THE MSE OF AN ADAPTIVE RIDGE ESTIMATOR IN A LINEAR REGRESSION MODEL WITH SPHERICALLY SYMMETRIC ERROR

Kazumasa Mori and Hiroshi Kurata

Received November 29, 2009; revised March 8, 2010

Abstract. This paper considers a linear regression model with possible multicollinearity. When the matrix $\mathbf{A}^t\mathbf{A}$ is nearly singular, the least squares estimator (LSE) gets unstable. Typical solutions for this problem include the generalized ridge estimator due to Hoerl and Kennard(1970a,b) and its derivatives. Among them, we focus on an adaptive ridge estimator discussed by Wang and Chow(1990) under normality. We assume the error term $\mathbf{e}$ is distributied as a spherically symmetric distributiuon and derive a sufficient condition so that the estimator is superior to the LSE under mean squared error (MSE) and quadratic loss. Several numerical examples are also given.

1 Introduction Let us consider the following linear regression model

(1.1) $\mathbf{y} = \mathbf{A}\boldsymbol{\beta} + \mathbf{e} \quad \text{with} \quad \mathrm{E}(\mathbf{e}) = \mathbf{0} \quad \text{and} \quad \mathrm{V}(\mathbf{e}) = \sigma^2 I,$

where $\mathbf{y} : n \times 1$, $\mathbf{A} : n \times p$, rank $\mathbf{A} = p$, $\boldsymbol{\beta} : p \times 1$, and $\mathbf{e}$ is an vector of random errors. By the Gauss-Markov theorem, the least squares estimator (LSE)

(1.2) $\hat{\boldsymbol{\beta}} = (\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t\mathbf{y}$

is the best linear unbiased estimator of $\boldsymbol{\beta}$, which has the covariance matrix

$$\mathrm{V}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{A}^t\mathbf{A})^{-1}.$$

When the column vectors of $\mathbf{A}$ are approximately linearly dependent, which is often the case in practice, the matrix $\mathbf{A}^t\mathbf{A}$ is nearly singular and the estimate is unstable. This problem is known as the one of multicollinearity. Various estimators that modify the LSE have been proposed so far from various points of view. A typical example is the generalized ridge estimator due to Hoerl and Kennard (1970a,b), which modifies the LSE by replacing $\mathbf{A}^t\mathbf{A}$ with a more stable matrix. To state it precisely, let a spectral decomposition of $\mathbf{A}^t\mathbf{A}$ be

$$\mathbf{A}^t\mathbf{A} = \Phi\Lambda\Phi^t,$$

where Φ is a $p \times p$ orthogonal matrix and

$$\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_p) \quad \text{with} \quad \lambda_1 \geq \cdots \geq \lambda_p.$$

Then the generalized ridge estimator can be written as

(1.3) $\hat{\boldsymbol{\beta}}(\mathbf{K}) = (\mathbf{A}^t\mathbf{A} + \Phi\mathbf{K}\Phi^t)^{-1}\mathbf{A}^t\mathbf{y},$

2000 *Mathematics Subject Classification.* 62J07, 62C20.
Key words and phrases. ridge estimator, shperically symmetric distribution, mean squared error, quadratic loss,minimax estimator.

where $\mathbf{K} = \mathrm{diag}(k_1, \cdots, k_p)$ is a nonrandom diagonal matrix suitably chosen. The simplest choice for $\mathbf{K}$ is $\mathbf{K} = k\mathbf{I}$ ($k > 0$). This choice yields the (original) ridge estimator $\hat{\boldsymbol{\beta}}(k\mathbf{I}) = (\mathbf{A}^t\mathbf{A} + k\mathbf{I})^{-1}\mathbf{A}^t\mathbf{y}$, which has been widely employed in applications. While the generalized ridge estimator thus defined is biased unless $\mathbf{K} = \mathbf{O}$, it can be superior to the LSE in terms of mean squared error (MSE), which is a typical criterion. The MSE of an estimator $\mathbf{b}$ is defined as $\mathrm{E}[(\mathbf{b} - \boldsymbol{\beta})^t(\mathbf{b} - \boldsymbol{\beta})]$. In fact, the MSE of $\hat{\boldsymbol{\beta}}(\mathbf{K})$ attains its minimum at

$$k_i = \sigma^2/\beta_i^2, \quad i = 1, \cdots, p,$$

which is smaller than that of the LSE. However, in most cases, the quantities σ^2/β_i^2 are unknown and $\hat{\boldsymbol{\beta}}(\mathbf{K})$ is not feasible.

Many authors have proposed feasible versions of $\hat{\boldsymbol{\beta}}(\mathbf{K})$ by replacing $\mathbf{K}$ with an appropriate estimator. Among others, the estimator studied by Vinod and Ullah (1980) and Ullah Vinod and Kadiyala (1980) is basic, which is defined as $\hat{\boldsymbol{\beta}}(\mathbf{K})$ with

$$(1.4)\quad k_i = \frac{f_1\hat{\sigma}^2}{\hat{\beta}_i^2 - f_2\hat{\sigma}^2/\lambda_i} = \tilde{k}_{i(f_1,f_2)}$$

and is called an adaptive generalized ridge estimator. Here, $\hat{\beta}_i$ is the i-th element of the LSE $\hat{\boldsymbol{\beta}}$, f_i's are nonrandom constants and

$$\hat{\sigma}^2 = ||\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}}||^2/m \quad \text{with} \quad m = n - p.$$

In the above two papers, the asymptotic evaluation of the MSE of $\hat{\boldsymbol{\beta}}(\mathbf{K})$ was given under the assumption that the error term is distributed as the normal distribution $\mathrm{N}_n(0, \sigma^2 I)$. They derived a region of (f_1, f_2), on which $\hat{\boldsymbol{\beta}}(\mathbf{K})$ is asymptotically superior to the LSE in terms of MSE. This result was further strengthened by Wang and Chow (1990), where they obtained a finite-sample domination result. More precisely, they considered the adaptive generalized ridge estimator $\hat{\boldsymbol{\beta}}(\mathbf{K})$ with the following $\mathbf{K}$ matrix, in which $\hat{\beta}_i^2$ in (1.4) is replaced with $\hat{\boldsymbol{\beta}}^t\hat{\boldsymbol{\beta}}$:

$$(1.5)\quad k_i = \frac{l_1\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}^t\hat{\boldsymbol{\beta}} - l_2\hat{\sigma}^2/\lambda_i} = \hat{k}_{i(l_1,l_2)},$$

where l_1 and l_2 are nonrandom constants, and derived the following sufficient condition for $\hat{\boldsymbol{\beta}}(\mathbf{K})$ to dominate the LSE in terms of MSE under normality,

$$(1.6)\quad \begin{aligned} &(i)\ \ 0 \le l_1 \le \frac{2(n-p)}{n-p+2}\left(\lambda_p^2 \sum_{i=1}^{p} \lambda_i^{-2} - 2\right) \\ &(ii)\ \ l_1 \ge l_2. \end{aligned}$$

The aim of this paper is to extend the result of Wang and Chow (1990) to the case where the error term $\mathbf{e}$ is distributed as a spherically symmetric distribution. More specifically, we assume that $\mathbf{e}$ has the probability density function (pdf) of the form

$$p(\mathbf{e}) = \sigma^{-n} f(\mathbf{e}^t\mathbf{e}/\sigma^2) \quad \text{for some} \quad f : [0, \infty) \to [0, \infty),$$

(see, for example, Muirhead (1982)), and derive a sufficient condition so that $\hat{\boldsymbol{\beta}}(\mathbf{K})$ is superior to the LSE. As an effeciency criterion, we apopt the quadratic loss function in addtion to the MSE. Since the result of Wang and Chow(1990) is limited to the MSE, our result is an extension of Wang and Chow (1990) from the view point of both distributional assumption and efficiency criterion. Since the LSE is not only best linear unbiased, but also minimax under the quadratic loss, it follows from our result that the adaptive ridge estimator is also minimax. Maruyama and Strawderman (2005) considered another but similar class of biased estimators and derived a sufficient condition for the estimators to be minimax. Their class contains a class of generalized Bayes minimax estimators under normality. See also Firinguetti (1999), in which the finite-sample efficiency of an adaptive generalized ridge estimator is studied focusing on the evaluation of moments.

Our main theorem is stated in Section 2, and the proof in Section 3. In Section 4, we give several numerical examples.

2 Main Result Denote by L_j $(j = 0, 1)$ the following quadratic loss function

$$(2.1)\quad \mathrm{L}_j(\mathbf{b}, \boldsymbol{\beta}, \sigma^2) = (\mathbf{b} - \boldsymbol{\beta})^t (\mathbf{A}^t\mathbf{A})^j (\mathbf{b} - \boldsymbol{\beta})/\sigma^2$$

and by R_j the corresponding risk functions

$$(2.2)\quad \mathrm{R}_j(\mathbf{b}, \boldsymbol{\beta}, \sigma^2) = \mathrm{E}[\mathrm{L}_j(\mathbf{b}, \boldsymbol{\beta}, \sigma^2)],$$

where $(\mathbf{A}^t\mathbf{A})^0$ is interpreted as the identity matrix. The risk functions of the LSE $\hat{\boldsymbol{\beta}}$ are easily calculated as $\mathrm{R}_0(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}, \sigma^2)) = \mathrm{tr}[(\mathbf{A}^t\mathbf{A})^{-1}]$ and $\mathrm{R}_1(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}, \sigma^2)) = p$, respectively. It is obvious that comparison by L_0 is equivalent to that by the MSEs. While L_0 is a kind of distance between $\mathbf{b}$ and $\boldsymbol{\beta}$, note that L_1 can be rewritten as $\mathrm{L}_1(\mathbf{b}, \boldsymbol{\beta}, \sigma^2) = ||\mathbf{A}\mathbf{b} - \mathbf{A}\boldsymbol{\beta}||^2/\sigma^2$.

The main result below gives a region of (l_1, l_2), on which the inequality

$$(2.3)\quad \mathrm{R}_j(\hat{\boldsymbol{\beta}}(\mathbf{K}), \boldsymbol{\beta}, \sigma^2) \le \mathrm{R}_j(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}, \sigma^2)$$

holds uniformly for $\boldsymbol{\beta}$ and $\sigma^2 (j = 0, 1)$.

Theorem. *In the model (1.1), if l_1, l_2 in (1.5) satisfy*

$$(i)\quad 0 \le l_1 \le \frac{2(n-p)}{n-p+2} \left(\frac{\lambda_p^2}{\lambda_1^j} \sum_{i=1}^{p} \lambda_i^{j-2} - 2 \right),$$

$$(ii)\quad l_1 \ge l_2,$$

then the inequality (2.3) holds. That is, the corresponding adaptive generalized ridge estimator dominates the LSE under the loss function L_j.

The proof is given in the next section. The conditions (i) and (ii) with $j = 0$ is the same as (1.6), in other words the result of Wang and Chow(1990). Thus their result remains valid even under spherically symmetric error. As is stated in the previous section, the adaptive ridge estimator satisfying the conditions (i) and (ii) is a minimax estimator under both L_0 and L_1.

3 Technical Details We begin with reducing the model to the canonical form adopted by Maruyama and Strawderman (2005). Let $\mathbf{Q}$ be an $n \times n$ orthogonal matrix such that

$$\mathbf{Q}\mathbf{A} = \begin{pmatrix} \Lambda^{1/2}\Phi^t \\ \mathbf{O} \end{pmatrix} \quad \text{or equivalently,} \quad \mathbf{A} = \mathbf{Q}^t \begin{pmatrix} \Lambda^{1/2} \\ \mathbf{O} \end{pmatrix} \Phi^t.$$

Needless to say, the latter expression gives a singular value decomposition of $\mathbf{A}$. Let Λ_* be the following $n \times n$ diagonal matrix:

$$\Lambda_* = \mathrm{diag}(\lambda_1, \cdots, \lambda_p, 1, \cdots, 1).$$

Using $\mathbf{Q}$ and Λ_*, we transform $\mathbf{y}$ into

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{z} \end{pmatrix} = \Lambda_*^{1/2}\mathbf{Q}\mathbf{y}.$$

Then the vector $(\mathbf{x}^t, \mathbf{z}^t)^t$ has the joint pdf of the form

$$\sigma^{-n}|\Lambda|^{1/2} f[\{(\mathbf{x}-\boldsymbol{\alpha})^t\Lambda(\mathbf{x}-\boldsymbol{\alpha}) + \mathbf{z}^t\mathbf{z}\}/\sigma^2],$$

where $\boldsymbol{\alpha} = \Phi^t\boldsymbol{\beta}$. In other words, $(\mathbf{x}^t, \mathbf{z}^t)^t$ is distributed as an elliptically symmetric distribution. It is important to note that

$$\mathbf{x} = \Phi^t\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\alpha}},$$

$$\mathbf{z}^t\mathbf{z} = m\hat{\sigma}^2 = (\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}})^t(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}}) = \mathrm{S}. \tag{3.1}$$

Thus the problem of estimating $\boldsymbol{\beta}$ is transformed into that of estimating $\boldsymbol{\alpha}$ based on $\hat{\boldsymbol{\alpha}}$ and S.

By using these quantities, we can rewrite respectively the LSE $\hat{\boldsymbol{\beta}}$ as $\hat{\boldsymbol{\alpha}}$ and the adaptive generalized ridge estimator $\hat{\boldsymbol{\beta}}(\mathbf{K})$ as

$$\hat{\boldsymbol{\alpha}}(\mathbf{K}) = (\Lambda + \mathbf{K})^{-1}\Lambda\hat{\boldsymbol{\alpha}}$$

with

$$\mathbf{K} = \mathrm{diag}(k_1, \cdots, k_p) \quad \text{and} \quad k_i = \frac{l_1\hat{\sigma}^2}{\hat{\boldsymbol{\alpha}}^t\hat{\boldsymbol{\alpha}} - l_2\hat{\sigma}^2/\lambda_i} = \hat{k}_{(l_1,l_2)}.$$

Correspondingly, the loss functions are also rewritten as

$$L_j(\mathbf{a}, \boldsymbol{\alpha}, \sigma^2) = (\mathbf{a} - \boldsymbol{\alpha})^t\Lambda^j(\mathbf{a} - \boldsymbol{\alpha})/\sigma^2 (j = 0, 1).$$

In the proof below, we use two identities that were obtained by Kubokawa and Srivastava (1999,2001) to extend the Stein and chi-square identities to the case of spherically symmetric distirubition. To state their results, let

$$F(x) = \frac{1}{2}\int_x^\infty f(t)dt,$$

and define

$$\mathrm{E}^f[h(\mathbf{x}, \mathbf{z})] = \int\int h(\mathbf{x}, \mathbf{z})\sigma^{-n}|\Lambda|^{1/2} f[\{(\mathbf{x}-\boldsymbol{\alpha})^t\Lambda(\mathbf{x}-\boldsymbol{\alpha}) + \mathbf{z}^t\mathbf{z}\}/\sigma^2]\mathbf{dxdz}$$

$$\mathrm{E}^F[h(\mathbf{x}, \mathbf{z})] = \int\int h(\mathbf{x}, \mathbf{z})\sigma^{-n}|\Lambda|^{1/2} F[\{(\mathbf{x}-\boldsymbol{\alpha})^t\Lambda(\mathbf{x}-\boldsymbol{\alpha}) + \mathbf{z}^t\mathbf{z}\}/\sigma^2]\mathbf{dxdz}$$

for an integrable function $h(\cdot)$.

Lemma 1. *(Kubokawa and Srivastava (2001)) Let* h *be a differentiable function such that the expectations below exist. Then the following identity holds:*

$$\mathrm{E}^f[(\hat{\alpha}_i - \alpha_i)h(\hat{\alpha}_i)] = \frac{\sigma^2}{\lambda_i}\mathrm{E}^F\left[\frac{\partial}{\partial\hat{\alpha}_i}h(\hat{\alpha}_i)\right]. \tag{3.2}$$

Lemma 2. *(Kubokawa and Srivastava (1999)) Let g be a differentiable function such that the expectations below exist. Then the following identity holds:*

$$\text{(3.3)}\quad \mathrm{E}^f[\mathrm{S}g(\mathrm{S})] = \sigma^2\mathrm{E}^F\left[mg(\mathrm{S}) + 2\mathrm{S}\frac{\partial}{\partial \mathrm{S}}g(\mathrm{S})\right],$$

where the quantity $\mathrm{S} = \mathbf{z}^t\mathbf{z}$ *is defined in (3.1).*

Proof of Theorem Let $\hat{\boldsymbol{\alpha}}_i(k_i)$ be the i-th element of $\hat{\boldsymbol{\alpha}}(\mathbf{K})$. Then we have

$$\hat{\alpha}_i(k_i) = \left(\frac{\lambda_i}{\lambda_i + k_i}\right)\hat{\alpha}_i = \left(1 - \frac{l_1\hat{\sigma}^2\lambda_i^{-1}}{\hat{\boldsymbol{\alpha}}^t\hat{\boldsymbol{\alpha}} + (l_1 - l_2)\hat{\sigma}^2\lambda_i^{-1}}\right)\hat{\alpha}_i = \hat{\alpha}_i - \frac{l_1\hat{\sigma}^2\lambda_i^{-1}}{g_i}\hat{\alpha}_i,$$

where $g_i = \hat{\boldsymbol{\alpha}}^t\hat{\boldsymbol{\alpha}} + (l_1 - l_2)\hat{\sigma}^2\lambda_i^{-1}$. Hence the risk function (2.2) is expressed as

$$\text{(3.4)}\quad \mathrm{R}_j(\hat{\boldsymbol{\alpha}}(\mathbf{K}), \boldsymbol{\alpha}, \sigma^2) \;=\; \frac{1}{\sigma^2}\sum_{i=1}^{p}\mathrm{E}^f\left[\lambda_i^j(\hat{\alpha}_i(k_i) - \alpha_i)^2\right],$$

which is further expanded as

(3.5)

$$\mathrm{E}^f\left[\lambda_i^j(\hat{\alpha}_i(k_i) - \alpha_i)^2\right] = \mathrm{E}^f\left[\lambda_i^j(\hat{\alpha}_i - \alpha_i)^2\right] + \lambda_i^j\frac{l_1^2}{\lambda_i^2}\mathrm{E}^f\left[\frac{\hat{\sigma}^2\hat{\alpha}_i}{g_i}\right]^2 - 2\lambda_i^j\frac{l_1}{\lambda_i}\mathrm{E}^f\left[(\hat{\alpha}_i - \alpha_i)\frac{\hat{\sigma}^2\hat{\alpha}_i}{g_i}\right].$$

Applying (3.2) to the second term of (3.5) yields

$$\begin{aligned}
\lambda_i^j\frac{l_1^2}{\lambda_i^2}\mathrm{E}^f\left[\frac{\hat{\sigma}^2\hat{\alpha}_i}{g_i}\right]^2 &= \lambda_i^j\frac{l_1^2}{\lambda_i^2}\mathrm{E}^f[\mathrm{S}g(\mathrm{S})] \quad \text{with } g(\mathrm{S}) = \frac{\mathrm{S}\hat{\alpha}_i^2}{m^2g_i^2} \\
&= \lambda_i^j\frac{l_1^2}{\lambda_i^2}\sigma^2\mathrm{E}^F\left[m\left(\frac{\mathrm{S}\hat{\alpha}_i^2}{m^2g_i^2}\right) + 2\mathrm{S}\left\{\frac{\hat{\alpha}_i^2}{m^2}\left(\frac{1}{g_i^2} - \frac{2\mathrm{S}(l_1 - l_2)}{m\lambda_ig_i^3}\right)\right\}\right] \\
\text{(3.6)}\qquad &= \lambda_i^j\frac{l_1^2}{\lambda_i^2}\sigma^2\mathrm{E}^F\left[\frac{\hat{\sigma}^2\hat{\alpha}_i^2 + 2m^{-1}\hat{\sigma}^2\hat{\alpha}_i^2}{g_i^2} - \frac{4\hat{\sigma}^4\hat{\alpha}_i^2(l_1 - l_2)}{m\lambda_ig_i^3}\right].
\end{aligned}$$

Note that g_i depends on $\mathrm{S} = m\hat{\sigma}^2$. Apply (3.3) to the third term of (3.5),

$$\text{(3.7)}\quad \mathrm{E}^f\left[(\hat{\alpha}_i - \alpha_i)\frac{\hat{\sigma}^2\hat{\alpha}_i}{g_i}\right] = \frac{\sigma^2}{\lambda_i}\mathrm{E}^F\left[\frac{\partial}{\partial\hat{\alpha}_i}\frac{\hat{\sigma}^2\hat{\alpha}_i}{g_i}\right] = \frac{\sigma^2}{\lambda_i}\mathrm{E}^F\left[\frac{\hat{\sigma}^2}{g_i} - \frac{2\hat{\sigma}^2\hat{\alpha}_i^2}{g_i^2}\right].$$

From (3.5),(3.6),(3.7), we get

$$\begin{aligned}
&\mathrm{E}^f\left[\lambda_i^j(\hat{\alpha}_i(k_i) - \alpha_i)^2\right] \\
&= \mathrm{E}^f\left[\lambda_i^j(\hat{\alpha}_i - \alpha_i)^2\right] + \lambda_i^j\frac{l_1^2}{\lambda_i^2}\sigma^2\mathrm{E}^F\left[\frac{\hat{\sigma}^2\hat{\alpha}_i^2 + 2m^{-1}\hat{\sigma}^2\hat{\alpha}_i^2}{g_i^2} - \frac{4\hat{\sigma}^4\hat{\alpha}_i^2(l_1 - l_2)}{m\lambda_ig_i^3}\right] - 2\lambda_i^j\frac{l_1}{\lambda_i}\sigma^2\mathrm{E}^F\left[\frac{\hat{\sigma}^2}{g_i} - \frac{2\hat{\sigma}^2\hat{\alpha}_i^2}{g_i^2}\right] \\
&= \mathrm{E}^f\left[\lambda_i^j(\hat{\alpha}_i - \alpha_i)^2\right] + l_1\sigma^2\mathrm{E}^F\left[\frac{l_1\hat{\sigma}^2\hat{\alpha}_i^2 + 2l_1m^{-1}\hat{\sigma}^2\hat{\alpha}_i^2 + 4\hat{\sigma}^2\hat{\alpha}_i^2}{\lambda_i^{-j+2}g_i^2} - \frac{4l_1\hat{\sigma}^4\hat{\alpha}_i^2(l_1 - l_2)}{m\lambda_i^{-j+3}g_i^3} - \frac{2\hat{\sigma}^2}{\lambda_i^{-j+2}g_i}\right].
\end{aligned}$$

(3.8)

Note that $\dfrac{l_1\hat{\sigma}^2\hat{\alpha}_i^2 + 2l_1m^{-1}\hat{\sigma}^2\hat{\alpha}_i^2 + 4\hat{\sigma}^2\hat{\alpha}_i^2}{\lambda_i^2 g_i^2}$ is decreasing in λ_i. Then,

$$\begin{aligned}
&\frac{l_1\hat{\sigma}^2\hat{\alpha}_i^2 + 2l_1m^{-1}\hat{\sigma}^2\hat{\alpha}_i^2 + 4\hat{\sigma}^2\hat{\alpha}_i^2}{\lambda_i^{-j+2} g_i^2} = \lambda_i^j\left[\frac{l_1\hat{\sigma}^2\hat{\alpha}_i^2 + 2l_1m^{-1}\hat{\sigma}^2\hat{\alpha}_i^2 + 4\hat{\sigma}^2\hat{\alpha}_i^2}{\lambda_i^2 g_i^2}\right] \\
&\le \lambda_i^j\left[\frac{l_1\hat{\sigma}^2\hat{\alpha}_i^2 + 2l_1m^{-1}\hat{\sigma}^2\hat{\alpha}_i^2 + 4\hat{\sigma}^2\hat{\alpha}_i^2}{\lambda_p^2 g^2}\right] \le \lambda_1^j\left[\frac{l_1\hat{\sigma}^2\hat{\alpha}_i^2 + 2l_1m^{-1}\hat{\sigma}^2\hat{\alpha}_i^2 + 4\hat{\sigma}^2\hat{\alpha}_i^2)}{\lambda_p^2 g^2}\right],
\end{aligned} \tag{3.9}$$

where $g = \hat{\boldsymbol{\alpha}}^t\hat{\boldsymbol{\alpha}} + (l_1 - l_2)\hat{\sigma}^2\lambda_p^{-1}$.
Since $l_1 \ge l_2$, we have $g_i = \hat{\boldsymbol{\alpha}}^t\hat{\boldsymbol{\alpha}} + (l_1 - l_2)\hat{\sigma}^2\lambda_i^{-1} \le \hat{\boldsymbol{\alpha}}^t\hat{\boldsymbol{\alpha}} + (l_1 - l_2)\hat{\sigma}^2\lambda_p^{-1} = g$, and hence

$$-\frac{2\hat{\sigma}^2}{\lambda_i^{-j+2} g_i} \le -\frac{2\hat{\sigma}^2}{\lambda_i^{-j+2} g}. \tag{3.10}$$

By substituting (3.9) and (3.10) into (3.8), and by noting $\dfrac{4l_1\hat{\sigma}^4\hat{\alpha}_i^2(l_1 - l_2)}{m\lambda_i^{-j+3} g_i^3} \ge 0$, we have

$$\begin{aligned}
\mathrm{E}^f\left[\lambda_i^j(\hat{\alpha}_i(k_i) - \alpha_i)^2\right] \;&\le\; \mathrm{E}^f\left[\lambda_i^j(\hat{\alpha}_i - \alpha_i)^2\right] \\
&\quad + l_1\sigma^2\mathrm{E}^F\left[\frac{\lambda_1^j(l_1\hat{\sigma}^2\hat{\alpha}_i^2 + 2l_1m^{-1}\hat{\sigma}^2\hat{\alpha}_i^2 + 4\hat{\sigma}^2\hat{\alpha}_i^2)}{\lambda_p^2 g^2} - \frac{2\hat{\sigma}^2}{\lambda_i^{-j+2} g}\right],
\end{aligned}$$

from which it follows

$$\begin{aligned}
\mathrm{R}(\hat{\boldsymbol{\alpha}}(\mathbf{K}), \boldsymbol{\alpha}, \sigma^2) \;&\le\; \mathrm{R}(\hat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}, \sigma^2) + \frac{1}{\sigma^2}l_1\sigma^2\mathrm{E}^F\left[\frac{\lambda_1^j(l_1\hat{\sigma}^2 + 2l_1m^{-1}\hat{\sigma}^2 + 4\hat{\sigma}^2)}{\lambda_p^2 g^2}\hat{\boldsymbol{\alpha}}^t\hat{\boldsymbol{\alpha}} - \frac{2\hat{\sigma}^2}{g}\sum_{i=1}^p \lambda_i^{j-2}\right] \\
&\le\; \mathrm{R}(\hat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}, \sigma^2) + l_1\mathrm{E}^F\left[\frac{\hat{\sigma}^2}{g}\left\{\frac{\lambda_1^j(l_1 + 2l_1m^{-1} + 4)}{\lambda_p^2} - 2\sum_{i=1}^p \lambda_i^{j-2}\right\}\right],
\end{aligned} \tag{3.11}$$

where $\dfrac{\hat{\boldsymbol{\alpha}}^t\hat{\boldsymbol{\alpha}}}{g} \le 1$ is used in the last line.

Since the second term of the right hand side of (3.11) can be decomposed as

$$\begin{aligned}
&\mathrm{E}^F\left[\frac{\hat{\sigma}^2}{g}\left\{\frac{\lambda_1^j(l_1 + 2l_1m^{-1} + 4)}{\lambda_p^2} - 2\sum_{i=1}^p \lambda_i^{j-2}\right\}\right] \\
&= \left[\frac{\lambda_1^j(l_1 + 2l_1m^{-1} + 4)}{\lambda_p^2} - 2\sum_{i=1}^p \lambda_i^{j-2}\right]\mathrm{E}^F\left[\frac{\hat{\sigma}^2}{g}\right],
\end{aligned}$$

a sufficient condition for the inequality (2.3) is

$$\left[\frac{\lambda_1^j(l_1 + 2l_1m^{-1} + 4)}{\lambda_p^2} - 2\sum_{i=1}^p \lambda_i^{j-2}\right] < 0$$

which is implied by the conditions (i) and (ii). This completes the proof. □

4 Numerical Studies In this section, the MSEs of the LSE and several adaptive generalized ridge estimators are compared numerically. Simulations are done in the case in which $p = 6$, $n = 20$, $\Lambda = (1, 0.01, 0.0019, 0.0017, 0.0015, 0.001)$ and $\beta_i = 0, 0.5, 1, 1.5, 2$ for each i . The entries of Λ indicate the presence of multicollinearity. As a distribution of the error term $\mathbf{e}$, we adopt the multivariate standard normal distribution and the multivariate t-distribution with degrees of freedom 10.

Tables (4.1) and (4.2) show the relative efficiency measured by $\mathrm{R}_0(\hat{\boldsymbol{\beta}}(\mathbf{K}), \boldsymbol{\beta}, \sigma)/\mathrm{R}_0(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}, \sigma)$ under these two distributions. In the simulation, the values of l_1 are chosen in such a way that $l_1 = l_1^*$, $0.75l_1^*$ $0.5l_1^*, 0$ and $0.25l_1^*$, where l_1^* denotes the upper bound of the admissible value of l_1 in the condition (i). As for the constant l_2, we set $l_2 = l_1, 0.5l_1, 0, -0.5l_1, -l_1, -1.5l_1$, $-2l_1$. The relative efficiency is simulated via 10,000 replications. We can observe from the results that

(a) To select constant l_1, larger value within limitation of condtion (i) is better for improvement of relative effeciency.

(b) Remarkable improvement is not shown in terms of selecting l_2. Better choice may be selecting larger values limitted in condition (ii), which improve or keep relative performance.

Acknowledgement The authors thank the associate editor and the two anonymous referees for their valuable comments. Thanks also go to Professor Kiyoshi Inoue at Waseda University for his helpful comments. Kurata's portion of this research is financially supported by Grant-in-Aid for Scientific Research (A)(20243016) and (C)(21500272).

References

[1] Firinguetti, L. (1990), A generalized ridge regression estimator and its finite sample properties, Comm. Statist. Theory Methods,**28**,1217-1229.

[2] Hoerl, A.E. and Kennard, R.W. (1970a), Ridge regression: Biased estimation for non-orthogonal problems, Technometrics, **12**,55-67.

[3] Hoerl, A.E. and Kennard, R.W. (1970b), Ridge regression:Applications to non-orthogonal problems, Technometrics, **12**, 69-82.

[4] Kubokawa, T. and Srivastava, M.S. (1999), Robust improvement in estimation of a covariance matrix in an elliptically contoured distribution, Ann. Statist., **27**,600-609.

[5] Kubokawa, T. and Srivastava, M.S. (2001), Robust improvement in estimation of a mean matrix in an elliptically contoured distribution, J. Multivariate Anal., **76**,138-152.

[6] Maruyama, Y. and Strawderman, W.E. (2005), A new class of generalized Bayes minimax ridge regression estimators, Ann. Statist., **33**,1753-1770.

[7] Muirhead, R.B. (1982), *Aspects of multivariate statistical theory*, John Wiley & Sons, Inc., New York.

[8] Ullah, A., Vinod, H.D., and Kadiyala, R.K. (1980), A family of improved shrinkage factors for the ordinary ridge estimator, *Proceedings of the Econometric Society European Meeting, 1979*, North-Holland Publishing Company.

[9] Vinod, H.D., and Ullah A. (1980), *Recent advances in regression models*, Dekker, New York.

[10] Wang S.G. and Chow, S.C. (1990), A note on adaptive generalized ridge regression estimator, Statist. Probab. Lett., **10**, 17-21.

β_i	l_2	l_1: $0.25l_1^*$	$0.5l_1^*$	$0.75l_1^*$	l_1^*
0	$-2l_1$	0.967	0.941	0.919	0.899
	$-1.5l_1$	0.967	0.939	0.916	0.895
	$-l_1$	0.966	0.938	0.913	0.891
	$-0.5l_1$	0.966	0.936	0.909	0.885
	0	0.965	0.934	0.906	0.880
	$0.5l_1$	0.964	0.931	0.901	0.873
	l_1	0.963	0.932	0.904	0.882
0.5	$-2l_1$	0.967	0.941	0.918	0.899
	$-1.5l_1$	0.966	0.939	0.916	0.895
	$-l_1$	0.966	0.937	0.912	0.890
	$-0.5l_1$	0.965	0.935	0.909	0.885
	0	0.964	0.933	0.905	0.879
	$0.5l_1$	0.964	0.931	0.900	0.872
	l_1	0.963	0.928	0.895	0.864
1	$-2l_1$	0.967	0.941	0.919	0.899
	$-1.5l_1$	0.966	0.939	0.916	0.895
	$-l_1$	0.966	0.937	0.912	0.890
	$-0.5l_1$	0.965	0.935	0.909	0.885
	0	0.964	0.933	0.904	0.879
	$0.5l_1$	0.963	0.930	0.900	0.871
	l_1	0.962	0.927	0.895	0.865
1.5	$-2l_1$	0.966	0.939	0.916	0.897
	$-1.5l_1$	0.966	0.938	0.913	0.892
	$-l_1$	0.965	0.936	0.910	0.887
	$-0.5l_1$	0.964	0.934	0.906	0.882
	0	0.964	0.931	0.902	0.876
	$0.5l_1$	0.963	0.929	0.898	0.868
	l_1	0.962	0.926	0.892	0.860
2	$-2l_1$	0.967	0.940	0.918	0.898
	$-1.5l_1$	0.966	0.939	0.915	0.894
	$-l_1$	0.966	0.937	0.912	0.889
	$-0.5l_1$	0.965	0.935	0.908	0.884
	0	0.964	0.933	0.904	0.878
	$0.5l_1$	0.964	0.931	0.900	0.871
	l_1	0.963	0.928	0.895	0.864

Table 4.1: Relative risk in case of $\mathbf{N}_6$

β_i	l_2	l_1: $0.25l_1^*$	$0.5l_1^*$	$0.75l_1^*$	l_1^*
0	$-2l_1$	0.968	0.942	0.920	0.901
	$-1.5l_1$	0.968	0.941	0.918	0.897
	$-l_1$	0.967	0.939	0.915	0.893
	$-0.5l_1$	0.966	0.937	0.911	0.888
	0	0.966	0.935	0.908	0.882
	$0.5l_1$	0.965	0.933	0.903	0.876
	l_1	0.964	0.931	0.899	0.870
0.5	$-2l_1$	0.968	0.941	0.919	0.900
	$-1.5l_1$	0.967	0.940	0.917	0.896
	$-l_1$	0.966	0.938	0.913	0.891
	$-0.5l_1$	0.966	0.936	0.910	0.886
	0	0.965	0.934	0.906	0.880
	$0.5l_1$	0.965	0.932	0.902	0.874
	l_1	0.964	0.929	0.897	0.866
1	$-2l_1$	0.966	0.939	0.917	0.897
	$-1.5l_1$	0.966	0.938	0.914	0.893
	$-l_1$	0.965	0.936	0.910	0.888
	$-0.5l_1$	0.964	0.934	0.907	0.882
	0	0.963	0.931	0.902	0.876
	$0.5l_1$	0.963	0.929	0.898	0.869
	l_1	0.962	0.926	0.893	0.863
1.5	$-2l_1$	0.969	0.944	0.922	0.903
	$-1.5l_1$	0.968	0.942	0.920	0.900
	$-l_1$	0.968	0.940	0.917	0.895
	$-0.5l_1$	0.967	0.939	0.913	0.890
	0	0.966	0.937	0.910	0.885
	$0.5l_1$	0.966	0.934	0.905	0.878
	l_1	0.965	0.932	0.901	0.872
2	$-2l_1$	0.968	0.943	0.921	0.902
	$-1.5l_1$	0.967	0.941	0.918	0.898
	$-l_1$	0.967	0.939	0.915	0.894
	$-0.5l_1$	0.966	0.937	0.912	0.889
	0	0.965	0.935	0.908	0.883
	$0.5l_1$	0.965	0.933	0.904	0.877
	l_1	0.964	0.933	0.908	0.887

Table 4.2: Relative risk in case of $\mathbf{t}_6$

Kazumasa Mori
Department of Advanced Social and International Studies, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo, 153-8902 JAPAN
E-mail: morikzm@nifty.com

Hiroshi Kurata
Department of Advanced Social and International Studies, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo, 153-8902 JAPAN
E-mail: kurata@waka.c.u-tokyo.ac.jp