The Annals of Statistics
1974, Vol. 2, No. 5, 935-949

COEFFICIENT ERRORS CAUSED BY USING THE WRONG COVARIANCE MATRIX IN THE GENERAL LINEAR MODEL

BY OTTO NEALL STRAND

Wave Propagation Laboratory, Boulder, Colorado

A method is derived to place an approximate bound on the mean-square error incurred by using an incorrect covariance matrix in the Gauss-Markov estimator of the coefficient vector in the full-rank general linear model. The bound thus obtained is a function of the incorrect covariance matrix $\tilde{S}$ actually used, the Frobenius norm of $S - \tilde{S}$, where S is the correct covariance matrix, and the basis matrix ϕ. This estimate can therefore be computed from known or easily-approximated data in the usual regression problem. All mathematics related to the method is derived, and numerical examples are presented.

1. Introduction. In this paper we are concerned with the full-rank general linear model

$$y = \phi\alpha + e \tag{1.1}$$

where

y is an $n \times 1$ vector of real measurements,
ϕ is an $n \times p$ real matrix of rank p, $p < n$,
α is an unknown $p \times 1$ real vector of coefficients,
e is an $n \times 1$ random real column vector of measurement errors,

such that

$$E(e) = 0 \quad \text{and} \tag{1.2}$$

$$E(ee^T) = S\,. \tag{1.3}$$

Here S is assumed strictly positive definite, $E(\cdot)$ is the expected-value operator, and the superscript ^{T} denotes matrix transposition.

It is well known that the optimum linear unbiased estimate of α (i.e., that estimate having the smallest expected mean-square error) is achieved when the residuals are weighted in accordance with S^{-1}. However, in practice S is not known, but one only has an approximation $\tilde{S}$ to S and perhaps a reasonable bound on the departure of $\tilde{S}$ from S. If one uses $\tilde{S}$ in place of S, one will ordinarily incur an error in the estimated coefficient vector α. In this paper we develop a method of placing an approximate bound on the expected mean-square value of this error. This bound is a function of the basis matrix ϕ, the

Received April 1973; revised January 1974.
AMS 1970 *subject classifications*. Primary 15A60; Secondary 62F10.
Key words and phrases. Covariance matrix, expected mean-square error, incorrect covariances, matrix trace expressions.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*. www.jstor.org

approximate covariance matrix $\tilde{S}$ and a norm of the departure of $\tilde{S}$ from the true covariance matrix S. Several authors [1], [3], [4], [5], [6], [8], [10], have recently investigated questions closely related to this. However, although these papers are of theoretical interest, they do not provide readily usable estimates of the error to be expected in α for a given size error in S. The present paper differs from these others in that it attacks the mean-square error expression directly to provide an estimate of the maximum possible expected mean-square error in terms of quantities that are often available; that is, in terms of ϕ, $\tilde{S}$ and a bound on the Frobenius norm of $S - \tilde{S}$.

In the following sections we begin by obtaining covariance and mean-square error expressions (Section 2), show how to maximize the mean-square error in α for a given covariance-matrix error norm (Section 3), show how to determine the required maximal eigenvalue (Section 4), apply the theory to regression errors for unit-weight least squares (Section 5), and finally, present numerical illustrations (Section 6).

2. Covariance and mean-square error results. The results in the first part of this section are well known and are easily derived by taking expected values and traces. We therefore do not include derivations. In this paper we assume that both $\tilde{S}$ and S are *strictly* positive definite and that $\tilde{S}$ and ϕ are known. The *optimum* or *Gauss–Markov* estimate of α is given by $\hat{\alpha}$, where

$$\hat{\alpha} = (\phi^T S^{-1} \phi)^{-1} \phi^T S^{-1} y \,. \tag{2.1}$$

Furthermore,

$$E(\hat{\alpha}) = \alpha \,, \tag{2.2}$$

so that α is unbiased. The covariance matrix of $\hat{\alpha}$ is

$$S_{\hat{\alpha}-\alpha} = (\phi^T S^{-1} \phi)^{-1} \,. \tag{2.3}$$

(Whenever convenient, we denote the covariance matrix of a random vector v by S_v.) Now suppose that, instead of using the correct covariance matrix S, we use the incorrect covariance matrix $\tilde{S}$ in the estimation. Then we form the corresponding estimate

$$\tilde{\alpha} = (\phi^T \tilde{S}^{-1} \phi)^{-1} \phi^T \tilde{S}^{-1} y \,, \tag{2.4}$$

where the only difference between (2.1) and (2.4) is that S is replaced by $\tilde{S}$ in the latter. Again we have

$$E(\tilde{\alpha}) = \alpha \tag{2.5}$$

and this time

$$S_{\tilde{\alpha}-\hat{\alpha}} = V - (\phi^T S^{-1} \phi)^{-1} \tag{2.6}$$

where

$$V = (\phi^T \tilde{S}^{-1} \phi)^{-1} \phi^T \tilde{S}^{-1} S \tilde{S}^{-1} \phi (\phi^T \tilde{S}^{-1} \phi)^{-1} = S_{\tilde{\alpha}-\alpha} \,. \tag{2.7}$$

It follows from (2.6) and (2.7) that

$$S_{\tilde{\alpha}-\alpha} = S_{\tilde{\alpha}-\hat{\alpha}} + S_{\hat{\alpha}-\alpha} \,. \tag{2.8}$$

By taking traces we obtain

$$E[(\tilde{\alpha} - \alpha)^T(\tilde{\alpha} - \alpha)] = E[(\tilde{\alpha} - \hat{\alpha})^T(\tilde{\alpha} - \hat{\alpha})] + E[(\hat{\alpha} - \alpha)^T(\hat{\alpha} - \alpha)] \,. \tag{2.9}$$

It is also covenient to write

$$\operatorname{tr} S_{\tilde{\alpha}-\hat{\alpha}} = \operatorname{tr} P_1 S - \operatorname{tr} (\phi^T S^{-1} \phi)^{-1} \tag{2.10}$$

and

$$\operatorname{tr} S_{\tilde{\alpha}-\alpha} = \operatorname{tr} P_1 S \tag{2.11}$$

where

$$P_1 = \tilde{S}^{-1}\phi(\phi^T \tilde{S}^{-1}\phi)^{-2}\phi^T \tilde{S}^{-1} \tag{2.12}$$

and we note that P_1 involves only the basis matrix ϕ, which is assumed known, and the known incorrect covariance matrix $\tilde{S}$. Throughout this paper we shall be concerned with $\operatorname{tr} S_{\tilde{\alpha}-\hat{\alpha}}$ (as given by (2.10)) as a measure of the departure of $\tilde{\alpha}$ from the optimum solution $\hat{\alpha}$.

It is apparent from (2.10) that the nonlinear dependence of the term $\operatorname{tr} [(\phi^T S^{-1} \phi)^{-1}]$ on the (unknown) correct covariance matrix S presents difficulties which must be overcome. For this purpose we let

$$S = \tilde{S} + \Delta S \,, \tag{2.13}$$

where ΔS is a symmetric error matrix. Further define the Frobenius (or Euclidean) norm of any square matrix A by

$$||A||^2 = \operatorname{tr} A^T A \,. \tag{2.14}$$

Equation (2.14) defines the only type of matrix norm to be used in this paper.

Since S itself is not known, we will assume that a number $\varepsilon > 0$ is known such that

$$||\Delta S|| \leqq \varepsilon \,. \tag{2.15}$$

An important restriction is made at the outset: the number ε must be sufficiently small that for every ΔS satisfying (2.15), the matrix S in (2.13) is strictly positive definite. Since the eigenvalues of $\tilde{S}$ are all positive by hypothesis, and since the eigenvalues of $\tilde{S} + \Delta S$ are continuous functions of ΔS, and furthermore all $\tilde{S} + \Delta S$ are symmetric (having real eigenvalues), it follows that as ΔS is continuously varied in compliance with (2.15), the eigenvalues of $\tilde{S} + \Delta S$ move continuouly along the real axis and that the set of admissible ε has the form $\{\varepsilon \,|\, 0 < \varepsilon < \varepsilon_0 \text{ for some } \varepsilon_0 > 0\}$. What we do not permit is for any eigenvalue of $\tilde{S} + \Delta S$ to touch or cross the origin as ΔS varies arbitrarily over all symmetric matrices satisfying (2.15). In what follows we will assume that this restriction is satisfied unless a specific statement to the contrary is made. It will be discovered that many of the arguments in this paper depend on the interaction of P_1 (see (2.12)) and P_2, defined below:

$$P_2 = \tilde{S}^{-1} - \tilde{S}^{-1}\phi(\phi^T\tilde{S}^{-1}\phi)^{-1}\phi^T\tilde{S}^{-1} . \tag{2.16}$$

Note that P_2 also involves only known quantities. The following identities follow readily from (2.12) and (2.16):

$$\begin{aligned} &P_1^T = P_1 \quad \text{and} \quad P_2^T = P_2 \\ &\phi^T P_2 = P_2\phi = 0 \\ &\phi^T P_1 \phi = I_p = p \times p \quad \text{identity} \\ &\operatorname{rank} P_1 = p \\ &P_2\tilde{S}P_1 = P_1\tilde{S}P_2 = 0 \\ &P_2\tilde{S} \quad \text{and} \quad \tilde{S}P_2 \quad \text{are idempotent,} \\ &\text{i.e.,} \quad (P_2\tilde{S})^2 = P_2\tilde{S} \quad \text{and} \quad (\tilde{S}P_2)^2 = \tilde{S}P_2 \\ &\phi(\phi^T\tilde{S}^{-1}\phi)^{-1}\phi^T = \tilde{S} - \tilde{S}P_2\tilde{S} . \end{aligned} \tag{2.17}$$

In the special case $\tilde{S} = I$, some of the relationships (2.17) may be written as

$$\begin{aligned} &P_2 P_1 = P_1 P_2 = 0 \\ &P_2^j = P_2 , \qquad j = 1, 2, 3, \cdots \\ &\phi(\phi^T\phi)^{-1}\phi^T = I - P_2 . \end{aligned} \tag{2.18}$$

It follows from the continuity of the eigenvalues of S as functions of the coefficients ΔS_{ij} that the set $E = \{\Delta S \mid S \text{ is strictly positive definite}\}$ is an open set in the sense that if $\Delta S \in E$ then for each $i \leqq j$ the element $\Delta S_{ij} \equiv \Delta S_{ji}$ can be varied within an open interval without causing ΔS to depart from E. Thus every $\Delta S \in E$ is an interior element.

If $\Delta S \in E$, then the matrix $I + P_2\Delta S$ is nonsingular. This will be established by proving that if v is an n-tuple, then

$$(I + P_2\Delta S)v = 0 \quad \text{implies} \quad v = 0 . \tag{2.19}$$

Suppose (2.19) is satisfied; then by (2.13),

$$[(I - P_2\tilde{S}) + P_2 S]v = 0 . \tag{2.20}$$

Since $P_2\tilde{S}$ is idempotent (see 2.17) we may multiply (2.20) on the left by $P_2\tilde{S}$ to get

$$P_2\tilde{S}P_2 S v = 0 . \tag{2.21}$$

Since $\tilde{S}P_2$ is also idempotent, we may multiply (2.21) on the left by $\tilde{S}$ to get $\tilde{S}P_2 S v = 0$, from which

$$P_2 S v = 0 , \quad \text{since} \quad \tilde{S} \quad \text{is nonsingular.} \tag{2.22}$$

Substituting (2.22) into (2.20) gives

$$v = P_2\tilde{S}v . \tag{2.23}$$

We calculate $v^T S v = v^T S P_2 \tilde{S} v = (P_2 S v)^T \tilde{S} v = 0$. Since S is strictly positive definite by hypothesis, this implies $v = 0$.

We are now ready to state and prove a theorem that greatly simplifies the study of the expected mean-square departure of $\tilde{\alpha}$ from $\hat{\alpha}$.

THEOREM 2.1. *If $\tilde{S} + \Delta S = S$ is a strictly positive definite matrix, then*

$$E[(\tilde{\alpha} - \hat{\alpha})^T(\tilde{\alpha} - \hat{\alpha})] = \operatorname{tr} P_1 S - \operatorname{tr} (\phi^T S^{-1} \phi)^{-1} \equiv \operatorname{tr} [P_1 \Delta S P_2 \Delta S (I + P_2 \Delta S)^{-1}], \tag{2.24}$$

where P_1 is given by (2.12), *P_2 is given by* (2.16), *and ΔS is given by* (2.13). *The equality of the second and third members of* (2.24) *holds whether S is strictly positive definite or not.*

PROOF. We first note the useful identity

$$(I + R)^{-1} = \sum_{j=0}^{k} (-R)^j + T_k \tag{2.25}$$

where $k = 0, 1, 2, \cdots, T_k = (-R)^{k+1}(I + R)^{-1} \equiv (I + R)^{-1}(-R)^{k+1}$, R is any square matrix for which $I + R$ is nonsingular. Formula (2.25) is easily established by pre- or post-multiplication by $I + R$. We calculate from (2.13)

$$(\phi^T S^{-1} \phi)^{-1} = (\phi^T \tilde{S}^{-1} \phi)^{-1}(I - A)^{-1} \tag{2.26}$$

where

$$A = \phi^T \tilde{S}^{-1} \Delta S(\ + \tilde{S}^{-1} \Delta S)^{-1} \tilde{S}^{-1} \phi (\phi^T \tilde{S}^{-1} \phi)^{-1}, \tag{2.27}$$

and from (2.16)

$$(I + \tilde{S}^{-1} \Delta S)^{-1} = (I + P_2 \Delta S)^{-1}(I - B) \tag{2.28}$$

where

$$B = \tilde{S}^{-1} \phi (\phi^T \tilde{S}^{-1} \phi)^{-1} \phi^T \tilde{S}^{-1} \Delta S (I + \tilde{S}^{-1} \Delta S)^{-1}. \tag{2.29}$$

Applying (2.25) to (2.26) with $k = 0$ and $R = -A$ gives

$$(\phi^T S^{-1} \phi)^{-1} = (\phi^T \tilde{S}^{-1} \phi)^{-1} + (\phi^T \tilde{S}^{-1} \phi)^{-1} A (I - A)^{-1}. \tag{2.30}$$

Substituting (2.28) and (2.29) into (2.27) gives

$$A = \phi^T \tilde{S}^{-1} \Delta S (I + P_2 \Delta S)^{-1} \tilde{S}^{-1} \phi (\phi^T \tilde{S}^{-1} \phi)^{-1} [I - A]. \tag{2.31}$$

Substituting (2.31) into (2.30) and taking traces gives

$$\operatorname{tr} (\phi^T S^{-1} \phi)^{-1} = \operatorname{tr} [P_1 \tilde{S} + P_1 \Delta S (I + P_2 \Delta S)^{-1}]. \tag{2.32}$$

Applying (2.25) with $R = P_2 \Delta S$ and $k = 0$, and noting (2.13), shows that (2.32) is equivalent to (2.24).

Note that the proof just given holds for any symmetric matrix ΔS for which the indicated inverses exist. In fact, the identity of the second and third members of (2.24) is valid even when the resulting matrix S is not positive definite, and this fact was demonstrated numerically for randomly-chosen symmetric matrices ΔS. Expression (2.24) is the basis for the analysis of this paper.

The objective is to find a maximum for the mean-square error expression (2.24) under the assumption that $||\Delta S||$ is fixed.

3. Maximization of mean-square error for a given covariance-matrix error norm. We begin by maximizing the expression (2.24) under the assumption that

$$||\Delta S||^2 = \operatorname{tr}\{(\Delta S)^T \Delta S\} = \varepsilon^2 . \tag{3.1}$$

Since all admissible ΔS must be symmetric, we require the further constraint

$$\Delta S_{ij} = \Delta S_{ji} \quad \text{for} \quad i < j\,, \quad \text{where} \quad \Delta S = (\Delta S_{ij})\,. \tag{3.2}$$

In accordance with the method of Lagrange, we require that the quantity $G(\Delta S)$ be stationary, where

$$\begin{aligned} G(\Delta S) &= \operatorname{tr}\{P_1 \Delta S P_2 \Delta S (I + P_2 \Delta S)^{-1}\} \\ &\quad - \lambda[\operatorname{tr}\{(\Delta S)^T \Delta S\} - \varepsilon^2] - \textstyle\sum_{j=2}^{n} \sum_{i=1}^{j-1} \mu_{ji}(\Delta S_{ji} - \Delta S_{ij}) \end{aligned} \tag{3.3}$$

and λ and the μ_{ij} are Lagrange multipliers. We may differentiate the first trace in (3.3) as follows. If we denote this trace by $f(\Delta S)$, then

$$\begin{aligned} &f(\Delta S + H) - f(\Delta S) \\ &\quad = \operatorname{tr}[P_2 \Delta S (I + P_2 \Delta S)^{-1} P_1 H + (I + P_2 \Delta S)^{-1} P_1 \Delta S P_2 H \\ &\qquad - (I + P_2 \Delta S)^{-1} P_1 \Delta S P_2 \Delta S (I + P_2 \Delta S)^{-1} P_2 H + O(H^2)]\,. \end{aligned}$$

Differentiating the linear terms with respect to H (note the well-known formula $(d/dH)[\operatorname{tr} BH] = B^T$), simplifying and applying the condition $\Delta S = \Delta S^T$ gives

$$\begin{aligned} &(d/d\Delta S) \operatorname{tr}\{P_1 \Delta S P_2 \Delta S (I + P_2 \Delta S)^{-1}\} \\ &\qquad = (I + P_2 \Delta S)^{-1} Q(\Delta S)(I + \Delta S P_2)^{-1} \end{aligned} \tag{3.4}$$

where

$$Q(\Delta S) = P_1 \Delta S P_2 + P_2 \Delta S P_1 + P_2 \Delta S P_1 \Delta S P_2\,. \tag{3.5}$$

Also,

$$(d/d\Delta S) \operatorname{tr}\{(\Delta S)^T \Delta S\} = 2\Delta S \tag{3.6}$$

and

$$(d/d\Delta S)[\textstyle\sum_{j=2}^{n} \sum_{i=1}^{j-1} \mu_{ji}(\Delta S_{ji} - \Delta S_{ij})] = M = (M_{ij})$$

where

$$\begin{aligned} M_{ij} &= -\mu_{ji} \quad & i < j \\ &= 0 & i = j \\ &= \mu_{ij} & i > j\,. \end{aligned} \tag{3.7}$$

Thus $M = -M^T$ so that M is skew symmetric. The stationarity condition $(d/d\Delta S)[G(\Delta S)] = 0$ and the symmetry of ΔS together imply that M is also symmetric; consequently $M = 0$ and all μ_{ij} are zero. We therefore obtain the equation

$$T_{\mathrm{NL}}(\Delta S) = \lambda \Delta S \tag{3.8}$$

where the constraint (3.2) has been imposed, the subscript NL denotes "nonlinear," and

$$2T_{\mathrm{NL}}(\Delta S) = (I + P_2\Delta S)^{-1}Q(\Delta S)(I + \Delta SP_2)^{-1} \tag{3.9}$$

where $Q(\Delta S)$ is defined by (3.5).

Up to this point no approximations have been made. Two formidable difficulties become apparent. First, it is not clear that solving (3.8) to maximize (2.24) using the equality constraint $||\Delta S|| = \varepsilon$ will necessarily suffice for the inequality constraint $||\Delta S|| \leqq \varepsilon$. Furthermore, solving the nonlinear eigenvalue problem (3.8) would itself be a formidable task. We therefore resort to an approximation which consists of maximizing the leading term of (2.24). Applying identity (2.25) with $k = 0$ and $R = P_2\Delta S$ to (2.24) gives

$$E\{(\tilde{\alpha} - \hat{\alpha})^T(\tilde{\alpha} - \hat{\alpha})\} = \operatorname{tr}\{P_1\Delta SP_2\Delta S[I - P_2\Delta S(I + P_2\Delta S)^{-1}]\} \tag{3.10}$$

so that the error committed by using a second-degree approximation is

$$\text{Error} = -\operatorname{tr}\{P_1\Delta S(P_2\Delta S)^2(I + P_2\Delta S)^{-1}\}\,. \tag{3.11}$$

Since (3.11) is of degree 3 or higher in $P_2\Delta S$, it can be ignored for sufficiently small ΔS. Of course, if (3.11) could be bounded in a satisfactory manner, a rigorous upper bound to (2.24) could be obtained. Otherwise the bound is only approximate. However, even a rough upper bound on the expected mean-square error is often useful. The situation is similar to that occurring when one uses differentials to estimate errors in a function of several variables.

Dropping third and higher-degree terms from (2.24) gives the simpler problem

$$\operatorname{tr}\{P_1\Delta SP_2\Delta S\} = \text{maximum} \tag{3.12}$$

under the constraints $||\Delta S|| = \varepsilon$ and $\Delta S = \Delta S^T$ as in (3.2). It will be seen that the resulting maximum will also suffice for the inequality constraint. Carrying out the differentiation gives the following equation corresponding to (3.8):

$$\tfrac{1}{2}(P_1\Delta SP_2 + P_2\Delta SP_1) = \lambda\Delta S\,. \tag{3.13}$$

We present some facts about expressions (2.24) and (3.12) and the solutions of (3.8) and (3.13).

First, if S if strictly positive definite, then

$$\operatorname{tr}\{P_1\Delta SP_2\Delta S\} \geqq 0 \quad \text{for all symmetric matrices} \quad \Delta S\,. \tag{3.14}$$

This follows from (2.12) and the fact that P_2 is nonnegative definite by writing

$$\operatorname{tr}\{P_1\Delta SP_2\Delta S\} = \operatorname{tr}\{[\Delta S\tilde{S}^{-1}\phi(\phi^T\tilde{S}^{-1}\phi)^{-1}]^T P_2[\Delta S\tilde{S}^{-1}\phi(\phi^T\tilde{S}^{-1}\phi)^{-1}]\} \geqq 0\,. \tag{3.15}$$

The right-hand side of the exact expression (2.24) is also nonnegative, as it is a mean-square error.

It follows readily from (2.24) that the two types of symmetric matrices

$$\Delta S = \phi A\phi^T \tag{3.16}$$

where A is any symmetric $p \times p$ matrix and

$$\Delta S = k \cdot \tilde{S} \tag{3.17}$$

where k is any constant for which S is positive definite, are such that

$$E\{(\tilde{\alpha} - \hat{\alpha})^T(\tilde{\alpha} - \hat{\alpha})\} = 0 \,. \tag{3.18}$$

These are special cases in which the expected mean-square error is not increased by an error, ΔS, in the covariance matrix. If $\tilde{S} = I$, such cases are easily characterized.

THEOREM 3.1. *If $\tilde{S} = I$ and S is strictly positive definite, then* (3.18) *holds if and only if*

$$P_1 \Delta S P_2 = 0 \tag{3.19}$$

or equivalently

$$\phi^T \Delta S P_2 = 0 \,. \tag{3.20}$$

In case ΔS satisfies (3.19), then both (2.24) and the leading term $\operatorname{tr}\{P_1 \Delta S P_2 \Delta S\}$ vanish, and both (3.8) and (3.13) are satisfied with $\lambda = 0$.

PROOF. Suppose (3.18) holds. Then, since $\operatorname{tr}\{P_1 \Delta S P_2 \Delta S(I + P_2 \Delta S)^{-1}\}$ is zero, it is an unconstrained minimum, and according to the discussion in Section 2, ΔS is an interior point of the set E. Thus ΔS satisfies (3.8) with $\lambda = 0$. Recalling that $I + P_2 \Delta S$ and $I + \Delta S P_2 = (I + P_2 \Delta S)^T$ are nonsingular matrices and cancelling the inverses gives

$$Q(\Delta S) = 0 \,. \qquad \text{(See (3.5).)} \tag{3.21}$$

Pre-multiplying by $I - P_2$ and noting (2.18) gives $P_1 \Delta S P_2 = 0$, and (3.13) is satisfied with $\lambda = 0$. Conversely, if $P_1 \Delta S P_2 = 0$, then both (2.24) and the leading term vanish and both (3.8) and (3.13) are satisfied with $\lambda = 0$.

We note that ΔS as given by (3.16) and (3.17) both satisfy (3.19) when $\tilde{S} = I$. It is known [8], that the most general assignment of ΔS for which (3.18) holds (when $\tilde{S} = I$) has the form

$$\Delta S = \phi A \phi^T + x B x^T + r \cdot I \tag{3.22}$$

where r is a scalar, A and B are arbitrary symmetric matrices and x is an n-tuple such that $x^T \phi = 0$. It is easily verified that (3.22) satisfies (3.19).

If the simplified equation (3.13) is satisfied, and $\tilde{S} = I$, then the last member of (2.24) reduces to (3.12) even if $\lambda \neq 0$. Because of this, one might expect that the corresponding solution of (3.13) would solve (3.8). Unfortunately, this is not the case. To see this, note that if X is a symmetric solution to (3.13), $\lambda \neq 0$ and S is strictly positive definite, then

$$P_2 X P_2 = 0 \,. \tag{3.23}$$

If $X \neq 0$ also satisfies (3.8), we have

$$P_1 X P_2 + P_2 X P_1 + P_2 X P_1 X P_2 = 2\lambda(I + P_2 X) X (I + X P_2) \,. \tag{3.24}$$

Pre-multiplication by P_2 gives

$$P_2XP_1 = 2\lambda P_2(I + P_2X)X \tag{3.25}$$

and post-multiplication of (3.25) by P_2 gives $P_2(I + P_2X)XP_2 = 0$ or

$$P_2XP_2 = -P_2X^2P_2 . \tag{3.26}$$

According to (3.23) this implies

$$P_2X^2P_2 = (P_2X)(P_2X)^T = 0 , \tag{3.27}$$

from which $P_2X = 0$ and therefore $P_2XP_1 = P_1XP_2 = 0$ which, according to (3.13), is a contradiction unless $\lambda = 0$. Thus, even when $\tilde{S} = I$, a solution of (3.13) only approximates the maximum expected mean-square error in α.

We continue the discussion of the approximate problem (3.12). Multiplying by $\Delta S \equiv \Delta S^T$ and taking the trace gives

$$\operatorname{tr}\{P_1XP_2X\} = \lambda\varepsilon^2 \tag{3.28}$$

if X is a solution of (3.12) under the stated constraints. Thus (3.12) will be maximized by substituting the maximal λ for which (3.13) has a solution into (3.28). It follows from (3.15) that the desired solution for λ must be positive. Since (3.28) defines an increasing function of ε^2, this value of λ will then maximize (3.12) under the constraint $||\Delta S||^2 \leqq \varepsilon^2$. In the next section we derive a method of solving (3.13). Because of the fact that the unknown element, ΔS, is a *matrix* and not an n-tuple, (3.13) is not (as it stands) a vector-matrix characteristic equation. However, it will be discovered in the next section that (3.13) represents a characteristic equation in a space whose elements are $n \times n$ symmetric matrices.

4. Determination of maximal operator eigenvalue. Let $\mathscr{H}$ be the set of all $n \times n$ real symmetric matrices. We denote elements of $\mathscr{H}$ by capital letters $X, Y, \cdots$. (These are elements such as ΔS in (3.13).) Clearly $\mathscr{H}$ is a linear manifold. We define the *inner product* of elements of $\mathscr{H}$ by the formula

$$(X, Y) = \operatorname{tr} XY^T = \operatorname{tr} XY = \operatorname{tr} YX . \tag{4.1}$$

With the definition (4.1) $\mathscr{H}$ is a finite-dimensional Hilbert space. Taking as a basis the set of matrices having ones in two symmetrically-placed locations and zeros elsewhere, together with matrices having a single one on the diagonal and zeros elsewhere, shows that $\mathscr{H}$ has dimension $n(n + 1)/2$.

Consider the operator T defined by the left-hand side of (3.13).

$$TX \equiv \tfrac{1}{2}[P_1XP_2 + P_2XP_1] , \qquad X \in \mathscr{H} . \tag{4.2}$$

Since TX is again a real symmetric $n \times n$ matrix, T maps $\mathscr{H}$ into $\mathscr{H}$. It is easily verified that T is linear. (It is also true that T_{NL} (see (3.8)) maps $\mathscr{H}$ into $\mathscr{H}$, although the mapping is nonlinear.) Now if $X, Y \in \mathscr{H}$, then

$$\begin{aligned}(TX, Y) &= \operatorname{tr}[\tfrac{1}{2}(P_1XP_2 + P_2XP_1)Y] \\ &= \operatorname{tr}\{X[\tfrac{1}{2}(P_2YP_1 + P_1YP_2)]\} = (X, TY) .\end{aligned} \tag{4.3}$$

Therefore T is self adjoint. Since $(TX, X) = \operatorname{tr}(P_1XP_2X) \geqq 0$ for all $X \in \mathscr{H}$ (see (3.15)), it follows that T is nonnegative definite. Thus T is a *compact, self-adjoint nonnegative definite* operator defined on $\mathscr{H}$.

Equation (3.13) may be written in the form

$$TX = \lambda X \tag{4.4}$$

and what we require to solve (3.13) for λ to use in (3.28) is a *maximal eigenvalue* of the operator T. We may invoke the theory of such operators [9], and state that there exists a maximal positive λ which provides a nontrivial solution to (4.4) (and hence (3.13)) and that any matrix X which corresponds to the maximal λ in (4.4) will cause

$$(TX, X) = \operatorname{tr}[P_1XP_2X] \tag{4.5}$$

to be a maximum for a given $||X|| = ||\Delta S|| = \varepsilon$.

We develop a method for finding the required maximal λ. It follows from the theory of operators such as T [9] that any element $X_0 \in \mathscr{H}$ can be written in the form

$$X_0 = \alpha_1H_1 + \alpha_2H_2 + \cdots + \alpha_kH_k + \alpha_{k+1}H_{k+1} + \cdots + \alpha_pH_p + Q \tag{4.6}$$

where

$$\begin{aligned} TH_i &= \lambda H_i\,, & i &= 1, 2, \cdots k \\ TH_i &= \lambda_{i-k}H_i\,, & i &= k+1, \cdots p \\ TQ &= 0\,, & &\text{all } H_i \text{ are orthonormal,} \\ & & &\text{and } \lambda > \lambda_1 \geqq \lambda_2 \cdots \geqq \lambda_{p-k} > 0\,. \end{aligned} \tag{4.7}$$

Here we deliberately single out the subspace of eigenelements associated with the maximal eigenvalue λ. Numerical experiments have shown that this subspace may be multidimensional. Although this does not affect the validity of the theory, it must be provided for in the derivation. We develop a slight generalization of the *power* method of von Mises [7], similar to a method often employed for matrix eigenvalue problems. Assume that X_0 is such that

$$\textstyle\sum_{i=1}^{k} \alpha_i^2 \neq 0 \qquad \text{in (4.6)} \tag{4.8}$$

and repeatedly apply the operator T to X_0. For $j \geqq 1$ the result is

$$\begin{aligned} T^jX_0 = \lambda^j \Big[&(\alpha_1H_1 + \alpha_2H_2 + \cdots + \alpha_kH_k) \\ &+ \left(\frac{\lambda_1}{\lambda}\right)^j \alpha_{k+1}H_{k+1} + \cdots + \left(\frac{\lambda_{p-k}}{\lambda}\right)^j \alpha_pH_p \Big]. \end{aligned} \tag{4.9}$$

The *Rayleigh quotient* for the operator T with argument $V \in \mathscr{H}$ is:

$$R = (TV, V)/(V, V)\,. \tag{4.10}$$

Letting R_j be the Rayleigh quotient for T with argument T^jX_0 gives

$$R_j = (T^{j+1}X_0, T^jX_0)/(T^jX_0, T^jX_0)\,. \tag{4.11}$$

Evaluating (4.11) with the help of (4.9) and the orthonormality of the H_i gives

$$R_j = a_j/b_j \qquad \text{where}$$

$$(4.12) \qquad a_j = \lambda^{2j+1}[\textstyle\sum_{i=1}^{k} \alpha_i^2 + \sum_{i=1}^{p-k} \alpha_{k+i}^2(\lambda_i/\lambda)^{2j+1}]$$

$$b_j = \lambda^{2j}[\textstyle\sum_{i=1}^{k} \alpha_i^2 + \sum_{i=1}^{p-k} \alpha_{k+i}^2(\lambda_i/\lambda)^{2j}] \, .$$

Since $0 < \lambda_i < \lambda$, $i = 1, 2, \cdots, (p - k)$, it follows from (4.12) and (4.8) that:

$$(4.13) \qquad \lim_{j\to\infty} R_j = \lambda \, .$$

Furthermore, the limit

$$(4.14) \qquad X = \lim_{j\to\infty} T^j X_0/||T^j X_0||$$

is an element such that $TX = \lambda X$. (See (4.9).) In practical computations we take X_0 as a symmetric matrix whose entries are obtained from a random number generator. We cease applying T to X_0 as soon as R_j and R_{j+1} agree to the desired precision. As a check, the corresponding element $T^j X_0/||T^j X_0||$ then approximates an element X for which the maximum in (3.12) is actually achieved. Substituting λ into (3.28) gives the desired bound on the expected mean-square error in the estimate of α. In summary, we have the following approximate error bound, for which (3.12) is maximized,

$$(4.15) \qquad E[(\hat{\alpha} - \tilde{\alpha})^T(\tilde{\alpha} - \hat{\alpha})] \leqq \lambda\varepsilon^2$$

where S and $\tilde{S}$ are strictly positive definite, and $||\Delta S|| = [\text{tr}\, \Delta S^2]^{\frac{1}{2}} \leqq \varepsilon$.

Finally, note that any solution of (3.13) (whether $\tilde{S} = I$ or not) is orthogonal in $\mathscr{H}$ to ΔS as given by either (3.16) or (3.17). To see this, calculate (assuming 3.13)

$$\text{tr}\,\{\phi A \phi^T \Delta S\} = \tfrac{1}{2}\, \text{tr}\,\{\phi A \phi^T [P_2 \Delta S P_1 + P_1 \Delta S P_2]\} = 0$$

and

$$\text{tr}\,\{k\tilde{S}\Delta S\} = \frac{k}{2}\, \text{tr}\,\{P_2 \tilde{S} P_1 + P_1 \tilde{S} P_2\} = 0 \, .$$

In Section 6 we shall give more details on the numerical computation of λ and X. First, however, we present an interesting special case in which λ can be calculated explicitly.

5. Regression errors caused by using equal-weight least squares instead of Gauss-Markov estimation in the linear model. In many applications of the full-rank general linear model it is convenient to estimate coefficients by simply requiring that the equally-weighted sum of the squares of the residuals be a minimum. This procedure is often used even when the data are correlated and have unequal variances, so that the results thus obtained are not optimum. However, the non-optimum estimate obtained in this way is entirely satisfactory if the resulting errors are sufficiently small. In this paper we apply the theory already developed to establish a remarkably simple approximate upper bound on such errors. So far we have been mainly concerned with coefficient errors, i.e., errors in α. However, in many applications it is more meaningful to study errors in $\phi\alpha$, as these errors often represent the error incurred when the estimate is actually used. For example, in the case of polynomial regression one is usually much more interested in the quality of the fit than in the coefficient errors, as the latter are not invariant under scaling. In this section errors in $\phi\alpha$ will be called *regression errors*, in contrast to errors in α, which have been called *coefficient errors*. The use of unit-weight least squares amounts to assuming $\tilde{S} = I$. (The modification in case $\tilde{S} = \sigma^2 I$ is immediate.) We therefore state the main result as follows.

THEOREM 5.1. *If* $\tilde{S} = I$ *and* $||\Delta S|| = ||S - I|| \leqq \varepsilon$, *then the regression error vector* $\phi(\tilde{\alpha} - \hat{\alpha})$ *is such that*

$$E\{[\phi(\tilde{\alpha} - \hat{\alpha})]^T[\phi(\tilde{\alpha} - \hat{\alpha})]\} \leqq \varepsilon^2/2 + O(\varepsilon^3) . \tag{5.1}$$

PROOF. First note that, in accordance with the propagation of covariances and formula (2.6),

$$\begin{aligned} E\{[\phi(\tilde{\alpha} - \hat{\alpha})]^T[\phi(\tilde{\alpha} - \hat{\alpha})]\} &= \operatorname{tr} \phi S_{\tilde{\alpha}-\hat{\alpha}} \phi^T \\ &= \operatorname{tr} \phi^T S \phi (\phi^T \phi)^{-1} - \operatorname{tr} \phi^T \phi (\phi^T S^{-1} \phi)^{-1} . \end{aligned} \tag{5.2}$$

If A is any nonsingular $p \times p$ matrix and one replaces ϕ by ϕA in (5.2), he finds that the expected mean-square regression error (5.2) is unchanged. Since $(\phi^T\phi)^{-1}$ is strictly positive definite, there exists [2] a nonsingular $p \times p$ matrix W such that $(\phi^T\phi)^{-1} = W^T W$. If we put $\psi = \phi W^T$, then $\psi^T\psi = I$. However, if we replace ϕ by ψ, so that $\phi^T\phi = I$, then coefficient errors and regression errors are identical. Therefore we may obtain the desired expected mean-square regression error from (2.10) by simply imposing the condition $\phi^T\phi = I$. If this is done, (5.2) then reduces to (2.10), and the coefficient-error theory may be used to majorize the error as given by (5.2). Thus we assume $\phi^T\phi = I$ and proceed to solve (3.13) for λ. Multiplying (3.13) on the left by P_1 and on the right by P_2, noting that both P_1 and P_2 are now idempotent, gives

$$\tfrac{1}{2} P_1 \Delta S P_2 = \lambda P_1 \Delta S P_2 . \tag{5.3}$$

According to Theorem 3.1, the maximal λ will occur when $P_1 \Delta S P_2 \neq 0$ unless the mean-square error (5.2) is identically zero for all admissible ΔS. In the first case, $\lambda = \frac{1}{2}$ and Theorem 5.1 is proved. In the second case, Theorem 5.1 is trivially true. Although it is not needed for the proof, one can construct an example showing that the second case cannot occur for $1 \leqq p \leqq n - 1$.

In the following section we indicate the nature of the numerical verifications that have been obtained.

6. Numerical illustrations. In the numerical work reported here all computations were performed in double precision. Whenever no particular assumption is made regarding a variable (such as ϕ, ΔS, etc.), we assigned the values by the use of a Gaussian random-number generator. The resulting values were rounded in the computer before they were used in such a way that the values given below are *exactly* those actually used to at least single-precision accuracy, or to about 11 significant figures. It is realized that the determination of λ to an accuracy of two digits would suffice for most applications; however, the more accurate values are given to show the ease of obtaining the λ to great accuracy by the method of Section 4. The exactness of the fundamental formula (2.24) was verified numerically by many computations.

EXAMPLE 1. (See 1.1.) $n = 4$, $p = 3$, $\varepsilon = .01$, $\tilde{S} = I$,

$$\phi = \begin{bmatrix} .008 & .845 & -.362 \\ .388 & .903 & -.796 \\ -.286 & -.520 & -.691 \\ -.373 & -1.387 & -.157 \end{bmatrix} . \tag{6.1}$$

Three random choices of X_0 were chosen and λ was computed by the method of Section 4 as

$$\lambda = 6.26055201920083 . \tag{6.2}$$

This exact value was obtained after 13 to 14 iterations from each of the three randomly-chosen X_0. In this particular case the eigenspace associated with λ appears to be one-dimensional, i.e., $k = 1$ in (4.6). The normalized eigenmatrix obtained was

$$X = \begin{bmatrix} .50630972 & -.12855536 & -.18828182 & .27563970 \\ -.12855536 & -.13497855 & .28344970 & -.27880171 \\ -.18828182 & .28344970 & -.26125714 & .19105482 \\ .27563970 & -.27880171 & .19105482 & -.11007403 \end{bmatrix} . \tag{6.3}$$

Either this matrix or its negative was obtained exact to the number of decimals given for all three random choices of X_0. The corresponding error bound (4.15) or (3.28) is:

$$E\{(\tilde{\alpha} - \hat{\alpha})^T(\tilde{\alpha} - \hat{\alpha})\} \leqq \varepsilon^2\lambda = 6.26\cdots \times 10^{-4} . \tag{6.4}$$

Note that for other small assumed values of ε we need only multiply λ by ε^2. Thus for $\varepsilon = .1$ we have a bound of $6.26\cdots \times 10^{-2}$ corresponding to (6.4). A random sample of five symmetric error matrices ΔS for which $||\Delta S|| = .01$ was taken and the actual expected mean-square error for each was calculated by (2.10) and (2.24). The maximum actual expected mean-square error achieved by any ΔS was about 15 % of the bound (6.4). It will be noticed throughout the numerical examples that the chances of obtaining a matrix ΔS which approximates the bound are not great. This is apparently related to the extensive nature of the space for which the increase in expected mean-square error is zero. (See (3.22), for instance.) The bound of (6.4) was actually achieved by $\varepsilon^2 X$, where X is given by (6.3).

EXAMPLE 2. $n = 11$, $p = 4$, $\varepsilon = .01$, $\tilde{S} = I$. The matrix ϕ is not reproduced here, but was obtained from the random-number generator. With three randomly-chosen starting elements X_0, the maximal eigenvalue λ was obtained to an accuracy of at least 15 significant figures after 58 to 66 iterations. The X obtained was different in each case, indicating $k > 1$ in (4.6). One of the X's was substituted into (2.10) and (2.24) to verify that the maximum, $\lambda\varepsilon^2$, was actually achieved. In a random sample of 30 symmetric matrices ΔS for which $||\Delta S|| = .01$, it was found that the largest actual expected mean-square error was about 35 % of the bound.

Cases were computed in which $\tilde{S} \neq I$ was determined by calculating $A^T A$, where A is a randomly-chosen triangular matrix. In many of these cases it was seen that calculating the bound $\varepsilon^2\lambda$ gave a value that was within a couple of percent of the expected mean-square error obtained when X was substituted into (2.24). This indicated that a reasonable bound could probably be obtained by dropping 3rd and higher degree terms in ΔS from (2.24).

As a check of the results of Section 5, a sample of 200 symmetric matrices ΔS was obtained by using the random-number generator. The resulting matrices were normalized to have $||\Delta S|| = 10^{-3}$ and $||\Delta S|| = 10^{-2}$ for each ΔS. The basis ϕ was chosen from a table of orthogonal polynomials such that $\phi^T\phi = I$, $p = 3$, $n = 11$, and (5.2) was computed for each ΔS (both normalizations). The largest actual relative mean-square regression error (5.2) was 70.4 % of $||\Delta S||^2/2$.

7. Summary. A rational procedure has been developed for finding an approximate bound on the expected mean-square error caused by using an incorrect covariance matrix to estimate the coefficient vector in the general linear model. A new expression (Theorem 2.1) for this mean-square error was derived. The matrices P_1 and P_2 ((2.12) and (2.16), respectively) were defined, and it was found that their properties greatly aided in the derivation. The method of Lagrange was employed to maximize the expected mean-square error under the constraints that $||\Delta S|| = \varepsilon$ and that ΔS was symmetric. A necessary and sufficient condition that $E[(\tilde{\alpha} - \hat{\alpha})^T(\tilde{\alpha} - \hat{\alpha})] = 0$ was derived (Theorem 3.1). A second-degree approximation to the nonlinear expected mean-square error was introduced. In Section 4 a method was derived to solve for the required eigenvalue for the approximation; it was found that the theory of self-adjoint, nonnegative definite, compact operators played an essential role in the derivation, which involved the (finite-dimensional) Hilbert space whose elements are symmetric matrices. In Section 5 we derived a simple approximate upper bound for the expected mean-square regression error caused by wrongly using unit-weight least squares; it was possible to obtain the explicit solution $\lambda = \frac{1}{2}$ in this case. In Section 6 numerical results were presented to illustrate the practical computation of the error bounds.

REFERENCES

[1] Cleveland, W. S. (1971). Projection with the wrong inner product and its application to regression with correlated errors and linear filtering of time series. *Ann. Math. Statist.* **42** 616–624.

[2] FADDEEVA, V. N. (1959). *Computational Methods of Linear Algebra.* Dover, New York. 81–85.

[3] GOLUB, G. H. (1963). Comparison of the variance of minimum variance and weighted least squares regression coefficients. *Ann. Math. Statist.* **34** 984–991.

[4] GRENANDER, U. (1954). On the estimation of the regression coefficients in the case of an auto-correlated disturbance. *Ann. Math. Statist.* **25** 252–272.

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

[6] MAGNESS, T. A. and MCGUIRE, J. B. (1962). Comparison of least squares and minimum variance estimates of regression parameters. *Ann. Math. Statist.* **33** 462–470.

[7] RALSTON, A. (1965). *A First Course in Numerical Analysis.* McGraw-Hill, New York. 474–480.

[8] RAO, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Fifth Berkeley Symp. Math. Statist. Prob.* **1** 355–371. Univ. of California Press.

[9] RIESZ, F. and SZ-NAGY, B. (1955). *Functional Analysis.* Ungar, New York. Chapter VI.

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

ENVIRONMENTAL RESEARCH LABS
U. S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND
ATMOSPHERIC ADMINISTRATION
BOULDER, COLORADO 80302