The Relative Efficiency Based on the MSE in Generalized Ridge Estimate

Chao Yuan, Bao Guang Tian

***Abstract*—**A relative efficiency is defined as Ridge Estimate in the general linear model. The relative efficiency is based on the Mean square error. In this paper, we put forward a parameter of Ridge Estimate and discussions are made on the relative efficiency between the ridge estimation and the General Ridge Estimate. Eventually, this paper proves that the estimation is better than the general ridge estimate, which is based on the MSE.

***Keywords*—**Ridge estimate, generalized ridge estimate, MSE, relative efficiency.

I. Introduction

THE generalized Markov-Gauss model [9] is described by

$$Y = X\beta + \varepsilon \tag{1}$$

where Y is the n observation vector, X is a $n\times p$ column full rank design matrix , β is an unknown parameter vector. $E(\varepsilon)=0$, ε is the $n\times 1$ observation vector. $Cov(\varepsilon)=\sigma^2\Sigma$, Σ is the $n\times n$ positive definite co variance matrix and the rank is $r(\Sigma)=m\le n$. σ^2 is unknown parameter.

In the light of the least square unified theory of estimation [1], the BLU of β is

$$Q(\beta)=\|y-X\beta\|$$

$$X'\Sigma^{-1}X\beta = X'\Sigma^{-1}y$$

$$\min Q(\beta)=\min(y-X\beta)'\Sigma^{-1}(y-X\beta)$$

the generalized least eigenvalue of β is the $\hat{\beta}_{GLS}$.

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

and

$$Cov\hat{\beta}=\sigma^2(X'\Sigma^{-1}X)^{-1}$$

$$MSE\hat{\beta}=\sigma^2 tr(X'\Sigma^{-1}X)^{-1}=\sum_{i=1}^{p}\frac{\sigma^2}{\lambda}$$

C. Yuan and B. Tian are with the Department of Mathematics, Qingdao University of Science & Technology, Shandong 266061, China, e-mail: (675593912@qq.com).

where $\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_p>0$, λ_i represents the characteristic value of $X'\Sigma^{-1}X$.

In 1970, Hoerl and Kennard [2], [3] put forward the Ridge Estimate. In practical application, it is proved that the Ridge Estimate improves the performance of the LSE. Reference [4] raised a new relative efficiency and proved the Ridge Estimate was better than LSE. In this paper, the method is further applied to the new Ridge Estimate.

II. The Definition and Lemma

As is well known, $\hat{\beta}_{GLS}$ is the BLUE for regression parameters. However, there is the multi-collinearity in model, the distance is relatively large between $\hat{\beta}_{GLS}$ and β . Meanwhile, the estimate variance of $\hat{\beta}_{GLS}$ will become large. Therefore, an estimate C which is equal with the sum of squares of the residuals can be selected, $\hat{\beta}_{GLS}$ is defined by solving the minimum estimated distance in the C . That is

$$\min\{\beta'\beta+\tfrac{1}{k}[(y-X\beta)'\Sigma^{-1}(y-X\beta)-c]\} \tag{3}$$

In model (3), it is based on the choice of different k that we get the sum of squares about different residual. Then we select the best estimate value by LSE. As a result;

$$\min\{\beta'\beta+\tfrac{1}{k}[(y-X\beta)'\Sigma^{-1}(y-X\beta)-c]\} \tag{4}$$

Theorem 1. In model (1), the Trenkler [4] Ridge Estimate (RE) is

$$k>0,$$

$$\hat{\beta}_{RE}=(X'\Sigma^{-1}X+kI_r)^{-1}X'\Sigma^{-1}y=T_k\hat{\beta}_{GLS}$$

where

$$T_k=(I_r+ks^{-1})^{-1}=s_k^{-1}s=ss_k^{-1}$$

$$s=X'\Sigma^{-1}X$$

$$s_k=X'\Sigma^{-1}X+kI_r$$

Theorem 2. In model (1), the estimated value of the unknown parameter β is

$$\hat{\beta}(K) = (X'\Sigma^{-1}X + QKQ')^{-1}X'\Sigma^{-1}y \tag{5}$$

which is the generalized RE of β where

$$Q'X'\Sigma^{-1}XQ = \wedge = diag(\lambda_1, \lambda_2, \cdots, \lambda_p)$$

$$K = diag(k_1, k_2, \cdots k_p)$$

$k_1, k_2, \cdots, k_p$ are p ridge parameters. Consider the following: Assume that $K = diag(k_1, k_2, \cdots k_p)$, where

$$k_i = \varphi(k'_i)$$

and

$$1 \le i \le p\,, c > 0\,.$$

$$\varphi(k'_i) = \begin{cases} -c, & k'_i < c \\ k'_i, & |k'_i| \le c \\ c, & k'_i > c \end{cases} \tag{6}$$

Then this paper makes some discussions on the efficiency of ridge estimator and the general ridge estimator based on mean square error.

Theorem 3. The relative efficiency is defined as $\hat{\beta}$ being relative to β^*, which is

$$e(\hat{\beta}, \beta^*) = 1 - \frac{MSE(\hat{\beta})}{MSE(\beta^*)} \tag{7}$$

If β^* is BLUE, then $e(\hat{\beta}, \beta^*) = e(\hat{\beta})$.

According to Theorem 3, if $0 < e(\hat{\beta}, \beta^*) \le 1$, then $\hat{\beta}$ is better than β^*. And $e(\hat{\beta}, \beta^*)$ is larger, the efficiency is higher. So, let's prove this result, which is

$$0 < e(\hat{\beta}(k)) < e(\beta^*(K)) \le 1$$

if

$$\theta = Q'\beta, \theta = (\theta_1, \theta_2, \cdots, \theta_p)',$$

then

$$\hat{\theta} = \Lambda^{-1}Q'X'\Sigma^{-1}y \tag{8}$$

$$\hat{\theta}(K) = (\Lambda + K)^{-1}Q'X'\Sigma^{-1}y \tag{9}$$

Lemma 1[6] $MSE(\hat{\theta}(K)) = \sum_{i=1}^{p} f_i(k_i)$, where

$$f_i(k_i) = \frac{\sigma_2\lambda_i + \theta_i^2 k_i^2}{(\lambda_i + k_i)^2}.$$

Proof. According to the result (8),

$$\hat{\theta} = \Lambda^{-1}Q'X'\Sigma^{-1}y$$

$$\hat{\theta}(K) = (\Lambda + K)^{-1}\Lambda\hat{\theta} = A\hat{\theta}$$

where

$$A = (\Lambda + K)^{-1}\Lambda \;= diag\left[\frac{\lambda_1}{\lambda_1 + k_1}, \frac{\lambda_2}{\lambda_2 + k_2}, \cdots, \frac{\lambda_p}{\lambda_p + k_p}\right]$$

and

$$MSE(\hat{\theta}(K)) = trCov(\theta(K)) + \left\|E(\hat{\theta}(K) - \theta)\right\|^2$$

$$trCov(\hat{\theta}(K)) = trCov(A\hat{\theta}) = \sigma^2 tr[A(X'\Sigma^{-1}X)^{-1}A'] = \sigma^2\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i + k_i)^2}$$

$$E(\hat{\theta}(K)) = E(A\hat{\theta}) = AE\hat{\theta} = A\hat{\theta}$$

then

$$\left\|E(\hat{\theta}(K) - \theta)\right\|^2 = \left\|(A - I)\theta\right\|^2 = \sum_{i=1}^{p}\frac{\theta_i^2 k_i^2}{(\lambda_i + k_i)^2}$$

Above all,

$$MSE(\hat{\theta}(K)) = \sum_{i=1}^{p} f_i(k_i)\,.$$

Lemma 2 [7] If $\beta = Q\theta$ and Q is orthogonal matrix, then

$$MSE(\beta) = MSE(\theta)$$

Proof. According to the condition $\beta = Q\theta$, then

$$MSE(\beta) = MSE(Q\theta)$$

$$MSE(\hat{\beta}) = E(\hat{\beta} - \beta)'(\hat{\beta} - \beta) = E(Q\hat{\theta} - Q\theta)'(Q\hat{\theta} - Q\theta) = E(\hat{\theta} - \theta)'Q'Q(\hat{\theta} - \theta) = E(\hat{\theta} - \theta)'(\hat{\theta} - \theta) = MSE(\theta)$$

Above all,

$$MSE(\beta) = MSE(\theta)\,.$$

Lemma 3. [7]

$$f(k) = \frac{\sigma^2\lambda + \theta^2 k^2}{(\lambda + k)^2}$$

If $k = \frac{\sigma^2}{\theta^2}$, meanwhile, the value of $f(k)$ is the least value, and where $k > 0, \lambda > 0$.

Proof. According to the condition,

$$f'(k) = \frac{2\theta^2 k(\lambda+k)^2 - 2(\lambda+k)(\sigma^2\lambda + \theta^2 k^2)}{(\lambda+k)^4}$$
$$= \frac{2\theta^2 k(\lambda+k) - 2(\sigma^2\lambda + \theta^2 k^2)}{(\lambda+k)^3} = \frac{2\lambda(\theta^2 k - \sigma^2)}{(\lambda+k)^3}$$

Assume that

$$f'(k) = 0 \,, k = \frac{\sigma^2}{\theta^2} \,,$$

then $f(k)$ has an extreme value.

If

$$0 \le k < \frac{\sigma^2}{\theta^2} \,,$$

then

$$f'(k) = \frac{2\lambda(\theta^2 k - \sigma^2)}{(\lambda+k)^3} < \frac{2\lambda(\theta^2 \frac{\sigma^2}{\theta^2} - \sigma^2)}{(\lambda+k)^3} = 0 \,.$$

If

$$k > \frac{\sigma^2}{\theta^2} \,,$$

then

$$f'(k) > 0 \,.$$

Above all, if $k = \frac{\sigma^2}{\theta^2}$, meanwhile, the value of $f(k)$ is the least value, and where $k > 0, \lambda > 0$.

Lemma 4. [8] In the model of (1), when $k > 0$, then

$$MSE\hat{\beta}(k) = \sum_{i=1}^{p} \frac{\sigma^2\lambda_i + \theta_i^2 k^2}{(\lambda_i + k)^2}$$

and

$$0 < e(\hat{\beta}(k)) \le 1 \,.$$

Lemma 5. [4] In the model of (1), when

$$0 < k \le \frac{2\sigma^2\lambda_p}{\theta_p^2\lambda_1 - \sigma^2} \,,$$

then

$$e(\beta(K)) \ge e_0 > 0 \tag{10}$$

and

$$e_0 = \frac{2k\lambda_p + k^2(1 - \theta_p^2\lambda_1/\sigma^2)}{(\lambda_1 + k)^2} \,,$$

$$\theta = Q'\beta \,,$$

β_p is the max vector component of θ module.

III. The Main results

Lemma. In the model of (1), when

$$k > 0 \,, K = diag(k_1, k_2, \cdots k_p)$$

then

$$0 \le e(\hat{\beta}(K), \hat{\beta}(k)) \le 1 \,.$$

Proof. When $k > 0$,

i. If $k = \sigma^2/\theta_i^2$, $i = 1, 2, \cdots, p$ assume that

$$k'_i = k, i = 1, 2, \cdots, p \,, c > \sigma^2/\theta^2 \,,$$

then

$$k_i = \varphi(k_i') = k_i', \qquad |k_i'| \le c \tag{11}$$

$$k_i = \varphi(k_i') = c, \qquad k_i' > c \tag{12}$$

Discussion is made on the condition (1), $e(\hat{\beta}(K), \hat{\beta}(k)) = 0$ is obvious. Discussions are made on the condition (2),

$$MSE(\hat{\beta}(k)) - MSE(\hat{\beta}(K)) = \sum_{i=1}^{p} f_i(k) - \sum_{i=1}^{p} f_i(c)$$
$$= \sum_{i=1}^{p} f_i(k) - \sum_{i=1}^{p} f_i(c)$$

According to the Lemma 2,

$$MSE(\hat{\beta}(k)) = MSE(\hat{\beta}(K)) \,,$$

$$MSE(\hat{\beta}(k)) - MSE(\hat{\beta}(K)) = 0$$

$$e(\hat{\beta}(K), \hat{\beta}(k)) = 1$$

ii. Assume that there is an i ,which lets $k \ne \sigma^2/\theta_i^2$.

If

$$k_1 = \sigma^2/\theta_i^2 \,, k_i' = k \,,$$

then

$$MSE(\hat{\beta}(k)) - MSE(\hat{\beta}(K)) = \sum_{i=1}^{p} f_i(k) - \sum_{i=1}^{p} f_i(k_i)$$
$$= f_i(k) - f_1(k_1)$$

Discussions are made on the condition (1), this paper gets the result according to the Lemma 3,which is

$$f_i(k) - f_1(k_1) > 0$$

$$MSE(\hat{\beta}(k)) - MSE(\hat{\beta}(K)) > 0$$

thus

$$e(\hat{\beta}(K), \hat{\beta}(k)) > 0$$

Discussions are made on the condition (2),

$$MSE(\hat{\beta}(k)) - MSE(\hat{\beta}(K)) = \sum_{i=1}^{p} f_i(k) - \sum_{i=1}^{p} f_i(k_i)$$

$$= f_1(k) - f_1(k_1) + \sum_{i=2}^{p} f_i(k) - \sum_{i=2}^{p} f_i(c)$$

According to the Lemma3, when $k > c > \sigma^2 / \theta^2$, then

$$f_1(k) - f_1(k_1) > 0,$$

$$\sum_{i=2}^{p} f_i(k) - \sum_{i=2}^{p} f_i(c) > 0$$

and

$$MSE(\hat{\beta}(k)) - MSE(\hat{\beta}(K)) > 0$$

thus

$$e(\hat{\beta}(K), \hat{\beta}(k)) > 0$$

Above all,

$$0 \le e(\hat{\beta}(K), \hat{\beta}(k)) \le 1 .$$

Inference. In the model of (1), when $k > 0$, thena

$$0 < e(\hat{\beta}(k)) \le e(\hat{\beta}(K)) \le 1 .$$

Proof. According to the Lemma 4 and 5, the result is obvious.

IV. Conclusions

Owing to the determining theorem of unbiased estimator, the estimation of LSE in linear model is used. But there is the multi-collinearity in model, which is very difficult to obtain solution of the problem. So, people like to use the Biased estimate, for example the RE. And this estimation may be not exact solution of the problem, which will take some losses. But this paper proves that the new estimation is better than the general RE, which is based on the MSE.

On account of the relative efficiency of linear mode based on mean square error, further studies can focus on these limitations and make improvements.

1. Discussions are made on the relative efficiency of the Generalized RE and the RE, which is based on the ridge parameter that is proposed in this paper.
2. Further studies can focus on the relation with the generalized relative coefficient and efficiency.
3. Try to find another better parameter estimation in the RE model.

References

[1] A.Y. Liu, S.G. Wang. A new relative efficiency of least square estimation in linear model. Applied probability and statistics, 1989, 15(2), pp.97-104.

[2] Hoerl, A.E., Kennard, R.W. Ridge regression: Biased estimation for non-orthogonal problems (J). Technometrics, 1970, 12:55-88.

[3] Hoerl, A.E., Kennard, R.W. Ridge regression: Application for non-orthogonal problems (J). Technometrics, 1970, 12:69-72.

[4] Q. Tang. The Relative Efficiency of General Ridge Estimate (J). Journal of Chengdu University (Natural science), 1991, 10 (2): 42-451

[5] Trenkler, G. On the performance of biased estimators in the linear regression model with correlated or heteroscedastic errors (J). Journal of Econometrics 1984, 25:179-190.

[6] P.H. Wang. The Relative Efficiency of Generalized Ridge Estimate (J), Journal of Chengdu University (Natural science), 2001, 20 (2).

[7] P.H. Wang. The Efficiency of Generalized Ridge Estimate (J). Journal of Quan Zhou University (Natural science) 1998, 16 (2): 13-151.

[8] A.Y.L. Huang, G.J. Chen. The relative efficiency of parameter estimation in linear model (J). Applied probability and statistics, 1998, 14 (2): 159-1641.

[9] C. Yuan, B.G. Tian, A new relative efficiency based on the least eigenvalue in generalized linear model (J). International science index, 2016, 9 (10).