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 $A'A$ 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 $e$ is distributed as a spherically symmetric distribution 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

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

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

\[(1.2) \ \hat{\beta} = (A' A)^{-1} A' y \]

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

\[V(\hat{\beta}) = \sigma^2 (A' A)^{-1}. \]

When the column vectors of $A$ are approximately linearly dependent, which is often the case in practice, the matrix $A'A$ 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 $A'A$ with a more stable matrix. To state it precisely, let a spectral decomposition of $A'A$ be

\[A' A = \Phi \Lambda \Phi', \]

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

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

Then the generalized ridge estimator can be written as

\[(1.3) \ \hat{\beta}(K) = (A' A + \Phi K \Phi')^{-1} A' y, \]

\[2000 \ Mathematics \ Subject \ Classification. \quad 62J07, 62C20. \]

\[Key \ words \ and \ phrases. \quad \text{ridge estimator, spherically symmetric distribution, mean squared error, quadratic loss, minimax estimator.} \]
where \( K = \text{diag}(k_1, \ldots, k_p) \) is a nonrandom diagonal matrix suitably chosen. The simplest choice for \( K \) is \( K = k \mathbf{I} \) \((k > 0)\). This choice yields the (original) ridge estimator \( \hat{\beta}(k \mathbf{I}) = (\mathbf{A}' \mathbf{A} + k \mathbf{I})^{-1} \mathbf{A}' \mathbf{y} \), which has been widely employed in applications. While the generalized ridge estimator thus defined is biased unless \( K = 0 \), 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 \( E[(\mathbf{b} - \beta)'(\mathbf{b} - \beta)] \). In fact, the MSE of \( \hat{\beta}(K) \) attains its minimum at
\[
 k_i = \frac{\sigma^2}{\beta_i^2}, \quad i = 1, \ldots, p,
\]
which is smaller than that of the LSE. However, in most cases, the quantities \( \sigma^2/\beta_i^2 \) are unknown and \( \hat{\beta}(K) \) is not feasible.

Many authors have proposed feasible versions of \( \hat{\beta}(K) \) by replacing \( K \) 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 \( \hat{\beta}(K) \) with
\[
(1.4) \quad k_i = \frac{f_1 \hat{\sigma}^2}{\hat{\beta}_i^2 - f_2 \hat{\sigma}^2/\lambda_i} = \hat{k}_{i(f_1,f_2)}
\]
and is called an adaptive generalized ridge estimator. Here, \( \hat{\beta}_i \) is the \( i \)-th element of the LSE \( \mathbf{\hat{\beta}} \), \( f_i \)'s are nonrandom constants and
\[
\hat{\sigma}^2 = ||\mathbf{y} - \mathbf{A} \mathbf{\hat{\beta}}||^2/m \quad \text{with} \quad m = n - p.
\]
In the above two papers, the asymptotic evaluation of the MSE of \( \hat{\beta}(K) \) was given under the assumption that the error term is distributed as the normal distribution \( N_n(0, \sigma^2 \mathbf{I}) \). They derived a region of \((f_1, f_2)\), on which \( \hat{\beta}(K) \) 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 \( \hat{\beta}(K) \) with the following \( K \) matrix, in which \( \hat{\beta}_i^2 \) in (1.4) is replaced with \( \hat{\beta}_i \hat{\beta}' \):
\[
(1.5) \quad k_i = \frac{l_1 \hat{\sigma}^2}{\hat{\beta}_i \hat{\beta}' - l_2 \hat{\sigma}^2/\lambda_i} = \hat{k}_{i(l_1,l_2)},
\]
where \( l_1 \) and \( l_2 \) are nonrandom constants, and derived the following sufficient condition for \( \hat{\beta}(K) \) to dominate the LSE in terms of MSE under normality,
\[
(i) \quad 0 \leq l_1 \leq \frac{2(n - p)}{n - p + 2} \left( \lambda_p^2 \sum_{i=1}^{p} \lambda_i^{-2} - 2 \right)
\]
\[
(1.6) \quad (ii) \quad l_1 \geq l_2.
\]

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
\[
p(\mathbf{e}) = \sigma^{-n} f(\mathbf{e}' \mathbf{e}/\sigma^2) \quad \text{for some} \quad f : [0, \infty) \to [0, \infty),
\]
(see, for example, Muirhead (1982)), and derive a sufficient condition so that \( \hat{\beta}(K) \) is superior to the LSE. As an efficiency criterion, we adopt the quadratic loss function in addition 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

\[
L_j(b, \beta, \sigma^2) = (b - \beta)^t(A^tA)^j(b - \beta)/\sigma^2
\]

and by \( R_j \) the corresponding risk functions

\[
R_j(b, \beta, \sigma^2) = E[L_j(b, \beta, \sigma^2)]
\]

where \((A^tA)^0\) is interpreted as the identity matrix. The risk functions of the LSE \( \hat{\beta} \) are easily calculated as \( R_0(\hat{\beta}, \beta, \sigma^2) = \text{tr}[(A^tA)^{-1}] \) and \( R_1(\hat{\beta}, \beta, \sigma^2) = p \), 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 \( b \) and \( \beta \), note that \( L_1 \) can be rewritten as \( |b A^t - A^t \beta|/\sigma^2 \).

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

\[
R_j(\hat{\beta}(K), \beta, \sigma^2) \leq R_j(\hat{\beta}, \beta, \sigma^2)
\]

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

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

\[
(i) \quad 0 \leq l_1 \leq \frac{2(n - p)}{p} \left( \frac{\lambda_{\hat{\beta}}^2}{\lambda_{\hat{\beta}}^2} \sum_{i=1}^{p} \lambda_{i}^{2} - 2 \right),
\]

\[
(ii) \quad l_1 \geq l_2,
\]

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 \( Q \) be an \( n \times n \) orthogonal matrix such that

\[
QA = \left( \begin{array}{c} \Lambda^{1/2} \Phi^t \\ O \end{array} \right) \quad \text{or equivalently,} \quad A = Q^t \left( \begin{array}{c} \Lambda^{1/2} \\ O \end{array} \right) \Phi^t.
\]

Needless to say, the latter expression gives a singular value decomposition of \( A \). Let \( \Lambda \) be the following \( n \times n \) diagonal matrix:

\[
\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_p, 1, \cdots, 1).
\]
Using $Q$ and $\Lambda$, we transform $y$ into
\[ \begin{pmatrix} x \\ z \end{pmatrix} = \Lambda^{1/2} Q y. \]
Then the vector $(x', z')^t$ has the joint pdf of the form
\[ \sigma^{-n} |\Lambda|^{1/2} f\left[\{(x - \alpha)^t \Lambda (x - \alpha) + z' z\}/\sigma^2\right], \]
where $\alpha = \Phi^t \beta$. In other words, $(x', z')^t$ is distributed as an elliptically symmetric distribution. It is important to note that
\[ x = \Phi^t \hat{\beta} = \hat{\alpha}, \]
(3.1) $z' z = m \hat{\sigma}^2 = (y - A^t \hat{\beta})^t (y - A^t \hat{\beta}) = S.$
Thus the problem of estimating $\beta$ is transformed into that of estimating $\alpha$ based on $\hat{\alpha}$ and $S$.
By using these quantities, we can rewrite respectively the LSE $\hat{\beta}$ as $\hat{\alpha}$ and the adaptive generalized ridge estimator $\hat{\beta}(K)$ as
\[ \hat{\alpha}(K) = (\Lambda + K)^{-1} \Lambda \hat{\alpha} \]
with
\[ K = \text{diag}(k_1, \cdots, k_p) \quad \text{and} \quad k_i = \frac{l_i \hat{\sigma}^2}{\hat{\alpha}^t \hat{\alpha} - l_2 \hat{\sigma}^2/\lambda_i} = \hat{k}(l_1, l_2). \]
Correspondingly, the loss functions are also rewritten as
\[ L_j(a, \alpha, \sigma^2) = (a - \alpha)^t \Lambda^t (a - \alpha)/\sigma^2 (j = 0, 1). \]
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 distribution. To state their results, let
\[ F(x) = \frac{1}{2} \int_{-\infty}^{\infty} f(t) dt, \]
and define
\[ E^F[h(x, z)] = \int \int h(x, z) \sigma^{-n} |\Lambda|^{1/2} f\left[\{(x - \alpha)^t \Lambda (x - \alpha) + z' z\}/\sigma^2\right] dxdz \]
\[ E^F[h(x, z)] = \int \int h(x, z) \sigma^{-n} |\Lambda|^{1/2} F\left[\{(x - \alpha)^t \Lambda (x - \alpha) + z' z\}/\sigma^2\right] dxdz \]
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:

\[ E^F[(\hat{\alpha}_i - \alpha_i) h(\hat{\alpha}_i)] = \frac{\sigma^2}{\lambda_i} E^F \left[ \frac{\partial}{\partial \hat{\alpha}_i} h(\hat{\alpha}_i) \right]. \]
Lemma 2. (Kubokawa and Srivastava (1999)) Let $g$ be a differentiable function such that the expectations below exist. Then the following identity holds:

\[
E[f(S)] = \sigma^2 EF \left[ m g(S) + 2S \frac{\partial}{\partial S} g(S) \right],
\]

where the quantity $S = z'z$ is defined in (3.1).

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

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

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

\[
R_g(\hat{\alpha}(K), \alpha, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^{p} E[f] \left[ \lambda_i^2 (\hat{\alpha}_i(k_i) - \alpha_i)^2 \right],
\]

which is further expanded as

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

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

\[
\lambda_i^2 \frac{\sigma^2}{\lambda_i^2} E[f \left[ \hat{\sigma}^2 \hat{\alpha}_i \right] = \lambda_i^2 \frac{\sigma^2}{\lambda_i^2} E[f \left[ Sg(S) \right] \text{ with } g(S) = \frac{S\hat{\sigma}^2}{m^2 g_i^2}
\]

\[
= \lambda_i^2 \frac{\sigma^2}{\lambda_i^2} EF \left[ m \left( \frac{S\hat{\sigma}^2}{m^2 g_i^2} \right) + 2S \left\{ \frac{\hat{\alpha}^2}{m^2} \left( \frac{1}{g_i} - \frac{2S(l_1 - l_2)}{m\lambda_i g_i^2} \right) \right\} \right]
\]

\[
= \lambda_i^2 \sigma^2 \frac{\sigma^2}{\lambda_i^2} EF \left[ \frac{\hat{\sigma}^2 \hat{\alpha}_i^2 + 2m^{-1} \hat{\sigma}^2 \hat{\alpha}_i^2}{g_i^2} - \frac{4\hat{\sigma}^2 \hat{\alpha}_i^2 (l_1 - l_2)}{m\lambda_i g_i^2} \right].
\]

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

\[
E[f] \left[ \frac{\hat{\sigma}^2 \hat{\alpha}_i}{g_i} \right] = \frac{\sigma^2}{\lambda_i} EF \left[ \frac{\partial}{\partial \hat{\alpha}_i} \frac{\hat{\sigma}^2 \hat{\alpha}_i}{g_i} \right] = \frac{\sigma^2}{\lambda_i} EF \left[ \frac{\hat{\sigma}^2 \hat{\alpha}_i}{g_i} - \frac{2\hat{\sigma}^2 \hat{\alpha}_i^2}{g_i^2} \right].
\]

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

\[
E[f] \left[ \lambda_i^2 (\hat{\alpha}_i(k_i) - \alpha_i)^2 \right]
\]

\[
= E[f] \left[ \lambda_i^2 (\hat{\alpha}_i - \alpha_i)^2 \right] + \lambda_i^2 \sigma^2 EF \left[ \frac{\hat{\sigma}^2 \hat{\alpha}_i^2 + 2m^{-1} \hat{\sigma}^2 \hat{\alpha}_i^2}{g_i^2} - \frac{4\hat{\sigma}^2 \hat{\alpha}_i^2 (l_1 - l_2)}{m\lambda_i g_i^2} \right] - 2\lambda_i^2 \frac{l_1}{\lambda_i} \sigma^2 EF \left[ \frac{\hat{\sigma}^2}{g_i} - \frac{2\hat{\sigma}^2 \hat{\alpha}_i^2}{g_i^2} \right]
\]

\[
\quad = E[f] \left[ \lambda_i^2 (\hat{\alpha}_i - \alpha_i)^2 \right] + l_1 \sigma^2 EF \left[ \frac{l_1 \hat{\sigma}^2 \hat{\alpha}_i^2 + 2l_1 m^{-1} \hat{\sigma}^2 \hat{\alpha}_i^2 + 4\hat{\sigma}^2 \hat{\alpha}_i^2}{\lambda_i^{j+2} g_i^2} - \frac{4l_1 \hat{\sigma}^4 \hat{\alpha}_i^2 (l_1 - l_2)}{m\lambda_i^{j+3} g_i^2} \right] - 2\hat{\sigma}^2 \frac{\alpha_i^2}{\lambda_i^{j+2} g_i^2}.
\]

(3.8)
Note that \( \frac{l_1 \hat{\sigma}^2 \hat{\alpha}_i^2 + 2l_1 m^{-1} \hat{\sigma}^2 \hat{\alpha}_i^2 + 4 \hat{\sigma}^2 \hat{\alpha}_i^2}{\lambda_i^2 g_i} \) is decreasing in \( \lambda_i \). Then,

\[
\frac{l_1 \hat{\sigma}^2 \hat{\alpha}_i^2 + 2l_1 m^{-1} \hat{\sigma}^2 \hat{\alpha}_i^2 + 4 \hat{\sigma}^2 \hat{\alpha}_i^2}{\lambda_i^{-j+2} g_i} = \lambda_i^2 \left[ \frac{l_1 \hat{\sigma}^2 \hat{\alpha}_i^2 + 2l_1 m^{-1} \hat{\sigma}^2 \hat{\alpha}_i^2 + 4 \hat{\sigma}^2 \hat{\alpha}_i^2}{\lambda_i^2 g_i} \right] \leq \lambda_i^2 \left[ \frac{l_1 \hat{\sigma}^2 \hat{\alpha}_i^2 + 2l_1 m^{-1} \hat{\sigma}^2 \hat{\alpha}_i^2 + 4 \hat{\sigma}^2 \hat{\alpha}_i^2}{\lambda_i^2 g_i} \right],
\]
(3.9)

where \( g = \hat{\alpha}' \hat{\alpha} + (l_1 - l_2) \hat{\sigma}^2 \lambda_p^{-1} \).

Since \( l_1 \geq l_2 \), we have \( g_i = \hat{\alpha}' \hat{\alpha} + (l_1 - l_2) \hat{\sigma}^2 \lambda_i^{-1} \leq \hat{\alpha}' \hat{\alpha} + (l_1 - l_2) \hat{\sigma}^2 \lambda_p^{-1} = g \), and hence

\[
(3.10) \quad \frac{2 \hat{\sigma}^2}{\lambda_i^{-j+2} g_i} \leq - \frac{2 \hat{\sigma}^2}{\lambda_i^{-j+2} g}.
\]

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

\[
\begin{align*}
\mathbb{E}^f \left[ \lambda_i^2 (\hat{\alpha}_i(k_i) - \alpha_i)^2 \right] & \leq \mathbb{E}^f \left[ \lambda_i^2 (\hat{\alpha}_i - \alpha_i)^2 \right] + l_1 \sigma^2 \mathbb{E}^f \left[ \frac{\lambda_i \frac{l_1 \hat{\sigma}^2 \hat{\alpha}_i^2 + 2l_1 m^{-1} \hat{\sigma}^2 \hat{\alpha}_i^2 + 4 \hat{\sigma}^2 \hat{\alpha}_i^2}{\lambda_i^2 g^2} - \frac{2 \hat{\sigma}^2 e}{\lambda_i^{j+2} g} \sum_{i=1}^{p} \lambda_i^{-j-2}} \right],
\end{align*}
\]

from which it follows

\[
\begin{align*}
\mathcal{R}(\hat{\alpha}(K), \alpha, \sigma^2) & \leq \mathcal{R}(\hat{\alpha}, \alpha, \sigma^2) + \frac{1}{\sigma^2} l_1 \sigma^2 \mathbb{E}^f \left[ \frac{\lambda_i (l_1 \hat{\sigma}^2 + 2l_1 m^{-1} \hat{\sigma}^2 + 4 \hat{\sigma}^2) - \frac{2 \hat{\sigma}^2 e}{\lambda_i^2 g} \sum_{i=1}^{p} \lambda_i^{-j-2}} \right] \sum_{i=1}^{p} \lambda_i^{j+2} g_i^{-2},
\end{align*}
\]
(3.11)

where \( \frac{\hat{\alpha}' \hat{\alpha}}{g} \leq 1 \) is used in the last line.

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

\[
\begin{align*}
\mathbb{E}^f \left[ \frac{\hat{\sigma}^2}{g} \left\{ \frac{\lambda_i (l_1 + 2l_1 m^{-1} + 4)}{\lambda_i^2} - 2 \sum_{i=1}^{p} \lambda_i^{-j-2} \right\} \right] & = \left\{ \frac{\lambda_i (l_1 + 2l_1 m^{-1} + 4)}{\lambda_i^2} - 2 \sum_{i=1}^{p} \lambda_i^{-j-2} \right\} \mathbb{E}^f \left[ \frac{\hat{\sigma}^2}{g} \right],
\end{align*}
\]

a sufficient condition for the inequality (2.3) is

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

which is implied by the conditions (i) and (ii). This completes the proof. \( \Box \)
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 \( p = 6, n = 20, \Lambda = (1, 0.01, 0.0019, 0.0015, 0.001) \) and \( \beta_i = 0, 0.5, 1, 1.5, 2 \) for each \( i \). The entries of \( \Lambda \) indicate the presence of multicollinearity. As a distribution of the error term \( e \), 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 \( R_0(\hat{\beta}(K), \beta, \sigma)/R_0(\hat{\beta}, \beta, \sigma) \) 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 condition (i) is better for improvement of relative efficiency.

(b) Remarkable improvement is not shown in terms of selecting \( l_2 \). Better choice may be selecting larger values limited 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

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

Table 4.1: Relative risk in case of $N_0$

<table>
<thead>
<tr>
<th>$\beta_i$</th>
<th>$l_2$</th>
<th>$0.25l_i^2$</th>
<th>$0.5l_i^2$</th>
<th>$0.75l_i^2$</th>
<th>$l_i^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$-2l_i$</td>
<td>0.968</td>
<td>0.942</td>
<td>0.920</td>
<td>0.901</td>
</tr>
<tr>
<td></td>
<td>$-1.5l_i$</td>
<td>0.968</td>
<td>0.941</td>
<td>0.918</td>
<td>0.897</td>
</tr>
<tr>
<td></td>
<td>$-l_i$</td>
<td>0.967</td>
<td>0.939</td>
<td>0.915</td>
<td>0.893</td>
</tr>
<tr>
<td></td>
<td>$-0.5l_i$</td>
<td>0.966</td>
<td>0.937</td>
<td>0.911</td>
<td>0.888</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.965</td>
<td>0.935</td>
<td>0.908</td>
<td>0.882</td>
</tr>
<tr>
<td></td>
<td>$0.5l_i$</td>
<td>0.965</td>
<td>0.933</td>
<td>0.903</td>
<td>0.876</td>
</tr>
<tr>
<td></td>
<td>$l_i$</td>
<td>0.964</td>
<td>0.931</td>
<td>0.899</td>
<td>0.870</td>
</tr>
<tr>
<td>0.5</td>
<td>$-2l_i$</td>
<td>0.968</td>
<td>0.941</td>
<td>0.919</td>
<td>0.900</td>
</tr>
<tr>
<td></td>
<td>$-1.5l_i$</td>
<td>0.967</td>
<td>0.940</td>
<td>0.917</td>
<td>0.896</td>
</tr>
<tr>
<td></td>
<td>$-l_i$</td>
<td>0.966</td>
<td>0.938</td>
<td>0.913</td>
<td>0.891</td>
</tr>
<tr>
<td></td>
<td>$-0.5l_i$</td>
<td>0.965</td>
<td>0.936</td>
<td>0.910</td>
<td>0.886</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.964</td>
<td>0.934</td>
<td>0.906</td>
<td>0.880</td>
</tr>
<tr>
<td></td>
<td>$0.5l_i$</td>
<td>0.965</td>
<td>0.932</td>
<td>0.902</td>
<td>0.874</td>
</tr>
<tr>
<td></td>
<td>$l_i$</td>
<td>0.964</td>
<td>0.929</td>
<td>0.897</td>
<td>0.866</td>
</tr>
<tr>
<td>1</td>
<td>$-2l_i$</td>
<td>0.966</td>
<td>0.939</td>
<td>0.917</td>
<td>0.897</td>
</tr>
<tr>
<td></td>
<td>$-1.5l_i$</td>
<td>0.966</td>
<td>0.938</td>
<td>0.914</td>
<td>0.893</td>
</tr>
<tr>
<td></td>
<td>$-l_i$</td>
<td>0.965</td>
<td>0.936</td>
<td>0.910</td>
<td>0.888</td>
</tr>
<tr>
<td></td>
<td>$-0.5l_i$</td>
<td>0.964</td>
<td>0.934</td>
<td>0.907</td>
<td>0.882</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.963</td>
<td>0.931</td>
<td>0.902</td>
<td>0.876</td>
</tr>
<tr>
<td></td>
<td>$0.5l_i$</td>
<td>0.963</td>
<td>0.929</td>
<td>0.898</td>
<td>0.869</td>
</tr>
<tr>
<td></td>
<td>$l_i$</td>
<td>0.962</td>
<td>0.926</td>
<td>0.893</td>
<td>0.863</td>
</tr>
<tr>
<td>1.5</td>
<td>$-2l_i$</td>
<td>0.966</td>
<td>0.939</td>
<td>0.916</td>
<td>0.897</td>
</tr>
<tr>
<td></td>
<td>$-1.5l_i$</td>
<td>0.966</td>
<td>0.938</td>
<td>0.913</td>
<td>0.892</td>
</tr>
<tr>
<td></td>
<td>$-l_i$</td>
<td>0.965</td>
<td>0.936</td>
<td>0.910</td>
<td>0.887</td>
</tr>
<tr>
<td></td>
<td>$-0.5l_i$</td>
<td>0.964</td>
<td>0.934</td>
<td>0.906</td>
<td>0.882</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.964</td>
<td>0.931</td>
<td>0.902</td>
<td>0.876</td>
</tr>
<tr>
<td></td>
<td>$0.5l_i$</td>
<td>0.963</td>
<td>0.929</td>
<td>0.898</td>
<td>0.868</td>
</tr>
<tr>
<td></td>
<td>$l_i$</td>
<td>0.962</td>
<td>0.926</td>
<td>0.892</td>
<td>0.860</td>
</tr>
<tr>
<td>2</td>
<td>$-2l_i$</td>
<td>0.967</td>
<td>0.940</td>
<td>0.918</td>
<td>0.898</td>
</tr>
<tr>
<td></td>
<td>$-1.5l_i$</td>
<td>0.966</td>
<td>0.939</td>
<td>0.915</td>
<td>0.894</td>
</tr>
<tr>
<td></td>
<td>$-l_i$</td>
<td>0.966</td>
<td>0.937</td>
<td>0.912</td>
<td>0.889</td>
</tr>
<tr>
<td></td>
<td>$-0.5l_i$</td>
<td>0.965</td>
<td>0.935</td>
<td>0.908</td>
<td>0.884</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.964</td>
<td>0.933</td>
<td>0.904</td>
<td>0.878</td>
</tr>
<tr>
<td></td>
<td>$0.5l_i$</td>
<td>0.964</td>
<td>0.931</td>
<td>0.900</td>
<td>0.871</td>
</tr>
<tr>
<td></td>
<td>$l_i$</td>
<td>0.963</td>
<td>0.928</td>
<td>0.895</td>
<td>0.864</td>
</tr>
</tbody>
</table>

Table 4.2: Relative risk in case of $t_0$
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