SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Editors : P. C. MAHALANOBIS, C. R. RAO

SERIES A, VOL. 33 | DECEMBER 1971 | PART 4

UNIFIED THEORY OF LINEAR ESTIMATION*

By C. RADHAKRISHNA RAO

Indian Statistical Institute

SUMMARY : We consider a general Gauss-Markoff model $(Y, X\beta, \sigma^2 V)$, where $E(Y) = X\beta$, $D(Y) = \sigma^2 V$. There may be deficiency in $R(X)$, the rank of X and V may be singular.

Two unified approaches to the problem of finding BLUE's (minimum variance linear unbiased estimators) have been suggested. One is a direct approach where the problem of inference on the unknown β is reduced to the numerical evaluation of the inverse of a partitioned matrix. The second is an analogue of least squares, where the matrix used in defining the quadratic form in $(Y - X\beta)$ to be minimized is a g-inverse of $(V + XUX')$ in all situations, whether V is nonsingular or not, where U is arbitrary subject to a condition.

Complete robustness of BLUE's under different alternatives for V has been examined.

A study of BLE's (minimum mean square estimators) without demanding unbiasedness is initiated and a case has been made for further examination.

The unified approach is made possible through recent advances in the calculus of generalized inverse of matrices (see the recent book by Rao and Mitra, 1971a).

1. INTRODUCTION

It is indeed a great honour for a statistician to be called upon to preside over a conference of econometricians. I have accepted this position only with the hope of learning from you about the statistical aspects of econometric problems and examining the adequacy of the existing statistical methods to meet your needs. Most of the developments in statistical methodology have been greatly influenced by problems in biology and technology, but not to the same extent by problems in economics although statistics had its origin in the compilation and interpretation of economic data. It is also a historical fact that journals like *Biometrika*, *Biometrics* and *Technometrics* have been started and run by statisticians while *Econometrica* has remained outside the orbit of statisticians. But some of the recent contributions in *Econometrica* have convinced me that there is an equally rich field for economists and statisticians to do collaborative research.

I want to take this opportunity of discussing with you some of the recent developments in the estimation of parameters in the Gauss-Markoff model leading to a *unified theory*, which may be of interest in econometric work.

*Presidential address prepared for the Annual Conference of the Econometric Society, 1971.

1.1. *The Gauss-Markoff model*

Notation. In this paper we use bold face letters to represent matrices and vectors, like, $\boldsymbol{X}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{\beta}$, etc. The vector space generated by the columns of a matrix $\boldsymbol{X}$ is represented by $\mathcal{M}(\boldsymbol{X})$. The vector space orthogonal to $\mathcal{M}(\boldsymbol{A})$ is denoted by $\mathcal{M}(\boldsymbol{A}^{\perp})$ where $\boldsymbol{A}^{\perp}$ is a matrix of maximum rank with its columns orthogonal to the columns of $\boldsymbol{X}$.

If $\boldsymbol{V}$ is a n.n.d. (non negative definite) matrix, the expression

$$\| \boldsymbol{p} \| = (\boldsymbol{p}'\boldsymbol{V}\boldsymbol{p})^{\frac{1}{2}}$$

where $\boldsymbol{p}$ is a vector, is called the V-norm of $\boldsymbol{p}$. Similarly the value $\hat{\boldsymbol{\beta}}$ which minimises

$$(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{M}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}),$$

or more generally which satisfies the normal equation with no restriction on $\boldsymbol{M}$,

$$\boldsymbol{X}'\boldsymbol{M}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{M}\boldsymbol{Y}$$

where $\boldsymbol{Y}$ and $\boldsymbol{\beta}$ are vectors, is called a M-least squares solution of $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}$.

The symbols E, V, C and D will be used for expectation, variance, covariance and dispersion (variance and covariance) of random variables. $N_p(\boldsymbol{m}, \boldsymbol{\Sigma})$ denotes a p-variate normal distribution of a p-vector variable with mean $\boldsymbol{m}$ and dispersion matrix $\boldsymbol{\Sigma}$. χ^2_f stands for a chi-square variable on f degrees of freedom. $\boldsymbol{Y} \sim N_p$ means that the random variable $\boldsymbol{Y}$ has a p-variate normal distribution.

$(\boldsymbol{X} : \boldsymbol{V})$ denotes a partitioned matrix and $R(\boldsymbol{X})$, the rank of matrix $\boldsymbol{X}$. A matrix with all zero entries is denoted by $\boldsymbol{0}$.

BLUE (best linear unbiased estimator) stands for a linear unbiased estimator with minimum variance, LUE for a linear unbiased estimator and BLE for a linear estimator with minimum mean square error.

Gauss-Markoff model. The Gauss-Markoff model may be represented by the triplet $(\boldsymbol{Y}, \boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{V})$ where $\boldsymbol{Y}$ is n-vector of observations, $\boldsymbol{X}$ is $n \times m$ design matrix such that

$$E(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}, \; D(\boldsymbol{Y}) = \sigma^2\boldsymbol{V}. \qquad \ldots \quad (1.1.1)$$

In (1.1.1), $\boldsymbol{\beta}$ is an unknown m-vector parameter, σ^2 is an unknown scale parameter and $\boldsymbol{V}$ is a non-negative definite (n.n.d.) matrix, which may be known, partly known or unknown. In econometric applications, $\boldsymbol{X}$ is called the matrix of explanatory variables and $\boldsymbol{Y}$, the vector of the dependent variable. The problem is one of estimating $\boldsymbol{\beta}$ by linear functions of $\boldsymbol{Y}$, and σ^2 by a quadratic function of $\boldsymbol{Y}$, without making any assumption about the actual distribution of $\boldsymbol{Y}$. In our discussion we assume that $\boldsymbol{V}$ is completely known.[1]

[1]Some developments are taking place in the simultaneous estimation of $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{V}$. See the recent papers by Rao (1970, 1971a, 1971b, 1972) on Minque theory and J.N.K. Rao and K. Subrahmaniam (1971) on simultaneous estimation.

Gauss introduced the method of *least squares* in 1821, and showed that it provides the BLUE (best linear unbiased estimator) of $\boldsymbol{\beta}$ under the assumption (see Gauss, 1855; Plackett, 1949),

$$R(\boldsymbol{X}) = m \text{ and } \boldsymbol{V} = \boldsymbol{I} \text{ or diagonal.} \quad \dots \quad (1.1.2)$$

Later, Markoff (1912) and David and Neyman (1938) made a systematic presentation of the theory under the same assumptions (1.1.2), while Aitken (1934) extended the method to the case

$$R(\boldsymbol{X}) = m \text{ and } |\boldsymbol{V}| \neq 0 \quad \dots \quad (1.1.3)$$

where $\boldsymbol{V}$ is not necessarily diagonal.

Bose (1943) was the first to consider the situation

$$R(\boldsymbol{X}) = r < m, \ \boldsymbol{V} = \boldsymbol{I} \quad \dots \quad (1.1.4)$$

where only certain and not all linear functions of $\boldsymbol{\beta}$ are unbiasedly estimable. The deficiency in the rank of $\boldsymbol{X}$ is described as *multicollinearity* in econometric work. Rao (1945a) showed that the method of least squares is applicable even when $R(\boldsymbol{X}) < m$ and $\boldsymbol{V}$ is nonsingular.

One may also consider given linear restrictions (also called *constraints*) on the $\boldsymbol{\beta}$ parameter such as $\boldsymbol{R\beta} = \boldsymbol{c}$ in which case the Gauss-Markoff model can be written as

$$(\boldsymbol{Y}, \boldsymbol{X\beta} \,|\, \boldsymbol{R\beta} = \boldsymbol{c}, \sigma^2\boldsymbol{V}). \quad \dots \quad (1.1.5)$$

The general theory in such a case has been fully worked out (see Rao, 1945b) when

$$R(\boldsymbol{X}' : \boldsymbol{R}') \leqslant m, \ |\boldsymbol{V}| \neq 0.$$

The most general situation is the set up $(\boldsymbol{Y}, \boldsymbol{X\beta}, \sigma^2\boldsymbol{V})$ where

$$R(\boldsymbol{X}) \leqslant m \text{ and } \boldsymbol{V} \text{ possibly singular.} \quad \dots \quad (1.1.6)$$

The model (1.1.5) with constraints on $\boldsymbol{\beta}$ is, indeed, a special case of (1.1.6) since we can define

$$\boldsymbol{Y}_e' = (\boldsymbol{Y}' : \boldsymbol{c}'), \ \boldsymbol{X}_e' = (\boldsymbol{X}' : \boldsymbol{R}'),$$

$$\boldsymbol{V}_e = \begin{pmatrix} \boldsymbol{V} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, \quad \dots \quad (1.1.7)$$

and write (1.1.5) as $(\boldsymbol{Y}_e, \boldsymbol{X}_e\boldsymbol{\beta}, \sigma^2\boldsymbol{V}_e)$ recognising that the dispersion matrix of the constant vector $\boldsymbol{c}$ is null. In (1.1.7) and elsewhere $\boldsymbol{0}$ stands for null matrices.

There has been no satisfactory approach for the general case (1.1.6). The methods described by Mitra and Rao (1968), Theil (1971), Zyskind and Martin (1969) are all a little involved and no attempts were made to provide a unified treatment valid for all situations. Rao and Mitra (1971b) suggested one unified method of least squares which holds good whether $\boldsymbol{V}$ is singular or not. I would like to describe to you two approaches to the problem providing a *unified theory* of linear estimation without any assumptions on $\boldsymbol{X}$ or $\boldsymbol{V}$ and also on linear hypotheses which may be required to be tested on the basis of (1.1.6). This has been possible due to recent developments in the theory and application of inversion of singular and rectangular matrices which has been developed in a number of papers (Rao, 1955, 1962, 1967), and fully described in a recent book by Rao and Mitra (1971a). (See also Rao and Mitra, 1971c).

1.2. *Some lemmas on g-inverse*

In this section we give a brief introduction to g-inverse of matrices and prove some new lemmas useful in discussions on estimation from linear models.

Definition 1 : Let A be an $m \times n$ matrix. A g-inverse of A is an $n \times m$ matrix, denoted by A^-, satisfying the condition

$$AA^-A = A. \qquad \ldots \quad (1.2.1)$$

An immediate application of A^- is in solving a consistent linear equation $Ax = q$, whose solution can be expressed as $x = A^-q$. Indeed the entire class of solutions to $Ax = q$ can be written as

$$x = A^-q+(I-A^-A)z \qquad \ldots \quad (1.2.2)$$

where I is the identity matrix and z is an arbitrary vector.

Definition 2 : A matrix, denoted by $A^-_{m(V)}$, is said to be a minimum V-norm inverse of A if $\hat{x} = A^-_{m(V)} q$ is a solution of a consistent equation $Ax = q$ with the the smallest V-norm (being defined as $\sqrt{x'Vx}$) where V is an n.n.d. matrix. $\hat{x}$ is called the minimum V-norm solution of $Ax = q$.

Defination 3 : Let $Ax = q$ be a not necessarily consistent equation. Then a matrix, denoted by $A^-_{l(W)}$ is said to be a W-least squares inverse of A if $\hat{x} = A^-_{l(W)} q$ minimises the quadratic form

$$(Ax-q)' \; W(Ax-q) \qquad \ldots \quad (1.2.3)$$

where W is an n.n.d. matrix. $\hat{x}$ is called a W-least squares solution of $Ax = q$.

Definition 4 : A matrix denoted by A^+_{WV} is said to be a minimum V-norm W-least squares inverse of A if $x = A^+_{WV}\, q$ is a W-least squares solution of $Ax = q$ with a minimum V-norm.

The existence of A^-, $A^-_{m(V)}$, $A^-_{l(W)}$ and A^+_{WV}, explicit expressions for them and computational procedures are discussed in Rao and Mitra (1971a). Some of these inverses are not unique in which case the classes are represented by $\{A^-\}$, $\{A^-_{m(V)}\}$, ..., etc. When we write an equation such as $A^- = B^-$, it means that corresponding to a given choice of A^- there is a choice of B^- equal to A^-.

The following results are used in the development of a unified theory of linear estimation.

Theorem 1.1 : *Let the matrices V, $V+A^*UA$ and T^- be n.n.d. Then*

(i) $A^-_{m(V)} = A^-_{m(V+A^*UA)}$, *$U$ arbitrary,* ... (1.2.4)

(ii) $(A^*)^-_{l(T^-)} \in \{[A^-_{m(V)}]^*\}$ ... (1.2.5)

where $T = V+A^*UA$ such that $\mathcal{M}(A^*) \subset \mathcal{M}(T)$.

(iii) $(A^*)^-_{l(V^{-1})} = [A^-_{m(V)}]^*$ *if* $|V| \neq 0$, ... (1.2.6)

where A^* represents the conjugate transpose of A.

The results are easily established using the characterizations of these inverses given in Rao and Mitra (1971a). Theorem 1.1 is important since it demonstrates that the problem of finding a minimum norm solution of a linear equation is algebraically equivalent to the problem of finding a least squares solution of a similar linear equation. (This is the reason why the least squares solution provides minimum variance estimators in the Gauss-Markoff model).

Let V be an n.n.d. (non negative definite) matrix of order n and X be another matrix of order $n \times m$. Further let

$$\begin{pmatrix} V & X \\ X' & 0 \end{pmatrix}^- = \begin{pmatrix} C_1 & C_2 \\ C_3 & -C_4 \end{pmatrix} \qquad \dots \quad (1.2.7)$$

be any choice of g-inverse where X' denotes the transpose of X.

Theorem 1.2 : *Let V, C_1, C_2, C_3, C_4 and X be as defined in* (1.2.7). *Then the following hold* :

(i) $$\begin{pmatrix} V & X \\ X' & 0 \end{pmatrix}^- = \begin{pmatrix} C_1 & C_3' \\ C_2' & -C_4 \end{pmatrix} \qquad \dots \quad (1.2.8)$$

is another choice of g-inverse.

(ii) $XC_3X = X, \quad XC_2'X = X.$ ... (1.2.9)

i.e., C_3 and C_2' are g-inverses of X.

(iii) $X'C_1X = 0, \quad VC_1X = 0, \quad X'C_1V = 0.$ ... (1.2.10)

(iv) $VC_2X' = XC_2'V = XC_4X' = XC_4'X' = VC_3'X' = XC_3V.$ ... (1.2.11)

(v) $VC_1VC_1V = VC_1V, \ \mathrm{Tr}\ VC_1 = R(V:X) - R(X).$ ... (1.2.12)

(vi) $\begin{pmatrix} C_1 \\ C_3 \end{pmatrix}$ *is a g-inverse of* $(V:X)$. ... (1.2.13)

Note that C_3' and C_2 are in fact minimum V-norm g-inverses of X'.

The results are easily established by considering the equation

$$\begin{pmatrix} V & X \\ X' & 0 \end{pmatrix}\begin{pmatrix} C_1 & C_2 \\ C_3 & -C_4 \end{pmatrix}\begin{pmatrix} V & X \\ X' & 0 \end{pmatrix} = \begin{pmatrix} V & X \\ X' & 0 \end{pmatrix}.$$

2. Is there a new problem to be solved ?

Let us define

$$(\boldsymbol{Y}, \boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}), \quad R(\boldsymbol{X}) = m \qquad \ldots \quad (2.1)$$

i.e., when the rank of $\boldsymbol{X}$ is full as the Standard Gauss-Markoff model or SGM model in short. We show that all the other cases (1.1.3)-(1.1.6) can be reduced to SGM to which the method of least squares propounded by Gauss (1821) is applicable. Thus, in a sense, there is no new problem to be solved.

(i) Suppose the model is

$$(\boldsymbol{Y}, \boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}), \quad R(\boldsymbol{X}) = r < m \qquad \ldots \quad (2.2)$$

with some deficiency in the rank of $\boldsymbol{X}$ (i.e., multicollinearity of $\boldsymbol{X}$). Consider the singular value decompostion of $\boldsymbol{X}$

$$\boldsymbol{X} = \boldsymbol{P}\Lambda\boldsymbol{Q}', \qquad \ldots \quad (2.3)$$

where Λ is a $r \times r$ diagonal matrix of singular values of $\boldsymbol{X}$ and $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I}$. Introducing a new r-vector parameter $\boldsymbol{\gamma} = \Lambda\boldsymbol{Q}'\boldsymbol{\beta}$ the model (2.2) reduces to

$$(\boldsymbol{Y}, \boldsymbol{P}\boldsymbol{\gamma}, \ \sigma^2\boldsymbol{I}) \qquad \ldots \quad (2.4)$$

which is the SGM model. Indeed, the reduction of (2.2) to (2.4) through (2.3) leads to interesting results. The Gauss least squares estimator of $\boldsymbol{\gamma}$ is $\hat{\boldsymbol{\gamma}} = \boldsymbol{P}'\boldsymbol{Y}$ and the estimate of any estimable linear parametric function $\boldsymbol{p}'\boldsymbol{\beta}$ is then $\boldsymbol{p}'\boldsymbol{Q}\Lambda^{-1}\hat{\boldsymbol{\gamma}}$ $= \boldsymbol{p}'\boldsymbol{Q}\Lambda^{-1}\boldsymbol{P}'\boldsymbol{Y}$. The variance of the estimator is $V(\boldsymbol{p}'\boldsymbol{Q}\Lambda^{-1}\hat{\boldsymbol{\gamma}}) = \sigma^2\boldsymbol{p}'\boldsymbol{Q}\Lambda^{-2}\boldsymbol{Q}'\boldsymbol{p}$ and an unbiased estimator of σ^2 is $(n-r)^{-1}(\boldsymbol{Y}'\boldsymbol{Y}-\hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}})$. For a direct least squares approach for (2.2) see Rao (1965).

(ii) Let us consider the model

$$(\boldsymbol{Y}, \boldsymbol{X}\boldsymbol{\beta} \,|\, \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{c}, \ \sigma^2\boldsymbol{I}), \quad R(\boldsymbol{X}' : \boldsymbol{R}') = m. \qquad \ldots \quad (2.5)$$

Using the constraints $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{c}$ we may eliminate some of the parameters and obtain a reduced model

$$(\boldsymbol{Y}, \boldsymbol{X}_1\boldsymbol{\beta}_1, \sigma^2\boldsymbol{I}) \qquad \ldots \quad (2.6)$$

which is in SGM. In practice the reduction need not be actually done since

$$\min_{\boldsymbol{\beta}_1}(\boldsymbol{Y}-\boldsymbol{X}_1\boldsymbol{\beta}_1)'(\boldsymbol{Y}-\boldsymbol{X}_1\boldsymbol{\beta}_1) = \min_{\boldsymbol{R}\boldsymbol{\beta}=\boldsymbol{c}} (\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}) \qquad \ldots \quad (2.7)$$

and an estimable parametric function $\boldsymbol{p}_1'\boldsymbol{\beta}_1$ can be written as $\boldsymbol{p}'\boldsymbol{\beta}$. Then $\boldsymbol{p}'\hat{\boldsymbol{\beta}} = \boldsymbol{p}_1'\hat{\boldsymbol{\beta}}_1$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_1$ are the values of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_1$ minimising the right and left hand sides of (2.7). If $R(\boldsymbol{X}' : \boldsymbol{R}') < m$, then (2.6) is of the type (2.2).

(iii) Let $\boldsymbol{V}^{\frac{1}{2}}$ be a square root of the p.d. (positive definite) matrix $\boldsymbol{V}$. Then the model

$$(\boldsymbol{Y}, \boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{V}), \ R(\boldsymbol{X}) = m, \ |\boldsymbol{V}| \neq 0 \qquad \ldots \quad (2.8)$$

considered by Aitken (1934) can be reduced by the transformation

$$\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{Y} = \boldsymbol{Y}_t, \ \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{X} = \boldsymbol{X}_t \text{ to } (\boldsymbol{Y}_t, \boldsymbol{X}_t\,\boldsymbol{\beta}, \sigma^2\boldsymbol{I}) \qquad \ldots \quad (2.9)$$

which is in SGM. An application of Gauss least squares method to (2.9) gives what is called Aitken's generalized least squares estimator of β.

(iv) If in (2.8), $R(X) = r < m$, then the reduced model (2.9) is of the type (2.2) and the formulae derived for (2.2) become applicable.

(v) Finally we consider the GGM-model (General Gauss-Markoff model),

$$(Y, X\beta, \sigma^2 V), R(X) = r \leqslant m \text{ and } R(V) = s < n. \qquad \dots \quad (2.10)$$

Let $F_1, \dots, F_s$ be eigen vectors corresponding to the non-zero eigen values $\lambda_1^2, \dots, \lambda_s^2$ of V and $N_1, \dots, N_u$ be $n-s = u$ eigen vectors corresponding to the zero root, and define the matrices

$$F = (\lambda_1^{-1}F_1 : \dots : \lambda_s^{-1}F_s), \quad N = (N_1 : \dots : N_u). \qquad \dots \quad (2.11)$$

Then make the transformation

$$Y_1 = F'Y, Y_2 = N'Y, X_1 = F'X, X_2 = N'X \qquad \dots \quad (2.12)$$

leading to the reduced model

$$(Y_1, X_1\beta \mid X_2\beta = Y_2, \sigma^2 I) \qquad \dots \quad (2.13)$$

which is of the type (2.5) with $R(X_1' : X_2') \leqslant m$, and hence the formulae derived for (2.5) can be used. (Observe that Y_2 is a constant vector, since $D(Y_2) = \sigma^2 N'VN = 0$, so that $Y_2 = X_2\beta$ is in the nature of constraints on β). The method described involves only the spectral decomposition of V and a straight forward application of the least squares theory with constraints on the β parameter—the approach first suggested by Mitra and Rao, (1968).

3. Unified theory—direct approach

We have seen that no new problem arises when departures from the standard linear model (SGM) are considered such as constraints on parameters, deficiency in the rank of the design matrix (multicollinearity of the explanatory variables), correlations in the series of the dependent variable, and collinearity leading to a singular dispersion matrix of the vector of dependent variables. It has been the practice among certain authors to call the formulae and methods appropriate for one type of the above departures from the SGM as "generalized least squares" and more than one type of departures as "generalization of generalized least squares" although all situations can essentially be reduced to the original Gauss-Markoff model.

The discussion of section 2 does not, however, preclude us from attempting a unified theory (a single method to cover all situations) of linear estimation applicable to the general model, $(Y, X\beta, \sigma^2 V)$ where the rank of X is *possibly* deficient and the matrix V is *possibly* singular. While the discussion of section 2 did not make any *direct* reference to generalized inverse of a matrix and depended only on the classical spectral and singular value decomposition of matrices, the unified theory developed in sections 3 and 4 depends to some extent on the calculus of generalized inverses. Section 3 deals with a direct approach to the problem by minimizing the variance of a linear unbiased estimator while section 4 links it up with the algebraically equivalent least squares theory.

It will be seen that in the unified method no attempt need be made in the beginning to find out to which type (among 1.1.1–1.1.6) a given model belongs and to reduce it by preliminary computations to the SGM. Such knowledge may be useful in the final stages of computations and interpretation of results. The unified theory appears to be somewhat simpler as it lays down a *common numerical* procedure for all situations and avoids the complicated algebraic notations generally used in the discussion of the GGM model.

3.1. *Consistency of the linear model*

We consider the GGM model

$$(Y, X\beta, \sigma^2 V) \qquad \ldots \quad (3.1.1)$$

without any assumptions on X or V. As observed earlier the model (3.1.1) includes the case of any given constraints on β.

When V is singular, there are some natural restrictions on the dependent variable Y, which may be tested to make sure that there is no obvious inconsistency in the model. One such restriction is

$$L'X = 0, \quad L'V = 0 \Longrightarrow L'Y = 0 \qquad \ldots \quad (3.1.2)$$

where L is a vector, or in other words $Y \epsilon \mathcal{M}(V : X)$. This condition is automatically satisfied if V is nonsingular.

Further

$$L'V = 0 \Longrightarrow L'(Y - X\beta) = 0 \qquad \ldots \quad (3.1.3)$$

i.e., the vector Y and the parameter β must be such that $Y - X\beta \epsilon \mathcal{M}(V)$, which is again automatically satisfied if V is nonsingular. Thus singularity of V implies some restrictions on both Y and the unknown parameter β.

3.2. *The basic theorem of linear estimation*

Let us consider the GGM model (3.1.1) and one choice of g-inverse as in Theorem 1.2,

$$\begin{pmatrix} V & X \\ X' & 0 \end{pmatrix}^{-} = \begin{pmatrix} C_1 & C_2 \\ C_3 & -C_4 \end{pmatrix} \qquad \ldots \quad (3.2.1)$$

where V and X are as defined in the model (3.1.1). Once a g-inverse is computed by a suitable procedure, we seem to have a *Pandora box* supplying all the ingredients needed for obtaining the BLUE's, their variances and covariances, an unbiased estimate of σ^2, and constructing test criteria without any further computations except for a few matrix multiplications. Theorem 3.1 provides the basic results. Thus the problem of inference from a linear model is reduced to the *numerical problem* of finding an inverse (or g-inverse) of the symmetric matrix given in (3.2.1).

Theorem 3.1 : *Let* C_1, C_2, C_3, C_4 *be as defined in* (3.2.1.). *Then the following hold* :

(i) [Use of C_2 or C_3]. *The BLUE of an estimable parametric function* $p'\beta$ *is* $p'\hat{\beta}$, *where*

$$\hat{\beta} = C_2'Y \quad \text{or} \quad \hat{\beta} = C_3Y. \qquad \dots \quad (3.2.2)$$

(ii) [Use of C_4]. *The dispersion matrix of* $\hat{\beta}$ *is* $\sigma^2 C_4$ *in the sense that*

$$V(p'\hat{\beta}) = \sigma^2 p' C_4 p$$

$$\text{cov}\,(p'\hat{\beta}, q'\hat{\beta}) = \sigma^2 p' C_4 q = \sigma^2 q' C_4 p \qquad \dots \quad (3.2.3)$$

whenever $p'\beta$ *and* $q'\beta$ *are estimable.*

(iii) [Use of C_1]. *An unbiased estimator of* σ^2 *is*

$$\hat{\sigma}^2 = f^{-1}\ Y'C_1Y \qquad \dots \quad (3.2.4)$$

where $f = R(V : X) - R(X)$.

Proof of (i) : If $L'Y$ is an unbiased estimator of $p'\beta$, then $X'L = p$. Subject to this condition $V(L'Y) = \sigma^2 L'VL$ or $L'VL$ has to be minimized to obtain the BLUE of $p'\beta$.

Let L_* be an optimum choice and L be any other vector such that $X'L = X'L_*$. Then

$$L'VL = (L - L_* + L_*)'V(L - L_* + L_*)$$
$$= (L - L_*)'V(L - L_*) + L_*'VL_* + 2L_*'V(L - L_*) \geqslant L_*'VL_*$$

iff $L_*'V(L - L_*) = 0$ whenever $X'(L - L_*) = 0$, i.e., $VL_* = -XK_*$ for a suitable K_*. Then L_* and K_* satisfy the equations

$$VL_* + XK_* = 0,$$
$$X'L_* \qquad = p. \qquad \dots \quad (3.2.5)$$

We observe that the equations (3.2.5) admit a solution and any two solutions L_{1*} and L_{2*} satisfy the condition $V(L_{1*} - L_{2*}) = 0$. Since (3.2.5) is consistent, a solution is given by $L_* = C_2 p$, $K_* = -C_4 p$, or $L_* = C_3' p$, $K_* = -C_4 p$. Then the BLUE of $p'\beta$ is $L_*'Y = p'C_2'Y = p'C_3Y$.

Proof of (ii) : We use the fact $p = X'M$ for some M. Then

$$V(p'C_2'Y) = \sigma^2 M'(XC_2'V)C_2X'M$$
$$= \sigma^2 M'XC_4(X'C_2X')M \text{ using (1.2.11)}$$
$$= \sigma^2 M'XC_4X'M \qquad \text{using (1.2.9)}$$
$$= \sigma^2 p'C_4 p.$$

Similarly

$$\text{cov}\,(p'C_2'Y, q'C_2'Y) = \sigma^2 p'C_4 q = \sigma^2 q'C_4 p.$$

Proof of (iii) : Since $X'C_1V = 0$ and $X'C_1X = 0$ (using (1.2.10)),

$$Y'C_1Y = (Y - X\beta)'C_1(Y - X\beta)$$

$$E[(Y - X\beta)'C_1(Y - X\beta)] = \sigma^2 \text{Tr}\, C_1[E\{(Y - X\beta)(Y - X\beta)'\}]$$
$$= \sigma^2 \text{Tr}\, C_1 V = \sigma^2[R(V : X) - R(X)]$$

using (1.2.12).

Corollary 3.1.1 : *If $\sigma^2 D$ is the dispersion matrix of LUE's of a set of estimable parametric functions and $\sigma^2 F$ that of their BLUE's then $D-F$ is n.n.d.*

Let $p_1'\beta, \ldots, p_k'\beta$ be k estimable parametric functions and consider the estimation of the single function $p'\beta$ where $p = m_1 p_1 + \ldots + m_k p_k$. If t_i is a LUE (linear unbiased estimator) and T_i is the BLUE of $p_i'\beta$, then $\Sigma m_i t_i$ is the LUE and $\Sigma m_i T_i$ is a BLUE of $p'\beta$. By definition

$$V(\Sigma m_i t_i) \geqslant V(\Sigma m_i T_i) \quad \text{or} \quad \sigma^2 m' D m \geqslant \sigma^2 m' F m$$

for all $m' = (m_1, \ldots, m_k)$ which implies $D-F$ is n.n.d.

Corollary 3.1.2 : *If D and F are as defined in Corollary 3.1.1., then*

(i) $Tr\, D \geqslant Tr\, F$.

(ii) $|D| \geqslant |F|$.

(iii) $Tr\, QD \geqslant Tr\, QF$ *where Q is any n.n.d. matrix, and*

(iv) $\lambda_{max}(D) \geqslant \lambda_{max}(F)$

where $\lambda_{max}(G)$ denotes the maximum eigen value of G.

The results of Corollary 3.1.2 are simple consequences of $D-F$ being n.n.d. Thus the BLUE's are optimal according to any of the criteria (i)–(iv) of Corollary 3.1.2. Result (iii) implies that if e is the vector of errors in BLUE's and g is the vector of errors in LUE's then $E(e'Qe) \leqslant E(g'Qg)$ for any n.n.d. matrix Q.

Theorem 3.2 : *Let $P'\hat{\beta}$ be the vector of BLUE's of a set of k estimable parametric functions $P'\beta$, $R_0^2 = Y'C_1Y$ and f be as defined in Theorem 3.1. If $Y \sim N_n(X\beta, \sigma^2 V)$, then :*

(i) *$P'\hat{\beta}$ and $Y'C_1Y$ are independently distributed, with*

$$P'\hat{\beta} \sim N_k(P'\beta, \sigma^2 D), \quad \ldots \quad (3.2.6)$$

and

$$Y'C_1Y \sim \sigma^2 \chi_f^2. \quad \ldots \quad (3.2.7)$$

where $D = P'C_4P$.

(ii) *Let $P'\beta = w$ be the null hypothesis. The null hypothesis is consistent iff*

$$DD^-u = u \quad \ldots \quad (3.2.8)$$

where $u = P'\hat{\beta} - w$. If the hypothesis is consistent then,

$$F = \frac{u'D^-u}{h} \div \frac{R_0^2}{f}, \quad h = R(D) \quad \ldots \quad (3.2.9)$$

has a central F distribution on h and f degrees of freedom when the hypothesis is true, and a non-central F distribution when the hypothesis is wrong.

Proof of (i) : The result (3.2.6) is easy to establish. (3.2.7) follows since $Y'C_1Y = (Y-X\beta)'C_1(Y-X\beta)$, and by (1.2.12)

$$VC_1VC_1V = VC_1V$$

which is an NAS condition for a χ^2 distribution. (See Rao, 1965a, p. 443, and also Rao and Mitra, 1971a). The degrees of freedom of the χ^2 is

$$\text{Tr } VC_1 = R(V:X) - R(X) = f, \text{ using (1.2.12).}$$

Since $P'\beta$ is estimable, $P' = QX$ for some Q. Then $P'\hat{\beta} = QXC_3Y$. The condition for independence of $Y'C_1Y$ and QXC_3Y is

$$VC_1 VC_3'X'Q' QXC_3V = 0$$

which is true since

$$VC_1VC_3'X' = VC_1XC_3V = 0,$$

using (1.2.11) and (1.2.10).

Proof of (ii) : The hypothesis $P'\beta = w$ is consistent if for any vector m.

$$V[m'(P'\hat{\beta}-w)] = 0 \Longrightarrow m'(P'\hat{\beta}-w) = 0$$

i.e. $m'Dm = 0 \Longrightarrow m'u = 0$ or $u \in \mathcal{M}(D)$, for which a NAS condition is $DD^-u = u$, for any g-inverse D^- of D.

Since $D(u) = \sigma^2 D$ and $DD^-D = D$

$$\frac{u'D^-u}{\sigma^2} \sim \chi^2_h, \quad h = R(D).$$

using the result (viii) on p. 443 of Rao (1965a). R_0^2 is distributed as χ^2_f independently of u. Hence the result (3.2.9) follows.

In Theorem 3.2, the numerator of the F statistic for testing the linear hypothesis $P'\beta = w$ was obtained in the form $u'D^-u$ which involved the estimation of deviations in individual hypotheses, computation of their dispersion matrix and its inverse.

Theorem 3.3, provides an alternative method of computing the numerator, as in the theory of least squares.

Theorem 3.3 : *Let*

$$\begin{pmatrix} V & X \\ 0 & P \end{pmatrix}^- = \begin{pmatrix} E_1 \\ E_3 \end{pmatrix}$$

be one choice of the g-inverse of the matrix involved. Then

$$u'D^-u = Y'E_1 \begin{pmatrix} Y \\ w \end{pmatrix} - Y'C_1Y. \qquad \ldots \quad (3.2.10)$$

For consistency of the hypothesis $P'\beta = w$ the test is

$$\begin{pmatrix} V & X \\ 0 & P \end{pmatrix} \begin{pmatrix} E_1 \\ E_3 \end{pmatrix} \begin{pmatrix} Y \\ w \end{pmatrix} = \begin{pmatrix} Y \\ w \end{pmatrix}. \qquad \ldots \quad (3.2.11)$$

Note. If the object is only to estimate linear parametric functions and test hypotheses we do not need the matrix C_4 as defined in (3.2.1). The matrices needed are C_1, C_3 and E_1, E_3. Then C_1, C_3 and E_1, E_3 can be obtained from g-inverses as indicated in (3.2.12).

$$(V:X)^- = \begin{pmatrix} C_1 \\ C_3 \end{pmatrix}, \quad \begin{pmatrix} V & X \\ 0 & P \end{pmatrix}^- = \begin{pmatrix} E_1 \\ E_3 \end{pmatrix}. \qquad \ldots \quad (3.2.12)$$

4. Unified theory—least squares approach

In section 3, the problem of BLUE is formulated as the minimization of a quadratic form $L'VL$ (the variance) subject to the restriction $X'L = p$ (unbiasedness). This is equivalent to finding a minimum V-norm solution of the equation $X'L = p$, V-norm being defined as $\sqrt{L'VL}$. The solution is provided by a minimum V-norm g-inverse of X' (defined in section 1.2),

$$L = (X')^-_{m(V)}\, p$$

giving the BLUE of $p'\beta$ as

$$p'[(X')^-_{m(V)}]'\ Y \quad \dots \quad (4.1)$$

for any choice of g-inverse. Now we use the relationship between a minimum V-norm inverse and a least squares inverse established in Theorem 1.1,

$$(X^-_{l(T^-)})' \in \{(X')^-_{m(V)}\} \quad \dots \quad (4.2)$$

where T^- is an n.n.d. g-inverse of $T = V+XUX'$ for any choice of the matrix U, subject only to the condition $\mathcal{M}(X) \subset \mathcal{M}(T)$. (For instance if $\mathcal{M}(X) \subset \mathcal{M}(V)$ or $|V| \neq 0$, then U can be chosen to be a null matrix). Using (4.2) the BLUE of $p'\beta$, given in (4.1), can be written as

$$p'X^-_{l(T^-)}\, Y = p'\hat{\beta} \quad \dots \quad (4.3)$$

where $\hat{\beta}$ minimizes

$$(Y-X\beta)'\ T^-(Y-X\beta) \quad \dots \quad (4.4)$$

i.e., $\hat{\beta}$ is the T^- least squares solution of the equation $Y = X\beta$.

The algebraic relationship in (4.2) holds the main key to the question, "why least squares theory for obtaining BLUE's ?" It also shows that the choice of the matrix in defining the nrom (4.4) is a function of V and X in general. Thus the principle of least squares does not seem to be a primitive postulate but an algebraic artifact to obtain minimum variance estimators.

A unified theory of linear estimation through the method of least squares is contained in Theorem 4.1 where the matrix U is taken as $k^2 I$, with an arbitrary $k \neq 0$. However, the basic result of the unified theory of least squares is contained in Theorem 4.3.

Theorem 4.1 : *Let $(Y, X\beta, \sigma^2 V)$ be a GGM model (i.e., where V may or may not be non singular and X may or may not have deficiency in rank). Further let $T = V+k^2XX'$, $k \neq 0$ and T^- be any g-inverse of T. Then the following hold :*

(i) *The BLUE of $p'\beta$ is $p'\hat{\beta}$ where $\hat{\beta}$ minimizes $(Y-X\beta)'\ T^-(Y-X\beta)$.*

(ii) $$V(p'\hat{\beta}) = \sigma^2 p'[(X'T^-X)^- - k^2 I]p \quad \dots \quad (4.5)$$

$$cov(p'\hat{\beta}, q'\hat{\beta}) = \sigma^2 p'[(X'T^-X)^- - k^2 I]q \quad \dots \quad (4.6)$$

(iii) *An unbiased estimator of σ^2 is*

$$\hat{\sigma}^2 = f^{-1}(Y-X\hat{\beta})'\ T^-(Y-X\hat{\beta})$$
$$= f^{-1}(Y'T^-Y-\hat{\beta}'\ X'T^-Y), \quad f = R(V : X)-R(X). \quad \dots \quad (4.7)$$

The result (i) of Theorem 4.1 is established in (4.4) for an n.n.d. g-inverse of T. However, we note that all the expressions involving T^- are invariant for any choice of g-inverse and hence the result as stated in the Theorem is true. The results (ii) and (iii) are easy to establish.

Note 1. It is interesting to note that all the results quoted in Theorem 4.1 are independent of the actual value of $k \neq 0$ used in defining T.

Note 2. If $\mathcal{M}(X) \subset \mathcal{M}(V)$ or V is nonsingular, then k *can* be chosen to be zero in which case, $T^- = V^-$,

$$V(p'\hat{\beta}) = \sigma^2 p'(X'V^-X)^- p$$

$$\text{cov}\,(p'\hat{\beta}, q'\hat{\beta}) = \sigma^2 p'(X'V^-X)^- q$$

and the rest of the formulae are as in Theorem 4.1. Note that the expressions (4.5) and (4.6) for variances and covariances when $k \neq 0$ contain an extra term and do not conform to the usual formulae associated with the least squares theory.

Note 3. The unbiased estimator of σ^2 in (iii) is the same as that given in (iii) of Theorem 3.1.

Note 4. Linear hypotheses can be tested as mentioned in Theorem 3.2, estimating the deviations $u = P'\hat{\beta} - w$ and their dispersion matrix $\sigma^2 D$ using the formulae (4.5) and (4.6) and setting up the test criterion (consistency and F tests) as in (iii) of Theorem 3.2.

Note 5. It is unfortunate that the numerator in the F-statistic, $u'D^- u$ for testing the hypothesis $P'\beta = w$, cannot be obtained as

$$\min_{P'\beta = w} (Y - X\beta)'\, T^-(Y - X\beta) - \min_{\beta} (Y - X\beta)'\, T^-(Y - X\beta) \qquad \ldots \quad (4.8)$$

as in the usual least squares theory. However, it can be exhibited as a difference by choosing a different matrix instead of T^- in the first expression of (4.8).

Note 6. Although V may be singular, there is a possibility of $V + k^2XX'$ being nonsingular in which case T^- would be the regular inverse.

Theorem 4.1 shows that the computational procedure for obtaining BLUE's is that of least squares (of Aitken's type and the *same* for all situations 1.1.2–1.1.6 described in section 1). The matrix to be used is an inverse of $(V + k^2XX')$ (unlike that of Aitken's procedure) where $k \neq 0$ in all situations. The formulae for variances and covariances are also the *same* for all cases (but the expressions are slightly different from those in the usual least squares theory). Thus we have a general formulation of the least squares theory in Theorem 4.1. The most general formulation is, however, given in Theorem 4.3.

In Theorem 4.1 it is shown that the BLUE's can be obtained from $(V + k^2XX')^-$—least squares solution of $Y = X\beta$. The introduction of an arbitrary constant k is itself an indication that the choice of the matrix M in the expression (for the norm of $Y - X\beta$)

$$(Y - X\beta)'M(Y - X\beta) \qquad \ldots \quad (4.9)$$

to be minimized is not unique. Let us therefore characterize the class of all possible matrices M in defining the expression (4.9) whose minimization leads to BLUE's.

In order to introduce greater generality in the problem we define $\hat{\beta}$ to be a M-least squares solution of $Y = X\beta$ if it satisfies what may be called the normal equation

$$X'MX\beta = X'MY, \quad \dots \quad (4.10)$$

where M is not necessarily symmetric. Theorems 4.2 and 4.3 provide a complete characterization of M such that a solution $\hat{\beta}$ of (4.10) provides the minimum V-variance estimator (BLUE) of an estimable parametric function $p'\beta$ in the form $p'\hat{\beta}$.

Theorem 4.2 : *Let* $(Y, X\beta, \sigma^2 V)$ *be a GGM model and* $\hat{\beta}$ *be any solution of* (4.10) *where* M *is such that* (4.10) *is consistent. If* $p'\hat{\beta}$ *is the BLUE of* $p'\beta$, *whenever* $p'\beta$ *is estimable, then it is NAS that* M *is of the form*

$$M = (V+XUX')^{-}+K. \quad \dots \quad (4.11)$$

and

$$R(X'MX) = R(X'), \quad \dots \quad (4.12)$$

where U *and* K *are arbitrary subject to conditions*

$$\mathcal{M}(V : X) = \mathcal{M}(V+XU'X') = \mathcal{M}(V+XUX'), \quad \dots \quad (4.13)$$

and K *statisfies the equations*

$$VKX = VK'X = 0, \; X'KX = 0. \quad \dots \quad (4.14)$$

By assumption the equation (4.10) is consistent in which case $\hat{\beta} = (X'MX)^{-}X'MY$ is a solution. Let $p = X'L$, in which case $p'\beta$ is estimable. Then

$$E(p'\hat{\beta}) = p'\beta \Longrightarrow L'X(X'MX)^{-}(X'MX) = L'X.$$

Since L can be arbitrary

$$X(X'MX)^{-}X'MX = X$$

which $\Longrightarrow R(X'MX) = R(X')$. Thus (4.12) is proved.

If $p'\hat{\beta}$ is the BLUE of $p'\beta$ for all p of the form $X'L$, then it is NAS that $X'MY$ are the BLUE's of their expected values. Applying the lemma on p. 257 in Rao (1965) an NAS condition for $X'MY$ to be BLUE's is

$$X'MVZ = 0 \quad \dots \quad (4.15)$$

where Z is a matrix of maximum rank such that $X'Z = 0$.

The equation (4.15) $\Longrightarrow VM'X = XQ$ for some Q. Then there exists a matrix U such that

$$(V+XU'X')M'X = X. \quad \dots \quad (4.16)$$

The equation (4.16) together with (4.12) $\Longrightarrow$ (4.13). Now let

$$M = (V+XUX')^{-}+K \quad \dots \quad (4.17)$$

choosing any g-inverse of $(V+XUX')$. Substituting (4.17) for M in (4.16) we have

$$(V+XU'X')K'X = 0 \quad \dots \quad (4.18)$$

$$\Longrightarrow X'KX = 0 \text{ since } \mathcal{M}(X) = \mathcal{M}(V+XU'X'). \quad \dots \quad (4.19)$$

Then (4.18) and (4.19) $\Longrightarrow VK'X = 0$. From (4.15) and (4.19) we find $VKX = 0$, which proves (4.11) with (4.14).

Note 1. It is seen that when $U = k^2I$ and $K = 0$ we have the choice of M as in Theorem 4.1, which is indeed simple and does not involve any computaton. But Theorem 4.2 characterizes the complete class of the M matrices which lead to the BLUE's via least squares estimators of β.

Note 2. Let $\hat{\beta}$ be a solution of (4.10) with any M as characterized in Theorem 4.2. Then

$$V(p'\hat{\beta}) = \sigma^2 p'(X'MX)^- p - \sigma^2 p'Up. \quad \ldots \quad (4.20)$$

$$\text{cov}(p'\hat{\beta}, q'\hat{\beta}) = \sigma^2 p'(X'MX)^- q - \sigma^2 p'Uq. \quad \ldots \quad (4.21)$$

Note 3. We have seen in Theorem 4.1 that an unbiased estimator of σ^2 is given by

$$\hat{\sigma}^2 = f^{-1}(Y - X\hat{\beta})'(V + k^2XX')^-(Y - X\hat{\beta}). \quad \ldots \quad (4.22)$$

Is the formula (4.22) applicable when we use any matrix M satisfying the conditions of Theorem 4.2 instead of $(V + k^2XX')^-$? The answer is not in the affirmative in general. The exact form of M for such a purpose is given in Theorem 4.3. which contains the *basic result* of the unified theory of least squares.

Theorem 4.3 : *Let* $(Y, X\beta, \sigma^2V)$ *be a GGM model and* $\hat{\beta}$ *be any solution of* $X'MX\beta = X'MY$. *If* $p'\hat{\beta}$ *is the BLUE of any estimable parametric function* $p'\beta$ *and further if an unbiased estimator of* σ^2 *is*

$$\hat{\sigma}^2 = f^{-1}(Y - X\hat{\beta})'M(Y - X\hat{\beta}), \quad \ldots \quad (4.23)$$

$$f = R(V : X) - R(X),$$

then it is NAS that M *is a g-inverse of* $V + XUX'$ *where* U *is a matrix such that* $\mathcal{M}(V : X) = \mathcal{M}(V + XUX')$ *as in Theorem* 4.1.

If $p'\hat{\beta}$ is the BLUE of any estimable parametric function $p'\beta$, we have already seen in Theorem 4.2 that M is of the form

$$M = (V + XUX')^- + K \quad \ldots \quad (4.24)$$

where (4.14) holds. If in addition we want an unbiased estimator of σ^2 as given in (4.23) then it can be shown that $VKV = 0$ giving the result

$$(V + XUX')K(V + XUX') = 0.$$

Hence choosing M as in (4.24)

$$(V + XUX')M(V + XUX') = V + XUX'$$

which shows that M is a g-inverse of $V + XUX'$.

Theorem 4.3 provides the basic result for a unified theory of linear estimation via least squares. Given V and X, we choose a matrix U such that

$$\mathcal{M}(V : X) = \mathcal{M}(V+XUX').$$

Then the usual least squares theory is applied using any g-inverse of $(V+XUX')$ in defining the normal equation

$$X'(V+XUX')^{-}X\beta = X'(V+XUX')^{-}Y.$$

The variances and covariances of estimators are as in (4.20, 4.21) and an unbiased estimator of σ^2 is as in (4.23). It may be noted that we need not choose a n.n.d. g-inverse of $(V+XUX')$ or even a symmetric g-inverse in setting up the normal equations and estimating σ^2, etc., as all the expressions involved are invariant for any choice of $(V+XUX')^{-}$. If in (4.10), $X'MX$ is symmetrical then U in (4.16) can be chosen to be symmetrical. Then (4.26) is obtained by minimizing $(Y-X\beta)'(V+XUX')'$ $(Y-X\beta)$.

5. An important lemma

In the previous sections we found the BLUE's of estimable parametric functions corresponding to a given dispersion matrix V. In this section we raise the question of identification of V given the class of BLUE's of all estimable functions. It appears that the correspondence is not one to one and there is no unique V given the class of BLUE's.

To specify the class of BLUE's it is enough to know the estimators of the parametric functions $X\beta$. Let CY be the BLUE's of $X\beta$, where without loss of generality $R(C) \leqslant R(X)$.

Theorem 5.1 provides the properties of C and its relationship with V. We obtain more general results than those given earlier by the author (Rao, 1965b, 1967b), and Mitra and Rao (1969) in similar investigations.

Theorem 5.1 : *Let $(Y, X\beta, \sigma^2 V)$ be a GGM model and CY be the BLUE of $X\beta$, where $R(C) \leqslant R(X)$. Then it is NAS that*

(i) $CX = X$, *so that* $R(C) = R(X)$,

(ii) C *is idempotent,*

(iii) $CV = (CV)'$ *or* CV *is symmetric, and*

(iv) $C = XD$ *where* D' *is a minimum* V*-norm inverse of* X'.

Results (i) and (ii) follow from the condition of unbiasedness of CY for $X\beta$ and the condition $R(C) \leqslant R(X)$. Result (iii) follows by expressing the condition that the covariance between CY and any linear function $L'Y$, such that $L'X = 0$, is zero. Result (iv) follows from (i) and (iii).

Result (iii) shows that there is no unique V corresponding to a given C and *any* V satisfying the equation $CV = (CV)'$ gives CY as the BLUE of $X\beta$. Theorems 5·2–5·5 examine the question a little further.

Theorem 5.2 : *Consider a general model $(Y, X\beta, \sigma^2 V_0)$ which is consistent in the sense of section* 3.1. *If the BLUE of every estimable parametric function under $(Y, X\beta, \sigma^2 V_0)$ is also the BLUE under $(Y, X\beta, \sigma^2 V)$ where V preserves the consistency of V_0, then it is NAS that*

$$\mathcal{M}(VZ) \subset \mathcal{M}(V_0Z) \qquad \dots \quad (5.1)$$

where $Z = X^{\perp}$.

We say that V preserves the consistency of V_0 if

$$L'V_0 = 0,\ L'X = 0 \Longrightarrow L'V = 0$$

which is a reasonable requirement for $L'V_0 = 0$, $L'X = 0 \Longrightarrow L'Y = 0$ as V_0 is chosen to be consistent. Observe that an NAS condition that $L'Y$ is the BLUE under $(Y, X\beta, \sigma^2 V_0)$ is $L'V_0Z = 0$. Further any L satisfying the equation $L'V_0Z = 0$ provides a BLUE. Thus $L \in \mathcal{M}[(V_0Z)^{\perp}]$. If $L'Y$ is also the BLUE under $(Y, X\beta, \sigma^2 V)$, then $L'VZ = 0$. If this is true for all $L \in \mathcal{M}[(V_0Z)^{\perp}]$, then

$$Z'V(V_0Z)^{\perp} = 0 \Longrightarrow \mathcal{M}(VZ) \subset \mathcal{M}(V_0Z).$$

Theorem 5.3 : *Let X, Z, V_0 and V be defined as in Theorem 5.2 and let the columns of X and V_0Z span the whole n dimensional space. Then it is NAS that*

$$V = XAX' + V_0Z\,BZ'V_0 \qquad \dots \quad (5.2)$$

where A and B are arbitrary symmetric matrices.

Let $V = FF'$ and write $F = XC + V_0ZD$. Then

$$V = XCC'X' + V_0ZDD'Z'V_0 + XCD'Z'V_0 + V_0ZDC'X'.$$

Applying the condition $Z'V(V_0Z)^{\perp} = 0$, we have

$$Z'V_0ZDC'X'(V_0Z)^{\perp} = 0 \Longleftrightarrow V_0ZDC'X' = 0$$

leading to the representation (5.2) of V.

Corollary 5.3.1 : *The expression* (5.2) *can also be written as*

$$V = XAX' + V_0\,ZBZ'V_0 + V_0. \qquad \dots \quad (5.3)$$

Corollary 5.3.2 : *If $V_0 = I$, then V can be represented in the forms*

$$V = XAX' + ZBZ',$$

$$V = XAX' + ZBZ' + I. \qquad \dots \quad (5.4)$$

Theorem 5.4 : *If the condition* (5.1) *is satisfied then it is NAS that*

$$V = V_0 + V_0ZU + WX' \qquad \dots \quad (5.5)$$

where U and W are arbitrary subject to the condition that $V_0\,ZU + WX'$ is symmetric.

We can write $V = V_0 + T$ in which case $Z'T(V_0Z)^{\perp} = 0$. A general solution for $T = V_0ZU + WX'$. (Note that the expression (5.5) is not in a satisfactory form since the symmetry of V is not utilized.)

Note 1 : If $V_0 = I$, then (5.1) reduces to $\mathcal{M}(VZ) \subset \mathcal{M}(Z)$, which is the same as the condition $X'VZ = 0$ derived in my earlier paper, Rao (1967).

Note 2 : It may be seen that V satisfying (5.1) preserves consistency with respect to V_0, i.e.,

$$L'V_0 = 0, \quad L'X = 0 \Longrightarrow L'V = 0$$

But consistency of V may not be true with respect to V_0. For instance if $V_0 = I$, then we can find a singular V such that $\mathcal{M}(VZ) \subset \mathcal{M}(Z)$. For such a V, the BLUE's with respect to I are also BLUE's with respect to V. A singular V implies that there exists a nonnull vector L such that $V(L'Y|V) = 0$. But $V(L'Y|I) \neq 0$.

Note 3 : If in Theorem 5.2 we demand that the reciprocal relation must also be true, i.e., V_0 preserves the consistency of V then $\mathcal{M}(VZ) = \mathcal{M}(V_0Z)$.

Now we investigate the condition under which we can find linear estimators for all estimable parametric functions which are the BLUE's with respect to two possible alternative models $(Y, X\beta, \sigma^2V)$ and $(Y, X\beta, \sigma^2V_0)$. We restate a theorem proved in Rao (1968) which provides the desired answer.

Theorem 5.5 : *Let $(Y, X\beta, \sigma^2V)$ and $(Y, X\beta, \sigma^2V_0)$ be two different models. If for every estimable parametric function we can specify a linear function of Y which is the BLUE whichever model is true, then each of the following equivalent conditions is NAS.*

(i) $\mathcal{M}(0 : 0 : X)' \subset \mathcal{M}(VZ : V_0Z : X)'$.

(ii) *$\mathcal{M}(VZ : V_0Z)$ and $\mathcal{M}(X)$ are virtually disjoint.*

(iii) $\mathcal{M}\begin{pmatrix} Z'V \\ Z'V_0 \end{pmatrix} \subset \mathcal{M}\begin{pmatrix} Z'VZ \\ Z'V_0Z \end{pmatrix}$.

Note the subtle difference between the Theorems 5.1 and 5.5. In Theorem 5.1 we demanded that consistency under V_0 should be preserved under V. Theorem 5.5 does not require this condition, but it says that whenever V_0 and V satisfy any one of the three conditions (i), (ii) and (iii), we can *specify* linear functions of Y which are BLUE's *whether V or V_0 is true*. The relationship between V and V_0 is symmetrical in Theorem 5.5.

If $L'Y$ is the BLUE of $p'\beta$ whether V or V_0 is true then L must satisfy the equations

$$VL = XM_1, \; V_0L = XM_2, \; X'L = p. \qquad \dots \quad (5.6)$$

The equations (5.6) are equivalent to

$$Z'VL = 0, \; Z'V_0L = 0, \; X'L = p.$$

We need this to be true for all p of the form Xq, for which any one of the three conditions (i), (ii) and (iii) of Theorem 5.5 is NAS.

Note 1 : If the columns of X and V_0Z span the whole n dimensional space, then V satisfying any one of the conditions of Theorem 5.5 can be expressed as (5.2) and (5.3).

Note 2 : Given a particular pair V and V_0 satisfying any one of the conditions of Theorem 5.5, the common BLUE of $p'\beta$ is $L'Y$ where L is any solution of the equations

$$Z'VL = 0 = Z'V_0L, \; X'L = p.$$

6. Best linear estimation (BLE)

Not much work is done on BLE compared to that on BLUE. References to earlier work on BLE can be found in papers by Hoerl and Kennard (1970a, 1970b) who introduced what are called ridge regression (biased) estimators. In this section we approach the problem of BLE in a direct manner.

To make matters simple and to compare BLE's with BLUE's let us consider the model $(Y, X\beta, \sigma^2 V)$ with $R(X) = m$ (i.e., of full rank) and $|V| \neq 0$. Let $L'Y$ be an estimator of $p'\beta$. The mean square error of $L'Y$ is

$$E(L'Y-p'\beta)^2 = \sigma^2 L'VL+(X'L-p)'\beta\beta'(XL-p) \qquad \ldots \quad (6.1)$$

which involves both the unknown parameters σ^2 and β, and as it stands is not a suitable criterion for minimizing. Then we have the following possibilities.

(i) Choose an apriori value of $\sigma^{-1}\beta$ say b, based on previous knowledge, and set up the criterion as $\sigma^2 S$ where

$$S = L'VL+(X'L-p)'W(X'L-p), \qquad \ldots \quad (6.2)$$

with $W = bb'$.

(ii) If β is considered to have an apriori distribution with a dispersion matrix $\sigma^2 W$ where W is known, then the criterion is $\sigma^2 S$ where S is of the same form as in (6.2).

(iii) We observe that the expression (6.1) is the sum of two parts, one representing the variance and the second bias. In such a case the choice of W in the criterion (6.2) represents the relative weight we attach to bias compared to variance. Then W may be chosen taking into account the relative importance of bias and variance in the estimator.

If S, as in (6.2), is chosen as the criterion with an appropriate symmetric W, then Theorem 6.1 gives the optimum choice of L.

Theorem 6.1 : *The BLE of any function* $p'\beta$ *is* $p'\tilde{\beta}$ *where*

$$\tilde{\beta} = WX'(V+XWX')^{-1}Y. \qquad \ldots \quad (6.3)$$

The minimum value of S is attained when L satisfies the equation

$$(V+XWX')L = XWp$$

so that $L = (V+XWX')^{-1}XWp$ giving the estimator

$$L'Y = p'WX'(V+XWX')^{-1}Y = p'\tilde{\beta}. \qquad \ldots \quad (6.4)$$

The estimator $\tilde{\beta}$ may be called BLE of β.

The BLUE of β is (using Theorem 4.3)

$$\hat{\beta} = (X'T^{-1}X)^{-1}X'T^{-1}Y, \; T = V+XWX' \qquad \ldots \quad (6.5)$$

whereas the BLE of β is

$$\tilde{\beta} = WX'T^{-1}Y = G\hat{\beta}, \quad G = WX'T^{-1}X. \qquad \ldots \quad (6.6)$$

which establishes the relationship between $\tilde{\beta}$ and $\hat{\beta}$. The mean dispersion error of $\tilde{\beta}$ is

$$F = GDG'+(G-I)\beta\beta'(G-I)' \qquad \ldots \quad (6.7)$$

where D is the dispersion matrix of $\hat{\beta}$. The matrix $D-GDG'$ is n.n.d. so that there is a possibility of $D-F$ being n.n.d. for a certain range of values of β. Thus, if we have some knowledge about the domain in which β is expected to lie, we may be able to choose W suitably to ensure that the BLE's have uniformly smaller mean dispersion error than the BLUE's. Further investigation in this direction such as comparison of the estimator (6.3) with the ridge estimator of Hoerl and Kennard (1970a, 1970b) will be useful.

We have assumed that the ranks of X and V are full. We can obtain BLE's when these conditions are not satisfied by following the methods of sections 3 and 4. For instance in the formulae (6.3), an appropriate g-inverse may be used when the regular inverse does not exist.

7. Best linear minimum bias estimation (BLIMBE)

We consider the GGM model $(Y, X\beta, \sigma^2V)$, which as mentioned earlier includes the case of constraints on the β parameter.

When $R(X) \neq m$, not all linear parametric functions admit linear unbiased estimators (LUE). In fact an NAS condition that $p'\beta$ has a LUE is that $p \in \mathcal{M}(X')$. We raise two questions.

(i) What is the minimum restriction to be put on β such that every parametric function admits a LUE and hence the BLUE ?

(ii) In what sense can we find a best linear minimum bias estimator (BLIMBE) of $p'\beta$ if it does not admit a LUE ?

The answer to the first question is contained in Theorem 7.1.

Theorem 7.1 : *Let $R(X) = r < m$. The minimum restriction on β can be expressed in two alternative forms.*

(i) $R\beta = c$, $R(R) = m-r$ *and* $\mathcal{M}(R') \cap \mathcal{M}(X') = \{0\}$.

(ii) $\beta = d+TX'\gamma$, *where T is any matrix such that* $R(XTX') = R(X)$ and γ *is arbitrary.*

The first restriction is obvious and the second can be deduced from the first. Both the restrictions imply that the parameter β is contained in a hyperplane of dimension r.

To answer the second question we proceed as follows. The bias in $L'Y$ as an estimator of $p'\beta$ is $\beta'(X'L-p)$ which is zero if and only if $p \in \mathcal{M}(X')$. If not we will try to minimize the bias by choosing L in such a way that $\|X'L-p\|$, a suitably defined norm of the deviation $X'L-p$, is a minimum (see Chipman, 1964).

Theorem 7.2 : *Let the norm of a vector u be defined as $\|u\| = (u'Mu)^{\frac{1}{2}}$ where M is p.d. Then a LIMBE of $p'\beta$ is*

$$p'[(X')^-_{l(M)}]'\,Y = p'X^-_{m(M^{-1})}\,Y. \qquad \dots \quad (7.1)$$

The result of Theorem 7.2 follows from the definitions of least squares and minimum norm g-inverses and their equivalence relation as given in Theorem 1.1.

Note that if M is not p.d. but only p.s.d. then a LIMBE of $p'\beta$ is

$$p'[(X')^{-}_{l(M)}]'\ Y. \qquad \dots \quad (7.2)$$

The LIMBE may not be unique in which case we shall choose one for which the variance is a minimum. Such an estimator may be called BLIMBE.

Theorem 7.3 : *The BLIMBE of* $p'\beta$, *i.e., a linear function* $L'Y$ *such that* $L'VL$ *is a minimum in the class of* L *which minimizes the bias,*

$$\|X'L-p\|^2 = (X'L-p)'M(X'L-p), \text{ is } p'[(X')^{+}_{MV}]'\,Y,$$

where $(X')^{+}_{MV}$ *is the minimum* V*-norm* M*-least squares inverse of* X'.

The result follows from the definition of a V-norm M-least squares inverse of X'. Explicit expressions for such inverses in various situations are given in Rao and Mitra (1971a).

8. Concluding remarks

1. It is shown that the problem of linear estimation and inference from a GGM model $(Y, X\beta, \sigma^2 V)$ where multicollinearity may exist in the explanatory variable X and the dispersion matrix $\sigma^2 V$ may be singular (which includes the case of constraints on the β parameter) reduces to the *numerical* computation of a g-inverse of the symmetric matrix.

$$\begin{pmatrix} V & X \\ X' & 0 \end{pmatrix}. \qquad \dots \quad (8.1)$$

Theorem 3.1 shows that the computation of a g-inverse of (8.1) is like opening a *Pandora box*, giving all that is necessary for drawing inferences on β.

If need be, different situations such as singularity of V, deficiency in the rank of X etc., may be considered in computing a g-inverse of (8.1). But one need not think of distinguishing between the models in the beginning by these situations as the approach given in Section 3 is the same for all cases. Efforts should be made to obtain a suitable algorithm for computig a g-inverse of (8.1) in an efficient way. The computational methods developed by Golub and Kahan (1965) and Golub and Reinsch (1969) for singular value decomposition of a matrix and related problems might provide the answer.

An algebraic expression for the g-inverse of (8.1) in terms of g-inverses of $(V+XX')$ and $[X(V+XX')^{-}X]$ is given in Theorem 3.6.4. in Rao and Mitra (1971a, p. 66).

2. A general method of least squres is proposed which is simple and valid for all situations. The expression to be minimized in *all situations, whether* V *is singular or not,* is

$$(Y-X\beta)'M(Y-X\beta)$$

where M is any g-inverse of $V+XUX'$, U being any matrix such that $\mathcal{M}(V:X) = \mathcal{M}(V+XUX')$. A simple choice of $U = k^2 I$, $k \neq 0$ in which case $M = (V+k^2XX')^{-}$.

This generalizes Aitken's result when V is singular. It is observed that the least squares method is not based on a primitive postulate but it is an artifact for computing minimum variance linear unbiased estimators.

3. An important lemma is proved to examine complete robustness of BLUE's with respect to alternative dispersion matrices of the observation vector in the Gauss-Markoff model.

4. A case has been made for dropping the criterion of unbiasedness and using prior knowledge for uniformly improving the BLUE's. Such estimators are called BLE's.

5. Minimum variance minimum bias estimators are proposed in cases where unbiased estimators do not exist.

Appendix

The atom bomb and generalized inverse

The author was first led to the definition of a pseudo-inverse (now called generalized invrese or g-inverse) of a singular matrix in 1954-1955 when he undertook to carry out multivariate analysis of anthropometric data obtained on families of Hiroshima and Nagasaki to study the effects of radiation due to atom bomb explosions, on a request from Dr. W. J. Schull of the University of Michigan. The computation and use of a pseudo-inverse are given in a statistical report prepared by the author, which is incorporated in Publication No. 461 of the National Academy of Sciences, U.S.A., by Neel and Schull (1956). It may be of interest to the audience to know the circumstances under which the pseudo-inverse had to be introduced.

Let us consider the standard Gauss-Markoff model $(Y, X\beta, \sigma^2 I)$ with X of full rank, and write the normal equation as $S\beta = Q$. Since S is of full rank, $C = S^{-1}$ exists giving $\hat{\beta} = CQ$ as the solution of $S\beta = Q$. Then $V(p'\hat{\beta}) = \sigma^2 p'Cp$. The matrix C, called the C matrix of Fisher, thus plays an important role in computing the estimate of any parametric function $p'\beta$ as well as the variance of its estimate.

When the rank of X is not full, a situation which was first formulated by R. C. Bose, the author (Rao, 1945a) showed that the theory of least squares is still applicable, i.e., the normal equation $S\beta = Q$ is solvable and the BLUE of an *estimable* parametric function $p'\beta$ is $p'\hat{\beta}$ where $\hat{\beta}$ is *any* solution of $S\beta = Q$. To find $V(p'\hat{\beta})$ the following rule was suggested. If $p'\hat{\beta}$ can be expressed as $k'Q$, then $V(p'\hat{\beta}) = \sigma^2 k'p$, which works well in many problems but not convenient as the computations of k for each p is a little involved when the matrix X does not have a simple structure.

The atom bomb data had a highly non orthogonal design matrix X in addition to some deficiency in rank. Estimates of different contrasts had different precisions and I did not know what contrasts were of interest to the investigators for which I should provide estimates and also compute their standard errors (by the formula proposed in my 1945 paper). This led me to look for a matrix C in situations where S is singular, which can be used in the same way as the C matrix of Fisher in obtaining a solution to the normal equation and also in computing in a simple way the standard error of any desired contrast without any algebraic manipulation.

I established the existence of such a matrix C and also showed how to compute it when S is singular, which I called a pseudo-inverse of S, with the property that

(i) $\hat{\boldsymbol{\beta}} = CQ$ is a solution of $S\boldsymbol{\beta} = Q$,

(ii) $p'\hat{\boldsymbol{\beta}}$ is the BLUE of $p'\boldsymbol{\beta}$, and

(iii) $V(p'\hat{\boldsymbol{\beta}}) = \sigma^2 p'Cp$,

where $p'\boldsymbol{\beta}$ is any estimable function. In my report to the Michigan University I gave the C matrix explaining how it can be used to find the variance of any contrast they wished to examine by simple matrix multiplication. The above results were also given in Rao (1955).

Later, I discovered Penrose's 1955 paper published about the same time, where g-inverse was defined using four conditions out of which only two were satisfied by my pseudo-inverse. Apparently my work showed that in the discussion of least squares theory one needs a g-inverse only in the weak sense defined in section 1.2. This was further demonstrated in another paper (Rao, 1962)[4].

Part of the data collected from Hiroshima and Nagasaki was sent to the statistics department of the University of North Carolina, and I understand from Dr. M. Kastenbaum that the data had stimulated new research in the analysis of categorical data. It was also reported that the Japanese geneticists used the backyards of their shattered homes to conduct experiments on plants for studying the effects of radiation soon after the bomb was dropped, and numerous papers were published on the subject.

It is hard to believe that scientists have found in what has been described as the greatest human tragedy a source for providing material and stimulation for research in many directions.

References

AITKEN, A. C. (1934): On least squares and linear combination of observations. *Proc. Roy. Soc. Edinb.* A, 55, 42-47.

BOSE, R. C. (1947): The Design of Experiments. Presidential Address to Statistics Section, 34th Session of The Indian Science Congress Association.

CHIPMAN, J. S. (1964): On least squares with insufficient observations. *J. Amer. Statist. Assoc.* 59, 1078-1111.

DAVID, F. N. and NEYMAN, J. (1938): Extension of the Markoff theorem on least squares. *Statist. Res. Mem.*, 2, 105-116.

GAUSS, C. F. (1855). *Methode des moindres Carres* (Trans. J. Bertrand). Paris.

GOLUB, G. and KAHAN, W. (1965): Calculating the singular values and pseudoinverse of a matrix. *SIAM J. Numer. Anal. Ser.* B, 2, 205-224.

GOLUB, G. and REINSCH, C. (1969): Singular value decomposition and least squares solutions, *Technical Report*, CS, 133, Computer Science Department, Stanford.

HOERL, A. E. and Kennard, R. W. (1970a). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12, 55-68.

——— (1970b): Ridge regression: Applications to nonorthogonal problems. *Technometrics*, 12, 69-82.

[4]Many authors still insist on using what is called Moore-Penrose inverse when they can state their results in a more general form by using a weaker form of g-inverse. See for example the recent book by Theil (1971, foot note on p. 283) and a paper by Schönfeld (1971).

MARKOFF, A. A. (1912): *Warscheinlichkeitsrechnung*. (Trans. H. Liebmann). 2nd edition. Leipzig and Berlin.

MITRA, S. K. and RAO, C. R. (1968): Some results in estimation and tests of linear hypotheses under the Gauss-Markoff model. *Sankhyā*, A. **30**, 281-290.

——— (1969): Conditions for optimality and validity of simple least squares theory. *Ann. Math. Statist.* **40**, 1617-1624.

NEEL, J. V. and SCHULL, W. J. (1956): The effect of exposure to the Atom Bombs on Pregnancy Termination in Hiroshima and Nagasaki. Publication No. 461. National Academy of Sciences, U.S.A.

PLACKETT, R. L. (1949): A historical note on the method of least squares. *Biometrika*, **36**, 458-460.

RAO, C. RADHAKRISHNA (1945a). Generalization of Markoff's theorem and tests of linear hypotheses. *Sankhyā*, **7**, 9-16.

——— (1945b): Markoff's theorem with linear restrictions on parameters. *Sankhyā*, **7**, 16-19.

——— (1955): Analysis of dispersion for multiply classified data with unequal numbers in cells. *Sankhya*, **15**, 253-280.

——— (1962): A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. *J. Roy. Statist. Soc.*, B, **24**, 152-158.

——— (1965a): *Linear Statistical Inference and its Applications*, Wiley, New York.

——— (1965b): The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika*, **52**, 447-458.

——— (1966): Generalized inverse for matrices and its applications in mathematical statistics, *Research Papers in Statistics*, Festschrift for J. Neyman, Wiley, New York.

——— (1967a): Calculus of generalized inverse of matrices, Part I: General theory. *Sankhyā*, A, **29**, 317-342.

——— (1967b): Least squares theory using an estimated dispersion matrix and its application to measurement of signals *Proc. Fifth Berkeley Symposium on Math. Statist. and Prob.* **1**, 355-372.

——— (1968): A note on a previous lemma in the theory of least squares and some further results. *Sankhyā*, A, **30**, 245-253.

——— (1970): Estimation of heteroscedastic variances in linear models. *J. Amer. Statist. Assoc.*, **65**, 161-172.

——— (1971a): Estimation of variance and covariance components—MINQUE theory. *J. Multivariate Analysis*, **1**, 257-275.

——— (1971b): Minimum variance quadratic unbiased estimation of variance components. *J. Multivariate Analysis*, **1**, (in press).

——— (1972): Estimation of variance and covariance components. *J. Amer. Statist. Assoc.* (in press).

RAO, C. RADHAKRISHNA and Mitra, S. K. (1971a): *Generalized Inverse of Matrices and its Applications*, Wiley, New York.

——— (1971b): Further contributions to the theory of generalized inverse of matrices and its Applications. *Sankhyā*, A, **33**, 289–300.

——— (1971c): Theory and application of constrained inverse of matrices. (submitted for publication).

RAO, J. N. K. and SUBRAHMANIAM, K. (1971): Combining independent estimators and estimation in linear regression with unequal variances. *Biometrics*, **27**, 971–990.

SCHÖNFELD, PETER (1971): Best linear minimum bias estimation in linear regression. *Econometrica*, **39**, 532-544.

THEIL HENRI (1971): *Principles of Econometrics*, Wiley New York.

ZYSKIND, G. and MARTIN, F. B. (1969): On best linear estimation and a general Gauss-Markoff theorem in linear models with arbitrary negative covariance structure. *SIAM J. Appl. Math.*, **17**, 1190-1202.