On The *Ridge* Estimations With The Correlated Error Structure

Won, Byung Chool*

Abstracts : In this paper, we shall construct a *ridge* estimator in a multiple linear model with the correlated error structure. The existence of the biasing parameter satisfying the Mean Squared Error Criterion is also proved. Furthermore, we shall determine the value of shrinkage factors by the iteration method.

I. Introduction

Consider the generalized linear model (1.1)

$$Y = X\beta + \varepsilon \tag{1.1}$$

where it is assumed that $X = (X_1, X_2, \ldots, X_p)$ is a known matrix of rank $q \leq p$, Y is an n x 1 vector of observations and ε is the $n \times 1$ vector of errors such that $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \sigma^2 V_n$

The classical estimation procedure of model (1.1) is that of generalized least squares (GLS) in which b is chosen such that the residual sum of squares $\Phi(b) = (Y - Xb)'V_n^{-1}(Y - Xb)$ is minimized. In the case that $X'V_n^{-1}X$ is of full rank, $(X'V_n^{-1}X)^{-1}$ exists and the GLS estimators are given by

$$b_G = (X'V_n^{-1}X)^{-1}X'V_n^{-1}Y \tag{1.2}$$

However, in the case that $X'V_n^{-1}X$ is of rank $q < p$, alternate methods must be employed to obtain the estimators. In fact if $X'V_n^{-1}X$ is of full rank but at least one eigenvalue approaches zero, the GLS estimators are sensitive to a number of errors. Further, the variance of GLS estimator become large as the matrix $X'V_n^{-1}X$ approaches singularity. Although the Gauss-Markov

* Korea Atomic Energy Research Institute

Theorem assures us that in the class of all unbiased estimators, the GLS estimator among the estimable functions have minimum variance, we are faced with the unhappy circumstances and, hence procedure a large confidence intervals for the estimators.

One way to remedy this problem is to drop the requirement that b_G is is unbiased. Hoerl and Kennard (1970) have suggested that the ordinary least squares estimator may be replaced by the *ridge* estimator with positive biasing parameter. But Hoerl and Kennard's *ridge* estimator is derived from the assumption usually made concerning the linear regression model with the uncorrelated error structure.

In this article, we showed that if a biased estimator could be considered and if one would use a different criterion for estimator selection, namely the mean squared error criterion of an estimator, the *ridge*-type regression estimator could be shown to be superior to the GLS estimator.

Ⅱ. The Form of *Ridge* Estimator

Let Γ be a diagonal matrix of eigenvalue, λ_i, of $X'V_n^{-1}X$ and G be an orthogonal matrix of corresponding eigenvectors. Then we have $G'X'V_n^{-1}XG=\Gamma$ and $GG'=I_P$. If we write $X^*=XG$ and $\alpha=G'\beta$, then the linear model (1.1) may be written as

$$Y = X^*\alpha + \varepsilon \tag{2.1}$$

Then the GLS estimator a of α is given by

$$a = (X^{*\prime}V_n^{-1}X^*)^{-1}X^{*\prime}V_n^{-1}Y$$

$$= \Gamma^{-1}G'X'V_n^{-1}Y$$

$$\mathrm{Var}(a) = \sigma^2\Gamma^{-1}$$

Unfortunately if at least one or more eigenvalues approach to zero, the corresponding coordinate of an estimator has large varianc $\sigma^2\lambda_i^{-1}$. By allowing a small amount of bias, we can obtain a biased estimator that has variance less than any unbiased estimators. A number of procedure have have developed for obtaining biased estimators of regression coefficients.

The generalized *ridge* estimator (GRE) of α in (2.1) is obtained by augmenting the i^{th} diagonal elements of Γ by a positive constants k_i, i.e.,

$$\alpha_K = (\Gamma + K)^{-1}G'X'V_n^{-1}Y, \qquad K = \mathrm{Diag}(k_i) \tag{2.2}$$

Then the GRE of β in (1.1) may now be written as

$$\begin{aligned} b_K &= G\alpha_K \\ &= G(\Gamma + K)^{-1}G'X'V_n^{-1}Y \\ &= G\Delta G' b_G \end{aligned} \tag{2.3}$$

where $\Delta = (\Gamma + K)^{-1}\Gamma = \mathrm{Diag}(\delta_i)$ is diagonal matrix of shrinkage factors.

<u>Lemma 2.1</u> The generalized *ridge* estimator b_K is biased estimator and

$$\mathrm{Var}(b_K) = \sigma^2(X'V_n^{-1}X + K)^{-1}X'V_n^{-1}X(X'V_n^{-1}X + K)^{-1}$$

$$\mathrm{Bias}(b_K) = G\ \mathrm{Diag}(\delta_i - 1)G'\beta, \text{ where } \delta_i = \lambda_i(\lambda_i + k_i)^{-1},\ i=1,2,\ldots,p$$

Gauss in 1809 suggested mean squared error (MSE) as the most relevent criterion for choice among estimators. The MSE as defined in (2.4) is just the expected squared distance from b to β and frequently used to measure adequacy of an estimator.

$$\mathrm{MSE}\ (b) = E\ (b - \beta)'\ (b - \beta) \tag{2.4}$$

Also, more generally, it may be possible form a suitable weight sum of coefficients mean squared error WMSE $(b) = E(b - \beta)'W(b - \beta)$, where W is a non-negative definite matrix. The mean square error of b_K therefore is

$$\begin{aligned} \mathrm{MSE}(b_K) &= \mathrm{Tr}[\mathrm{Var}(b_K)] + [\mathrm{Bias}(b_K)]'[\mathrm{Bias}(b_K)] \\ &= \sigma^2\mathrm{Tr}[(X'V_n^{-1}X + K)^{-1}X'Vn^{-1}X(X'V_n^{-1}X + K)^{-1}] \\ &\quad + \alpha'\mathrm{Diag}(\delta_i - 1)^2\alpha \end{aligned}$$

$$= \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k_i)^2} + \sum_{i=1}^{p} \alpha_i^2 \left(\frac{\lambda_i}{\lambda_i + k_i} - 1\right)^2 \quad (2.5)$$

The first term on the right-hand side of (2.5) is the sum of variance of the parameter in b_K and the second term is the square of the bias. If $k_i > 0$, note that the bias in b_K increase with k_i.

Ⅲ. Mean Squared Error Comparisons

In using *ridge* regression, we would like to choose a value of biasing parameter such that the reduction in the variance term is greater than the increase in the squared bias. If this can be done, the mean squared error of the *ridge* estimator b_K will be less than that of GLS estimator b_G.

The purpose of this section is to establish a existence of a single biasing parameter k such that MSE of *ridge* estimator b_k ($K=kI_p$) is less than the MSE of the GLS estimator b_G. To show this, the following MSE matrix will be used.

$$\text{MtxMSE}(b) = E(b - \beta)(b - \beta)' \quad (3.1)$$

Lemma 3.1 The followings are equivalent.

(a) MtxMSE (b_1) - MtxMSE (b_2) is non-negative definite.

(b) WMSE (b_1) - WMSE (b_2) > 0 for all non-negative definite W

(Proof) See C.M. Theobald(1974)

Lemma 3.2 Let A be a positive definite n x n matrix and u be an n x 1 column vector (i.e. $u' \in E_n$). Then

$$\sup_{x' \in E_n} \frac{(u'x)^2}{x'Ax} = u'A^{-1}u \quad \text{and supremum is attained at } x=A^{-1}u.$$

(Proof) See C.R. Rao (1965)

For $K = kI_p$, next theorem says that there exists a non-zero k for which $WMSE(b_k)$ is less than that of generalized least squares estimator b.

<u>Theorem 3.3</u> There exists a $k_{max} > 0$ such that $MtxMSE(b) - MtxMSE(b_k)$ is non-negative definite where $0 < k < k_{max}$, and $k_{max} = \dfrac{2\sigma^2}{\beta'\beta}$

(Proof) Let $\Omega = MtxMSE(b) - MtxMSE(b_k)$ and η be any non-zero constant vector. Then

$$\Omega = \sigma^2(X'V_n^{-1}X)^{-1} - \sigma^2(X'V_n^{-1}X + kI_p)^{-1}X'V_n^{-1}X(X'V_n^{-1}X + kI_p)^{-1}$$
$$- k^2(X'V_n^{-1}X + kI_p)^{-1}\beta\beta'(X'V_n^{-1}X + kI_p)^{-1}$$

Put $\xi = (X'V_n^{-1}X + kI_p)^{-1}\eta$

$$\eta'\Omega\eta = \sigma^2\xi'[(X'V_n^{-1}X + kI_p)(X'V_n^{-1}X)^{-1}(X'V_n^{-1}X + kI_p)$$
$$- X'V_n^{-1}X - k^2\frac{\beta'\beta}{\sigma^2}]\xi$$
$$= \sigma^2\xi'[k^2(X'V_n^{-1}X)^{-1} + 2kI_p - k^2\frac{2\beta\beta'}{\sigma^2}]\xi$$

For any non-zero constant vector η,

$$\eta'\Omega\eta \geq 0 \quad \text{for } \eta \neq 0$$

if and only if
$$\frac{\xi'\beta\beta'\xi}{\xi'[(X'V_n^{-1}X)^{-1} + (2/k)I_p]\xi} \leq \sigma^2$$

if and only if
$$\theta(\xi) = \frac{(\xi'\beta)^2}{\sigma^2\xi'[(X'V_n^{-1}X)^{-1} + (2/k)I_p]\xi} \leq 1$$

By <u>Lemmma 3.2</u>

$$\sup_{\xi}\theta(\xi) = \frac{1}{\sigma^2}\beta'[(X'V_n^{-1}X)^{-1} + (2/k)I_p]^{-1}\beta \leq 1$$

if and only if
$$\beta'[(X'V_n^{-1}X)^{-1} + (2/k)I_p]^{-1}\beta \leq \sigma^2$$

In terms of eigenvalue decomposition,

if and only if $$\beta' G(\Gamma^{-1} + \frac{2}{k} I_p)^{-1} G' \beta \leq \sigma^2$$

if and only if $$\alpha' \frac{k}{2} [I_p - \mathrm{Diag}(\frac{k}{k + 2\lambda_i})] \alpha \leq \sigma^2$$

Since $I_p - \mathrm{Diag}(\frac{k}{k + \lambda_i})$ has positive diagonal elements less than 1,

$$\alpha' \frac{k}{2} I_p \alpha \leq \sigma^2 \quad \text{implies} \quad \frac{k}{2} \alpha' \alpha \leq \sigma^2$$

Hence $$k \leq \frac{2\sigma^2}{\alpha' \alpha} = \frac{2\sigma^2}{\beta' \beta} = k_{max}$$

Since k_{max} must be strictly positive for all σ and β provided that $\beta' \beta$ is bounded, this proves the existence theorem w.r.t MSE criterion.

Ⅳ. The Estimation Procedure of Biasing Parameter

In the previous section, the existence of k satisfying MSE criterion is proved by showing that $(0, k_{max})$ is non-empty. However, the range $(0, k_{max})$ contains infinite number of values of k, it is difficult to find the appropriate value of k in the acceptable range $(0, k_{max})$. In this section, we derive the analytic solution of k to the generalized *ridge* estimator.

From equation (2.5), the MSE of the i^{th} component of b_K is

$$\mathrm{MSE}\ (b_{K,i}) = \sigma^2 \delta_i^2 \lambda_i^{-1} + \alpha_i^2 (\delta_i - 1)^2 \qquad (4.1)$$

The necessary condition for a minimum of (4.1) requires that its derivative with respect to δ_i is to be zero. Then we obtain the minimum MSE value of δ_i as

$$\delta_i^M = \alpha_i^2 (\sigma^2 \lambda_i^{-1} + \alpha_i^2)^{-1} \qquad (4.2)$$

and the corresponding MSE value of k is

$$k_i^M = \sigma^2 \alpha_i^{-2}$$

Now starting with the initial value as $k_i^{(0)} = s^2(a_i)^{-2}$ where s^2 is the unbiased estimator of σ^2 and $\delta_i^{(0)} = F_i/(1 + F_i)$ where $F_i = \lambda_i(a_i)^2/s^2$, a supscript (j), j = 0,1,2..., is number of iterates, we obtain that the generalized *ridge* estimator $a_{K,i}^{(0)}$ of α_i is now $\delta_i^{(0)} a_i$.

At j-th iteration, we may consider the following sequences

$$\langle \delta_i^{(0)}, \delta_i^{(1)} = \frac{F_i}{(\delta_i^{(0)})^{-2} + F_i}, \ldots, \delta_i^{(j+1)} = \frac{F_i}{(\delta_i^{(j)})^{-2} + F_i}, \ldots \rangle \quad (4.3)$$

and the corresponding generalized *ridge* estimaotr is

$$\langle a_{K,i}^{(0)}, a_{K,i}^{(1)} = \delta_i^{(0)} a_{K,i}^{(0)}, \ldots, a_{K,i}^{(j+1)} = \delta_i^{(j)} a_{K,i}^{(j)}, \ldots \rangle \quad (4.4)$$

For terminating the sequence $\langle \delta_i^{(j+1)} \rangle$, the difference $\delta_i^{(j+1)} - \delta_i^{(j)}$ must be monotonically decreasing. That is, the derivative of $\delta_i^{(j+1)} - \delta_i^{(j)}$ w.r.t $\delta_i^{(j)}$ must be negative. Then $\delta_i^{(j)}$ satisfies the following equations.

$$2F_i(\delta_i^{(j)})^{-3} < F_i^2 + 2F_i(\delta_i^{(j)})^{-2} + (\delta_i^{(j)})^{-4} \quad (4.5)$$

Mutiplying by $(\delta_i^{(j)})^4$, F_i^{-2},

$$2F_i^{-1} \delta_i^{(j)} < (\delta_i^{(j)})^4 + 2F_i(\delta_i^{(j)})^2 + F_i^{-2} \quad (4.6)$$

Hence we have the following convergence condition.

<u>Theorem 4.1</u> The sequence in (4.3) converges to δ_i^* when the inequality holds.

$$2F_i^{-1} \delta_i^{(j)} < [(\delta_i^{(j)})^2 + F_i^{-1}]^2 \quad (4.7)$$

where $\delta_i^* = \frac{1}{2} \pm (\frac{1}{4} - F_i^{-1})^{1/2}$

Note that δ_i^* depends upon F_i whether $F_i \leq 4$ or $F_i > 4$

<u>Case 1</u> $F_i \leq 4$

Since the values of δ_i^* in <u>Theorem 4.1</u> are imaginary, omitting the imaginary part, δ_i^* equals to ½. But this solution does not satisfy the inequality in (4.7). In this case we will use δ_i as 0.

Case 2 $F_i > 4$

Although δ_i^* has two solution with negative and positive part, the solution with negative part does not satisfy the convergence condition in Theorem 4.1. Hence we have δ_i^* as $\frac{1}{2} + (\frac{1}{4} - F_i^{-1})^{1/2}$.

Hence we obtain the following results.

Theorem 4.2 The generalized *ridge* estimator $\alpha_K = (\alpha_{K,i})$ is written by

$$\alpha_{K,i} = \begin{cases} 0 & \text{if } F_i \leq 4 \\ [\frac{1}{2} + (\frac{1}{4} - F_i^{-1})^{1/2}]\alpha_i & \text{if } F_i > 4 \end{cases}$$

V. Concluding Remark

We have examined some properties of the *ridge* estimations in the linear regression model with multicollinearity. Inparticular, we have investigated bias and mean squared error of *ridge* estimator based on generalized least squares estimator. Also in addition to Hoerl & Kennard's *ridge* estimator based on ordinary least squares estimator, our *ridge* estimator emerging from this studies have considerable evidence indicating the superiority than generalized least squares estimator if multicollinearity is present.

Even though several rules for choosing k are proposed by many authors, thses rules are intended to aid an investigator confronted with a specfic regression problem to arrive at acceptable choice of k. But there is no known optimally mathmatical method of explicitly determining the value k in a given problem. That is, a Monte Carlo simulation failed to show any obvious superiority among these methods.

Also, in practice V_n is not known, one may have an approximation V_n^* to V_n and perhaps a reasonble bound on the departure of V_n^* from V_n. If one use V_n^* instead of V_n, we will usually incur an error in the estimated coefficient vector. Hence the explicit method of determining the approximation V_n^* is necessary in a given problem.

R E F E R E N C E S

1. Rao, C.R. "Linear Statistical Inference and Its Application", New York Wiley, 1965
2. Hoerl, A.E. & Kennard, R.W. "Ridge Regression : Biased Estimation for Nonorthogonal Problem", Technometrices 12, pp. 55-67, 1982
3. Theobald, C.M. "Generalization of Mean Squared Error Applied to Ridge Regression", J.Roy.Stat. Soc. Ser. B36, pp. 103-106, 1974
4. Lawless, J.F. and Wang, P. "A Simulation Study of Ridge and Other Regression Estimators", Comm. Stat. 5, pp. 307-323, 1976
5. Richard, F. Gunst and Robert, L. Mason, "Biased Estimation in Regression : An Evaluation Using Mean Squared Error", J.A.S.A. 72, pp. 616-626, 1977
6. Richard, F. Gunst and Robert, L. Mason, "Regression Analysis and Its Application : A Data-Oriented Approach", Marcel Dekker, Inc. N.Y. and Basel, 1980
7. Hrishikesh, D. Vinod and Aman Ullah, "Recent Advances in Regression Methods", Marcel Dekker, Inc. New York and Basel, 1981
8. Douglas, C. Montgomery and Elizabeth, A. Peck, "Introduction to Linear Regression Analysis", John Wiley & Sons, Inc. N.Y., 1982