HONAM MATHEMATICAL JOURNAL
Volume 9, Number 1. August 1987

On the Ridge Estimations with the Corrlated Error Structure

by

Byung Chool Won and Hae Kyung Kim

Department of Mathematics, Yonsei University, Seoul 120, Korea

Abstract

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.

1. Introduction

Consider the generalized linear model

$$Y = X\beta + \varepsilon \quad \cdots\cdots(1.1)$$

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

$$E(\varepsilon) = 0 \text{ and } E(\varepsilon\varepsilon') = \sigma^2 V_n \quad \cdots\cdots(1.2)$$

The classical estimation procedure for the generalized linear model is that of generalized least squares(GLS) in which $\hat{\beta}_G$ is chosen such that the residual sum of squares $\phi(\hat{\beta}_G) = (Y - X\hat{\beta}_G)' V_n^{-1} (Y - X\hat{\beta}_G)$ is minimized. The minimization methodology results in the well-known normal equations $X'V_n^{-1}X\hat{\beta}_G = X'V_n^{-1}Y$ which must be solved for $\hat{\beta}_G$. In the case that $X'V_n^{-1}X$ is of full rank, $(X'V_n^{-1}X)^{-1}$ exists and the GLS estimatora are given by

$$\hat{\beta}_G = (X'V_n^{-1}X)^{-1} X'V_n^{-1}Y \quad \cdots\cdots(1.3)$$

Received April 30, 1987, Accepted June 22, 1987.

However, in the case that $X'V_n^{-1}X$ is of rank $q<p$, alternate methods must be employed to obtain estimators. One such method is that which employs a generalized inverse. 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-Markoff Theorem assures us that in the class of all unbiased estimators, the GLS estimator of estimable functions have minimum variance; we are faced with the unhappy circumstances and, hence produce a large confidence intevals for the estimator.

One way to remedy this problem is to drop the requirement that the estimator of β is unbiased. Hoerl and kennard(1970) have suggested that the ordinary least squares estimator may be replaced by the ridge estimator $\hat{\beta}_G(k)$ with biasing parameter $k>0$. But Hoerl and kennard's ridge estimator is derived from the assumption usually made concerning the linear regression model with 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 aquared error criterion of an estimator, the ridge-type regression estimator could be shown to be superior to the GLS estimator.

2. The Form of Ridge Estimator

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

$$y=X^*\alpha+\varepsilon. \qquad (2.1)$$

Then the GLS estimator of α is given by

$$\begin{aligned}\hat{\alpha}_G&=(X^{*\prime}V_n^{-1}X^*)^{-1}X^{*\prime}V_n^{-1}y\\&=\Lambda^{-1}G'X'V_n^{-1}y \qquad (2.2)\end{aligned}$$

The variance of $\hat{\alpha}_G$ is then given by

$$\begin{aligned}\mathrm{Var}(\hat{\alpha}_G)&=\mathrm{Var}(\Lambda^{-1}G'X'V_n^{-1}y)\\&=\sigma^2\Lambda^{-1}.\end{aligned}$$

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

In this section, we consider some biased estimator which is a type of ridge estimator.

$$\text{Let } G=(G_1, G_2, \ldots, G_p) \text{ and } \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0.$$

Then G_i and λ_i satisfy the following equation.

$$(X'V_n^{-1}X)G_i = \lambda_i G_i$$

This equation implies

$$\begin{aligned}(X'V_n^{-1}X)G &= ((X'V_n^{-1}X)G_1, \ldots, (X'V_n^{-1}X)G_p)\\ &= (\lambda_1 G_1, \lambda_2 G_2, \ldots, \lambda_p G_p)\\ &= G\Lambda.\end{aligned}$$

From this expression, we obtain

$$\begin{aligned}X'V_n^{-1}X &= G\Lambda G'\\ &= \sum_{i=1}^{p} \lambda_i G_i G_i'.\end{aligned}$$

By the definition of matrix inverse,

$$(X'V_n^{-1}X)^{-1} = \sum_{i=1}^{p} \lambda_i^{-1} G_i G_i'$$

Hence we obtain

$$(X'V_n^{-1}X)^{-1} X'V_n^{-1}y = \sum_{i=1}^{p} \lambda_i^{-1} G_i G_i' X'V_n^{-1}y$$

This leads to the following result.

Theorem 2.1. The GLS estimator $\hat{\beta}_G$ is of the form

$$\hat{\beta}_G = \sum_{i=1}^{p} \lambda_i^{-1} c_i G_i, \quad \text{where } c_i = G_i' X' V_n^{-1} y \cdots\cdots\cdots\cdots\cdots\cdots (2.3)$$

The ridge type estimator is given by

$$\hat{\alpha}_G(k) = (\Lambda+K)^{-1}G'X'V_n^{-1}y \quad \cdots\cdots (2.4)$$

where $K = \text{Diag}(k_i)$ is a diagonal matrix of biasing parameter $k_i>0$. In terms of the original model, the generalized ridge estimator is defined by as followings

$$\begin{aligned}\hat{\beta}_G(K) &= G\hat{\alpha}_G(k) \\ &= G(\Lambda+K)^{-1}G'X'V_n^{-1}y \\ &= (X'V_n^{-1}X+GKG')^{-1}X'V_n^{-1}y \quad \cdots\cdots (2.5)\end{aligned}$$

Consequently from above, we obtain the following result.

Theorem 2.2. The generalized ridge estimator $\hat{\beta}_G(K)$ is of the form

$$\hat{\beta}_G(K) = \sum_{i=1}^{p}(\lambda_i+k_i)^{-1}c_iG_i$$

where $\quad c_i = G_i'X'V_n^{-1}y.$

Alternatively from the second equality of (2.5)

$$\begin{aligned}\hat{\beta}_G(K) &= G(\Lambda+K)^{-1}\Lambda G'G\Lambda^{-1}G'X'V_n^{-1}y \\ &= G\triangle G'\hat{\beta}_G \quad \cdots\cdots (2.6)\end{aligned}$$

where
$$\begin{aligned}\triangle &= (\Lambda+K)^{-1}\Lambda \\ &= \text{Diag}(\delta_i);\ \delta_i=\lambda_i(\lambda_i+k_i)^{-1},\ i=1,2,\ldots,p.\end{aligned}$$

is a diagonal matrix of shrinkage factors.

From the second equality of (2.6),

$$\begin{aligned}E(\hat{\beta}_G(K))-\beta &= G\triangle G'\beta-\beta \\ &= G(\triangle-I_p)G'\beta \quad \cdots\cdots (2.7)\end{aligned}$$

Theorem 2.3. $\hat{\beta}_G(K)$ is a biased estimator and

$$\text{Bias}(\hat{\beta}_G(k)) = -k(X'V_n^{-1}X+kI_p)^{-1} \quad \text{when all } k_i=k.$$

Proof. $$\begin{aligned}\mathrm{Diag}(\delta_i-1)&=\mathrm{Diag}\left(\frac{-k}{\lambda_i+k}\right)\\&=-k\ \mathrm{Diag}(\lambda_i+k)^{-1}\\&=-k(\mathrm{Diag}(\lambda_i)+kI_p)^{-1}\\&=-k(\Lambda+kI_p)^{-1}\end{aligned}$$

From the equality of (2.7),

$$\begin{aligned}\mathrm{Bias}(\hat{\beta}_G(k))&=-kG(\Lambda+kI_p)^{-1}G'\\&=-k(G\Lambda G'+kI_p)^{-1}\\&=-k(X'V_n^{-1}X+kI_p)^{-1}.///\end{aligned}$$

3. Mean Squared Error Comparison

Now in this section, we concentrate attention on the mean squared error(MSE) of the proposed estimators, where the MSE of an estimator $\hat{\beta}$ is defined by

$$MSE(\hat{\beta})=E(\hat{\beta}-\beta)'(\hat{\beta}-\beta) \qquad\cdots(3.1)$$

A frequent criticism of the use of the mean squared error as defined in (3.1) to measure the adequacy of an estimator is that it is only one member of a general class of measures. If the comparisons of the mean squared error are to be made on the basis of a single function, it may be possible to form a suitable weight sum of coefficient mean squared error or, more generally, to compare the values of

$$GMSE(\hat{\beta})=E(\hat{\beta}-\beta)'B(\hat{\beta}-\beta) \qquad\cdots(3.2)$$

where B is a non-negative definite matrix.

The purpose of this section is to establish a existence of a single biasing parameter k such that the MSE of the ridge estimator is less than the MSE of the GLS estimator. To show this, the follwing MSE matrix will be used. i.e.

$$MtxMSE(\hat{\beta})=E(\hat{\beta}-\beta)(\hat{\beta}-\beta)' \qquad\cdots(3.3)$$

Theobald(1974) showed that there exists a range of values of k such that $GMSE(\hat{\beta}_R) < GMSE(\hat{\beta}_{LS})$, where $\hat{\beta}_{LS}$ is the ordinary least squares estimator and $\hat{\beta}_R$ is the Hoerl and Kennard's ridge estimator using $\hat{\beta}_{LS}$. In this section, we will show that the ridge estimator of this thesis also has the Theobald's result. To show this, the following Lemmas are necessary. Let b_1 and b_2 be an estimator of β.

Lemma 3.1. The followings are equivalent.

i) $MtxMSE(b_1)-MtxMSE(b_2)$ is non-negative definite.

ii) $GMSE(b_1)-GMSE(b_2)>0$, for all non-negative definite B.

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

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

$$\sup_{x'\in E_n}\frac{(u'x)^2}{x'Ax}=u'A^{-1}u.$$

and supremum is attained at $x=A^{-1}u$.

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

For $K=kI_p$, we obtain

$$\mathrm{Var}(\hat{\beta}_G(k))=\sigma^2(x'V_n^{-1}X+kI_p)^{-1}X'V_n^{-1}X(X'V_n^{-1}+kI_p)^{-1}$$

and

$$\mathrm{Bias}(\hat{\beta}_G(k))=G\ \mathrm{Diag}(\delta_i-1)G'$$

The mean squared error of $\hat{\beta}_G(k)$ therefore is

$$\begin{aligned}MSE(\hat{\beta}_G(k))&=tr[\mathrm{Var}(\hat{\beta}_G(k))]+[\mathrm{Bias}(\hat{\beta}_G(k))]'[\mathrm{Bias}(\hat{\beta}_G(k))]\\&=\sigma^2tr[(X'V_n^{-1}X+kI_p)^{-1}X'V_n^{-1}X(X'V_n^{-1}X+kI_p)^{-1}]\\&\quad+\alpha'\,\mathrm{Diag}(\delta_i-1)^2\alpha.\\&=\sigma^2\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i+k)^2}+\sum_{i=1}^{p}\alpha_i^2\left(\frac{\lambda_i}{\lambda_i+k}-1\right)^2.\quad\cdots(3.4)\end{aligned}$$

The first term on the right-hand side of (3.4) is the sum of variances of the parameter in $\hat{\beta}_G(k)$ and the second term is the square of the bias. If $k>0$, note that the bias in $\hat{\beta}_G(k)$ increases with k. However, the variance decreases as k increases.

In using ridge regression, we would like to choose a value of k 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 $\hat{\beta}_G(k)$ will be less than that of generalized least squares estimator $\hat{\beta}_G$. Next theorem with Lemma 3.1 says that there exists a non-zero k for which $GMSE(\hat{\beta}_G(k))$ is less than the variance of $\hat{\beta}_G$.

Theorem 3.3. There exists a $k_{max}>0$ such that $MtxMSE(\hat{\beta}_G)-MtxMSE(\hat{\beta}_G(k))$ is non-negative definite where $0<k<k_{max}$.

Proof. Let $\nabla=MtxMSE(\hat{\beta}_G)-MtxMSE(\hat{\beta}_G(k))$ and η be any non-zero column-vector. Then

$$\begin{aligned}\nabla=&\sigma^2(X'V_n^{-1}X)^{-1}-\sigma^2(X'V_n^{-1}X+kI_p)^{-1}X'V_n^{-1}X(X'V_n^{-1}X+kI_p)^{-1}\\&-k^2(X'V_n^{-1}X+kI_p)^{-1}\beta\beta'(X'V_n^{-1}X+kI_p)^{-1}.\end{aligned}$$

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

$$\begin{aligned}\nabla&=\sigma^2\xi\Big[k^2(X'V_n^{-1}X+kI_p)(X'V_n^{-1}X)^{-1}(X'V_n^{-1}X+kI_p)-X'V_n^{-1}X-k^2\frac{\beta\beta'}{\sigma^2}\Big]\xi\\&=\sigma^2\xi\Big[k^2(X'V_n^{-1}X)^{-1}+2kI_p-k^2\frac{\beta\beta'}{\sigma^2}\Big]\xi.\end{aligned}$$

If ∇ is non-negative definite, then

$$\begin{aligned}&\eta'\nabla\eta\geq 0,\ \text{for}\ \eta\neq 0.\\ \text{iff}\quad&\sigma^2\xi\Big[k^2(X'V_n^{-1}X)^{-1}+2kI_p-k^2\frac{\beta\beta'}{\sigma^2}\Big]\xi\geq 0\\ \text{iff}\quad&\theta=\frac{k^2}{\sigma^2\xi'[k^2(X'V_n^{-1}X)^{-1}+2kI_p]\xi}\leq 1\end{aligned}$$

Now $\sup_{\xi}\theta\leq 1$ implies $\theta\leq 1$, for $\xi\neq 0$. Set $A=k^2(X'V_n^{-1}X)^{-1}+2kI_p$, $x=\xi$, and $u=\beta$. Then by Lemma 3.2,

$$\begin{aligned}&\sup_{\xi}\theta=\frac{1}{\sigma^2}k\beta'\Big[k^2(X'V_n^{-1}X)^{-1}+2kI_p\Big]^{-1}k\beta\leq 1.\\ &\text{implies}\quad \beta'\Big[(X'V_n^{-1}X)^{-1}+\frac{2}{k}I_p\Big]^{-1}\beta\leq\sigma^2.\end{aligned}$$

In terms of eigenvalue decomposition,

$$\begin{aligned}&\beta'(G\Lambda^{-1}G+\frac{2}{k}GI_pG')^{-1}\beta\\ &\text{implies}\quad \beta'G(\Lambda^{-1}+\frac{2}{k}I_p)^{-1}G'\beta\leq\sigma^2.\end{aligned}$$

After simplification, we get

$$\alpha'\frac{2}{k}\left(I_p-\text{Diag}\left(\frac{k}{k+2\lambda_i}\right)\right)\alpha\leq\sigma^2.$$

Since $I_p-\text{Diag}\left(\frac{k}{k+2\lambda_i}\right)$ 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 $$k\leq\frac{2\sigma^2}{\alpha'\alpha}=\frac{2\sigma^2}{\beta'\beta}.$$

Thus $$k_{max}=\frac{2\sigma^2}{\beta'\beta}.///$$

Note that $MtxMSE(\hat{\beta})=\text{Var}(\hat{\beta})+[\text{Bais}(\hat{\beta})][\text{Bias}(\hat{\beta})]'$.

For $K=kI_p$, by theorem 3.3, $GMSE(\hat{\beta}_G(k))$ is smaller than $GMSE(\hat{\beta}_G)$. Since k_{max} must be strictly positive for all σ^2 and β provided that $\beta'\beta$ is bounded, this proves the existence theorem for the most $MtxMSE$ criterion, and without any reference to multicollinearity. The existence theorem is true when multicollinearity is absent. However, in this cases there will be very little scope for reducing $MSE(\hat{\beta}_G,)$ and the positive k will be very close to zero. Indeed, the range $(0,\ k_{max})$ contains an infinitely number of values of k, hence to find a value of k is difficult if not im-possible.

4. Thc estimation procedure of biasing parameter

Let us define an acceptable range of k where in

$$GMSE(\hat{\beta}_G(k))<GMSE(\hat{\beta}_G) \quad\cdots\cdots(4.1)$$

In the previous section, the existence of k satisfying (4.1) is proved by showing that the acceptable range is non-empty. However, it is difficult to find the appropriate value of k in the acceptable range. In this section, first we describe the several selection rules of k that appears in literature and derive the Hemmerle's analytic solution to the generalized ridge estimator.

A. Hoerl and Kennard(1970) have suggested that an appropriate value of k may be determined by the inspection of ridge trace. The ridge trace is a plot of the element of $\hat{\beta}_G(k)$ versus k in the acceptable range. Its objective is to select a reasonably small values of k at which the ridge estimator $\hat{\beta}_G(k)$ is stable. However, choosing k by inspection of ridge trace is a subjective procedure requiring judgement on the part of the analyst, and it is difficiult for dealing with the generalized ridge estimator.

B. To find more analytical solution, Hoerl, Kennard and Baldwin(1975) have suggested that an appropriate choice for k is

$$\hat{k}=\frac{p\hat{\sigma}^2}{\hat{\beta}_G'\cdot\hat{\beta}_G} \quad\cdots\cdots (4.2)$$

where $\hat{\sigma}$ and $\hat{\beta}_G$ are founded from the generalized least squares solution. Note that $\hat{k}$ is the harmonic mean of k_i in (4.4).

C. In a subsequent paper, Hoerl and Kennard(197) proposed an iterative estimation procedure based upon (4.2). Specifically, they deduced the algorithm for terminating the sequence. Since this method is similar to that of Hemmerle's, we explain the Hemmerle' method in a later part of this section.

D. From equation (3.1) and (3.4),

$$E(\hat{\beta}_G-\beta)'(\hat{\beta}_G-\beta)=\sigma^2\sum_{i=1}^{p}\lambda_i^{-1}$$

and

$$E(\hat{\beta}_G'\hat{\beta}_G)=\beta'\beta+\sigma^2\sum_{i=1}^{p}\lambda_i^{-1}$$

If we put $Q=\hat{\beta}_G\hat{\beta}_G-\hat{\sigma}^2\sum_{i=1}^{p}\lambda_i^{-1}$, then Q is an unbiased estimator of $\beta\beta'$. MacDonald and Galarneau(1975) find a value of k such that the squared length of $\hat{\beta}_G(k)$ is an unbised estimator of $\beta\beta'$. That is,

If $Q>0$, choose k such that $\hat{\beta}_G'(k)\cdot\hat{\beta}_G(k)=Q$.
If $Q\leq 0$, choose k such that $k=0$ (or ∞).

Note that $\hat{\beta}_G(0)$ is the GLS estimator and $\hat{\beta}_G(\infty)$ is the zero vector.

We have defined the acceptable range of k in (4.1) which is relevant only for the ordinary ridge estimator $\hat{\beta}_G(k)$. For the generalized ridge estimator $\hat{\beta}_G(K)$, there, in general, a separate range for which k_i given by $0<k_i<k_{max,i}$.

From equation (3.4), note that

$$MSE(\hat{\beta}_G(K))=\sigma^2\sum_{i=1}^{p}\delta_i^2\lambda_i^{-1}+\sum_{i=1}^{p}\alpha_i^2(\delta_i-1)^2$$

where $\delta_i=(\lambda_i+k_i)^{-1}\lambda_i$.

Hopefully we minimize the MSE of $\hat{\beta}_G(K)$, and the MSE of the i-th component of $\hat{\beta}_G(K)$ is

$$MSE(\hat{\beta}_{G,i}(K))=\sigma^2\delta_i^2\lambda_i^{-1}+\alpha_i^2(\delta_i-1)^2. \qquad (4.3)$$

The necessary condition for a minimum of (4.3) requires that its derivatives with respet to δ_i be zero. Then we obtain the minimum MSE values of δ_i is

$$\delta_i^M=\frac{\lambda_i}{\sigma^2\alpha_i^{-2}+\lambda_i}.$$

From the definition of δ_i, the corresponding MSE values of k_i is

$$k_i^M=\frac{\sigma^2}{\alpha_i^2}.$$

Since k_i involves unknown parameter σ, α_i, we use k_i as an estimator

$$\hat{k}_i=\frac{\sigma^2}{\hat{\alpha}^2_{G,i}}. \qquad (4.4)$$

Hence, we obtain

$$\delta_i=\frac{F_i}{1+F_i}. \qquad (4.5)$$

where $F_i=\lambda_i\hat{\alpha}_{G,i}^2\sigma^{-2}$.

From the equation (2.7) and (4.5),

$$\hat{\alpha}_{G,i}(k)=\hat{\delta}_i\hat{\alpha}_{G,i}. \qquad (4.6)$$

Consider the following sequence of an estimator of α_i, δ_i,

$$\begin{array}{ll}
\hat{\alpha}_{G,i}, & \hat{\delta}_i^{(0)}=\dfrac{F_i}{1+F_i} \\
\hat{\delta}_i^{(0)}\hat{\alpha}_{G,i}, & \hat{\delta}_i^{(1)}=\dfrac{F_i}{(\hat{\delta}_i^{(0)-2})+F_i} \\
\vdots & \vdots \\
\hat{\delta}_i^{(j)}\cdot\hat{\alpha}_{G,i} & \hat{\delta}_i^{(j+1)}=\dfrac{F_i}{(\hat{\delta}_i^{(j)-2})+F_i} \\
\vdots & \vdots
\end{array}$$

In order to have algorithmic estimation, a procedure must be given for terminating the sequence. If the sequence $\hat{\delta}_i^{(j)}$ converges, the converging solution is known to solve the equation

$$\hat{\delta}_i^* = \frac{F_i}{F_i + (\hat{\delta}_i^*)^{-2}}$$

That is, $\hat{\delta}_i^{*2} - \hat{\delta}_i^* + F_i^{-1} = 0.$ ··························· (4.7)

This leads to a following result.

Theorem 4.1. Tha i-th component of $\hat{\alpha}_G(K)$ is defined by

$$\hat{\alpha}_{G,i}(k) = \hat{\delta}_i^* \hat{\alpha}_{G,i}$$

where
$$\hat{\delta}_i^* = \frac{1}{2} + \left(\frac{1}{4} - F_i^{-1}\right)^{\frac{1}{2}} \text{ or,} \qquad (4.7\text{-}1)$$
$$\hat{\delta}_i^* = \frac{1}{2} - \left(\frac{1}{2} - F_i^{-1}\right)^{\frac{1}{2}}. \qquad (4.7\text{-}2)$$

Next, consider the stability properties of $\hat{\delta}_i^*$. It seems reasonable to use the relative difference $[\hat{\delta}_i^{(j)}]$. For terminating the sequence $[\hat{\delta}_i^{(j)}]$, the relative difference $|\hat{\delta}_i^{(j+1)} - \hat{\delta}_i^{(j)}|$ must be monotonically decreasing. That is, the derivative of $|\hat{\delta}_i^{(j+1)} - \hat{\delta}_i^{(j)}|$ with respect to to $\hat{\delta}_i^{(j)}$ must be negative. Then $\hat{\delta}_i^{(j)}$ satisfies the following equation

$$2F_i(\hat{\delta}_i^{(j)-3}) < F_i^{-1} + 2F_i(\hat{\delta}_i^{(j)-2} + (\hat{\delta}_i^{(j)-4}).$$

After simplification,

$$2F_i^{-1}\hat{\delta}_i^{(j)} < (\hat{\delta}_i^{(j)})^4 + 2F_i^{-1}(\hat{\delta}_i^{(j)})^2 + F_i^{-2} \qquad (4.8)$$

whire we multiply by $(\hat{\delta}_i^{(j)})^4$ and F_i^{-2}. The inequality (4.8) simplifies to the following convergence condition

$$2F_i^{-1}\hat{\delta}_i^{(j)} < [(\hat{\delta}_i^{(j)})^2 + F_i^{-1}]^2 \qquad (4.9)$$

Hence we obtain the further result.

Theorem 4.2. The $\hat{\delta}_i^*$ in Theorem 4.1 satisfies the followings

$$2F_i^{-1}\hat{\delta}_i^* < [(\hat{\delta}_i^*)^2 + F_i^{-1}]^2.$$

Note that the value of $\hat{\delta}_i^*$ depond upon F_i whether $F_i>4$ or $F_i\leq 4$.

i) If $F_i\leq 4$, then both of the values of $\hat{\delta}_i^*$ are imaginary. Omitting the imaginary part, $\hat{\delta}_i^*$ equals to $\frac{1}{2}$. But this solution does not satisfy Theorem 4.2. In this case, we will use $\hat{\delta}_i^*$ as zero.

ii) If $F_i>4$, then $\hat{\delta}_i^*$ has two solution. But, the solution (4.7-2) does not satisfies Theorem 4.2, hence we obtain the solution (4.7-1). Consequently, from Theorem 4.1 and Theorem4.2, the following result is obtained.

Theorem 4.3. The i-th component of $\hat{\alpha}_G(K)$ is written by

$$\hat{\alpha}_{G,i}(K)=\begin{cases} 0 & \text{if } F_i\leq 4. \\ \left[\frac{1}{2}+\left(\frac{1}{4}-F_i^{-1}\right)^{\frac{1}{2}}\right]\hat{\alpha}_{G,i} & \text{if } F_i>4. \end{cases}$$

5. Conclusions.

A number of conclusions emerge from this study. A main one is that, with respect to the mean squared error criterion, ridge regression estimators are seen to be excellent. In particular, we have investigated the mean squared error of ridge-type estimator based on generalized least squares estimator.

Even though several rules for choosing k are proposed here, these rules are intended to aid an investigator confronted with a specific regression problem to arrive at an acceptable choice of k. There is no known mathematical method of explicitly determining the value of k in a given problem.

A final main concern is that our ridge-type estimator is achieved when the residuals are weighted in accordance with V_n^{-1}. However, in practice V_n is not known, but one only has an approximation $\hat{V}_n$ to V_n and perhaps a reasonable bound on the departure of $\hat{V}_n$ from V_n. If one use $\hat{V}_n$ instead of V_n, one will ordinarily incur an error in the estimated coefficient vector. Hence the explicit method of determining the approximation $\hat{V}_n$ is necessary in a given problem.

References

1. Hemmerle, W.J. and Carey, M.B. (1983), Some properties of generalized ridge estimators, *Comm. Statist.*, **12**, 239~253.

2. Hoerl, A.E. & Kennard, R.W. (1970), Ridge regression: Biased estimation for nonorthogonal problem, *Technometrices.*, ***12***, 55~67.
3. Hoerl, A.E., Kennard, R.W., and Baldwin, K.F., (1976), Ridge regression: Some simulations, *Comm. Statist.*, ***4***, 105~123.
4. Madonald, G.G., and Galarneau, D.I. (1975), A Monte Carlo evaluion of some ridge-type estimators, *J.A.S.A.*, ***15***, 463~481.
5. Rao, C.R. (1965), *Linear statistical infernce and its application*, New York, Wiley.
6. Theobald, C.M. (1974), Generalization of mean squared error applied to ridge regression, *J. Roy. Statist.*, Soc. Ser. B36, 103~106.