The Annals of Mathematical Statistics
1969, Vol. 40, No. 5, 1617–1624

CONDITIONS FOR OPTIMALITY AND VALIDITY OF SIMPLE LEAST SQUARES THEORY

BY SUJIT KUMAR MITRA AND C. RADHAKRISHNA RAO

Indian Statistical Institute

1. Notation and introduction. A matrix is denoted by a bold face letter such as $\mathbf{A}$, $\mathbf{X}$, $\mathbf{\Sigma}$ etc. For a matrix $\mathbf{X}$ of order $n \times m$

$R(\mathbf{X})$ represents the rank of $\mathbf{X}$.

$\mathfrak{M}(\mathbf{X})$ represents the linear space generated by the columns of $\mathbf{X}$.

$\mathbf{X}^-$ represents a g-inverse as defined by Rao (1962, 1966, 1967b).

$\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$ is the projection operator which projects arbitrary n-vectors onto $\mathfrak{M}(\mathbf{X})$.

$\mathbf{X}^\perp$ denotes a matrix of maximum rank such that $\mathbf{X}'\mathbf{X}^\perp = \mathbf{0}$.

$\mathbf{I}$ denotes an identity matrix. The order of $\mathbf{I}$ will usually not be explicitly mentioned but can always be determined from the context.

Consider the Gauss-Markoff model $(\mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{\Sigma})$ where $\mathbf{Y}$ is a vector of observations, $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ and $D(\mathbf{Y}) = \mathbf{\Sigma}$, $\mathbf{X}$ being a given matrix of order $n \times m$ and $\boldsymbol{\beta}$ a vector of unknown parameters. In the context of the discussion in the present paper, the model will be simply referred to as $(\mathbf{X}, \mathbf{\Sigma})$. The best linear unbiased estimator (BLUE) of an estimable parametric function $\mathbf{p}'\boldsymbol{\beta}$, where $\mathbf{p}$ is a vector, under the model $(\mathbf{X}, \mathbf{\Sigma})$ is a linear function $\mathbf{L}'\mathbf{Y}$ such that $E(\mathbf{L}'\mathbf{Y}) = \mathbf{p}'\boldsymbol{\beta}$ and $\mathbf{L}'\mathbf{\Sigma}\mathbf{L}$ is a minimum. It is well known that a BLUE under $(\mathbf{X}, \mathbf{\Sigma})$ can be obtained by the general method of least squares (see Rao, 1965, page 188 and Mitra and Rao, 1968).

The BLUE of $\mathbf{p}'\beta$ under $(\mathbf{X}_0, \mathbf{\Sigma}_0)$ is said to be $(\mathbf{X}, \mathbf{\Sigma})$ optimal if it is also the BLUE of $\mathbf{p}'\boldsymbol{\beta}$ under the model $(\mathbf{X}, \mathbf{\Sigma})$. The object of the present paper is to characterize the set of $(\mathbf{X}, \mathbf{\Sigma})$ such that for every estimable parametric function the BLUE under a given model $(\mathbf{X}_0, \sigma^2\mathbf{I})$ is $(\mathbf{X}, \mathbf{\Sigma})$-optimal. Further, the classes of $\mathbf{\Sigma}$ for which different statistical methods based on $(\mathbf{X}, \sigma^2\mathbf{I})$ remain valid have been obtained.

In previous papers Rao (1967a, 1968)[1] gave the necessary and sufficient contions for BLUE under $(\mathbf{X}, \mathbf{\Sigma}_0)$ to be $(\mathbf{X}, \mathbf{\Sigma})$-optimal, in which case the investigation was confined to the characterization of $\mathbf{\Sigma}$ only. Similar results, but not providing an exact representation of $\mathbf{\Sigma}$, were also obtained by Zyskind (1967), Watson (1967) and Kruskal (1968) in the special case of $\mathbf{\Sigma}_0 = \sigma^2\mathbf{I}$.

2. The main results.

LEMMA 2.1. *If for every estimable parametric function the* BLUE *under* $(\mathbf{X}_0, \sigma^2\mathbf{I})$ *is* $(\mathbf{X}, \sigma^2\mathbf{I})$*-optimal, it is necessary and sufficient that* $\mathbf{X}$ *is of the form*

$$\mathbf{X} = \mathbf{X}_0 + (\mathbf{I} - \mathbf{P}_{X_0})\mathbf{A}, \tag{2.1}$$

Received 29 October 1968; revised 19 February 1969.

[1] The results were first given in a lecture at the Fifth Berkeley Symposium in 1965.

where $\mathbf{A}$ *is any matrix such that*

$$\mathfrak{M}(\mathbf{A}') \cap \mathfrak{M}(\mathbf{X}_0') = \{\mathbf{0}\}, \tag{2.2}$$

a set consisting exclusively of the null vector, or equivalently

$$\mathbf{X} = \mathbf{X}_0 + (\mathbf{I} - \mathbf{P}_{X_0})\mathbf{F}[\mathbf{I} + (\mathbf{I} - \mathbf{P}_{X_0'})\mathbf{D}]^{\perp'}, \tag{2.3}$$

where $\mathbf{D}$ *and* $\mathbf{F}$ *are arbitrary.*

We note that from standard results in the Gauss-Markoff theory (see Rao, 1965, page 178) it suffices to investigate conditions under which the elements of the vector $\mathbf{X}_0'\mathbf{Y}$ have the same expectations and minimum variance under the models $(\mathbf{X}_0, \sigma^2\mathbf{I})$ and $(\mathbf{X}, \sigma^2\mathbf{I})$.

Sufficiency of (2.1) *and* (2.2). Let $\mathbf{X}$ be of the form given in (2.1) and (2.2). Then

$$E(\mathbf{X}_0'\mathbf{Y} \mid \mathbf{X}, \sigma^2\mathbf{I}) = \mathbf{X}_0'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_0'\mathbf{X}_0\boldsymbol{\beta} = E(\mathbf{X}_0'\mathbf{Y} \mid \mathbf{X}_0, \sigma^2\mathbf{I}),$$

since $\mathbf{I} - \mathbf{P}_{X_0}$ is symmetric and $(\mathbf{I} - \mathbf{P}_{X_0})\mathbf{X}_0 = \mathbf{0}$.

Let $\mathbf{Z} = \mathbf{X}^{\perp}$, in which case $E(\mathbf{Z}'\mathbf{Y} \mid \mathbf{X}, \sigma^2\mathbf{I}) = \mathbf{0}$. By definition $\mathbf{X}'\mathbf{Z} = \mathbf{0} = \mathbf{X}_0'\mathbf{Z} + \mathbf{A}'(\mathbf{I} - \mathbf{P}_{X_0})\mathbf{Z}$, i.e., $\mathbf{X}_0'\mathbf{Z} = -\mathbf{A}'(\mathbf{I} - \mathbf{P}_{X_0})\mathbf{Z}$, which is a contradiction in view of (2.2) unless $\mathbf{X}_0'\mathbf{Z} = \mathbf{0}$. Hence by Lemma (i), page 257 in Rao, 1965, sufficiency is established.

Necessity of (2.1) *and* (2.2). Let a BLUE under $(\mathbf{X}_0, \sigma^2\mathbf{I})$ be $(\mathbf{X}, \sigma^2\mathbf{I})$ optimal. Necessity of (2.1) is obvious for expectations of $\mathbf{X}_0'\mathbf{Y}$ to be the same under the two models. Now suppose that $\mathfrak{M}(\mathbf{A}'(\mathbf{I} - \mathbf{P}_{X_0})) \cap \mathfrak{M}(\mathbf{X}_0')$ contains a vector $\boldsymbol{\alpha} \neq \mathbf{0}$, in which case $\boldsymbol{\alpha} = \mathbf{A}'(\mathbf{I} - \mathbf{P}_{X_0})\boldsymbol{\mu} = \mathbf{X}_0'\mathbf{X}_0\boldsymbol{\lambda}$ for some vectors $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$. Let $\mathbf{z} = \mathbf{X}_0\boldsymbol{\lambda} - (\mathbf{I} - \mathbf{P}_{X_0})\boldsymbol{\mu}$. Check that $\mathbf{X}'\mathbf{z} = \mathbf{0}$ and observe that $\mathbf{X}_0'\mathbf{z} = \mathbf{X}_0'\mathbf{X}_0\boldsymbol{\lambda} = \boldsymbol{\alpha} \neq \mathbf{0}$. This shows that at least some elements of $\mathbf{X}_0'\mathbf{Y}$ are correlated with $\mathbf{z}'\mathbf{Y}$ which is a contradiction. Hence $\mathfrak{M}(\mathbf{A}'(\mathbf{I} - \mathbf{P}_{X_0})) \cap \mathfrak{M}(\mathbf{X}_0') = \{\mathbf{0}\}$. Necessity of (2.2) follows from the fact that, without loss of generality, one may take $\mathbf{A}$ to be $(\mathbf{I} - \mathbf{P}_{X_0})\mathbf{A}$.

Necessity and sufficiency of (2.3). Using the condition of unbiasedness

$$E(\mathbf{X}_0'\mathbf{Y} \mid \mathbf{X}_0, \sigma^2\mathbf{I}) = E(\mathbf{X}_0'\mathbf{Y} \mid \mathbf{X}, \sigma^2\mathbf{I}) \Rightarrow \mathbf{X} = \mathbf{X}_0 + (\mathbf{I} - \mathbf{P}_{X_0})\mathbf{G}.$$

Let $\mathbf{Z} = \mathbf{X}^{\perp}$ and C stand for covariance. Since $\mathbf{X}_0'\mathbf{Y}$ are BLUE's under $(\mathbf{X}, \sigma^2\mathbf{I})$,

$$C(\mathbf{Z}'\mathbf{Y}, \mathbf{X}_0'\mathbf{Y} \mid \mathbf{X}, \sigma^2\mathbf{I}) = \mathbf{0} \Rightarrow \mathbf{X}_0 = \mathbf{X}\mathbf{H} \quad \text{for some} \quad \mathbf{H}.$$

Multiplying both sides of $\mathbf{X} = \mathbf{X}_0 + \mathbf{X}_0^{\perp}\mathbf{G}$ by $\mathbf{H}$ and writing $\mathbf{X}_0 = \mathbf{X}\mathbf{H}$ we obtain

$$\mathbf{X}_0 = \mathbf{X}_0\mathbf{H} + \mathbf{X}_0^{\perp}\mathbf{G}\mathbf{H} \Rightarrow \mathbf{X}_0 = \mathbf{X}_0\mathbf{H} \quad \text{and} \quad \mathbf{G}\mathbf{H} = \mathbf{0}.$$

Observe that $\mathbf{X}_0 = \mathbf{X}_0\mathbf{H} \Rightarrow \mathbf{H} = \mathbf{I} + (\mathbf{I} - \mathbf{P}_{X_0'})\mathbf{D}$ for some $\mathbf{D}$ and since $\mathbf{G}\mathbf{H} = \mathbf{0}$,

$$\mathbf{G} = \mathbf{F}(\mathbf{H}^{\perp})' = \mathbf{F}[\mathbf{I} + (\mathbf{I} - \mathbf{P}_{X_0'})\mathbf{D}]^{\perp'}.$$

Hence

$$\mathbf{X} = \mathbf{X}_0 + (\mathbf{I} - \mathbf{P}_{X_0})\mathbf{F}[\mathbf{I} + (\mathbf{I} - \mathbf{P}_{X_0'})\mathbf{D}]^{\perp'},$$

where $\mathbf{F}$ and $\mathbf{D}$ are arbitrary.

COROLLARY 1. *A sufficient condition that for every estimable function the* BLUE *under* $(\mathbf{X}_0, \sigma^2\mathbf{I})$ *is* $(\mathbf{X}, \sigma^2\mathbf{I})$*-optimal is*

$$(2.4) \qquad \mathbf{X} = \mathbf{X}_0 + [\mathbf{I} - \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^-\mathbf{X}_0']\mathbf{B}[\mathbf{I} - \mathbf{X}_0'\mathbf{X}_0(\mathbf{X}_0'\mathbf{X})^-],$$

where $\mathbf{B}$ *is arbitrary.*

COROLLARY 2. *If* $R(\mathbf{X}_0)$ *is equal to the number of columns in* $\mathbf{X}_0$, *then the conditions of Lemma* 2.1 *reduce to* $\mathbf{X} = \mathbf{X}_0$.

PROOF.

$$(2.5) \qquad E(\mathbf{X}_0'\mathbf{Y} \mid \mathbf{X}, \sigma^2\mathbf{I}) = \mathbf{X}_0'\mathbf{X}_0\boldsymbol{\beta} \Rightarrow \mathbf{X} = \mathbf{X}_0 + \mathbf{X}_0^{\perp}\mathbf{G},$$

for some $\mathbf{G}$. Since $\mathbf{X}_0'\mathbf{Y}$ are BLUE's under $(\mathbf{X}, \sigma^2\mathbf{I})$,

$$(2.6) \qquad C(\mathbf{X}_0'\mathbf{Y}, \mathbf{Z}'\mathbf{Y}) = \mathbf{0} \Rightarrow \mathbf{X}_0'\mathbf{Z} = \mathbf{0} \Rightarrow \mathbf{X}_0 = \mathbf{X}\mathbf{D},$$

for some $\mathbf{D}$. From (2.5) and (2.6),

$$\mathbf{X}\mathbf{D} = \mathbf{X}_0\mathbf{D} + \mathbf{X}_0^{\perp}\mathbf{G}\mathbf{D} = \mathbf{X}_0.$$

Multiplying by $\mathbf{X}_0'$,

$$\mathbf{X}_0'\mathbf{X}_0 = \mathbf{X}_0'\mathbf{X}_0\mathbf{D} \Rightarrow \mathbf{D} = \mathbf{I},$$

since $\mathbf{X}_0'\mathbf{X}_0$ is non-singular. Then $\mathbf{X}_0 = \mathbf{X}$ from (2.6).

LEMMA 2.2. *For the* BLUE *under* $(\mathbf{X}_0, \sigma^2\mathbf{I})$ *to be* $(\mathbf{X}, \boldsymbol{\Sigma})$*-optimal for every estimable parametric function, it is necessary and sufficient that*

$$(2.7) \qquad \mathbf{X} = \mathbf{X}_0 + \mathbf{Z}_0\mathbf{G},$$

$$(2.8) \qquad \boldsymbol{\Sigma} = \mathbf{X}_0\mathbf{A}\mathbf{X}_0' + \mathbf{Z}_0\mathbf{B}\mathbf{Z}_0' + \mathbf{X}_0\mathbf{A}'\mathbf{G}'\mathbf{Z}_0' + \mathbf{Z}_0\mathbf{G}\mathbf{A}\mathbf{X}_0',$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{G}$ *are arbitrary except that* $\boldsymbol{\Sigma}$ *is non-negative definite and* $\mathbf{Z}_0$ *is written for* $\mathbf{X}_0^{\perp}$.

PROOF OF NECESSITY. As in Lemma 2.1, we consider the functions $\mathbf{X}_0'\mathbf{Y}$. The condition that $E(\mathbf{X}_0'\mathbf{Y})$ is the same for $(\mathbf{X}_0, \sigma^2\mathbf{I})$ and $(\mathbf{X}, \boldsymbol{\Sigma})$ implies that

$$(2.9) \qquad \mathbf{X} = \mathbf{X}_0 + \mathbf{Z}_0\mathbf{G},$$

for some $\mathbf{G}$. If $\mathbf{X}_0'\mathbf{Y}$ is optimal for $(\mathbf{X}, \boldsymbol{\Sigma})$, then

$$C(\mathbf{X}_0'\mathbf{Y}, \mathbf{Z}'\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\Sigma}) = \mathbf{X}_0'\boldsymbol{\Sigma}\mathbf{Z} = \mathbf{0},$$

where $\mathbf{Z} = \mathbf{X}^{\perp}$. We write

$$(2.10) \qquad \boldsymbol{\Sigma} = \mathbf{X}_0\boldsymbol{\Sigma}_1\mathbf{X}_0' + \mathbf{Z}_0\boldsymbol{\Sigma}_2\mathbf{Z}_0' + \mathbf{X}_0\boldsymbol{\Sigma}_3\mathbf{Z}_0' + \mathbf{Z}_0\boldsymbol{\Sigma}_3'\mathbf{X}_0',$$

where $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are symmetrical, as in Rao (1968, equation 2.8). Then

$$\mathbf{X}_0'\boldsymbol{\Sigma}\mathbf{Z} = \mathbf{X}_0'(\mathbf{X}_0\boldsymbol{\Sigma}_1\mathbf{X}_0' + \mathbf{X}_0\boldsymbol{\Sigma}_3\mathbf{Z}_0')\mathbf{Z} = \mathbf{0} \Rightarrow \mathbf{X}_0\boldsymbol{\Sigma}_1\mathbf{X}_0'\mathbf{X}_0 + \mathbf{Z}_0\boldsymbol{\Sigma}_3'\mathbf{X}_0'\mathbf{X}_0$$

$$= \mathbf{X}\mathbf{M} = \mathbf{X}_0\mathbf{M} + \mathbf{Z}_0\mathbf{G}\mathbf{M} \Rightarrow \mathbf{X}_0\boldsymbol{\Sigma}_1\mathbf{X}_0'\mathbf{X}_0 = \mathbf{X}_0\mathbf{M}, \qquad \mathbf{X}_0\boldsymbol{\Sigma}_1\mathbf{X}_0' = \mathbf{X}_0\mathbf{M}(\mathbf{X}_0'\mathbf{X}_0)^-\mathbf{X}_0',$$

$$\mathbf{Z}_0\boldsymbol{\Sigma}_3'\mathbf{X}_0'\mathbf{X}_0 = \mathbf{Z}_0\mathbf{G}\mathbf{M}, \qquad \mathbf{Z}_0\boldsymbol{\Sigma}_3'\mathbf{X}_0' = \mathbf{Z}_0\mathbf{G}\mathbf{M}(\mathbf{X}_0'\mathbf{X}_0)^-\mathbf{X}_0'.$$

Writing $\mathbf{A} = \mathbf{M}(\mathbf{X}_0'\mathbf{X}_0)^-$, $\mathbf{B} = \boldsymbol{\Sigma}_2$, $\boldsymbol{\Sigma}$ can be written in the form (2.8).

NOTE 1. In Corollary 2 to Lemma 2.1, it was noted that $\mathbf{X} = \mathbf{X}_0$ when the rank of $\mathbf{X}_0$ is full and $\mathbf{\Sigma}$ is restricted to the form $\sigma^2\mathbf{I}$. This result need not be true, when a general $\mathbf{\Sigma}$, as in (2.10), is considered.

NOTE 2. Lemma 2.2 asserts that, when $\mathbf{X}$ and $\mathbf{\Sigma}$ are as in (2.7, 2.8), a BLUE of an estimable parametric function under $(\mathbf{X}_0, \sigma^2\mathbf{I})$ is also a BLUE under $(\mathbf{X}, \mathbf{\Sigma})$. But this does not imply that the minimum variances attained in the two models are the same. To compare the variances let us first compute the dispersion matrices of $\mathbf{X}_0'\mathbf{Y}$ under the two models.

$$D(\mathbf{X}_0'\mathbf{Y} \mid \mathbf{X}_0, \sigma^2\mathbf{I}) = \sigma^2\mathbf{X}_0'\mathbf{X}_0,$$

$$D(\mathbf{X}_0'\mathbf{Y} \mid \mathbf{X}, \mathbf{\Sigma} \text{ as in } 2.8) = \mathbf{X}_0'\mathbf{X}_0\mathbf{A}\mathbf{X}_0'\mathbf{X}_0.$$

If the variances of a BLUE are to be equal, a necessary and sufficient condition is

$$\sigma^2\mathbf{X}_0'\mathbf{X}_0 = \mathbf{X}_0'\mathbf{X}_0\mathbf{A}\mathbf{X}_0'\mathbf{X}_0, \tag{2.11}$$

or $\mathbf{A} = \sigma^2(\mathbf{X}_0'\mathbf{X}_0)^-$, while $\mathbf{B}$ and $\mathbf{G}$ in (2.8) can be arbitrary.

LEMMA 2.3. *Let $\mathbf{P}$ be a matrix of order $m \times k$ such that $\mathfrak{M}(\mathbf{P}) \subset \mathfrak{M}(\mathbf{X}_0')$ and $R(\mathbf{P}) = k$. A necessary and sufficient condition for $\mathbf{L}'\mathbf{Y} = \mathbf{P}'(\mathbf{X}_0'\mathbf{X}_0)^-\mathbf{X}_0'\mathbf{Y}$ to be the* BLUEs *for $\mathbf{P}'\boldsymbol{\beta}$ under $(\mathbf{X}, \sigma^2\mathbf{I})$ is that*

$$\mathbf{X} = \mathbf{X}_0 + [\mathbf{I} - (\mathbf{L}\mathbf{L}^-)'][\mathbf{C}\mathbf{B}^- + \mathbf{D}(\mathbf{I} - \mathbf{B}\mathbf{B}^-)], \tag{2.12}$$

where $\mathbf{B} = (\mathbf{P}^-)'\mathbf{L}'\mathbf{L}$, $\mathbf{C} = \mathbf{L} - \mathbf{X}_0\mathbf{B}$ and $\mathbf{D}$ an arbitrary matrix of order $n \times m$.

PROOF OF SUFFICIENCY. The sufficiency part follows from the fact that $\mathbf{L}'\mathbf{X} = \mathbf{L}'\mathbf{X}_0 = \mathbf{P}'$ and $\mathbf{X}\mathbf{B} = \mathbf{X}_0\mathbf{B} + [\mathbf{I} - (\mathbf{L}\mathbf{L}^-)']\mathbf{C} = \mathbf{X}_0\mathbf{B} + \mathbf{C} = \mathbf{L}$ since $R(\mathbf{P}) = k \Rightarrow R(\mathbf{L}) = k \Rightarrow R(\mathbf{B}) = k \Rightarrow \mathbf{B}^-\mathbf{B} = \mathbf{I}$, and $\mathbf{L}'\mathbf{C} = \mathbf{L}'\mathbf{L} - \mathbf{L}'\mathbf{X}_0\mathbf{B} = \mathbf{L}'\mathbf{L} - \mathbf{P}'\mathbf{B} = \mathbf{0}$.

PROOF OF NECESSITY. To establish the necessity part note that $E(\mathbf{L}'\mathbf{Y} \mid \mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) = \mathbf{P}'\boldsymbol{\beta} = \mathbf{L}'(\mathbf{X} - \mathbf{X}_0) = \mathbf{0} \Rightarrow \mathbf{X} = \mathbf{X}_0 + [\mathbf{I} - (\mathbf{L}\mathbf{L}^-)']\mathbf{G}$. Further, the fact that $\mathbf{L}'\mathbf{Y}$ is BLUE for $\mathbf{P}'\boldsymbol{\beta}$ under $(\mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \Rightarrow \mathbf{L} = \mathbf{X}\mathbf{B}$ for some $\mathbf{B} \Rightarrow \mathbf{L}'\mathbf{L} = \mathbf{L}'\mathbf{X}\mathbf{B} = \mathbf{P}'\mathbf{B} \Rightarrow \mathbf{B} = (\mathbf{P}^-)'\mathbf{L}'\mathbf{L}$. The rest of the lemma follows from the general solution $(\mathbf{G})$ of the equation $\mathbf{A}\mathbf{G}\mathbf{B} = \mathbf{C}$, given in Theorem 2d of Rao (1967b) substituting $\mathbf{A} = \mathbf{I} - (\mathbf{L}\mathbf{L}^-)'$.

LEMMA 2.4. *Let us consider two alternative models $(\mathbf{X}, \sigma^2\mathbf{I})$ and $(\mathbf{X}, \mathbf{\Sigma})$ where $\mathbf{\Sigma}$ has the general representation $\mathbf{\Sigma} = \mathbf{X}\mathbf{\Sigma}_1\mathbf{X}' + \mathbf{Z}\mathbf{\Sigma}_2\mathbf{Z}' + \mathbf{X}\mathbf{\Sigma}_2\mathbf{Z}' + \mathbf{Z}\mathbf{\Sigma}_3'\mathbf{X}'$. Then the following are true.*

(a) $D(\mathbf{X}'\mathbf{Y} \mid \sigma^2\mathbf{I}) = D(\mathbf{X}'\mathbf{Y} \mid \mathbf{\Sigma}) \Rightarrow \mathbf{\Sigma}_1 = \sigma^2(\mathbf{X}'\mathbf{X})^-$.

(b) *Let $R_0^2 = \min_\beta (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{Y}'(\mathbf{I} - \mathbf{P}_X)\mathbf{Y}$. Then* $E[\mathbf{Y}'(\mathbf{I} - \mathbf{P}_X)\mathbf{Y} \mid \sigma^2\mathbf{I}] = E[\mathbf{Y}'(\mathbf{I} - \mathbf{P}_X)\mathbf{Y} \mid \mathbf{\Sigma}] \Rightarrow \text{trace } \mathbf{Z}\mathbf{\Sigma}_2\mathbf{Z}' = (n - r)\sigma^2$ *where r is the rank of $\mathbf{X}$ and n is the number of elements in $\mathbf{Y}$.*

(c) *Let $\mathbf{Y}$ have a multivariate normal distribution. Then the distribution R_0^2/σ^2 under $(\mathbf{X}, \mathbf{\Sigma})$ is χ^2 on $(n - r)$ degrees of freedom iff $(\mathbf{Z}\mathbf{\Sigma}_2\mathbf{Z}')/\sigma^2 = (\mathbf{I} - \mathbf{P}_X)$. If further $\mathbf{X}'\mathbf{Y}$ and R_0^2 are to be independently distributed, then $\mathbf{X}\mathbf{\Sigma}_3\mathbf{Z}' = \mathbf{Z}\mathbf{\Sigma}_3'\mathbf{X}' = \mathbf{0}$.*

PROOF. The result (a) is easily established. To prove (b), we observe that

$$E[\mathbf{Y}'(\mathbf{I}-\mathbf{P}_X)\mathbf{Y}\mid\mathbf{\Sigma}] = \operatorname{trace}[(\mathbf{I}-\mathbf{P}_X)\mathbf{\Sigma}] = \operatorname{trace}(\mathbf{Z\Sigma_2Z}' + \mathbf{Z\Sigma_3X}') = \operatorname{trace}\mathbf{Z\Sigma_2Z}'. \tag{2.13}$$

Since $E(\mathbf{Y}'(\mathbf{I}-\mathbf{P}_X)\mathbf{Y}\mid\sigma^2\mathbf{I}) = (n-r)\sigma^2$, the result (b) is established.

The necessary and sufficient condition for $\mathbf{Y}'(\mathbf{I}-\mathbf{P}_X)\mathbf{Y}/\sigma^2$ to have a χ^2-distribution is (using the condition on page 443 of Rao, 1965)

$$\begin{aligned}\mathbf{\Sigma}(\mathbf{I}-\mathbf{P}_X)\mathbf{\Sigma}(\mathbf{I}-\mathbf{P}_X)\mathbf{\Sigma} &= \sigma^2\mathbf{\Sigma}(\mathbf{I}-\mathbf{P}_X)\mathbf{\Sigma}\\ &\Leftrightarrow (\mathbf{Z\Sigma_2Z}' + \mathbf{X\Sigma_3Z}')(\mathbf{Z\Sigma_2Z}')(\mathbf{Z\Sigma_2Z}' + \mathbf{Z\Sigma_3}'\mathbf{X}')\\ &= \sigma^2(\mathbf{Z\Sigma_2Z}' + \mathbf{X\Sigma_3Z}')(\mathbf{Z\Sigma_2Z}' + \mathbf{Z\Sigma_3}'\mathbf{X}')\\ &\Leftrightarrow \mathbf{Z}'\mathbf{Z\Sigma_2Z}'\mathbf{Z\Sigma_2Z}'\mathbf{Z\Sigma_2Z}'\mathbf{Z} = \sigma^2\mathbf{Z}'\mathbf{Z\Sigma_2Z}'\mathbf{Z\Sigma_2Z}'\mathbf{Z}\\ &\Leftrightarrow (\mathbf{Z\Sigma_2Z}')(\mathbf{Z\Sigma_2Z}')(\mathbf{Z\Sigma_2Z}') = \sigma^2(\mathbf{Z\Sigma_2Z}')(\mathbf{Z\Sigma_2Z}')\\ &\Leftrightarrow (\mathbf{Z\Sigma_2Z}')/\sigma^2 \text{ is idempotent.}\end{aligned} \tag{2.14}$$

The df of χ^2 is (see page 443 Rao, 1965)

$$\operatorname{trace}(\mathbf{I}-\mathbf{P}_X)\mathbf{\Sigma}/\sigma^2 = \operatorname{trace}(\mathbf{Z\Sigma_2Z}' + \mathbf{Z\Sigma_3X}')/\sigma^2 = \operatorname{trace}\mathbf{Z\Sigma_2Z}'/\sigma^2. \tag{2.15}$$

If trace $\mathbf{Z\Sigma_2Z}' = (n-r)\sigma^2$, then we should have

$$\mathbf{Z\Sigma_2Z}' = \sigma^2(\mathbf{I}-\mathbf{P}_X). \tag{2.16}$$

If $\mathbf{Z}'\mathbf{Y}$ and $\mathbf{Y}'(\mathbf{I}-\mathbf{P}_X)\mathbf{Y}$ are to be independently distributed, then $\mathbf{X}'\mathbf{\Sigma}(\mathbf{I}-\mathbf{P}_X) = \mathbf{0} \Rightarrow \mathbf{X\Sigma_3Z}' = \mathbf{Z\Sigma_3}'\mathbf{X}' = \mathbf{0}$. Thus (c) is proved.

The following table gives the necessary and sufficient conditions on $\mathbf{\Sigma}$ for different procedures in the simple least squares theory (i.e. assuming $\mathbf{\Sigma} = \sigma^2\mathbf{I}$) to be optimal or valid.

Property	Representation of $\mathbf{\Sigma}$
(i) Every SLSE is BLUE	$\mathbf{X\Sigma_1X}' + \mathbf{Z\Sigma_2Z}'$
(ii) Expression for variance of a SLSE remains the same	$\sigma^2\mathbf{P}_X + \mathbf{Z\Sigma_2Z}' + \mathbf{X\Sigma_3Z}' + \mathbf{Z\Sigma_3}'\mathbf{X}'$
(iii) R_0^2/σ^2 is a χ^2 on $(n-r)$ df	$\mathbf{X\Sigma_1X}' + \sigma^2(\mathbf{I}-\mathbf{P}_X) + \mathbf{X\Sigma_3Z}' + \mathbf{Z\Sigma_3}'\mathbf{X}'$
(iv) Every SLSE and R_0^2 are independently distributed	$\mathbf{X\Sigma_1X}' + \mathbf{Z\Sigma_2Z}'$
(v) In addition to (iv), R_0^2/σ^2 is a χ^2 on $(n-r)$ df	$\mathbf{X\Sigma_1X}' + \sigma^2(\mathbf{I}-\mathbf{P}_X)$
(vi) In addition to (v), (ii) is satisfied	$\sigma^2\mathbf{I}$
(vii) In addition to (v), the variance of a particular SLSE, say that of $\mathbf{p}'\boldsymbol{\beta}$ is the same	$\sigma^2\mathbf{I} + \mathbf{XAX}'$ where $\mathbf{p}'\mathbf{Ap} = 0$

From the table we observe that while some of the procedures based on simple least squares theory are valid for a wider class of $\mathbf{\Sigma}$, for the full battery of procedures on estimation and testing to be applicable (case vi), it is necessary that $\mathbf{\Sigma} = \sigma^2\mathbf{I}$. Case (vii) shows that if the procedures are to be valid for particular parametric functions, then $\mathbf{\Sigma}$ can belong to a wider class. A well known example which falls in this category is the treatment of two way classification (mixed model) discussed in Rao (1965, page 216).

3. Tests of linear hypotheses when (2.1) **holds.** Consider a parametric function $\mathbf{p}'\boldsymbol{\beta}$ which is estimable under the model $(\mathbf{Y}, \mathbf{X}_0\boldsymbol{\beta}, \sigma^2\mathbf{I})$. It is clear that if the true model $(\mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ satisfies the conditions of Lemma 2.1, the expression for the BLUE of $\mathbf{p}'\boldsymbol{\beta}$ and for its variance as well remain unchanged in either case. Notice, however, that if R_0^2 denotes the residual sum of squares obtained under $(\mathbf{Y}, \mathbf{X}_0\boldsymbol{\beta}, \sigma^2\mathbf{I})$

$$\begin{aligned} E[\mathbf{R}_0^2 \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}] &= E[\mathrm{tr}\, \mathbf{Y}'(\mathbf{I} - \mathbf{P}_{X_0})\mathbf{Y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}] \\ &= E[\mathrm{tr}\, (\mathbf{I} - \mathbf{P}_{X_0})\mathbf{Y}\mathbf{Y}' \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}] \\ &= \mathrm{tr}\, (\mathbf{I} - \mathbf{P}_{X_0})(\sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}') \qquad (3.1) \\ &= (n - r)\sigma^2 + \boldsymbol{\beta}'\mathbf{X}'(\mathbf{I} - \mathbf{P}_{X_0})\mathbf{X}\boldsymbol{\beta} \\ &= (n - r)\sigma^2 + \boldsymbol{\beta}'\mathbf{A}'(\mathbf{I} - \mathbf{P}_{X_0})\mathbf{A}\boldsymbol{\beta} \end{aligned}$$

which is equal to $(n - r)\sigma^2$ if and only if $(\mathbf{I} - \mathbf{P}_{X_0})\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ that is $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_0\boldsymbol{\beta}$.

This shows that if the two models are distinct (a) $R_0^2/(n - r)$ overestimates σ^2 and (b) R_0^2/σ^2 is distributed as noncentral chi-square on $(n - r)$ df, with the noncentrality parameter $\lambda = \boldsymbol{\beta}'\mathbf{A}'(\mathbf{I} - \mathbf{P}_{X_0})\mathbf{A}\boldsymbol{\beta}/\sigma^2$, independently of the BLUE $\mathbf{p}'\hat{\boldsymbol{\beta}}$. Hence

$$(\mathbf{p}'\hat{\boldsymbol{\beta}} - \mathbf{p}'\boldsymbol{\beta})/[\mathbf{p}'(\mathbf{X}_0'\mathbf{X}_0)^{-}\mathbf{p}R_0^2/(n - r)]^{\frac{1}{2}}$$

is distributed as the ratio of a standard normal variable and an independent root mean (noncentral) chi-square on $(n - r)$ df. Also the usual F-statistic, computed for testing several linear hypotheses on unknown parameters, is, in the null case, distributed as the ratio of a (central) mean chi-square and an independent mean (noncentral) chi-square on $(n - r)$ df. For various values of the noncentrality parameter λ, the following table gives the true level attained by the F-test under the model $(\mathbf{X}, \sigma^2\mathbf{I})$ when 1 per cent and 5 per cent critical values under the model $(\mathbf{X}_0, \sigma^2\mathbf{I})$ are used.

Acknowledgment. The authors wish to thank P. Dasgupta of the Computer Science Unit, Indian Statistical Institute for the calculations in Table 1 carried out on IBM 1401 Computer.

TABLE 1

Table showing the actual level attained by the variance ratio test by using the tabulated critical value for 1% level of significance

n_1 = df for numerator; n_2 = df for denominator

n_1	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$
			$\alpha = 1\%, n_2 = 10$			
1	.0076	.0058	.0045	.0034	.0026	.0020
2	.0074	.0055	.0041	.0030	.0022	.0016
3	.0073	.0053	.0038	.0028	.0020	.0015
4	.0072	.0052	.0037	.0027	.0019	.0014
5	.0071	.0051	.0036	.0026	.0018	.0013
			$\alpha = 1\%, n_2 = 20$			
1	.0085	.0073	.0062	.0053	.0046	.0039
2	.0083	.0069	.0058	.0048	.0040	.0033
3	.0082	.0067	.0055	.0045	.0037	.0030
4	.0081	.0065	.0053	.0043	.0034	.0028
5	.0080	.0064	.0051	.0041	.0033	.0026
			$\alpha = 1\%, n_2 = 30$			
1	.0089	.0080	.0072	.0064	.0058	.0052
2	.0088	.0077	.0067	.0059	.0052	.0045
3	.0086	.0075	.0065	.0056	.0048	.0042
4	.0086	.0073	.0063	.0053	.0046	.0039
5	.0085	.0072	.0061	.0052	.0044	.0037
			$\alpha = 1\%, n_2 = 40$			
1	.0092	.0084	.0077	.0071	.0065	.0060
2	.0090	.0081	.0073	.0066	.0060	.0054
3	.0089	.0080	.0071	.0063	.0056	.0050
4	.0088	.0078	.0069	.0061	.0054	.0048
5	.0088	.0077	.0068	.0059	.0052	.0046
			$\alpha = 5\%, n_2 = 10$			
1	.0414	.0343	.0285	.0236	.0196	.0163
2	.0399	.0319	.0254	.0203	.0162	.0129
3	.0391	.0305	.0239	.0186	.0145	.0113
4	.0385	.0297	.0228	.0176	.0135	.0104
5	.0381	.0291	.0221	.0168	.0128	.0097
			$\alpha = 5\%, n_2 = 20$			
1	.0451	.0406	.0367	.0331	.0299	.0270
2	.0439	.0386	.0339	.0298	.0262	.0230
3	.0432	.0374	.0323	.0279	.0241	.0208
4	.0427	.0365	.0311	.0265	.0226	.0193
5	.0423	.0358	.0302	.0255	.0215	.0182

TABLE 1—*Continued*

n_1	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$
			$\alpha = 5\%, n_2 = 30$			
1	.0466	.0433	.0404	.0376	.0350	.0327
2	.0457	.0417	.0381	.0348	.0318	.0291
3	.0451	.0407	.0367	.0331	.0298	.0269
4	.0447	.0399	.0356	.0318	.0284	.0253
5	.0443	.0393	.0348	.0308	.0273	.0241
			$\alpha = 5\%, n_2 = 40$			
1	.0473	.0448	.0425	.0402	.0381	.0361
2	.0466	.0435	.0406	.0379	.0353	.0329
3	.0462	.0426	.0394	.0363	.0335	.0309
4	.0458	.0420	.0384	.0352	.0322	.0295
5	.0455	.0414	.0377	.0343	.0311	.0283

REFERENCES

[1] KRUSKAL, W. (1968). When are the Gauss-Markoff and least squares estimators identical? A coordinate free approach. *Ann. Math. Statist.* **39** 70–75.

[2] MITRA, S. K. and RAO, C. RADHAKRISHNA (1968). Some results in estimation and tests of linear hypotheses under the Gauss-Markoff model. *Sankhyā, Ser. A* **30** 281–290.

[3] RAO, C. RADHAKRISHNA (1962). A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. *J. Roy. Statist. Soc. Ser. B* **24** 152–158.

[4] RAO, C. RADHAKRISHNA (1965). *Linear Statistical Inference and its Applications.* Wiley, New York.

[5] RAO, C. RADHAKRISHNA (1966). Generalized inverse for matrices and its application in mathematical statistics. *Research Papers in Statistics* (Festchrift for J. Neyman). 263–280. Wiley, New York.

[6] RAO, C. RADHAKRISHNA (1967a). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 355–372 Univ. of California Press.

[7] RAO, C. RADHAKRISHNA (1967b). Calculus of generalized inverse of matrices, Part I: General theory. *Sankhyā Ser. A* **29** 317–342.

[8] RAO, C. RADKHAKRISHNA (1968). A note on a previous lemma in the theory of least squares and some further results. *Sankhyā Ser. A* **30** 245–252.

[9] WATSON, G. S. (1967). Linear least squares regression. *Ann. Math. Statist.* **38** 1679–1699.

[10] ZYSKIND, G. (1967). On canonical forms, negative covariance matrices and best and simple least squares linear estimator in linear models. *Ann. Math. Statist.* **38** 1092–1110.