COMPARISON OF LEAST SQUARES AND MINIMUM VARIANCE ESTIMATES OF REGRESSION PARAMETERS[1]

By T. A. Magness and J. B. McGuire

Space Technology Laboratories, Inc.

1. Introductory summary. The basic problem dealt with here is the estimation of linear regression parameters from a set of observations obscured by correlated noise. Two well-known solutions to this problem are minimum variance (or Markov, MV) and least squares (LS) estimation. Although MV is, by definition, an optimal method, LS possesses two distinct advantages which cause it to be used more frequently in practice: (1) computational simplicity and (2) the fact that it does not require knowledge of the correlation matrix of the noise, which in many cases is actually unknown. Therefore, a comparative study of these two methods to determine how much is lost by use of LS instead of MV is of value. In this connection, Grenander and Rosenblatt [1] have derived important asymptotic properties of LS and MV estimates when the noise is a stationary random process. The approach here is somewhat different from theirs, and no assumption regarding stationarity is made. The essence of this analysis is to re-formulate LS and MV in terms of the spectrum of the noise correlation matrix. This procedure offers some new insights into the nature of LS and MV and the differences between them. For example, it shows when they yield the same result and when LS performs worst compared with MV. It also exhibits the roles played by the maximum and minimum eigenvalues of the noise correlation matrix in setting bounds on the covariance matrices of both LS and MV estimates.

2. Special case: estimating a scalar parameter. We are concerned here with estimating the scalar parameter α in the linear regression equation

$$y = \phi\alpha + x, \tag{1}$$

where the observations y, the coefficients ϕ, and the noise x are column n-vectors. The noise x has zero mean and non-singular $n \times n$ correlation matrix ρ. For the present, we shall assume further that

$$\phi'\phi = 1 \qquad \text{and} \qquad Ex_i^2 = 1, \qquad i = 1, \cdots, n. \tag{2}$$

These do not constitute real restrictions, as seen later, but are a convenience in isolating the effects of correlated noise. Primes denote transposes of vectors and matrices.

We now introduce the spectrum of ρ: $\lambda_1, \cdots, \lambda_n$ are the n positive eigenvalues of ρ and $\psi_1, \cdots, \psi_n$ are the associated orthonormal eigenvectors. Thus $\rho\psi_i =$

Received July 12, 1961; revised October 30, 1961.

[1] This paper presents the results of one phase of research conducted under Contract JPL 950045 to the Jet Propulsion Laboratory, California Institute of Technology, sponsored bv the National Aeronautics and Space Administration, Contract NASw-6.

$\lambda_i\psi_i$ and $\psi_i'\psi_j = \delta_{ij}$, $i, j = 1, \cdots, n$. It follows that the ψ_i are eigenvectors of ρ^{-1} with eigenvalues $1/\lambda_i$. We can express the regression vector ϕ as a linear combination of the ψ_i,

$$\phi = \sum_{i=1}^{n} \beta_i\psi_i . \tag{3}$$

As a consequence of (2), $\sum_1^n \beta_i^2 = 1$.

Since the basic theory of LS and MV regression analysis is well documented (e.g., see [1] or Scheffé [2]), we shall merely re-state the formulas here:

Least squares. The least squares estimate of α in (1) is

$$a_{\mathrm{LS}} = (\phi'\phi)^{-1}\phi'y = \phi'y. \tag{4}$$

Note that this formula does not involve ρ. The variance of a_{LS} will, of course, depend on ρ and is given by

$$\sigma_{\mathrm{LS}}^2 = (\phi'\phi)^{-1}\phi'\rho\phi(\phi'\phi)^{-1} = \phi'\rho\phi. \tag{5}$$

substituting for ϕ from (3) leads to

$$\sigma_{\mathrm{LS}}^2 = \sum_1^n \beta_i^2\lambda_i . \tag{6}$$

Minimum variance. The minimum variance linear unbiased estimate of α (also called the Markov estimate) is

$$a_{\mathrm{MV}} = (\phi'\rho^{-1}\phi)^{-1}\phi'\rho^{-1}y. \tag{7}$$

The variance of a_{MV} is

$$\sigma_{\mathrm{MV}}^2 = (\phi'\rho^{-1}\phi)^{-1}. \tag{8}$$

By definition, $\sigma_{\mathrm{MV}}^2 \leqq \sigma_{\mathrm{LS}}^2$. Substituting for ϕ from (3) leads to

$$\sigma_{\mathrm{M}}^2 = \left(\sum_1^n \frac{\beta_i^2}{\lambda_i}\right)^{-1} . \tag{9}$$

The remainder of this section is devoted to a comparison of σ_{LS}^2 and σ_{MV}^2, as defined above, for different regression vectors ϕ when the noise correlation matrix ρ is held fixed. In this connection, we present first two special cases which it is instructive to consider. As a standard of comparison we shall include the estimation of α in the uncorrelated case, i.e., with $\rho = I$. In that event, LS and MV yield the same estimate having unit variance.

(a) When $\beta_i^2 = 1/n$, $i = 1, \cdots, n$, then, $\sigma_{\mathrm{LS}}^2 = (1/n)\sum_1^n \lambda_i = 1$. In this event, noise correlations do not "hurt" or "help" LS compared with the uncorrelated case. However, MV can do better, since $\sigma_{\mathrm{MV}}^2 = [(1/n)\sum_1^n (1/\lambda_i)]^{-1} \leqq 1$. Thus we have an example of a situation in which MV does better than LS by virtue of the fact that MV does better than the uncorrelated case.

(b) When $\beta_i^2 = \lambda_i/n$, $i = 1, \cdots, n$, then $\sigma_{\mathrm{LS}}^2 = (1/n)\sum_1^n \lambda_i^2 \geqq 1$. Here, noise correlations "hurt" the LS estimate. On the other hand $\sigma_{\mathrm{MV}}^2 = [\sum_1^n (1/n)]^{-1} = 1$, so that correlations do not help or hurt the MV estimate, compared with the uncorrelated case.

In both of these examples in which MV out-performs LS, the regression vector ϕ is a "mixture" of all of the eigenvectors of ρ. The theorem below shows what happens when ϕ is itself an eigenvector.

THEOREM 1. *The* MV *and* LS *estimates of* α *in* (1) *have equal variances if and only if* ϕ *is an eigenvector of* ρ, *in which case*

$$\sigma_{LS}^2 = \sigma_{MV}^2 = \lambda, \tag{10}$$

where λ *is the eigenvalue associated with* ϕ.

PROOF. If ϕ is an eigenvector of ρ, then direct substitution into (5) and (8) will verify (10). To prove the converse, let ϕ be any unit (regression) vector. Then one can always express

$$\rho\phi = \lambda\phi + \mu\phi^{\perp}, \tag{11}$$

where λ and μ are scalars and $\phi^{\perp}$ is a unit vector perpendicular to ϕ. Multiplying (11) on the left by ϕ' leads to

$$\sigma_{LS}^2 = \phi'\rho\phi = \lambda. \tag{12}$$

Since ρ is positive definite, $\lambda > 0$. Now solve (11) for ϕ and multiply on the left by ρ^{-1}:

$$\rho^{-1}\phi = (1/\lambda)\phi - (\mu/\lambda)\rho^{-1}\phi^{\perp}. \tag{13}$$

Multiplying (13) on the left by ϕ' and taking reciprocals leads to

$$\sigma_{MV}^2 = (\phi'\rho^{-1}\phi)^{-1} = [(1/\lambda) - (\mu/\lambda)\phi'\rho^{-1}\phi^{\perp}]^{-1}. \tag{14}$$

Now set $\sigma_{MV}^2 = \sigma_{LS}^2$. This implies either $\mu = 0$ or else $\phi^{\perp}$ is perpendicular to $\rho^{-1}\phi$. However, the second alternative implies the first. For, multiplying (13) on the left by $(\phi^{\perp})'$ leads to

$$(\phi^{\perp})'\rho^{-1}\phi = -(\mu/\lambda)(\phi^{\perp})'\rho^{-1}\phi^{\perp}. \tag{15}$$

The coefficient of $-\mu/\lambda$ in (15) is positive because ρ^{-1} is positive definite. Therefore, $(\phi^{\perp})'\rho^{-1}\phi = 0$ implies $\mu = 0$. Consequently, ϕ must be an eigenvector of ρ with eigenvalue $\lambda = \sigma_{LS}^2 = \sigma_{MV}^2$.

The maximum eigenvalue λ_{max} is the maximum value that either σ_{LS}^2 or σ_{MV}^2 can achieve, since it is the maximum value that σ_{LS}^2 in (6) can achieve, and it is achieved by both LS and MV when ϕ is an eigenvector going with λ_{max}. Similarly, the minimum eigenvalue λ_{min} is the minimum value that either σ_{LS}^2 or σ_{MV}^2 can achieve, since it is the minimum value that σ_{MV}^2 in (9) can achieve, and it is achieved by both LS and MV when ϕ is an eigenvector going with λ_{min}. To summarize, for any regression vector ϕ,

$$\lambda_{min} \leqq \sigma_{MV}^2 \leqq \sigma_{LS}^2 \leqq \lambda_{max}, \tag{16}$$

and in the limiting cases, LS and MV agree.

Consider now the ratio $\sigma_{\mathrm{LS}}^2/\sigma_{\mathrm{MV}}^2 = (\sum_1^n \beta_i^2\lambda_i)(\sum_1^n \beta_i^2/\lambda_i)$ as a function of ϕ, ρ being fixed. Its minimum value is, of course, one, which it assumes when ϕ is an eigenvector of ρ. The question arises, what is the maximum value of this ratio and when does it occur? The problem of finding the regression vector ϕ which maximizes this ratio may be formulated as follows:

Let $\lambda_1, \cdots, \lambda_n$ be a set of n positive numbers and let

$$F(\gamma_1, \cdots, \gamma_n) = \left(\sum_1^n \gamma_1 \lambda_i\right)\left(\sum_1^n \frac{\gamma_i}{\lambda_i}\right). \tag{17}$$

We wish to maximize F subject to $\gamma_i(=\beta_i^2) \geqq 0$ and $\sum_1^n \gamma_1 = 1$. First, consider the conditional maximum with respect to any pair of coordinates γ_p and γ_q when all other coordinates are set equal to zero. Subject to these constraints and substituting $\gamma_q = 1 - \gamma_p$, $F = (\gamma_p\lambda_p + (1 - \gamma_p)\lambda_q) \times (\gamma_p/\lambda_p + (1 - \gamma_p)/\lambda_q)$. If $\lambda_p = \lambda_q$, then $F \equiv 1$ and the problem is trivial. If $\lambda_p \neq \lambda_q$, then F takes on its maximum value $\frac{1}{4}(\lambda_p + \lambda_q)(1/\lambda_p + 1/\lambda_q)$ when $\gamma_p = \gamma_q = \frac{1}{2}$.

Next, introduce a third coordinate γ_k and let $\gamma_p = \gamma_q = \frac{1}{2} - \epsilon$, $\gamma_k = 2\epsilon$, where ϵ is small and positive. As a function of these three coordinates only,

$$F = [(\tfrac{1}{2} - \epsilon)(\lambda_p + \lambda_q) + 2\epsilon\lambda_k] \times [(\tfrac{1}{2} - \epsilon)(1/\lambda_p + 1/\lambda_q) + 2\epsilon/\lambda_k]. \tag{18}$$

Now expand (18) in powers of ϵ:

$$\begin{aligned} F &= \tfrac{1}{4}(\lambda_p + \lambda_q)(1/\lambda_p + 1/\lambda_q) \\ &\qquad + \epsilon \frac{(\lambda_p + \lambda_q)}{\lambda_p \lambda_q \lambda_k} (\lambda_p - \lambda_k)(\lambda_q - \lambda_k) + O(\epsilon^2). \end{aligned} \tag{19}$$

The first term on the right in (19) is just the conditional maximum in γ_p and γ_q. Hence we shall have a local maximum in the three coordinates γ_p, γ_q, γ_k at the point $\gamma_p = \gamma_q = \frac{1}{2}$, $\gamma_k = 0$ if the coefficient of ϵ is negative. This coefficient is negative if and only if $\lambda_p < \lambda_k < \lambda_q$ or else $\lambda_p > \lambda_k > \lambda_q$.

Let us assume temporarily that there is a unique maximum and a unique minimum among the λ's. Then by the above argument, F will have a unique local maximum at $\gamma_p = \gamma_q = \frac{1}{2}$ where $\lambda_p = \lambda_{\max}$ and $\lambda_q = \lambda_{\min}$. We assert this local maximum is the over-all maximum. Proof is by induction. First, let $n = 3$. Suppose our local maximum is not the over-all maximum, but the over-all maximum is attained at some vector γ^0. Then γ^0 must be an interior vector, since we have already investigated the boundaries in the 3-dimensional case. Pass a plane through γ^0 and our vector, $\gamma_p = \gamma_q = \frac{1}{2}$. Since F, when restricted to this plane and the surface defined by $\sum_1^n \gamma_i = 1$, is expressible as a second degree curve, it is impossible for it to have an interior maximum at γ^0 and an "inside" maximum at the boundary. Therefore there can be no interior maximum, either, and our conjecture is proved.

Now assume the conjecture is true for $n = N$ and consider the case $n = N + 1$. If $\lambda_p = \lambda_{\max}$ and $\lambda_q = \lambda_{\min}$, then $\gamma_p = \gamma_q = \frac{1}{2}$ yields a unique local maximum. If this is not the over-all maximum, then let the over-all maximum occur at γ^0.

By the same reasoning as used in the case $n = 3$, γ^0 cannot be an interior maximum. Moreover, γ^0 cannot occur on the boundary, since in that event the problem reduces to an N-dimensional case where, by hypothesis, our conjecture is true. Therefore $\gamma_p = \gamma_q = \frac{1}{2}$ yields a unique over-all maximum when $\lambda_{\max}$ and $\lambda_{\min}$ are unique.

On the other hand, if there is more than one $\lambda_{\max}$ and/or $\lambda_{\min}$, the maximum value of F is clearly the same as above but the point at which the maximum is attained is no longer unique. We have therefore proved the following.

THEOREM 2. *For given noise correlation matrix* ρ,

$$\sigma_{LS}^2/\sigma_{MV}^2 \leqq \tfrac{1}{4}(\lambda_{\max} + \lambda_{\min})(1/\lambda_{\max} + 1/\lambda_{\min}), \tag{20}$$

and this upper bound is attained when the regression vector ϕ *is of the form* $\phi \propto \psi_{\max} + \psi_{\min}$, *where* $\psi_{\max}$ *and* $\psi_{\min}$ *are (normalized) eigenvectors of* ρ *associated with the maximum and minimum eigenvalues,* $\lambda_{\max}$ *and* $\lambda_{\min}$.

Thus for the estimation of a scalar parameter, Theorem 2 puts an upper bound on how much is lost by using LS instead of MV.

While the article was in proof, it was brought to the authors' attention by Dr. G. H. Golub that the inequality on which Theorem 2 is based was proved by L. V. Kantorovich [3].

3. General case: estimating a vector parameter. We now turn to a more general regression equation in which the parameter to be estimated is a column p-vector γ:

$$z = \theta\gamma + w, \tag{21}$$

where the observations z and the noise w are column n-vectors, $n > p$, and where θ is a $n \times p$ matrix of regression coefficients. The p columns of θ are assumed to be linearly independent. The noise w has zero mean and non-singular $n \times n$ covariance matrix R. The estimation of γ in (21) includes, of course, the estimation of a scalar parameter when restrictions such as (2) are no longer imposed.

Least squares. We shall consider the so-called "weighted least squares" estimate of γ, g_{LS}, which minimizes the sum of squares of weighted residuals, each residual being weighted inversely as the standard deviation of the corresponding noise component. To compute g_{LS}, it is necessary to know (within a constant factor) the diagonal $n \times n$ matrix M whose elements are $(R_{ii})^{-\frac{1}{2}}$. Then g_{LS} is given by (see [2]):

$$g_{LS} = (\theta' M^2 \theta)^{-1} \theta' M^2 z. \tag{22}$$

The covariance matrix of g_{LS} is

$$G_{LS} = (\theta' M^2 \theta)^{-1} \theta' M^2 R M^2 \theta (\theta' M^2 \theta)^{-1}. \tag{23}$$

Minimum variance. To compute the minimum variance linear unbiased estimate of γ, g_{MV}, it is necessary to know R (within a constant factor; see [2]):

$$g_{MV} = (\theta' R^{-1} \theta)^{-1} \theta' R^{-1} z. \tag{24}$$

The covariance matrix of g_{MV} is

$$G_{\mathrm{MV}} = (\theta' R^{-1} \theta)^{-1}. \tag{25}$$

In order to study the relationship between G_{LS} and G_{MV}, we shall transform (21) into a "canonical" form analogous to (1). Multiplying (21) on the left by M and letting $y = Mz$, $x = Mw$, we have $y = M\theta\gamma + x$. The new noise x has covariance matrix $\rho = MRM$, which is the correlation matrix of w. (As in Section 2, λ_i and ψ_i will denote eigenvalues and eigenvectors of ρ.) Since we can always find a (non-singular) $p \times p$ matrix B which makes $B'\theta' M^2 \theta B = I$, we shall set $M\theta B = \phi$ and $B^{-1}\gamma = \alpha$ to obtain the canonical equation

$$y = \phi\alpha + x, \tag{26}$$

where $\phi'\phi = B'\theta' M^2 \theta B = I$. Estimating α in (26) is statistically equivalent to estimating γ in (21). Since (26) is the analogue of (1) for regression on a vector parameter, the corresponding LS and MV formulas are analogues of the formulas in Section 2. We shall prove the following result for $A_{\mathrm{LS}} = \phi'\rho\phi$ and $A_{\mathrm{MV}} = (\phi'\rho^{-1}\phi)^{-1}$, the covariance matrices of LS and MV estimates of α.

THEOREM 3. *The* LS *and* MV *estimates of* α *in* (26) *have identical covariance matrices if and only if the subspace spanned by the p columns of* ϕ *coincides with the space spanned by p of the eigenvectors of* ρ, *in which case both covariance matrices are similar to a diagonal matrix whose elements are the corresponding eigenvalues of* ρ.

Note that since $G_{\mathrm{LS}} = BA_{\mathrm{LS}}B'$ and $G_{\mathrm{MV}} = BA_{\mathrm{MV}}B'$, Theorem 3 provides necessary and sufficient conditions that G_{LS} and G_{MV} be identical.

PROOF OF THEOREM 3. Let the subspace spanned by the p columns of ϕ coincide with the space spanned by the p eigenvectors $\psi_{i_1}, \cdots, \psi_{i_p}$. If χ denotes the $n \times p$ matrix whose columns are $\psi_{i_1}, \cdots, \psi_{i_p}$, then by hypothesis there exists a $p \times p$ unitary matrix C such that $\chi = \phi C$. If we set $\alpha = C\alpha^*$, then the covariance matrices of LS and MV estimates of α^* are

$$A_{\mathrm{LS}}^* = C' A_{\mathrm{LS}} C = C'\phi'\rho\phi C = \chi'\rho\chi = \begin{bmatrix} \lambda_{i_1} & & \\ & \ddots & \\ & & \lambda_{i_p} \end{bmatrix}$$

and

$$A_{\mathrm{MV}}^* = C' A_{\mathrm{MV}} C = (C'\phi'\rho^{-1}\phi C)^{-1} = (\chi'\rho^{-1}\chi)^{-1} = \begin{bmatrix} \lambda_{i_1} & & \\ & \ddots & \\ & & \lambda_{i_p} \end{bmatrix}.$$

This implies

$$A_{\mathrm{LS}} = A_{\mathrm{MV}} = C \begin{bmatrix} \lambda_{i_1} & & \\ & \ddots & \\ & & \lambda_{i_p} \end{bmatrix} C',$$

which proves the first half of the theorem. Conversely, let $A_{\text{LS}} = A_{\text{MV}}$. Let C be a $p \times p$ unitary matrix which diagonalizes A_{LS} (i.e., $C'A_{\text{LS}}\,C$ is diagonal) and set $\alpha = C\alpha^*$. Then the covariance matrices of LS and MV estimates of α^*, denoted by A^*_{LS} and A^*_{MV}, are diagonal and equal. The regression matrix going with α^* is $\phi^* = \phi C$. Consider now any component α_i^* of α^*. The variance of the LS estimate of α_i^* equals $\phi_i'^*\rho\phi_i^*$, where ϕ_i^* is the ith column of ϕ^*. Similarly, since A^*_{MV} is diagonal, the variance of the MV estimate of α_i^* equals $(\phi_i^{*\prime}\rho^{-1}\phi_i^*)^{-1}$. By Theorem 1, there can be equality between these variances only if ϕ_i^* is an eigenvector of ρ, and in that event the value of $\phi_i^{*\prime}\rho\phi_i^*$ is the corresponding eigenvalue. Thus, the columns of ϕ^* are eigenvectors of ρ and the elements of A^*_{LS} are the corresponding eigenvalues. Since $\phi = \phi^*C'$ and $A_{\text{LS}} = A_{\text{MV}} = CA^*_{\text{LS}}C'$, the converse is proved.

We now turn our attention to a lemma which is the analogue of (16).

LEMMA. *If A_{LS} and A_{MV} denote the covariance matrices of* LS *and* MV *estimates of α in* (26), *then*

$$\lambda_{\min}I \leqq A_{\text{MV}} \leqq A_{\text{LS}} \leqq \lambda_{\max}I, \tag{27}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum eigenvalues of the noise correlation matrix ρ.

(We recall that one positive definite matrix is by definition less than or equal to another provided the second minus the first is non-negative definite. Also, we note that $A_{\text{MV}} \leqq A_{\text{LS}}$ by definition.)

PROOF. If α^* is related to α by a unitary transformation, $\alpha = C\alpha^*$, then the corresponding covariance matrices of LS and MV estimates of α^* are $A^*_{\text{LS}} = C'\phi'\rho\phi C$ and $A^*_{\text{MV}} = (C'\phi'\rho^{-1}\phi C)^{-1}$. Let us assume that C is chosen to make A^*_{LS} diagonal. Using the notation ψ for the $n \times n$ unitary matrix whose columns are the eigenvectors of ρ, $\Lambda = \psi'\rho\psi$ for the $n \times n$ diagonal matrix whose elements are the eigenvalues of ρ, and $\beta = C'\phi'\psi$, we have $A^*_{\text{LS}} = C'\phi'\psi\Lambda\psi'\phi C = \beta\Lambda\beta'$, diagonal. Since $\beta\beta' = I$, $\sum_{j=1}^n \beta_{ij}^2 = 1, i = 1, \cdots, p$, so that $\sum_{j=1}^n \beta_{ij}^2\lambda_j \leqq \lambda_{\max}$, $i = 1, \cdots, p$. Thus, every element of the diagonal matrix A^*_{LS} is less than or equal to $\lambda_{\max}$ and $A^*_{\text{LS}} \leqq \lambda_{\max}I$. This immediately implies $A_{\text{LS}} \leqq \lambda_{\max}I$, since α and α^* are related by a unitary transformation.

On the other hand, we could have chosen C to diagonalize A_{MV}. By an argument similar to the one above, this leads to the inequality $\lambda_{\min}I \leqq A^*_{\text{MV}} = C'A_{\text{MV}}C$, which implies $\lambda_{\min}I \leqq A_{\text{MV}}$.

Returning to the estimation of γ in (21), we recall that $G_{\text{LS}} = BA_{\text{LS}}B'$ and $G_{\text{MV}} = BA_{\text{MV}}B'$. We shall use the notation G_{UC} to denote G_{LS} (or G_{MV}) evaluated when $\rho = I$. G_{UC} thus corresponds to the covariance matrix of the LS (or MV) estimate of γ when the noise is uncorrelated. It follows that $G_{\text{UC}} = (\theta'M^2\theta)^{-1} = BB'$. Thus by multiplying (27) on the left by B and on the right by B', the inequalities are preserved and we have the following theorem.

THEOREM 4. *In the estimation of γ in* (21),

$$\lambda_{\min}G_{\text{UC}} \leqq G_{\text{MV}} \leqq G_{\text{LS}} \leqq \lambda_{\max}G_{\text{UC}}, \tag{28}$$

where G_{LS}, G_{MV} *and* G_{UC} *are covariance matrices of* LS, MV *and "uncorrelated case" estimates of* γ, *and where* λ_{max} *and* λ_{min} *are the maximum and minimum eigenvalues of the noise correlation matrix.*

In other words, if one computes the covariance matrix G_{UC}, which would apply to the LS estimate of γ in (21) if the noise w were uncorrelated, and if one multiplies G_{UC} by the minimum and the maximum eigenvalues of the noise correlation matrix, one obtains lower and upper bounds, respectively, on both G_{LS} and G_{MV}.

The remaining question is, for given noise correlation matrix ρ, how much greater than G_{MV} can G_{LS} be? It is seen below that the exact analogue of Theorem 2 holds for the general case of estimating a vector parameter. This places an upper bound on how much is lost by use of LS instead of MV.

THEOREM 5. *If* G_{LS} *and* G_{MV} *denote covariance matrices of* LS *and* MV *estimates of* γ *in* (21), *then*

$$G_{LS} \leqq \tfrac{1}{4}(\lambda_{max} + \lambda_{min})(1/\lambda_{max} + 1/\lambda_{min})G_{MV}, \tag{29}$$

where λ_{max} *and* λ_{min} *are the maximum and minimum eigenvalues of the noise correlation matrix.*

PROOF. We again refer back to the canonical regression equation (26). Letting ξ be any column p-vector of unit length, we wish to show that

$$\xi' A_{LS} \xi \leqq \tfrac{1}{4}(\lambda_{max} + \lambda_{min})(1/\lambda_{max} + 1/\lambda_{min})\xi' A_{MV} \xi. \tag{30}$$

To do this, we let $\phi\xi = \eta$ (η is thus a column n-vector of unit length) and consider the estimation of the *scalar* parameter ϵ in the hypothetical regression equation

$$y^* = \eta\epsilon + x^*, \tag{31}$$

where y^* denotes an n-vector of observations and the noise x^* has the same properties as x in (26). The variance of the LS estimate of ϵ in (31) is $\eta'\rho\eta = \xi'\phi'\rho\phi\xi = \xi' A_{LS}\xi$, and the variance of the MV estimate is

$$(\eta'\rho^{-1}\eta)^{-1} = (\xi'\phi'\rho^{-1}\phi\xi)^{-1} = (\xi' A_{MV}^{-1}\xi)^{-1}.$$

Therefore, from Theorem 2, we have

$$\xi' A_{LS} \leqq \tfrac{1}{4}(\lambda_{max} + \lambda_{min})(1/\lambda_{max} + 1/\lambda_{min})(\xi' A_{MV}^{-1}\xi)^{-1}.$$

Moreover,

$$(\xi' A_{MV}^{-1}\xi)^{-1} \leqq \xi' A_{MV}\xi. \tag{32}$$

For, consider another hypothetical regression equation

$$v = \xi\beta + u, \tag{33}$$

where v is a p-vector of observations, β denotes a scalar parameter to be estimated, and u is unbiased noise with $Euu' = A_{MV}$. Then $(\xi' A_{MV}^{-1}\xi)^{-1}$ and $\xi' A_{MV}\xi$ may be viewed as the variances of minimum variance and (unweighted) least squares estimates, respectively, of β in (33). Inequality (32) thus follows by definition of MV.

We have therefore established a chain of inequalities verifying (30), or equivalently,

$$A_{\mathrm{LS}} \leqq \tfrac{1}{4}(\lambda_{\max} + \lambda_{\min})(1/\lambda_{\max} + 1/\lambda_{\min})A_{\mathrm{MV}}\,. \tag{34}$$

Multiplying (34) on the left by B and on the right by B' preserves the inequality and proves the theorem, since $G_{\mathrm{LS}} = BA_{\mathrm{LS}}B'$ and $G_{\mathrm{MV}} = BA_{\mathrm{MV}}B'$.

EXAMPLE. To illustrate the strength of the inequality in Theorem 5, we include here an example of a regression problem in which equality is attained.

Let the regression equation be given in canonical form with

$$\rho = \begin{bmatrix} 1 & r & 0 & 0 \\ r & 1 & 0 & 0 \\ 0 & 0 & 1 & r \\ 0 & 0 & r & 1 \end{bmatrix}, \quad 0 < r < 1, \qquad \text{and} \qquad \phi = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Thus $n = 4$, $p = 2$ and $\phi'\phi = I$. Direct calculation shows that the covariance matrix of the LS estimate of α is $A_{\mathrm{LS}} = \phi'\rho\phi = I$. Similarly, the covariance matrix of the MV estimate of α can be computed as

$$A_{\mathrm{MV}} = (\phi'\rho^{-1}\phi)^{-1} = \begin{bmatrix} (1 - r^2)^{-1} & 0 \\ 0 & (1 - r^2)^{-1} \end{bmatrix} = (1 - r^2)^{-1}I.$$

Since the maximum and minimum eigenvalues of the noise correlation matrix are $\lambda_{\max} = 1 + r$ and $\lambda_{\min} = 1 - r$, we have $\frac{1}{4}(\lambda_{\max} + \lambda_{\min})(1/\lambda_{\max} + 1/\lambda_{\min}) = (1 - r^2)^{-1}$. Therefore equality in (34) is attained. It should be noted that an important feature of the matrix ρ in this example is that it is reducible. If ρ is irreducible and $n > p > 1$, we conjecture that equality in Theorem 5 cannot be attained.

REFERENCES

[1] GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York.

[2] SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York.

[3[KANTOROVICH, L. V. (1948). Funkcional'nyĭ analiz i prikladnaya matematika. *Uspehi Mat. Nauk* **3** (28) 89-185.