Tr. J. of Mathematics
22 (1998) , 77 – 84.

GENERALIZED INVERSE ESTIMATOR AND COMPARISON WITH LEAST SQUARES ESTIMATOR

S. Sakallıoğlu & F. Akdeniz

Abstract

Trenkler [13] described an iteration estimator. This estimator is defined as follows: for $0 < \gamma < 1/\lambda_i$ max

$$\hat{\beta}_{m,\gamma} = \gamma \sum_{i=0}^{m} (1 - \gamma X'X)^i X'y,$$

where λ_i are eigenvalues of $X'X$. In this paper a new estimator (generalized inverse estimator) is introduced based on the results of Tewarson [11]. A sufficient condition for the difference of mean square error matrices of least squares estimator and generalized inverse estimator to be positive definite (p.d.) is derived.

1. Introduction

Consider the linear regression model

$$y = X\beta + e, \tag{1}$$

where y is an $n\times 1$ vector of observations on the dependent variable, X is an $n\times p$ matrix and of full column rank, β is a $p \times 1$ parameter vector, $E(e) = 0$, and $Var(e) = \sigma^2 I$, and both β and σ^2 are unknown. The least squares estimator for β is

$$\hat{\beta} = (X'X)^{-1}X'y. \tag{2}$$

The two key properties of $\hat{\beta}$ are that it is unbiased: $E(\hat{\beta}) = \beta$, and that it has minimum variance among all linear unbiased estimators. The mean square error of $\hat{\beta}$ is

$$mse(\hat{\beta}) = \sigma^2 \sum_{i=1}^{p} \frac{1}{\lambda_i}, \tag{3}$$

where λ_i's are the eigenvalues of $X'X$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$. If the smallest eigenvalue of $X'X$ is very much smaller than 1, then a seriously ill-conditioned (or multicollinearity) problem arises. Thus, for ill-conditioned data, the least squares solution yields coefficients whose absolute values are too large and whose signs may actually reverse with negligible changes in the data. That is, in the case of multicollinearity the least squares estimator $\hat{\beta}$ can be poor in terms of various mean squared error criterion. Consequently a great deal of work has been done to construct alternatives to the least squares estimator when multicollinearity is present. To reduce effects of multicollinearity we define some biased estimators in the model (1).

Ridge Estimator: [4] $(k > 0)$

$$\hat{\beta}_k = (X'X + kI)^{-1} X'y. \tag{4}$$

Shrunken Estimator: [7] $(0 < s < 1)$

$$\hat{\beta}_s = s\hat{\beta}. \tag{5}$$

Principal Components Regression Estimator: [6]

$$\hat{\beta}_r = A_r^+ + X'y, \tag{6}$$

where A_r^+ is the Moore-Penrose generalized inverse of $X'X$ having prescribed rank r. For an extensive discussion of the theory of Moore-Penrose generalized inverses, we refer to the books by Albert [1], Ben Israel and Greville [2], and Rao and Mitra [9].

Iteration Estimator: i) [10, 13, 14, 15], $(0 < \gamma < 1/\lambda_{\max},\ \ m = 0, 1, \ldots)$

$$\hat{\beta}_{m,\gamma} = \gamma \sum_{i=0}^{m} (I - \gamma X'X)^i X'y. \tag{7}$$

This estimator is shown to have similar properties as ridge, shrunken, and principal component estimator. The estimator $\hat{\beta}_{m,\gamma}$ is based on the convergence of the sequence

$$X_{m,\gamma} = \gamma \sum_{i=0}^{m} (I - \gamma X'X)^i X'$$

(with limit $X^+ = (X'X)^{-1}X'$) when $m \to \infty$. The sequence $X_{m,\gamma}$ also converges when $X'X$ is singular. The matrix $X_{m,\gamma}$ can be found by iterative procedure

$$X_{0,\gamma} = \gamma X', \qquad X_{m+1,\gamma} = (I - \gamma X'X)X_{m,\gamma} + \gamma X'.$$

Thus, we get the sequence of estimators $\beta, \hat{\beta}^{(n)}$, which is defined by Öztürk as follows:

ii) [8], $(0 < h < 2/\lambda_{\max},\ n = 1, 2, \ldots)$

$$\hat{\beta}^{(n)} = (I - hX'X)\hat{\beta}^{(n-1)} + hX'y, \tag{8}$$

where $\hat{\beta}^{(0)}$ is fixed point in the parameter space E_p.

In [14], Trenkler compare the iteration estimator with least squares, ridge, shrunken and principal components estimator with respect to matrix-valued mean square error criterion.

Although these estimators are biased, some of them are in widespread use since both bias and total variance can be controlled to a large extent. Bias and total variance of an estimator $\tilde{\beta}$ are measured simultaneously by scalar-valued mean square error (mse):

$$\begin{aligned} mse(\tilde{\beta}) &= E(\tilde{\beta}-\beta)'(\tilde{\beta}-\beta) \\ &= V(\tilde{\beta}) + (bias\tilde{\beta})'(bias\tilde{\beta}), \end{aligned} \tag{9}$$

where $V(\tilde{\beta}) = tr(Var(\tilde{\beta}))$ denotes total variance.

But mse is only one measure of goodness of an estimator. Another is generalized scalar-valued mean square error (gmse):

$$mse_F(\tilde{\beta}) = E(\tilde{\beta}-\beta)'F(\tilde{\beta}-\beta), \tag{10}$$

where F is a nonnegative definite (n.n.d.) symmetric matrix of order $p \times p$. The matrix-valued mean square error for any estimator $\tilde{\beta}$ is defined as

$$\begin{aligned} MSE(\tilde{\beta}) &= E(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)' \\ &= Var(\tilde{\beta}) + (bias\tilde{\beta})(bias\tilde{\beta})'. \end{aligned} \tag{11}$$

For any estimators $j = 1, 2$ consider

$$MSE(\tilde{\beta}_j) = E(\tilde{\beta}_j - \beta)(\tilde{\beta}-\beta)'. \tag{12}$$

Theobald [12] proves that $mse_F(\tilde{\beta}_1) > mse_F(\tilde{\beta}_2)$ for all positive definite (p.d.) matrices F if and only if $MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2)$ is p.d.. Thus the superiority of $\tilde{\beta}_2$ over $\tilde{\beta}_1$ with respect to the mse criteria can be examined by comparing their MSE. If $MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2) \geq 0$ then $\tilde{\beta}_2$ can be considered better than $\tilde{\beta}_1$ in mse.

2. A New Estimator: (Generalized Inverse Estimator)

For

$$\delta_i = \sum_{j=1}^{q} c_j \lambda_i^j > 0 \quad (i = 1, 2, \ldots, p)$$

and $0 < h < 2/\delta_{\max}$ consider a new iteration estimator of β. This estimator can be written as $(n = 1, 2, \ldots)$

$$\hat{\beta}^{(n)} = (I - hGX)\hat{\beta}^{(n-1)} + hGy, \tag{13}$$

where λ_i are eigen values of $X'X$, $\hat{\beta}^{(0)} = hGy$, and

$$\begin{aligned} G &= [c_1 I_p + c_2 X'X + c_3 (X'X)^2 + \cdots + c_q (X'X)^{q-1}]X', \\ & \left|1 - c_1\lambda_i - c_2\lambda_i^2 - \cdots - c_q\lambda_i^q\right| = \left|1 - \sum_{j=1}^{q} c_j \lambda_i^j\right| < 1. \end{aligned} \tag{14}$$

The matrix G and condition (14) are the same as in Tewarson's Theorem 1 in [11].

The model (1) can be reduced to a canonical form by using $X = U\Omega V'$, the singular value decomposition of X, where U is a $(n \times n)$ orthogonal matrix, V is a $(p \times p)$ orthogonal matrix, $\Omega' = [\Lambda^{1/2}, 0]$, and $\Lambda^{1/2} = diag\{\lambda_i^{1/2}\}_{i=1}^{p}$. Then (1) becomes

$$y = Z\alpha + e, \tag{15}$$

where $Z = U\Omega = XV$ and $\alpha = V'\beta$. The least squares estimator of $\alpha, \hat{\alpha}$, is

$$\hat{\alpha} = (Z'Z)^{-1}Zy = \Lambda^{-1}Z'y. \tag{16}$$

In general,

$$\hat{\alpha} = Z^{+}y \tag{17}$$

where Z^{+} is the Moore-Penrose generalized inverse of Z.

Thus, the matrix G and generalized inverse estimator of $\alpha, \hat{\alpha}^{(n)}$ become

$$G = V[c_1 I_p + c_2\Lambda + c_3\Lambda^2 + \cdots + c_q\Lambda^{(q-1)}]\Omega' U'$$

and

$$\hat{\alpha}^{(n)} = V'\hat{\beta}^{(n)} = (I - hW\Lambda)\hat{\alpha}^{(n-1)} + hW\Lambda\hat{\alpha},$$

where $W = [c_1 I_p + c_2\Lambda + c_3\Lambda^2 + \cdots + c_q\Lambda^{(q-1)}]$.

Then, we obtain

$$\begin{aligned}
\hat{\alpha}^{(n)} &= (I - hW\Lambda)\hat{\alpha}^{(n-1)} + hW\Lambda\hat{\alpha} \\
&= (I - hW\Lambda)[(I - hW\Lambda)\hat{\alpha}^{(n-2)} + hW\Lambda\hat{\alpha}] + hW\Lambda\hat{\alpha} \\
&= (I - hW\Lambda)^2\hat{\alpha}^{(n-2)} + (I - hW\Lambda)hW\Lambda\hat{\alpha} + hW\Lambda\hat{\alpha} \\
&\vdots \\
&= (I - hW\Lambda)^n\hat{\alpha}^{(0)} + (I - hW\Lambda)^{n-1}hW\Lambda\hat{\alpha} + \cdots + (I - hW\Lambda)hW\Lambda\hat{\alpha} + hW\Lambda\hat{\alpha} \\
&= (I - hW\Lambda)^n\hat{\alpha}^{(0)} + \sum_{m=0}^{n-1}(I - hW\Lambda)^m hW\Lambda\hat{\alpha} \\
&= (I - hW\Lambda)^n\hat{\alpha}^{(0)} + \{I - (I - hW\Lambda)^n\}\hat{\alpha}.
\end{aligned} \tag{18}$$

If we take as an initial solution $\hat{\alpha}^{(0)} = 0$ then we get

$$\hat{\alpha}^{(n)} = \{I - (I - hW\Lambda)^n\}\hat{\alpha}. \tag{19}$$

Thus we have

$$E(\hat{\alpha}^{(n)}) = \alpha - (I - hW\Lambda)^n\alpha; \tag{20}$$

$$bias(\hat{\alpha}^{(n)}) = E(\hat{\alpha}^{(n)}) - \alpha = -(I - hW\Lambda)^n\alpha; \tag{21}$$

$$Var(\hat{\alpha}^{(n)}) = \sigma^2\{I-(I-hW\Lambda)^n\}^2\Lambda^{-1}; \tag{22}$$

$$\begin{aligned} mse(\hat{\alpha}^{(n)}) &= tr(Var(\hat{\alpha}^{(n)})) + (bias(\hat{\alpha}^{(n)}))'(bias(\hat{\alpha}^{(n)})) \\ &= \sigma^2\sum_{i=1}^{p}\{1-(1-hw_{ii}\lambda_i)^n\}^2\lambda_i^{-1} + \sum_{i=1}^{p}(1-hw_{ii}\lambda_i)^{2n}\alpha_i^2. \end{aligned} \tag{23}$$

3. Mean Square Error Comparisons of $\hat{\alpha}$ and $\hat{\alpha}^{(n)}$

In this section our objective is to compare the mean square error matrices. For this purpose consider the difference between $MSE(\hat{\alpha})$ and $MSE(\hat{\alpha}^{(n)})$ as

$$\begin{aligned} S = MSE(\hat{\alpha}) - MSE(\hat{\alpha}^{(n)}) &= \sigma^2\Lambda^{-1} - \sigma^2\{I-B\}^2\Lambda^{-1} - B\alpha\alpha' B \\ &= \sigma^2\{2B-B^2\}\Lambda^{-1} - B\alpha\alpha' B \\ &= T - B\alpha\alpha' B, \end{aligned} \tag{24}$$

where $B = (I-hW\Lambda)^n$ and $T = \sigma^2\{2B-B^2\}\Lambda^{-1}$. For

$$0 < \delta_i = \sum_{j=1}^{q} c_j\lambda_i^j < 1,$$

and $0 < h < 1/\delta_{\max}$, the i-th diagonal element of B, b_{ii}, is $0 < b_{ii} = [1-h\delta_i]^n < 1$, then the i-th diagonal element of T, t_{ii}, is

$$t_{ii} = (\sigma^2/\lambda_i)(2-b_{ii})b_{ii} > 0, \tag{25}$$

where $\lambda_i > 0$ because $X'X$ is a positive definite matrix. Since T is a diagonal matrix and all diagonal elements are positive, T is a positive definite matrix. Thus, using Farebrother's theorem in [5]: Let A be p.d. matrix, let c be a nonzero vector and let d be a positive scalar. Then $dA - cc'$ is p.d. iff $c'A^{-1}c$ is less than d. From this we obtain that $S > 0$ if and only if $\alpha' B' T^{-1} B\alpha < 1$ and then

$$\alpha' B' T^{-1} B\alpha = \sum_{i=1}^{p}[(\lambda_i b_{ii})/(2-b_{ii})]\alpha_i^2 < \sigma^2, \tag{26}$$

or

$$\alpha' diag\left(\frac{\lambda_i b_{ii}}{2-b_{ii}}\right)\alpha < \sigma^2. \tag{27}$$

Since as

$$n\to\infty \quad \lim\left(\frac{\lambda_i b_{ii}}{2-b_{ii}}\right) = 0 \quad \text{for} \quad i = 1,2,\ldots,p,$$

there exists an integer n_0 such that $MSE(\hat{\alpha}) - MSE(\hat{\alpha}^{(n)})$ is p.d. for all $n > n_0$.

Now, we may state the following theorem.

Theorem 3.1. *A sufficient condition for the generalized inverse estimator, $\hat{\alpha}^{(n)}$, to have smaller mse than the least squares estimator, $\hat{\alpha}$, is*

$$n > \max \left\{ \frac{\ln\left(\frac{2\sigma^2}{\sigma^2+\lambda_i\alpha_i^2}\right)}{\ln(1-hw_{ii}\lambda_i)} \right\} \qquad (i=1,2,\ldots,p). \tag{28}$$

where w_{ii} is the i-th diagonal element of W, and α_i is the i-th element of α.

Consequently under the conditions (27) or (28) the new iteration estimator $\hat{\beta}(n)$ (or $\hat{\alpha}^{(n)}$) is superior to $\hat{\beta}$ (or $\hat{\alpha}$).

Note that if we take $c_1 > 0$, $c_2 = c_3 = \cdots = c_q = 0$ the matrix G and condition (14) become $G = c_1 X'$, $|1 - c_1\lambda_i| < 1$, respectively, and we obtain $0 < c_1 < 2/\lambda_{\max}$. So we have seen that the generalized inverse estimator $\hat{\beta}^{(n)}$ is reduced to $\hat{\beta}_{m,\gamma}$, which is called a iteration estimator and is defined by Trenkler in [13].

4. Numerical Example

In this section, we used a particular model with a data set often used in examination of multicollinearity problems. The data (Hald (1952)) are from Daniel and Wood (1971, pp.100) [3]. For this data, we get the following results: the eigen values of $X'X$ are 2.235, 1.576, 0.186, 0.002, the least squares estimate of α is $\hat{\alpha} = (0.65696, -0.00831, 0.3028, 0.388)'$ and $mse(\hat{\alpha}) = 1.225$, $\hat{\sigma}^2 = 0.00196$. The condition number is 1117. So there is multicollinearity. Table 1 gives generalized inverse estimator $\hat{\alpha}^{(n)}$ of α for various values of c_1, c_2, n and also the values of $mse(\hat{\alpha}^{(n)})$. $q = 2$ and $h = 1$ are taken for simplicity of calculations.

The value n_0 of n in (28) is computed by using the unbiased estimates of α and σ^2. From the results in Table 1 we can say that $\hat{\alpha}^{(n)}$ is superior to $\hat{\alpha}$ for the selected values of n_0.

Table 1. Values of $\hat{\alpha}^{(n)}$ and $mse(\hat{\alpha}^{(n)})$ for various values of c_1, c_2, n.

c_1	c_2	n_0	$\hat{\alpha}^{(n)}$	$mse(\hat{\alpha}^{(n)})$
0.2	0.1	40	(0.65696, -0.00831, 0.24522, 0.00616)'	0.15833
0.2	0.1	45	(0.65696, -0.00831, 0.25601, 0.00692)'	0.15751
0.2	0.0	45	(0.65696, -0.00831, 0.24781, 0.00692)'	0.15787
0.1	0.15	70	(0.65696, -0.00831, 0.24663, 0.00539)'	0.15879
0.1	0.0	105	(0.65696, -0.00831, 0.24141, 0.00654)'	0.15831

5. Conclusions

Computationally, use of the generalized inverse estimator appears to be very attractive since no matrix inversion is required. So it can be reasonable to use the generalized inverse estimator. Furthermore, when multicollinearity exists the total variance $(tr(Var(\hat{\alpha}))$ of the least squares estimator increases but

$$V(\hat{\alpha}^{(n)}) = tr(Var(\hat{\alpha}^{(n)})) = \sigma^2 \sum_{i=1}^{p} \{1 - (1 - hw_{ii}\lambda_i)^n\}^2 \lambda_i^{-1}$$

tends to a finite limit when λ_p approaches zero. Therefore, when multicollinearity exists the generalized inverse estimator, $\hat{\alpha}^{(n)}$, is remarkably robust

References

[1] Albert, A. (1972): Regression and the Moore-Penrose Inverse. New York: Academic Press.

[2] Ben Israel, and Greville, T.N.E. (1974): Generalized Inverses. Theory and Appl. New York: Wiley.

[3] Daniel, C. and Wood, F.S. (1971): Fitting Equations to Data. John Wiley.

[4] Hoerl, A.E. and Kennard, R.W.: "Ridge Regression: Biased Estimation for Orthogonal Problems", Technometrics, **12**, 55-67, (1970).

[5] Farebrother, R.W.: "Further Results on the Mean Square Error of Ridge Regression", *J.R. Statist. Soc. B*, **38**, 248-250, (1976).

[6] Marquardt, D.W.: "Generalized Inverses, Ridge Regression, and Nonlinear Estimation", *Technometrics*, **12**, 591-612, (1970).

[7] Mayer, L.S. and Willke, T.A.: "On Biased Estimation in Linear Models", *Technometrics*, **15**, 497-508, (1973).

[8] Öztürk, F.: "A Discrete Shrinking Method as Alternative to Least Squares", *Commun. Fac. Sci. Univ.*, Ankara, **33**, 179-185, (1984).

[9] Rao, C.R. and Mitra, S.K. (1971): Generalized Inverse of Matrices and Its Appl. New York: Wiley.

[10] Terasvirta, T.: "Superiority Comparisons of Homogeneous Linear Estimators", *Commun. Statist.*, **11** (14), 1595-1601, (1982).

[11] Tewarson, R.P.: "An Iterative Method for Computing Generalized Inverses", *Intern. J. Computer Math. Section B*, **3**, 65-74 (1971).

[12] Theobald, C.M.: "Generalizations of Mean Square Error Applied to Ridge Regression", *J.R. Statist. Soc. B*, **36**, 103-106 (1974).

[13] Trenkler, G.: "An Iteration Estimator for the Linear Model", COMPSTAT, Physica-Verlag, 125-131 (1978).

[14] Trenkler, G.: "Generalized Mean Squared Error Comparisons of Biased Regression", *Commun. Statistics Theor. Meth.*, A9 **12**, 1247-1259 (1980).

[15] Trenkler, D. and Trenkler, G.: "A Simulation Study Comparing Some Biased Estimators in the Linear Model", *Computational Statistics, Quarterly*, **1**, 45-60 (1984).

S. SAKALLIOĞLU& F. AKDENİZ
Department of Mathematics
Çukurova University
01330 Adana - TURKEY

Received 26.08.1996