COMPARISON OF THE VARIANCE OF MINIMUM VARIANCE AND WEIGHTED LEAST SQUARES REGRESSION COEFFICIENTS[1]

By Gene H. Golub

Stanford University

0. Introduction. This paper compares the variance and generalized variance of minimum variance (MV) and weighted least squares (WLS) estimates of regression coefficients. Matrix inequalities originally developed to study the rate of convergence of various iterative methods of solving linear equations (cf. [4]) are used in making the comparisons. These inequalities are given in Section 1 and applied in Section 2. In Section 3, attention is focused on diagonal weight matrices, and an example is given in Section 4.

1. Matrix inequalities. Let A be a real positive definite matrix with

$$Ae_i = \lambda_i e_i \qquad i = 1, 2, \cdots, n$$

$$\|e_i\| = 1$$

and the eigenvalues λ_i satisfying $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_n > 0$, and let $\kappa = \lambda_1/\lambda_n$ be the *spectral condition number* of A. For any non-null vector x, define

$$\mu_k = \mu_k(x) = x'A^k x. \tag{1}$$

Then

$$1 \leqq \mu_{k+1}\mu_{k-1}/\mu_k^2 \leqq [(\kappa^{\frac{1}{2}} + \kappa^{-\frac{1}{2}})/2]^2. \tag{2}$$

For $k = 0$, the inequality on the right is the Kantorovich inequality [5], and equality is attained for $x = a(e_1 \pm e_n)(a \neq 0)$. This inequality was first derived in order to determine the rate of convergence of the method of steepest descent for solving linear equations [5]. Equality on the left is attained when $x = be_i$ ($b \neq 0, i = 1, 2, \cdots, n$).

Inequalities (2) can be generalized. Let

$$M_k = X'A^kX \quad \text{and} \quad \mu_k(X) = \det M_k . \tag{3}$$

If X has rank p, define the condition number by

$$\kappa_p = \lambda_1 \cdots \lambda_p/\lambda_{n-p+1} \cdots \lambda_n . \tag{4}$$

Inequality (2) becomes

$$1 \leqq \mu_{k+1}(X)\cdot\mu_{k-1}(X)/\mu_k^2(X) \leqq [(\kappa_p^{\frac{1}{2}} + \kappa_p^{-\frac{1}{2}})/2]^2. \tag{5}$$

Proofs for inequalities (2) and (5) are given by Schopf [8].

Received November 13, 1961; revised February 26, 1963.

[1] This work was supported in part by Space Technology Laboratories, Inc. and in part by Office of Naval Research Contract Nonr-225(37) at Stanford University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Now let x and y be non-null real orthogonal vectors. Then

$$(x'Ay)^2/(x'Ax)(y'Ay) \leqq [(\kappa - 1)/(\kappa + 1)]^2. \tag{6}$$

Equality in (6) is attained when

$$x = 2^{-\frac{1}{2}}(e_1 \pm e_n), \qquad y = 2^{-\frac{1}{2}}(e_1 \mp e_n). \tag{7}$$

Inequality (6) is attributed to Wielandt and has been generalized by Bauer and Householder [1].

2. Comparison of variances. Let $y = \Phi\alpha + \epsilon$ where Φ is an $n \times p$ matrix of rank p; α is a vector with p components which is to be estimated; and ϵ is a random vector of n components with $E\{\epsilon\} = 0$ and covariance matrix C. The minimum variance unbiased estimate of α is ([3], pp. 86–88) $\alpha^* = (\Phi'C^{-1}\Phi)^{-1}\Phi'C^{-1}y$, and the covariance matrix of α^* is

$$\Sigma_{\mathrm{MV}} = (\Phi'C^{-1}\Phi)^{-1}. \tag{8}$$

Frequently, C is not known or C^{-1} is not easily computed because n is very large, and consequently α is estimated by its *weighted least squares estimate* $\hat{\alpha} = (\Phi'W\Phi)^{-1}\Phi Wy$. The estimate of $\hat{\alpha}$ is unbiased and has a covariance matrix

$$\Sigma_{\mathrm{WLS}} = (\Phi'W\Phi)^{-1}\Phi'WCW\Phi(\Phi'W\Phi)^{-1}. \tag{9}$$

It is assumed that W is positive definite and symmetric so that $W = FF'$.

THEOREM 1. *Let $F'CFe_i = \lambda_i e_i$, $i = 1, 2, \cdots, n$ with the eigenvalues satisfying $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_n > 0$. For $\xi \neq 0$, let*

$$r(\xi) = [\xi'\Sigma_{\mathrm{WLS}}\xi - \xi'\Sigma_{\mathrm{MV}}\xi]/\xi'\Sigma_{\mathrm{WLS}}\xi. \tag{10}$$

Then

$$0 \leqq r(\xi) \leqq [(\kappa - 1)/(\kappa + 1)]^2 \tag{11}$$

where $\kappa = \lambda_1/\lambda_n$.

PROOF. The following computation is similar to one given by Bauer and Householder [1]. Let $\eta = F'\Phi(\Phi'W\Phi)^{-1}\xi$, and $\zeta = F^{-1}C^{-1}\Phi(\Phi'C^{-1}\Phi)^{-1}\xi$. Then a short calculation shows $r(\xi) = \eta'F'CF(\eta - \zeta)/\eta'F'CF\eta$.
Since

$$\eta'F'CF\zeta = \zeta'F'CF\zeta,$$

$$\eta'F'CF(\eta - \zeta) = (\eta - \zeta)'F'CF(\eta - \zeta),$$

then

$$0 \leqq r(\xi) \leqq [\eta'F'CF(\eta - \zeta)]^2/[\eta'F'CF\eta \cdot (\eta - \zeta)'F'CF(\eta - \zeta)].$$

Then since $(\eta - \zeta)'\eta = 0$, inequality (6) can be used, and the desired result follows.

Magness and McGuire [6] have obtained a similar result for $F'CF = R$, the correlation matrix.

Now by (7), equality on the right of (11) is attained when $\eta = 2^{-\frac{1}{2}}(e_1 \pm e_n)$, and $\zeta = \pm 2^{\frac{1}{2}} e_n$. Then for $p = 1$, equality is attained for

$$\Phi = a(F')^{-1}[e_1 \pm e_n](a \neq 0).$$

COROLLARY 2. *Let*

$$\sigma_{ij}^* = \{\Sigma_{\mathrm{MV}}\}_{ij}, \qquad \hat{\sigma}_{ij} = \{\Sigma_{\mathrm{WLS}}\}_{ij}.$$

Then $1 \leqq \hat{\sigma}_{ii}/\sigma_{ii}^* \leqq [(\kappa^{\frac{1}{2}} + \kappa^{-\frac{1}{2}})/2]^2$, $i = 1, 2, \cdots, p$.

PROOF. Let $\xi_j = 1$ for $j = i$, $= 0$ otherwise. The result follows from Theorem 1 and some simple manipulations.

The determinant of the covariance matrix of an estimate of a vector parameter is called the *generalized variance of the estimate.* It is possible to determine bounds similar to those given above for the generalized variance of α^* and $\hat{\alpha}$.

THEOREM 3. *Let* $\kappa_p = \lambda_1 \cdots \lambda_p / \lambda_{n-p+1} \cdots \lambda_n$ *where again the* λ_i*'s are the eigenvalues of* $F'\ C\ F$ *ordered decreasingly. Then,*

$$1 \leqq \det\Sigma_{\mathrm{WLS}}/\det\Sigma_{\mathrm{MV}} \leqq [(\kappa_p^{\frac{1}{2}} + \kappa_p^{-\frac{1}{2}})/2]^2.$$

PROOF. By (8) and (9)

$$\frac{\det \Sigma_{\mathrm{WLS}}}{\det \Sigma_{\mathrm{MV}}} = \frac{\det((\Phi' W\Phi)^{-1}\Phi' WCW\Phi(\Phi' W\Phi)^{-1})}{\det((\Phi' C^{-1}\Phi)^{-1})}$$

$$= \frac{\det(\Phi' WCW\Phi)\cdot\det(\Phi' C^{-1}\Phi)}{(\det(\Phi' W\Phi))^2}.$$

Let $\Psi = F'\Phi$ so that

$$\frac{\det \Sigma_{\mathrm{WLS}}}{\det \Sigma_{\mathrm{MV}}} = \frac{\det(\Psi' F' CF\Psi)\cdot\det(\Psi'(F'CF)^{-1}\Psi)}{(\det(\Psi'\Psi))^2}$$

The result follows immediately from (5) with $k = 0$.

Note that it follows immediately that α^* is the minimum generalized variance estimate of $\hat{\alpha}$.

3. Choice of F. As pointed out earlier, even though C may be known, it may be difficult to compute the MV estimate of α since it may be difficult to invert for large n. The question arises whether it is possible to choose W so that the WLS estimate best approximates the MV estimate in some sense. Note that the right hand side of inequality (11) is an increasing function of κ. It will be shown that for C belonging to a certain class of matrices, it is possible to minimize κ with respect to W being a diagonal matrix. The present results depend heavily upon those of Forsythe and Straus [2]. Let $\kappa(A) = \lambda_1/\lambda_n$ where λ_1 and λ_n are the largest and smallest eigenvalues, respectively, of the positive definite matrix A. Let $\mathfrak{J}$ be a class of regular linear transformations.

Define $A^T = T'AT$. Then, A is said to be *best conditioned with respect to* $\mathfrak{J}$ if $\kappa(A^T) \geqq \kappa(A)$ for all $T \varepsilon \mathfrak{J}$. The above definition and Theorem 4 and Lemma 5 below are given by Forsythe and Straus [2].

THEOREM 4. *Let $\mathfrak{D}$ be the class of regular* diagonal *transformations. A sufficient condition for DAD to be best conditioned with respect to $\mathfrak{D}$ is that for some pair of eigenvectors e_M, e_m belonging to λ_1 and λ_n, respectively,*

$$|e_{M,j}| = |e_{m,j}| \qquad j = 1, 2, \cdots, n. \tag{12}$$

Moreover, if λ_1 and λ_n are simple eigenvalues, (12) *is also necessary.*

If the rows and columns of a $(p + q)$ matrix can be rearranged so that the upper $p \times p$ and lower $q \times q$ submatrices are diagonal, then the matrix is said to have *Property* (A). Matrices with Property (A) occur frequently in the numerical solution of partial differential equations and have been discussed extensively by Young [9]. (It is not difficult to show that all tridiagonal matrices have Property (A).)

LEMMA 5. *Let S be a positive definite symmetric matrix with Property (A) and $s_{ii} = 1$, $i = 1, 2, \cdots, n$ then S is best conditioned with respect to $\mathfrak{D}$.*

THEOREM 6. *Let C be a covariance matrix with Property (A). Then if $\{D_0\}_{ii} = c_{ii}^{-\frac{1}{2}}$, $i = 1, 2, \cdots, n$, $\kappa(DCD) \geqq \kappa(D_0CD_0)$ for all D in $\mathfrak{D}$.*

PROOF. Consider any matrix $D_1 \varepsilon \mathfrak{D}$. Then

$$\kappa(D_1CD_1) = \kappa((D_1D_0^{-1})D_0CD_0(D_0^{-1}D_1)).$$

Since $D_0 \varepsilon \mathfrak{D}$, $D_0^{-1} \varepsilon \mathfrak{D}$, $D_1D_0^{-1} \varepsilon \mathfrak{D}$, and the result follows from Lemma 5.

4. An example. In this section, we shall investigate the weighted least squares which minimizes κ among all diagonal weightings for a particular model. The results of the previous section shall now be applied to an example given by Rosenblatt [7]. Consider the process $y_i = \alpha_1 + \alpha_2 i + \epsilon_i$ where the stationary residual ϵ_i is a first order autoregressive scheme with covariances

$$E\{\epsilon_{k+i}\,\epsilon_k\} = r_i = \rho^{|i|}/(1 - \rho^2) \qquad |\rho| < 1$$

and hence

$$c_{ij} = \{C\}_{ij} = \rho^{|i-j|}/(1 - \rho^2).$$

When the sample size is n,

$$\Phi' = \begin{pmatrix} 1, 1, \cdots, 1 \\ 1, 2, \cdots, n \end{pmatrix}.$$

C does *not* have Property (A) but C^{-1} does since it is tridiagonal. Specifically, $\{C^{-1}\}_{11} = \{C^{-1}\}_{nn} = 1$; $\{C^{-1}\}_{jj} = 1 + \rho^2$, $j = 2, \cdots, n - 1$; $\{C^{-1}\}_{ij} = -\rho$ for $|i - j| = 1$; and $\{C^{-1}\}_{ij} = 0$ for $|i - j| > 1$. Hence for C^{-1}, $\{D_0\}_{ii} = 1$, $i = 1, n$, and $\{D_0\}_{ii} = (1 + \rho^2)^{-\frac{1}{2}}$, $i = 2, 3, \cdots, n - 1$. Since $\kappa(D_0C^{-1}D_0) = \kappa(D_0^{-1}CD_0^{-1})$, $\kappa(DCD) \geqq \kappa(D_0^{-1}CD_0^{-1})$ for all $D \varepsilon \mathfrak{D}$. Thus the diagonal set of weights which minimize the condition number are $\{W\}_{11} = \{W\}_{nn} = 1/(1 + \rho^2)$, and $\{W\}_{ii} = 1$ for $i = 2, 3, \cdots, n - 1$.

In a similar fashion to Rosenblatt [7], the (i, j) elements of the covariance matrix of the least squares, minimum variance, and weighted least squares estimates are given in Table I for the sample sizes $n = 10, 20, 50$ and correlation

TABLE I

Elements of the covariance matrices of the least squares, minimum variance, and weighted least squares estimates of a linear regression, residual first-order autoregressive

ρ		(1,1)	(1,2) = (2,1)	(2,2)
		$N = 10$		
	(A)	6.56093	−.496608	.0902924
.900	(B)	5.97795	−.437550	.0795545
	(D)	6.87113	−.537556	.0977375
	(A)	.540686	−.0901623	.0163931
−.900	(B)	.167807	−.0249467	.00453577
	(D)	.181861	−.0274926	.00499866
	(A)	11.8998	−.577765	.105048
.950	(B)	11.1657	−.514981	.0936330
	(D)	12.2552	−.622432	.113170
	(A)	1.05540	−.183370	.0333401
−.950	(B)	.160585	−.0238995	.00434537
	(D)	.184222	−.0281945	.00512628
	(A)	54.4744	−1.06344	.193352
.990	(B)	51.3228	−.590179	.107305
	(D)	52.6078	−.699685	.127215
	(A)	5.46079	−.983913	.178893
−.990	(B)	.155138	−.0231084	.00420152
	(D)	.256793	−.0415910	.00756201
	(A)	162.634	−11.6406	2.11647
.995	(B)	101.342	−.600532	.109188
	(D)	102.655	−.709973	.129086
	(A)	11.0110	−1.99297	.362359
−.995	(B)	.154477	−.0230122	.00418404
	(D)	.353711	−.0592367	.0107703
	(A)	101567.	−18376.1	3341.11
.999	(B)	501.353	−.608977	.110723
	(D)	502.688	−.718303	.130600
	(A)	55.4506	−10.0728	1.83143
−.999	(B)	.153951	−.0229357	.00417013
	(D)	1.13381	−.201093	.0365623
		$N = 20$		
	(A)	6.27689	−.319749	.0304523
.900	(B)	5.38817	−.262533	.0250031
	(D)	6.55244	−.340300	.0324095

TABLE I—*Continued*

ρ		(1,1)	(1,2) = (2,1)	(2,2)
		$N = 20$		
	(A)	.138960	−.0113663	.00108251
−.900	(B)	.0676743	−.00506050	.000481952
	(D)	.0686342	−.00515030	.000490505
	(A)	12.1678	−.444678	.0423503
.950	(B)	10.7942	−.371205	.0353529
	(D)	12.5241	−.469556	.0447196
	(A)	.238654	−.0206939	.00197084
−.950	(B)	.0644835	−.00482484	.000459509
	(D)	.0657622	−.00494608	.000471055
	(A)	53.4937	−.613232	.0584030
.990	(B)	51.2070	−.508144	.0483946
	(D)	53.6624	−.615992	.0586659
	(A)	1.20489	−.112460	.0107105
−.990	(B)	.0620906	−.00464796	.000442663
	(D)	.0666665	−.00508374	.000484166
	(A)	110.392	−1.27635	.121556
.995	(B)	101.256	−.529778	.0504551
	(D)	103.834	−.637770	.0607400
	(A)	2.44655	−.230670	.0219686
−.995	(B)	.0618009	−.00462654	.000440623
	(D)	.0705464	−.00545945	.000519948
	(A)	9215.74	−830.367	79.0826
.999	(B)	501.289	−.547960	.0521867
	(D)	503.972	−.655814	.0624591
	(A)	12.4146	−1.17998	.112379
−.999	(B)	.0615705	−.00460951	.000439001
	(D)	.103689	−.00862083	.000821030
		$N = 50$		
	(A)	4.79468	−.124380	.00487763
.900	(B)	4.00795	−.0995044	.00390213
	(D)	4.93797	−.129147	.00506460
	(A)	.0337711	−.00106614	.0000418100
−.900	(B)	.0239681	−.000718468	.0000281757
	(D)	.0240267	−.000720649	.0000282612
	(A)	11.5363	−.251542	.00986438
.950	(B)	9.56262	−.196752	.00771575
	(D)	11.7987	−.259762	.0101868

TABLE I—*Continued*

ρ		(1,1)	(1,2) = (2,1)	(2,2)
		$N = 50$		
	(A)	.0442648	−.00145541	.0000570752
−.950	(B)	.0227864	−.000683224	.0000267935
	(D)	.0228348	−.000685072	.0000268660
	(A)	55.6734	−.504984	.0198033
.990	(B)	50.9067	−.415065	.0162770
	(D)	56.0755	−.515782	.0202268
	(A)	.175959	−.00654664	.000256731
−.990	(B)	.0219029	−.000656862	.0000257598
	(D)	.0220090	−.000661019	.0000259227
	(A)	106.959	−.571819	.0224243
.995	(B)	101.111	−.463739	.0181858
	(D)	106.975	−.566814	.0222280
	(A)	.359302	−.0137189	.000537997
−.995	(B)	.0217961	−.000653675	.0000256348
	(D)	.0219861	−.000661125	.0000259269
	(A)	1072.18	−22.7526	.892259
.999	(B)	501.240	−.508366	.0199359
	(D)	507.776	−.612102	.0240041
	(A)	1.85492	−.0723543	.00283742
−.999	(B)	.0217112	−.000651141	.0000255354
	(D)	.0225763	−.000685072	.0000268658

(A) Least squares, (B) Minimum variance, and (D) Weighted least squares (The asymptotic approximation of the covariance matrices (C) given by Rosenblatt are not included here.)

coefficients $\rho = \pm.9, \pm.95, \pm.995, \pm.999$. From the discussion in Grenander and Szegö ([3a], p. 71), it can be shown that $\kappa \rightarrow [(1 + |\rho|)/(1 - |\rho|)]^2$ as $n \rightarrow \infty$. Consequently, to show the improvement of choosing the optimum diagonal weighted least squares, we have chosen $|\rho| \geqq .9$. Note that the variances of the weighted least squares estimates are not uniformly smaller than the variances of the least squares estimates. However, for $\rho = .999$ there is a considerable reduction in the variance by simply weighting the first and last observation.

Acknowledgments. The author wishes to thank Dr. T. A. Magness for making the results of paper [6] available prior to its publication, Dr. E. M. Scheuer for helpful comments and Mr. P. Businger of the Stanford Computation Center for performing the computations in Table I.

REFERENCES

[1] Bauer, F. L. and Householder, A. S. (1960). Some inequalities involving the Euclidean condition of a matrix. *Numer. Math.* **2** 308–311.

[2] Forsythe, G. E. and Straus, E. G. (1955). On best conditioned matrices. *Proc. Amer. Math. Soc.* **6** 340–345.

[3] Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York.

[3a] Grenander, U. and Szegö, G. (1958). *Toeplitz Forms and Their Applications*. Univ. of California Press.

[4] Householder, A. S. and Bauer, F. L. (1960). On certain iterative methods for solving linear systems. *Numer. Math.* **2** 55–59.

[5] Kantorovich, L. V. (1948). Functional analysis and applied mathematics. (Russian) *Uspehi Math. Nauk* (*N.S.*) **3** 89–185.

[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] Rosenblatt, M. (1956). Some regression problems in time series analysis. *Proc. Third Berkeley Symp. Math. Statist. Prob.* 165–186. Univ. of California Press.

[8] Schopf, A. H. (1960). On the Kantorovich inequality. *Numer. Math.* **2** 344–346.

[9] Young, D. (1954). Iterative methods for solving partial difference equations of elliptic type. *Trans. Amer. Math. Soc.* **76** 92–111.