20)

(22) $$\hat{\phi} = \underline{b}'QBB^{-1}Ay - \underline{b}'QBB^{-1}R_1\,(R_1'B^{-1}R_1)^{-1}\,(R_1'B^{-1}A\underline{y} - \underline{\psi}_1)$$
$$= \underline{b}'QBB^{-1}A\underline{y} - \underline{b}'QR_1\,(R_1'B^{-1}R_1)^{-1}\,(R_1'B^{-1}A\underline{y} - \underline{\psi}_1)$$
$$= \underline{b}'QBB^{-1}A\underline{y}$$
$$= \underline{\ell}'B^{-1}A\underline{y} \qquad ;$$

but again we recognize that

(23) $$QB = Q[AA' + R_1R_1' + R_2R_2'] = QAA' \quad ;$$

and hence by an argument similar to that following (10), we observe that (21) holds if and only if $\underline{\ell}'\underline{p}$ is estimable with respect to the basic model (3). Thus, because of (10), (22), and the uniqueness of $A'(AA')^gA$, it follows that if $\underline{\ell}'\underline{p}$ is estimable with respect to (3), then its best estimate with respect to the model (14) coincides with its best estimate with respect to the model (3). Hence, we see that the non-estimable restrictions specified by (13) have no influence on the construction of b.l.u.e. of linear functions which are estimable with respect to (3). Their only effect is to widen the class of estimable functions from an n_o dimensional space spanned by the rows of A' to an $(n_o + s_1)$ dimensional space spanned by the rows of A' and R_1'. For the linear functions of the parameters belonging to this wider class (i. e., those estimable with respect to (14)), the b.l.u.e.'s are given by (20).

4. Models subject to estimable restrictions. In this section, we assume that the parameters of the model specified by (3) are known to satisfy $s_2 \leq n_o$ linear restrictions of the form

(24) $$L'\underline{p} = \underline{\gamma}$$

where $L' = L'(s_2 \times m)$ and $\underline{\gamma} = \underline{\gamma}(s_2 \times 1)$ are known. Without loss of generality, we may assume that the rows of L' are linearly independent. Finally, we assume that the restrictions (24) are estimable.

Definition: The restrictions specified by (24) are said to be estimable if the rows of L' are linearly dependent on the rows of A'; i.e., if the rows of L' can be written as linear combinations of the rows of A'.

Since the restrictions (24) are assumed to be estimable, there exists a matrix $K = K(s_2 \times n)$ of rank s_2 such that

(25) $$L' = KA'$$

We now can compactly write the model (3) subject to the restrictions (24) as

(26) $$\mathcal{E}(\underline{y}) = A'\underline{p},\ \mathrm{Var}(\underline{y}) = I_n \sigma^2$$
$$\underline{\gamma} = L'\underline{p} = KA'\underline{p} \qquad .$$

We first observe that the class of estimable functions associated with the model (26) is the same as the class of estimable functions associated with the model (3) since (25) implies that the vector space spanned by the rows of A' and the rows of L' is the same as the vector space spanned by the rows of A' alone. Hence, if we assume that $\phi = \underline{\ell}'\underline{p}$ is an estimable linear function of the parameters, then there exists $\underline{k}' = \underline{k}'\ (1 \times n)$ such that $\underline{\ell}' = \underline{k}'\ A'$. With $\underline{c}' = \underline{c}'\ (1 \times n)$ and $\underline{d}' = \underline{d}'(1 \times s_2)$, $\hat{\phi} = \underline{c}'\underline{y} + \underline{d}'\underline{\gamma}$ is the best estimate of ϕ provided

(27) $$\mathcal{E}(\hat{\phi}) = \underline{c}'A'\underline{p} + \underline{d}'KA'\underline{p} = \underline{k}'A'\underline{p} = \underline{\ell}'\underline{p}$$

and the variance of $\hat{\phi}$, which is given by $\underline{c}'\underline{c}\ \sigma^2$, is a minimum. We can determine $\hat{\phi}$ by the method of Lagrangian multipliers. Let

(28) $$v(\underline{c},\underline{d},\underline{\lambda}) = \underline{c}'\underline{c} - 2\underline{\lambda}'(A\underline{c} + AK'\underline{d} - A\underline{k})$$

where $\underline{\lambda}' = \underline{\lambda}'$ (1 x m) is the vector of Lagrangian multipliers. Differentiating with respect to the elements of $\underline{c}$, $\underline{d}$, $\underline{\lambda}$, we obtain the following system of equations for the stationary values of v

$$\underline{c} = A'\underline{\lambda}$$

$$(29) \qquad \underline{0} = KA'\underline{\lambda}$$

$$A\underline{c} + AK'\underline{d} = A\underline{k} \quad .$$

We may obtain a solution of the equations (29) as follows: First we observe that

$$A\underline{c} = AA'\underline{\lambda} \text{ implies } \underline{\lambda} = (AA')^g A\underline{c}$$

$$(30) \qquad \underline{0} = KA'(AA')^g A\underline{c}$$

$$KA'(AA')^g A\underline{c} + KA'(AA')^g AK'\underline{d} = KA'(AA')^g AK'\underline{d} = KA'(AA')^g A\underline{k}.$$

From Lemma 4, we recall that $A'(AA')^g$ is uniquely determined, symmetric, and idempotant; hence the third equation in (30) may be written as

$$(31) \qquad KA'(AA')^g\, AA'(AA')^g AK'\underline{d} = KA'(AA')^g A\underline{k} \qquad \text{or}$$

$$[KA'(AA')^g A]\ [KA'(AA')^g A]'\underline{d} = [KA'(AA')^g A]\underline{k} \qquad \text{or}$$

$$\tilde{\tilde{K}}\tilde{\tilde{K}}'\underline{d} = \tilde{\tilde{K}}\underline{k}$$

where $\tilde{\tilde{K}} = \tilde{\tilde{K}}(s_2 \text{ x m})$ is defined by $\tilde{\tilde{K}} = KA'(AA')^g A$; but from Lemmas 2 and 3, it follows that a solution to the equations (31) is given by

$$(32) \qquad \underline{d} = (\tilde{\tilde{K}}\tilde{\tilde{K}}')^g\, \tilde{\tilde{K}}\underline{k} \qquad ;$$

using (29), (30), and (32), we obtain

$$(33) \qquad \underline{c} = A'\underline{\lambda} = A'(AA')^g A\underline{c}$$

$$= A'(AA')^g\, (A\underline{k} - AK'\underline{d})$$

$$= A'(AA')^g A\underline{k} - \tilde{K}'\underline{d}$$

$$= A'(AA')^g A\underline{k} - \tilde{K}'(\tilde{K}\tilde{K}')^g \tilde{K}\underline{k} \quad .$$

Combining (32) and (33), we have that

$$(34) \qquad \hat{\phi} = \underline{k}'A'(AA')^g A\underline{y} - \underline{k}'\tilde{K}'(\tilde{K}\tilde{K}')^g(\tilde{K}\underline{y} - \underline{\gamma})$$

$$= \underline{\ell}'(AA')^g A\underline{y} - \underline{\ell}'(AA')^g L[L'(AA')^g L]^g (L'(AA')^g A\underline{y} - \underline{\gamma})$$

$$= \underline{\ell}'HA\underline{y} + \underline{\ell}'\,\underline{\gamma}^*$$

where $H = H\ (m \times m)$ and $\underline{\gamma}^* = \underline{\gamma}^*\ (m \times 1)$ are defined by

$$(35) \qquad H = (AA')^g - (AA')^g L[L'(AA')^g L]^g L'(AA')^g$$

$$\underline{\gamma}^* = (AA')^g L\ [L'(AA')^g L]^g \underline{\gamma} \quad .$$

Note that Lemma 4 implies that $A'HA = A'(AA')^g A - K'(KK')^g K$ is uniquely determined; and thus whenever $\underline{\ell}'\,\underline{p}$ is estimable, $\underline{\ell}'\,HA\underline{y}$ is uniquely determined and this causes $\underline{\ell}'\,\underline{\gamma}^*$ to become uniquely determined from (27).

To proceed further, we take $(AA')^g$ to be given by the matrix B^{-1} where B is defined by (6); that this is legitimate can be seen by observing that (11) is always a solution of (4) and by applying Lemma 2. We now re-write (34) and (35) with B^{-1} replacing $(AA')^g$ and obtain

$$(37) \qquad \hat{\phi} = \underline{\ell}'\,B^{-1}A\underline{y} - \underline{\ell}'B^{-1}L\,(L'B^{-1}L)^{-1}\,(L'B^{-1}A\underline{y} - \underline{\gamma})$$

$$= \underline{\ell}'HA\underline{y} + \underline{\ell}'\,\underline{\gamma}^*$$

where

$$(38) \qquad H = B^{-1} - B^{-1}L\,(L'B^{-1}L)^{-1}\,L'B^{-1}$$

$$\underline{\gamma}^* = B^{-1}L\,(L'B^{-1}L)^{-1}\underline{\gamma}$$

since L' is assumed to have the full rank s_2. We observe that the following relations are satisfied

$$L'H = O_{s_2 m}, \quad L'\underline{\gamma}^* = \underline{\gamma}$$

(39) $$HBH = H \quad .$$

With (37), (38), (39), we are now in a position to consider the partitioning of the class of estimable functions arising from the presence of the restrictions (24). If $\underline{\ell}'$ is a linear function of the rows of L' and can be written say as $\underline{\ell}' = \underline{d}^{*\prime}L'$ where $\underline{d}^{*\prime} = \underline{d}^{*\prime}\ (1 \times s_2)$, then the best estimate of $\underline{\ell}'\underline{p}$ is given by

(40) $$\underline{\ell}'HA\underline{y} + \underline{\ell}'\ \underline{\gamma}^* = \underline{d}^{*\prime}L'HA\underline{y} + \underline{d}^{*\prime}L'\underline{\gamma}^*$$
$$= \underline{d}^{*\prime}\underline{\gamma} \quad .$$

Thus, if $\underline{\ell}'$ is a linear function of the rows of L', we have that the best estimator of $\underline{\ell}'\underline{p}$ is the corresponding linear function of the $\underline{\gamma}$'s and the variance of this best estimator is zero.

On the other hand, let $Q = Q((n_o - s_2) \times m)$ be defined to be a matrix of rank $(n_o - s_2)$ whose rows are orthogonal to the columns of the matrix $R = R(m \times (m - n_o))$ used in defining B by (6) and to the columns of L; i.e., we have

(41) $$Q\ [R \quad L] = O_{n_o - s_2,\ m - n_o + s_2} \quad .$$

If $\underline{\ell}'$ is a linear function of the rows of QB and can be written say as

(42) $$\underline{\ell}' = \underline{q}'QB = \underline{q}'Q[AA' + RR'] = \underline{q}'QAA'$$

where $\underline{q}' = \underline{q}'(1 \times (n_o - s_2))$, then the best estimate of $\underline{\ell}'\underline{p}$ is given by

(43) $$\underline{\ell}'HA\underline{y} + \underline{\ell}'\ \underline{\gamma}^* = \underline{q}'QBB^{-1}A\underline{y} - \underline{q}'QBB^{-1}L(L'B^{-1}L)^{-1}(L'B^{-1}A\underline{y} - \underline{\gamma})$$
$$= \underline{q}'QBB^{-1}A\underline{y}$$
$$= \underline{\ell}'B^{-1}A\underline{y}$$
$$= \underline{\ell}'\hat{\underline{p}}$$

where $\hat{\underline{p}}$ is given by (11). However, (43) is the best estimate of such $\underline{\ell}'\underline{p}$ with respect to the model (3) in which there are no restrictions. Thus, one sees that the restrictions provide no information at all about linear functions of the parameters whose coefficient vectors lie in the vector space generated by the rows of $QB = QAA'$.

Next, we observe that the vector spaces generated by the rows of QB and the rows of L' are disjoint. For suppose there existed $\underline{d}^{*\prime}$ and $\underline{q}'$ such that $\underline{d}^{*\prime}L' = \underline{q}'QB$; then we would have $\underline{q}'QBQ'\underline{q} = \underline{d}^{*\prime}L'Q'\underline{q} = 0$ which implies $\underline{q}'Q = \underline{0}'$ since B is symmetric positive definite; but the rows of Q are independent which implies $\underline{q}' = \underline{0}'$ and this implies $\underline{d}^{*\prime} = \underline{0}'$. Since the only vector common to the two spaces is the null vector, we have that they are disjoint. Also, we have by assumption that L' has rank s_2; and by the non-singularity of B, that QB has rank $(n_o - s_2)$. Finally, since $L' = KA'$ and $QB = QAA'$, the vector spaces generated by the rows of L' and the rows of QB are contained in the vector space generated by the rows of A'. In conclusion the vector spaces generated by the rows of L' and the rows of QB are disjoint vector spaces which span the vector space generated by the rows of A', that is the vector space of coefficient vectors of estimable functions. As a result, if $\underline{\ell}'\underline{p}$ is estimable, then its coefficient vector $\underline{\ell}'$ may be uniquely written as

$$\underline{\ell}' = \underline{q}'QB + \underline{d}^{*\prime}L' \tag{44}$$

and its best estimate has the form

$$\widehat{\underline{\ell}'\underline{p}} = \underline{q}'QA\underline{y} + \underline{d}^{*\prime}\underline{\gamma} \tag{45}$$

We also note here that

(46) $$\mathrm{Var}(\hat{\underline{\ell}'\underline{p}}) = \underline{q}'QAA'Q'\underline{q}\sigma^2 = \underline{q}'QBQ'\underline{q}\sigma^2 \quad .$$

However, (46) is a difficult expression to compute since it involves Q and $\underline{q}'$ which are not easy to obtain. We may obtain an alternative expression for the variance of the best estimate of estimable $\underline{\ell}'\underline{p}$ as follows. First we observe that the best estimate of $\underline{\ell}'\underline{p}$ is given by (37) where $\underline{\ell}'$ may be written as $\underline{\ell}' = \underline{k}'A'$; hence we have that

(47) $$\mathrm{Var}(\hat{\underline{\ell}'\underline{p}}) = \underline{\ell}'HAA'H'\underline{\ell}\sigma^2 = \underline{k}'A'HAA'HA\underline{k}\sigma^2$$ where

(48) $$\begin{aligned} A'HA &= A'B^{-1}A - A'B^{-1}L\ (L'B^{-1}L)^{-1}\ L'B^{-1}A \\ &= A'B^{-1}A - A'B^{-1}AK'(KA'B^{-1}AA'B^{-1}AK')^{-1}\ KA'B^{-1}A \quad . \end{aligned}$$

If we let W = W (n x n) be defined by

(49) $$W = A'B^{-1}A$$

and recall from Lemma 4 that W is uniquely determined, symmetric, and idempotent, then we have that

(50) $$A'HA = W - WK'[KWK']^{-1}\ KW$$ and

(51) $$\begin{aligned} A'HAA'HA &= WW - WK'[KWK']^{-1}KWW - WWK'[KWK']^{-1}KW + WK'[KWK']^{-1}KWWK'[KWK']^{-1}KW \\ &= W - WK'[KWK']^{-1}KW - WK'[KWK']^{-1}KW + WK'[KWK']^{-1}KW \\ &= W - WK'[KWK']^{-1}KW \\ &= A'HA \end{aligned}$$

Hence, we may re-write (47) as

(52) $$\begin{aligned} \mathrm{Var}(\hat{\underline{\ell}'\underline{p}}) &= \underline{k}'A'HAA'HA\underline{k}\sigma^2 \\ &= \underline{k}'A'HA\underline{k}\sigma^2 \\ &= \underline{\ell}'H\underline{\ell}\sigma^2 \end{aligned}$$

where H is given by (38); a similar argument could be given to show that (52) is valid when H is given by the more general form (35) since A'HA is uniquely determined. Note that the expression given by (52) is never bigger than $\underline{\ell}'B^{-1}\underline{\ell}\sigma^2$ since $(L'B^{-1}L)^{-1}\sigma^2$ is symmetric positive definite; the two expressions are equal only if $\underline{\ell}'B^{-1}L = \underline{0}'$ which is the case if and only if $\underline{\ell}'$ lies in the vector space generated by the rows of QB. Thus, the variance of the best estimate of estimable $\underline{\ell}'\underline{p}$ corresponding to the model (26) is smaller than that corresponding to the model (3) unless $\underline{\ell}'$ lies in the vector space generated by the rows of QB in which case the two are equal.

Having studied the partitioning of the class of estimable functions, we now consider the induced partitioning on the set of linear functions of the observations. First of all there exists a matrix $E_1 = E_1((n - n_o) \times n)$ of rank $(n - n_o)$ whose rows are orthogonal to the rows of A; i.e., the rows of E_1 form a basis of the error space. The linear functions $E_1\underline{y}$ form a basis of the error set.

In the non-restricted model (3), the set of linear functions $L'\underline{p}$ are all estimable because of (25) and their best estimates are given by

(53) $$L'B^{-1}A\underline{y} = KA'B^{-1}A\underline{y} \quad .$$

If we define $E_2 = E_2(s_2 \times n)$ to be the matrix

(54) $$E_2 = L'B^{-1}A = KA'B^{-1}A = KW \quad ,$$

then we observe that

(55) $$E_2E_1' = KA'B^{-1}AE_1' = 0_{s_2,\, n - n_o}$$

and hence that the sets $E_1\underline{y}$ and $E_2\underline{y}$ are uncorrelated.

Finally, in either the non-restricted model (3) or the restricted model (26), the set of linear functions $QB\underline{p} = QAA'\underline{p}$ are all estimable and their best estimates are given by

$$QBB^{-1}A\underline{y} = QA\underline{y} \quad . \tag{56}$$

If we define $C = C((n_o - s_2) \times n)$ to be the matrix,

$$C = QA \quad , \tag{57}$$

then we observe that

$$CE_1' = QAE_1' = 0_{n_o - s_2,\, n - n_o} \quad \text{and} \tag{58}$$

$$\begin{aligned} CE_2' &= QAA'B^{-1}AK' \\ &= QBB^{-1}L \\ &= QL \\ &= 0_{n_o - s_2,\, s_2} \quad . \end{aligned} \tag{59}$$

Hence, the sets $E_1\underline{y}$, $E_2\underline{y}$, and $C\underline{y}$ are mutually uncorrelated and have variance-covariance matrices as follows

$$\begin{aligned} Var(E_1\underline{y}) &= E_1E_1'\sigma^2 \\ Var(E_2\underline{y}) &= KA'B^{-1}AA'B^{-1}AK'\sigma^2 = KA'B^{-1}AK'\sigma^2 = L'B^{-1}L\sigma^2 \\ Var(C\underline{y}) &= QAA'Q'\sigma^2 = QBQ'\sigma^2 \quad . \end{aligned} \tag{60}$$

From (60), we see that $E_1\underline{y}$, $E_2\underline{y}$, $C\underline{y}$ all have symmetric positive definite variance covariance matrices because B is symmetric positive definite and L' has the full rank s_2, Q has the full rank $(n_o - s_2)$, and E_1 has the full rank $(n - n_o)$. But this implies that E_2 has rank s_2 and C has rank $(n_o - s_2)$.

Thus, the sets $C\underline{y}$, $E_2\underline{y}$, and $E_1\underline{y}$ represent a partitioning of the set of linear functions of $y_1, y_2, \cdots, y_n$ into three mutually uncorrelated sets. We call the set $E_2\underline{y}$ <u>the induced error set</u> due to restrictions since the expected value of $E_2\underline{y}$ is $\underline{\gamma}$ which is known. We call $C\underline{y}$ <u>the reduced estimation set</u> due to restrictions.

Finally, let us consider least squares estimation in the model (26). We apply the method of Lagrangian multipliers to minimize

$$(61) \qquad S^2 = (\underline{y} - A'\underline{p})' (\underline{y} - A'\underline{p}) \text{ subject to } \underline{\gamma} = KA'\underline{p}.$$

Let

$$(62) \qquad v(\underline{p},\underline{\lambda}) = (\underline{y} - A'\underline{p})' (\underline{y} - A'\underline{p}) + 2\underline{\lambda}' (KA'\underline{p} - \underline{\gamma})$$

where $\underline{\lambda}' = \underline{\lambda}'(1 \times s_2)$ is the vector of Lagrangian multipliers. Differentiating with respect to the elements of $\underline{p}$ and $\underline{\lambda}$, we obtain the following system of equations for the stationary values of v

$$\begin{aligned} & -2(\underline{y} - A'\underline{p})'A' + 2\underline{\lambda}'KA' = \underline{0}' \\ (63) \qquad & +2(KA'\underline{p} - \underline{\gamma})' = \underline{0}' \end{aligned}$$

which may be re-written as

$$\begin{aligned} (64) \qquad AA'\underline{p} + AK'\underline{\lambda} &= A\underline{y} \\ KA'\underline{p} &= \underline{\gamma} \end{aligned}$$

The equations (64) may be solved by proceeding as follows. First

$$\begin{aligned} (65) \qquad KA'(AA')^g AA'\underline{p} + KA'(AA')^g AK'\underline{\lambda} &= KA'(AA')^g A\underline{y} \quad \text{or} \\ \check{K}A'\underline{p} + \check{K}\check{K}'\underline{\lambda} &= \check{K}\underline{y} \end{aligned}$$

where $\check{K}$ is defined in (31). Hence we have

$$(66) \qquad (\check{K}\check{K}')^g \check{K}A'\underline{p} + \underline{\lambda} = (\check{K}\check{K}')^g \check{K}\underline{y}$$

However, we also have

$$(67)\qquad \underline{\gamma} = KA'\underline{p} = -KA'(AA')^g AK'\underline{\lambda} + KA'(AA')^g A\underline{y}$$
$$= -\check{K}\check{K}'\underline{\lambda} + \check{K}\underline{y}$$
$$= \check{K}A'\underline{p}$$

Hence, a solution to (64) is given by

$$(68)\qquad \underline{\lambda} = (\check{K}\check{K}')^g\,(\check{K}\underline{y} - \underline{\gamma})$$
$$\hat{\underline{p}} = (AA')^g A\underline{y} - (AA')^g AK'(\check{K}\check{K}')^g(\check{K}\underline{y} - \underline{\gamma}) \quad .$$

We may rewrite the solution $\hat{\underline{p}}$ as

$$(69)\qquad \hat{\underline{p}} = [(AA')^g - (AA')^g L\,[L'(AA')^g L]^g\, L'(AA')^g]\, A\underline{y}$$
$$+ (AA')^g L\,[L'(AA')^g L]^g \underline{\gamma}$$
$$= HA\underline{y} + \underline{\gamma}^* \quad .$$

where H and $\underline{\gamma}^*$ are defined by (35).

Thus, the best estimate of estimable $\underline{\ell}'\underline{p}$ given by (34) may be re-written as

$$(70)\qquad \hat{\phi} = \underline{\ell}'HA\underline{y} + \underline{\ell}'\,\underline{\gamma}^* = \underline{\ell}'\hat{\underline{p}}$$

where $\hat{\underline{p}}$ is given by (69). Thus, the least squares estimators of estimable $\underline{\ell}'\underline{p}$ are their best linear unbiased estimators. In addition, let us find the value of Min S^2. To do this, let us first replace $(AA')^g$ by B^{-1}. Then

$$(71)\qquad \text{Min } S^2 = (\underline{y} - A'\hat{\underline{p}})'\,(\underline{y} - A'\hat{\underline{p}})$$
$$= (\underline{y} - A'HA\underline{y} - A'\,\underline{\gamma}^*)'(\underline{y} - A'HA\underline{y} - A'\,\underline{\gamma}^*)$$

$$\text{Min } S^2 = (\underline{y} - A'B^{-1}A\underline{y} + A'B^{-1}L(L'B^{-1}L)^{-1}\, L'B^{-1}A\underline{y} - A'B^{-1}L(L'B^{-1}L)\,\underline{\gamma})' \text{ [transpose same]}$$
$$= [(y - A'B^{-1}Ay) + A'B^{-1}L(L'B^{-1}L)^{-1}(L'B^{-1}A\underline{y} - \underline{\gamma})]' \text{ [transpose same]}$$
$$= [\;(I_n - A'B^{-1}A)\underline{y} + E_2'(E_2E_2')^{-1}\,(E_2\underline{y} - \underline{\gamma})]' \qquad \text{[transpose same]}$$

Since Lemma 4 implies that $A[I_n - A'B^{-1}A] = A - AA'B^{-1}A = 0$, we may take $I_n - A'B^{-1}A$ to be given by $E_1'(E_1E_1')^{-1} E_1$ (That this is all right follows also from Lemma 5) and hence find that

$$\text{(72)} \qquad \text{Min } S^2 = \underline{y}'E_1'(E_1E_1')^{-1} E_1\underline{y} + (E_2\underline{y} - \underline{\gamma})' (E_2E_2')^{-1} (E_2\underline{y} - \underline{\gamma})$$

$$= S_{e1}^2 + S_{e2}^2$$

where $S_{e1}^2 = \underline{y}'E_1'(E_1E_1')^{-1} E_1\underline{y}$ is called the sum of squares to error in the model (3), $S_{e2}^2 = (E_2\underline{y} - \underline{\gamma})' (E_2E_2')^{-1} (E_2\underline{y}-\underline{\gamma})$ is called the sum of squares due to induced error, and $S_e^2 = S_{e1}^2 + S_{e2}^2$ is the total sum of squares due to error in the model (26).

Lemma 5: Let $F = F(m \times n)$ and $G = G(m \times n)$ be two matrices satisfying the following conditions

(i) Rank F = Rank G = $n_o \leq m \leq n$,

(ii) There exists $P_c = P_c\ (n \times t)$ of rank $(n - n_o) \leq t$ such that $FP_c = GP_c = 0_{mt}$,

(iii) There exists $P_r = P_r(s \times m)$ of rank $(m - n_o) \leq s$ such that $P_rF = P_rG = 0_{sn}$,

(iv) There exists $S = S(n \times m)$ such that $FSF = F$, $GSG = G$,

Then $F = G$.

Proof: Now (i) and (ii) imply the existence of $C_1 = C_1(m \times m)$ such that $F = C_1G$ since the rows of F and the rows of G generate the same vector space-namely the vector space orthogonal to the space generated by the columns of P_c. Similarly, (i) and (iii) imply the existence of $C_2 = C_2$ (n x n) such that $G' = C_2'F'$ and hence $G = FC_2$. Thus, the following relation holds because of (iv)

$$F = C_1G = C_1GSG = FSG \tag{73}$$
$$= FSFC_2$$
$$= FC_2$$
$$= G$$

and this verifies the conclusion of the lemma. If in (71) and (72), we set

$$F = I_n - A'B^{-1}A, \; G = E_1' \, (E_1E_1')^{-1} \, E_1 \tag{74}$$
$$P_c = A', \; P_r = A, \; S = I_n \quad ,$$

then the conditions of Lemma 5 are satisfied and it follows that

$I_n - A'B^{-1}A = E_1' \, (E_1E_1')^{-1} \, E_1$.

We now define the sum of squares due to regression in the model (26) as the sum of squares due to the linear set $C\underline{y}$, and this is given by

$$S_r^2 = \underline{y}'C'(CC')^{-1} \, C\underline{y} \tag{75}$$
$$= \underline{y}'A'Q'(QBQ')^{-1}QA\underline{y} \quad .$$

We apply Lemma 5 in order to obtain a more readily computed formula for S_r^2. Let

$$F = A'Q'(QBQ')^{-1}QA, \; G = A'HA; \tag{76}$$

then the conditions of Lemma 5 are satisfied if

$$P_c = [E_1' \,\vdots\, E_2'] \; , \; P_r = P_c' \; , \; S = I_n \quad \text{because} \tag{77}$$

idempotency of G follows from (51) while idempotency of F follows from $QAA' = QB$; $FP_c = 0$ follows from (58) and (59) while $GP_c = 0$ follows from (49), (50), and (54); and Rank F = trace F = $n_o - s_2$ while

(78) $$\text{Rank } G = \text{trace } G = \text{trace } W - \text{trace } I_{s_2} = n_o - s_2 \quad .$$

Thus, we may re-write (75) as

(79) $$S_r^2 = \underline{y}'A'HA\underline{y} \quad .$$

The decomposition of the total sum of squares in the model (26) is not necessarily additive since

(80) $$\begin{aligned} \underline{y}'\underline{y} - S_r^2 - S_e^2 &= \underline{y}'\underline{y} - \underline{y}'A'HA\underline{y} - (\underline{y} - A'HA\underline{y} - A'\underline{\gamma}^*)'(\underline{y} - A'HA\underline{y} - A'\underline{\gamma}^*) \\ &= +2(A'\underline{\gamma}^*)'(\underline{y} - A'HA\underline{y}) - \underline{\gamma}^{*\prime}AA'\underline{\gamma}^* \\ &= 2\,\underline{\gamma}^{*\prime}A\underline{y} - 2\,\underline{\gamma}^{*\prime}AA'HA\underline{y} - \underline{\gamma}^{*\prime}AA'\underline{\gamma}^* \\ &= 2\,\underline{\gamma}^{*\prime}A\underline{y} - \underline{\gamma}^{*\prime}AA'\underline{\gamma}^* \\ &= 2\,\underline{\gamma}'(L'B^{-1}L)^{-1}E_2\underline{y} - \underline{\gamma}'(L'B^{-1}L)^{-1}\underline{\gamma} \end{aligned}$$

since $\underline{\gamma}^{*\prime}AA'HA = 0$ follows from (38), (54), (49), (50).

If we define S_d to be the departure from additivity, then

(81) $$S_d = \underline{y}'\underline{y} - S_r^2 - S_e^2 = \underline{\gamma}'(L'B^{-1}L)^{-1}(2E_2\underline{y} - \underline{\gamma}) \quad .$$

When we assume the $y_1, y_2, \ldots, y_n$ have a joint normal distribution with means and variances specified by the model (26), then our previous results tell us that $C\underline{y}$, $E_1\underline{y}$, $E_2\underline{y}$ are independent sets of random variables having the normal distributions $N(CA'\underline{p}, CC'\sigma^2)$, $N(\underline{0}, (E_1E_1')\sigma^2)$, $N(L'\underline{p}, L'B^{-1}L\sigma^2)$ respectively. Thus, we have that S_r^2, S_{e1}^2, S_{e2}^2 all are independently distributed with S_r^2 being distributed as non-central $\chi^{2'}(n_o - s_2, \frac{\underline{p}'AA'HAA'\underline{p}}{2\sigma^2}), S_{e1}^2$ being distributed as central $\chi^2(n - n_o)$, and S_{e2}^2 being distributed as central $\chi^2(\ s_2\ )$; also $S_e^2 = S_{e1}^2 + S_{e2}^2$ is distributed as central $\chi^2\ (n + s_2 - n_o)$. We also note that S_d is distributed as $N(\underline{\gamma}'(L'B^{-1}L)^{-1}\underline{\gamma}, 4\underline{\gamma}'(L'B^{-1}L)^{-1}\underline{\gamma})$ and is independent of S_r^2 and S_{e1}^2. We thus have the following analysis of variance for the model (26):

	Source	Degrees of Freedom	Sum of Squares	Mean Square	Variance Ratio
	Regression	$n_o - s_2$	$S_r^2 = y'A'HA\underline{y}$	$S_r^2/(n_o - s_2)$	$\frac{(n - n_o + s_2)S_r^2}{(n_o - s_2)S_e^2}$
(82)	Error	$n-n_o+s_2$	$S_e^2 = \underline{y}'\underline{y}-S_r^2-S_d$	$S_e^2/(n-n_o+s_2)$	
	Deviation		$S_d = 2\underline{\gamma}^{*\prime}A\underline{y}-\underline{\gamma}^{*\prime}AA'\underline{\gamma}^*$		
	Total	n	$\underline{y}'\underline{y}$		

Thus, we see that the presence of estimable restrictions does not alter the class of estimable functions of the parameters. However, by allowing us to estimate certain functions with zero variance, they permit the construction of estimates at least as good as and often better than those obtainable in the unrestricted model. Finally, they permit us to obtain a better estimate of error with more degrees of freedom.

5. Models subject to restrictions. Here, we assume that the parameters of the model specified by (3) are known to satisfy s linear restrictions of the form

(83) $$\Pi'\underline{p} = \underline{\psi}$$

where $\Pi' = \Pi'(s \times m)$ and $\underline{\psi} = \underline{\psi}(s \times 1)$ are known. Without loss of generality, we may assume that the rows of Π' are linearly independent. To deal with the restricted model specified by

(84) $$\mathcal{E}(\underline{y}) = A'\underline{p}, \ \mathrm{Var}\ (\underline{y}) = I_n\sigma^2$$
$$\Pi'\underline{p} = \underline{\psi} \ ,$$

we need to partition Π' and $\underline{\psi}$ as

(85) $$\Pi' = \begin{bmatrix} L' \\ R_1' \end{bmatrix}, \ \underline{\psi} = \begin{bmatrix} \underline{\gamma} \\ \underline{\psi}_1 \end{bmatrix}$$

where $L' = L'(s_2 \times m)$, $R_1' = R_1' \ (s_1 \times m)$, $\underline{\gamma} = \underline{\gamma}(s_2 \times 1)$, and $\underline{\psi}_1 = \underline{\psi}_1(s_1 \times 1)$ such that $s_2 \leq n_o$ and $s_1 = s - s_2 \leq m - n_o$, and the rows of L' lie in the vector space generated by the rows of A' while the rows of R_1' are linearly independent of the rows of A'; i.e., we assume the restrictions (83) can be partitioned into an estimable part like (24) and a non-estimable part like (13).

From Sections 3 and 4, we see that the class of estimable functions of the parameters can be partitioned into three disjoint subsets, namely the one generated by the rows of R_1', the one generated by the rows of L', and the one generated by the rows of QAA' where $Q = Q((n_o - s_2) \times m)$ is defined to be a matrix of rank $(n_o - s_2)$ such that

(86) $$Q[\Pi \quad R_2] = 0_{n_o - s_2, \ m - n_o + s_2}$$

where $R_2' = R_2'((m - n_o - s_1) \times m)$ is any matrix of rank $m - n_o - s_1$ whose rows are independent of the rows of A' and the rows of R_1'.

The class of linear functions of the observations may be partitioned in the same way as in Section 4 and the analysis of variance given by (82) also applies here. This holds true because as long as we deal with estimable functions (with respect to (3)), non-estimable restrictions have no effect; as before, the only effect of non-estimable restrictions is to widen the class of linear functions of the parameters which can be estimated unbiasedly. Finally, we give an expression for the best estimate $\hat{\phi}$ of $\phi = \underline{\ell}'\underline{p}$ such that $\underline{\ell}'$ can be written as $\underline{\ell}' = \underline{k}'A' + \underline{d}'R_1'$.

(87) $$\hat{\phi} = \underline{c}'\underline{y} + \underline{d}_1' \underline{\Psi}_1 + \underline{d}_2' \underline{\gamma}$$

where $\underline{c}$, $\underline{d}_1$, $\underline{d}_2$ are given by the solution of

$$\underline{c} = A'B^{-1}(\underline{\ell} - R_1\underline{d}_1 - L\underline{d}_2)$$

(88) $$\begin{bmatrix} R_1'B^{-1}R_1 & R_1'B^{-1}L \\ L'B^{-1}R_1 & L'B^{-1}L \end{bmatrix} \begin{bmatrix} \underline{d}_1 \\ \underline{d}_2 \end{bmatrix} = \begin{bmatrix} R_1'B^{-1}\underline{\ell} \\ L'B^{-1}\underline{\ell} \end{bmatrix}$$

where $B = AA' + R_1R_1' + R_2R_2'$.

6. The non-restricted model with a general but non-singular variance-covariance matrix. Let $y_1, y_2, \ldots, y_n$ be random variables whose expected values are given by (1) and whose variance-covariance matrix is non-singular and is known except for a multiplicative constant σ^2. If we let $V = V(n \times n)$ be the matrix defined by

(89) $$\text{Var}\,(\underline{y}) = V\sigma^2 = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{bmatrix} \sigma^2 ,$$

then the model can be compactly written as

(90) $$E(\underline{y}) = A'\underline{p}, \quad Var(\underline{y}) = V\sigma^2 \quad .$$

The set of estimable functions of the parameters is the same here as for the model specified by (3) since the design matrix A' is the same for both cases. Also, the error spaces associated with the models (3) and (90) are identical because both are generated by the set of (1 x n) vectors which are orthogonal to the rows of A. However, an important distinction can be made between the estimation spaces corresponding to these models. By definition, a linear function of $y_1, y_2, \ldots, y_n$ is an estimating function if it is uncorrelated with all linear functions belonging to error. Hence, $\underline{c}'\underline{y}$ is an estimating function if and only if

(91) $$\underline{c}'V\underline{e} = 0$$

for every $\underline{e}'\underline{y}$ belonging to error. As before, the set of all estimating functions is called the estimation set.

Lemma 6: A non-null linear function of $y_1, y_2, \ldots, y_n$ cannot, at the same time, belong to the estimation set and the error set.

Proof: Let $\underline{f}_1'\underline{y}, \underline{f}_2'\underline{y}, \ldots, \underline{f}'_{n_e}\underline{y}$ be n_e linearly independent non-null linear functions belonging to error (where as before n_e is the rank of the error space). Let $F = F(n_e \times n)$ be given by

(92) $$F = \begin{bmatrix} \underline{f}_1' \\ \underline{f}_2' \\ \cdots \\ \underline{f}_{n_e}' \end{bmatrix} \quad ;$$

then Var $(F\underline{y}) = FVF'\sigma^2$. There exists an orthogonal matrix $P = P(n_e \times n_e)$ such that

(93) $$PFVF'P' = D_\lambda$$

where D_λ is a diagonal matrix with positive diagonal elements $\lambda_1, \lambda_2, \ldots, \lambda_{n_e}$ since FVF' is symmetric positive definite. Defining $E = E(n_e \times n)$ and $\underline{z} = \underline{z}(n_e \times 1)$ by

(94) $$PF = E = \begin{bmatrix} \underline{e}_1' \\ \underline{e}_2' \\ \cdots \\ \underline{e}_{n_e}' \end{bmatrix}, \quad \underline{z} = PF\underline{y} = E\underline{y} = \begin{bmatrix} \underline{e}_1'\underline{y} \\ \underline{e}_2'\underline{y} \\ \cdots \\ \underline{e}_{n_e}'\underline{y} \end{bmatrix}$$

we have that $\underline{e}_1'\underline{y}, \underline{e}_2'\underline{y}, \ldots, \underline{e}_{n_e}'\underline{y}$ are uncorrelated linear functions of the y's belonging to error; they also form a basis of the error set. Hence, if $\underline{e}'\underline{y}$ is a non-null linear function belonging to error, then it can be expressed as

(95) $$\underline{e}'\underline{y} = \phi_1\underline{e}_1'\underline{y} + \phi_2\underline{e}_2'\underline{y} + \cdots + \phi_{n_e}\underline{e}'_{n_e}\underline{y} = \underline{\phi}'E\underline{y}.$$

If $\underline{e}'\underline{y}$ also belongs to the estimation set, then

(96) $$\underline{0}'_{n_e} = \text{Cov}\,(\underline{e}'\underline{y}, E\underline{y}) = \underline{\phi}'EVE'\sigma^2 = \underline{\phi}'D_\lambda\sigma^2 \qquad ;$$

but (96) implies that $\underline{\phi}' = \underline{0}'_{n_e}$ since all the λ's are positive. As a result, the only vector common to the estimation space and the error space is the null vector.

We now investigate the structure of the estimation space. Because of (91) and (95), the linear function $\underline{c}'\underline{y}$ is an estimating function if and only if $\underline{c}$ satisfies the equations

(97) $$E V \underline{c} = \underline{0}$$.

Since V is non-singular and E has the full rank n_e, the above represent a set of n_e independent linear homogeneous equations to determine $\underline{c}$. As a result, the equations (97) have $n_o = n - n_e$ independent solutions. One set of solutions to (96), which is of interest, is

(98) $$C = [\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_m] = V^{-1}A'$$.

Note that $EVC = EVV^{-1}A' = EA' = 0_{n_e m}$ since the rows of E and A are orthogonal by construction; also note that C has rank n_o since V^{-1} is non-singular. Thus, $C\underline{y}$ generates the estimation space. We can now prove the fundamental theorem of linear estimation in the model specified by (90).

<u>Theorem 2</u>: If $\underline{\ell}'\underline{p}$ is any estimable function of the parameters $p_1, p_2, \ldots, p_m$, then

(i) there exists a unique linear function $\underline{c}'\underline{y}$ of $y_1, y_2, \ldots, y_n$ such that $\underline{c}'$ belongs to the estimation space and $\mathcal{E}(\underline{c}'\underline{y}) = \underline{\ell}'\,\underline{p}$, also, the variance of $\underline{c}'\underline{y}$ is smaller than the variance of any other linear unbiased estimator of $\underline{\ell}'\underline{p}$;

(ii) the best estimator of $\underline{\ell}'\underline{p}$ may be written as either

(a) $\underline{q}'AV^{-1}\underline{y}$ where $\underline{q}'\,(AV^{-1}A') = \underline{\ell}'$

(b) $\underline{\ell}'\hat{\underline{p}}$ where $\hat{\underline{p}}$ is a value of $\underline{p}$ which minimizes the sum of squares $S^2 = (\underline{y} - A'\underline{p})'\,V^{-1}(\underline{y} - A'\underline{p})$; i.e., $\hat{\underline{p}}$ satisfies the equations $AV^{-1}A'\underline{p} = AV^{-1}\underline{y}$;

(iii) the variance of the best estimator of $\underline{\ell}'\underline{p}$ is uniquely determined and given by $\underline{\ell}'(AV^{-1}A')^g\underline{\ell}$.

Proof: It follows from Lemma 6, (96), and (98) that the n linear functions $\underline{e}_1'\underline{y}$, $\underline{e}_2'\underline{y}$, ..., $\underline{e}_{n_e}'\underline{y}$ and $\underline{c}_{i_1}'\underline{y}$, $\underline{c}_{i_2}'\underline{y}$ ..., $\underline{c}_{i_{n_o}}'\underline{y}$ are independent, where $\underline{c}_{i_1}', \underline{c}_{i_2}'$, ..., $\underline{c}_{i_{n_o}}'$ represent n_o independent rows of C, and that any arbitrary linear function of $y_1, y_2, \ldots, y_n$ may be expressed in a unique manner as a linear combination of them. Since $\underline{\ell}'\underline{p}$ is estimable, there exists a linear function $\underline{d}'\underline{y}$ such that $\mathcal{E}(\underline{d}'\underline{y}) = \underline{\ell}'\underline{p}$. From what has been indicated previously we can uniquely write

(99) $$\underline{d}'\underline{y} = \underline{c}'\underline{y} + \underline{e}'\underline{y}$$

where $\underline{c}'\underline{y}$ is in the estimation set and $\underline{e}'\underline{y}$ is in the error set. As a result,

(100) $$\mathcal{E}(\underline{c}'\underline{y}) = \mathcal{E}(\underline{d}'\underline{y}) - \mathcal{E}(\underline{e}'\underline{y}) = \underline{\ell}'\underline{p} - 0 = \underline{\ell}'\underline{p}.$$

Suppose $\underline{c}_o'\underline{y}$ is another linear function in the estimation set which is an unbiased estimator of $\underline{\ell}'\underline{p}$. Then, $\mathcal{E}\{(\underline{c}_o - \underline{c})'\underline{y}\} = 0$ and hence $(\underline{c}_o - \underline{c})'\underline{y}$ belongs to the error set; but it is also in the estimation set since both $\underline{c}_o'\underline{y}$ and $\underline{c}'\underline{y}$ are in the estimation set. From Lemma 6, it follows that $\underline{c}_o - \underline{c} = \underline{0}_n$ or $\underline{c}_o = \underline{c}$; hence $\underline{c}'\underline{y}$ is uniquely defined.
Next we observe that

(101) $$\mathrm{Var}(\underline{d}'\underline{y}) = \mathrm{Var}(\underline{c}'\underline{y}) + \mathrm{Var}(\underline{e}'\underline{y}) \geq \mathrm{Var}(\underline{c}'\underline{y})$$

since $\underline{c}'\underline{y}$ and $\underline{e}'\underline{y}$ are uncorrelated by construction and $\mathrm{Var}(\underline{e}'\underline{y})$ is positive except when $\underline{e}' = \underline{0}_n'$. This completes the proof of (i).

From (i) and (98), it follows that the best estimate of $\underline{\ell}'\underline{p}$ can be written as a linear function of $\underline{c}_1'\underline{y}$, $\underline{c}_2'\underline{y}$, ... $\underline{c}_m'\underline{y}$; i.e.,

(102) $$\widehat{\underline{\ell}'\underline{p}} = \underline{q}'C'\underline{y} = \underline{q}'AV^{-1}\underline{y} \quad .$$

But (102) and the fact that $\underline{\ell}'\underline{p} = \mathcal{E}(\underline{q}'AV^{-1}\underline{y}) = \underline{q}'AV^{-1}A'\underline{p}$ imply that $\underline{q}'$ satisfies

(103) $$\underline{\ell}' = \underline{q}'\,AV^{-1}A' \quad .$$

Alternatively, if $\hat{\underline{p}}$ is a solution to the equations

(104) $$(AV^{-1}A')\underline{p} = AV^{-1}\underline{y}$$

and if $\underline{\ell}'\underline{p}$ is estimable with $\underline{\ell}'$ satisfying (103), then

(105) $$\begin{aligned}\underline{\ell}'\hat{\underline{p}} &= \underline{q}'AV^{-1}A'\hat{\underline{p}}\\ &= \underline{q}'AV^{-1}\underline{y}\\ &= \widehat{\underline{\ell}'\underline{p}} = \text{best estimate of } \underline{\ell}'\underline{p} \quad .\end{aligned}$$

The equations (104) are called the generalized least squares (normal) equations. Also, the equations

(106) $$\frac{\partial S^2}{\partial \underline{p}} = -2(\underline{y} - A'\underline{p})'\,V^{-1}A' = \underline{0}'_m$$

which simplify to (104) determine the stationary values at which S^2 attains its minimum.

Finally, we note that

(107) $$\mathrm{Var}(\widehat{\underline{\ell}'\underline{p}}) = \underline{q}'AV^{-1}VV^{-1}A'\underline{q}\sigma^2 = \underline{q}'AV^{-1}A'\underline{q}\sigma^2 = \underline{\ell}'\underline{q}\sigma^2$$

Since the equations (103) are consistent when $\underline{\ell}'\underline{p}$ is estimable, a solution to them is given by

(108) $$\underline{q} = (AV^{-1}A')^g\underline{\ell} \quad .$$

Substituting (108) into (107), we obtain

(109) $$\mathrm{Var}(\widehat{\underline{\ell}'\underline{p}}) = \underline{\ell}'\,(AV^{-1}A')^g\underline{\ell}\sigma^2 \quad .$$

Since $\underline{\ell}'\underline{p}$ is estimable and V is non-singular, symmetric, and positive definite, Lemma 4 implies that (109) is uniquely determined. This completes the proof of the theorem.

Finally, we define the sum of squares due to regression in the model (90) by

$$(110) \qquad S_r^2 = \hat{\underline{p}}'AV^{-1}\underline{y} = \underline{y}'V^{-1}A'(AV^{-1}A')^g AV^{-1}\underline{y}$$

and the sum of squares due to error by

$$(111) \qquad S_e^2 = \underline{y}'V^{-1}\underline{y} - S_r^2 = \text{Min } S^2 \quad .$$

The above follow by procedures analagous to those given in Bose [1950] for determining the sum of squares due to a set of linear functions of the y's. Alternatively, one can use the transformation

$$(112) \qquad \underline{y}^* = T\underline{y} \quad ,$$

where T is a non-singular matrix such that $TVT' = I_n$, and consider ordinary least squares estimation in the model

$$(113) \qquad \mathcal{E}\{\underline{y}^*\} = TA'\underline{p} \ , \ \text{Var}(\underline{y}^*) = I_n\sigma^2 \quad .$$

For the model (113), the sum of squares due to regression is given by

$$(114) \quad S_r^2 = \underline{y}^{*\prime}T\ A'(AT'TA')^g AT'\underline{y}^* = \underline{y}'V^{-1}A'(AV^{-1}A')^g AV^{-1}\underline{y}$$

because $T'T = V^{-1}$ and

$$(115) \qquad S_e^2 = \underline{y}^{*\prime}\underline{y}^* - S_r^2 = \underline{y}'V^{-1}\underline{y} - S_r^2 = \text{Min } S^2 \quad .$$

The expressions (114) and (115) in terms of the y*'s are well known and are derived in Bose [1950].

Last of all, it is worthwhile to observe that the partitioning of the sum of squares in the model (90) is not additive in the sense that

(116) $$S_r^2 + S_e^2 = \underline{y}'V^{-1}\underline{y} \quad .$$

If we define S_d^2 by

(117) $$S_d^2 = \underline{y}'\underline{y} - \underline{y}'V^{-1}\underline{y} \quad ,$$

then it follows that S_d^2 represents the amount by which S_r^2 and S_e^2 fail to add to the total sum of squares $\underline{y}'\underline{y}$. This lack of additivity is due to the failure of V to be an identity matrix and the resulting lack of orthogonality between the estimation set and the error set.

In conclusion, we observe that the effect of the variance covariance matrix V on the linear model is exerted through the transformation of the estimation set from $A\underline{y}$ to $A\Sigma^{-1}\underline{y}$. As a result of this, best estimates of estimable linear functions and sums of squares are suitably modified. However, the class of estimable functions of the parameters and the class of linear functions of the observations belonging to error remain unchanged.

7. The model of Section 6 subject to restrictions. Suppose we have the model

(118) $$\mathcal{E}(\underline{y}) = A'\underline{p}, \; Var(\underline{y}) = V\sigma^2$$
$$\Pi'\underline{p} = \underline{\psi}$$

where V is given by (89) and the restrictions are given by (83). Let us make the transformation (112). Then

(119) $$\mathcal{E}(\underline{y}^*) = TA'\underline{p}, \; Var(\underline{y}) = I_n\sigma^2$$
$$\Pi'\underline{p} = \underline{\psi} \quad .$$

The results of Section 5 may now be applied to (119). This completes the analysis of this case.

8. The general linear model. Let $y_1, y_2, \ldots, y_n$ be random variables whose expected values are given by (1) and whose variance-covariance matrix is given by (89); here, however, we don't require V to be non-singular. The model can be compactly written as

$$\mathcal{E}(\underline{y}) = A'\underline{p},\ \mathrm{Var}(\underline{y}) = V\sigma^2 \tag{120}$$

where we assume that V has rank r .

The set of estimable functions of the parameters is the same here as for the model specified by (3). Also, the error space associated with the models (3) and (120) are identical because both are generated by the set of (1 x n) vectors which are orthogonal to the rows of A. However, the same distinction can be made between the estimation spaces corresponding to these models as was made in Section 6. Hence, $\underline{c}'\underline{y}$ is an estimating function if and only if

$$\underline{c}'V\underline{e} = 0 \tag{121}$$

for every $\underline{c}'\underline{y}$ belonging to error. Again the set of all estimating functions is called the estimation set. Lemma 6 may be generalized here to

Lemma 7: A non-null linear function of $y_1, y_2, \ldots, y_n$ having positive variance cannot, at the same time, belong to the estimation set and the subset of functions in the error set having positive variances; i.e., the only linear functions of the y's which are common to the estimation set and the error set are those having zero variances.

Proof: Let $\underline{f}_1'\underline{y}, \ldots, \underline{f}_{n_e}'\underline{y}$ be n_e linearly independent non-null linear functions belonging to error and let F be given by (92). Re-define P to be an orthogonal matrix such that

(122) $$PFVF'P' = \begin{bmatrix} D_\lambda & 0 \\ 0 & 0 \end{bmatrix}$$

where $D_\lambda = D_\lambda(r_e \times r_e)$ is a diagonal matrix with positive diagonal elements $\lambda_1, \lambda_2, \ldots, \lambda_{r_e}$ where $r_e \leq n_e$. Let E and $\underline{z}$ be again defined by

(123) $$PF = E = \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} = \begin{bmatrix} E_1(r_e \times n) \\ E_2((n_e - r_e) \times n) \end{bmatrix} = \begin{bmatrix} \underline{e}'_1 \\ \vdots \\ \underline{e}'_{r_e} \\ \cdots \\ \underline{e}'_{n_e} \end{bmatrix}, \quad \underline{z} = PF\underline{y} = \begin{bmatrix} E_1\underline{y} \\ E_2\underline{y} \end{bmatrix}.$$

Then, $\underline{e}'_1\underline{y}, \ldots, \underline{e}'_{n_e}\underline{y}$ are uncorrelated linear functions of the y's belonging to error which form a basis of the error set; also, $\underline{e}'_1\underline{y}, \ldots, \underline{e}'_{r_e}$ represents a subset of the error set in which the elements all have positive variances. If $\underline{e}'\underline{y}$ is a non-null linear function belonging to error, then it can be written as

(124) $$\underline{e}'\underline{y} = \phi_1\underline{e}'_1\underline{y} + \phi_2\underline{e}'_2\underline{y} + \ldots + \phi_{r_e}\underline{e}'_{r_e}\underline{y} + \ldots + \phi_{n_e}\underline{e}'_{n_e}\underline{y} = \underline{\phi}'E\underline{y}.$$

Since

(125) $$\mathrm{Var}(\underline{e}'\underline{y}) = \mathrm{Var}(\underline{\phi}'E\underline{y}) = \underline{\phi}'PFVF'P'\underline{\phi}\sigma^2 = \sum_{j=1}^{r_e} \lambda_j\phi_j^2\sigma^2 ,$$

we have that ϕ_j^2 must be positive for at least one, $j = 1, 2, \ldots, r_e$ if $\underline{e}'\underline{y}$ is to have positive variance. If $\underline{e}'\underline{y}$ also belongs to the estimation set, then

(126) $$\begin{aligned} 0 &= \mathrm{Cov}(\underline{e}'\underline{y}, \underline{e}'_j\underline{y}) && j = 1, 2, \ldots, n_e \\ &= \underline{\phi}'EV\underline{e}_j\,\sigma^2 && j = 1, 2, \ldots, n_e \quad ; \\ &= \phi_j\underline{e}'_jV\underline{e}_j && j = 1, 2, \ldots, n_e \end{aligned}$$

but (126) implies that $\phi_j = 0 \quad j = 1, 2, \ldots, r_e$ since $\underline{e}_j' V \underline{e}_j = \lambda_j > 0$ $j = 1, 2, \ldots, r_e$. However, this implies that $\underline{e}'\underline{y}$ has zero variance. As a result, the only linear functions of the y's which belong to both the estimation set and the error set are those which have zero variance.

Let us now examine more closely the linear set $E_2\underline{y}$. This set consists of linear functions of the y's which are identically zero (i.e., their expected values are zero and their variances are zero). The matrix E_2 is such that

(127) $$AE_2' = 0_{m, n_e - r_e} , \qquad VE_2' = 0_{n, n_e - r_e} ;$$

i.e., the rows of E_2 represent a basis of the vector space orthogonal to both the rows of A and V. Moreover, since the relations

(128) $$\underline{0} = E_2\underline{y} = E_2A'\underline{p} = 0\underline{p}$$

provide no information at all about the parameters, the linear set $E_2\underline{y}$ can be ignored.

Next let us consider the linear set $E_1\underline{y}$. This represents the true error set since all elements of it have their expected values equal to zero and their variances positive. The rank of this true error set is r_e.

Since V is of rank $r \geq r_e$, then there exist $(n - r)$ independent $(n \times 1)$ vectors which are orthogonal to the rows of V; moreover, $(n - r) - (n_e - r_e) = (n - n_e) - (r - r_e) = n_o - r_o$ (where $r_o \leq n_o$ = Rank A) of these may be taken to be orthogonal to the rows of E.

Hence, if we let $R_1' = R_1'((n_o - r_o) \times n)$ be an orthogonal completion to the vector space generated by the rows of V and the rows of E and if we define $\underline{\psi}_1 = \underline{\psi}_1((n_o - r_o) \times 1)$ by $\underline{\psi}_1 = R_1'\underline{y}$, then we see that

(129) $$\mathrm{Var}(R_1'\underline{y}) = R_1'VR_1\sigma^2 = 0_{n_o - r_o,\, n_o - r_o} \quad .$$

From (129), it follows that

(130) $$\underline{\psi}_1 = R_1'\underline{y} = R_1'\,\mathcal{E}(\underline{y}) = R_1'A'\underline{p} \quad .$$

As a result, we see that the parameters satisfy the $(n_o - r_o)$ restrictions specified by (130).

Finally, if we let $T = T(r_o \times n)$ be an orthogonal completion to the vector space generated by the rows of E and the rows of R_1' and if we define $\underline{y}^*$ by

(131) $$\underline{y}^* = T\underline{y} \quad ,$$

then

(132) $$\mathcal{E}(\underline{y}^*) = TA'\underline{p}, \quad \mathrm{Var}(\underline{y}^*) = TVT'\sigma^2$$

where TVT' is symmetric positive definite of rank r_o.

In summary, we have partitioned the set of linear functions of $y_1, y_2, \ldots, y_n$ into four disjoint and exhaustive sets $T\underline{y}$, $R_1'\underline{y}$, $E_1\underline{y}$, $E_2\underline{y}$ each with a meaningful interpretation. Using this, we consider the following transformation

(133) $$\begin{bmatrix} \underline{y}^* \\ \underline{e}^* \end{bmatrix} = \begin{bmatrix} T \\ E_1 \end{bmatrix} \underline{y} \quad .$$

The transformed model is

(134) $$\mathcal{E}\begin{bmatrix} \underline{y}^* \\ \underline{e}^* \end{bmatrix} = \begin{bmatrix} TA'\underline{p} \\ \underline{0}_{r_e} \end{bmatrix}, \quad Var\begin{bmatrix} \underline{y}^* \\ \underline{e}^* \end{bmatrix} = \begin{bmatrix} TVT' & TVE_1' \\ E_1VT' & D_\lambda \end{bmatrix} \sigma^2 .$$

$$\underline{\Psi}_1 = R_1'A'\underline{p}$$

where the rows of $R_1'A'$ are independent of the rows of TA'.
The model (134) is characterized by a non-singular variance-covariance matrix and the presence of non-estimable restrictions. It may be analyzed by the methods of Section 7.

The computations may be simplified somewhat by using the following modification to the above approach. Let E_2 be defined as before by the equations

(135) $$VE_2' = 0_{n, n_e - r_e} \; ; \quad AE_2' = 0_{m, n_e - r_e}$$

let E_1 be defined by

(136) $$AE_1' = 0_{m, r_e} \; ; \quad E_2E_1' = 0_{n_e - r_e, r_e}$$

let R_1' be defined by

(137) $$VR_1 = 0_{n, n_o - r_o} \; ; \quad E_1R_1 = 0_{r_e, n_o - r_o}$$

and let T be defined by

(138) $$ET' = 0_{n_e, r_o}$$

$$R_1'T' = 0_{n_o - r_o, r_o}$$

Then the transformed model is

(139) $$\mathcal{E}\begin{bmatrix} T\underline{y} \\ E_1\underline{y} \end{bmatrix} = \begin{bmatrix} TA'\underline{p} \\ \underline{0}_{r_e} \end{bmatrix}, \mathrm{Var}\begin{bmatrix} T\underline{y} \\ E_1\underline{y} \end{bmatrix} = \begin{bmatrix} T\,V\,T' & T\,V\,E_1' \\ E_1VT' & E_1VE_1' \end{bmatrix} \sigma^2$$

$$\underline{\psi}_1 = R_1'A'\underline{p}$$

which can be analyzed by the methods of Section 7. The computations leading to E_2, E_1, R_1', T can be formidable in which case this type of analysis probably should be avoided. In any event, it does demonstrate that the effect of singularities in V appears in the form of the non-informative identities (128) and, more importantly, the 'non-estimable' restrictions (130).

Finally, if there should be any restrictions associated with the model (120), then they can be applied to either of the transformed models (134) or (139) and dealt with accordingly.

8.A. Solution of Equations (121). We restrict $\underline{e}'$ to be such that $\underline{e}'\underline{y}$ has positive variance. Hence $\underline{e}'$ lies in vector space generated by E_1. Now we may write $E_1 = U_1V$. Consider the equations

(140) $$E_1V\underline{c} = \underline{0}.$$

Two sets of solutions to (140), which are of interest, are the columns of

(141) $$C = [\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_m] = (V^2)^g VA' \quad \text{and} \quad R_1 ;$$

by Lemma 4, we have

(142) $$E_1 VC = U_1 V^2 (V^2)^g VA' = U_1 VA' = E_1 A' = 0; \quad \text{also } VR_1 = 0$$

Note that the computations involved with this approach may be as extensive as those associated with the transformed models (139).

In any event, this result may be of use in generalizing Theorem 2.

BIBLIOGRAPHY

Bose, R. C. [1950], Least Square Aspects of the Analysis of Variance, Mimeographed series, No. 9, North Carolina University.

Lucas, H. L. [1965], Lecture Notes - Statistics 691, North Carolina State University.

Plackett, R. L. [1960], Principles of Regression Analysis, Clarendon Press, Oxford.

Rao, C. R. [1945a], Generalization of Markov's theorem and tests of linear hypotheses, Sankhya 7, 9-15.

Rao, C. R. [1945b], Markov's theorem with linear restrictions on parameters, Sankhya 7, 16-19.

Rao, C. R. [1946], On the linear combination of observations and the general theory of least squares, Sankhya 7, 237-256.

Rao, C. R. [1965], Linear Statistical Inference and Its Applications, John Wiley and Sons, Inc., New York.