CIRJE-F-25

Estimating Risk and Mean Squared Error Matrix in Stein Estimation

T. Kubokawa and M.S. Srivastava
University of Tokyo and University of Toronto

October 1998

Estimating Risk and Mean Squared Error Matrix in Stein Estimation

T. Kubokawa and M.S. Srivastava
University of Tokyo and *University of Toronto*

It is well known that the uniformly minimum variance unbiased (UMVU) estimators of the risk and the mean squared error (MSE) matrix proposed in the literature for Stein estimators can take negative values with positive probability. In this paper, improved truncated estimators of the risk, risk difference, and MSE matrix are proposed and shown to be better than the UMVU estimators in terms of mean squared error.

Key words and phrases: Inadmissibility, quadratic loss, uniformly minimum variance unbiased estimators, risk, risk difference, truncated estimators of risk.

AMS (1991) subject classifications: Primary 62F11, Secondary 62F10, 62J07

1 Introduction

Stein estimators or more generally shrinkage estimators have found many applications. Efron and Morris (1975) used it in predicting batting average of baseball players and estimating the Toxoplamosis prevalence rates. Fay and Herriot (1979) used it in estimating income for small places, Battese, Harter and Fuller (1985) used it in prediction of county crop areas and Tsutakawa, Shoop and Marienfied (1985) used it to estimate cancer mortality rates. More recently Breiman and Friedman (1997) showed that the shrinkage estimators performs better in multivariate regression model with many independent variables. To assess the performance of these estimators, however, researchers often resort to Monte Carlo methods. Thus, it would be desirable to provide estimates of the expected risk / risk difference of the existing and Stein estimators etc. Although uniformly minimum variance unbiased (UMVU) estimators are available for many cases, they have the shortcoming of taking negative values with positive probability. See Stein (1973), Efron and Morris (1976), Bilodeau and Srivastava (1988), Carter, Srivastava, Srivastava and Ullah (1990) for UMVU estimators.

Venter and Steel (1990) considered a truncated estimator for the risk of the positive-part Stein estimator and showed numerically its superiority over UMVU estimators. But a systematic theoretical study is lacking although some related estimating problems of loss functions have been considered by Johnston (1987), Rukhin (1987) and Lu and Berger (1989).

The objective of this paper is to present estimators of the risk, risk difference and mean squared error matrix etc, which are better than the UMVU estimators in terms of the mean squared error (MSE) matrix and take positive values with probability one.

In Section 2, we consider the estimation of a risk function of the Stein type shrinkage estimators with respect to a scale-invariant loss function, and propose positive truncated estimators improving on the UMVU estimator of the risk. Also the problem of estimating the MSE matrix divided by a dispersion parameter is treated, and improved positive procedures are derived. In Section 3, we propose estimators for the difference of the MSE / MSE matrices of shrinkage estimators and the maximum likelihood estimator. In Section 4, we discuss the problem of estimating the MSE and MSE matrix of the shrinkage estimators.

2 Improved Non-Negative Estimators of the Risk

2.1 Notations and unbiased estimators of the risk

Let $\boldsymbol{X} = (X_1, \ldots, X_p)'$ be a p-dimensional random vector distributed as normal with mean vector $\boldsymbol{\theta}$ and covariance matrix $\sigma^2 \boldsymbol{I}$, denoted by $\mathcal{N}_p(\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$. Let S be distributed as $\sigma^2 \chi_n^2$, where χ_n^2 denotes a chi-square distribution with n degrees of freedom. It will be assumed that $\boldsymbol{X}$ and S are independently distributed and that $p \geq 3$. Consider the problem of estimating $\boldsymbol{\theta}$ by an estimator $\boldsymbol{\delta}$ with an invariant loss function given by

$$L_I(\boldsymbol{\theta}, \boldsymbol{\delta}) = ||\boldsymbol{\delta} - \boldsymbol{\theta}||^2 / \sigma^2,$$

where $||\boldsymbol{z}||^2 = \boldsymbol{z}'\boldsymbol{z}$ for column vector $\boldsymbol{z} \in \boldsymbol{R}^p$. The risk function of $\boldsymbol{\delta}$ is given by

$$\begin{aligned} R_I(\boldsymbol{\delta}) &\equiv E[L_I(\boldsymbol{\theta}, \boldsymbol{\delta})] \\ &= (\sigma^2)^{-1} E[L(\boldsymbol{\theta}, \boldsymbol{\delta})] \equiv (\sigma^2)^{-1} R(\boldsymbol{\delta}), \text{ say.} \end{aligned} \tag{2.1}$$

Stein-type estimators that have smaller risks than the maximum likelihood (ML) and UMVU estimator $\boldsymbol{X}$ is given by

$$\boldsymbol{\delta}_g = \left\{1 - \frac{g(W)}{W}\right\} \boldsymbol{X}, \quad \text{for} \quad W = \frac{||\boldsymbol{X}||^2}{S}$$

where the shrinkage function $g(w)$ satisfies the conditions

(a) $g(w)$ is absolutely continuous and non-decreasing,

(b) $0 < g(w) \leq 2(p-2)/(n+1)$.

The estimator proposed by James and Stein (1961), denoted by $\boldsymbol{\delta}^{JS}$ is obtained by putting

$$g(w) = k \equiv (p-2)/(n+2)$$

and the positive rule estimator denoted by $\boldsymbol{\delta}^{S+}$ is obtained by choosing $g(w) = \min\{k, w\}$. The risk function for the ML estimator $\boldsymbol{\delta} = \boldsymbol{X}$ is given by

$$R_I(\boldsymbol{X}) = p,$$

and the risk function for the Stein-type estimators $\boldsymbol{\delta}_g$ is given by

$$R_I(\boldsymbol{\delta}_g) = p - E[a(W)]$$

where

$$a(W) = 2(p-2)\frac{g(W)}{W} - (n+2)\frac{g^2(W)}{W} + 4g'(W) + 4g(W)g'(W). \tag{2.2}$$

Clearly, then the UMVU estimator of $R_I(\boldsymbol{\delta}_g)$ is given by

$$\widehat{R}_I^{UB}(\boldsymbol{\delta}_g) = p - a(W), \tag{2.3}$$

which can take negative values with positive probability. For example, consider the case when $g(W) = k = (p-2)/(n+2)$, giving the James-Stein estimator. For this case,

$$a(W) = \frac{(p-2)^2}{(n+2)W}$$

and hence for $W < (p-2)^2/\{p(n+2)\}$, $p - a(W) < 0$. Since W can be less than $(p-2)^2/\{p(n+2)\}$ with positive probability, the UMVU estimator will be negative with positive probability.

The first objective of this section is to propose an improved positive estimator of $R_I(\boldsymbol{\delta})$ that has smaller mean squared error than the UMVU estimator. These results are generalized to the situation where the covariance matrix of the random vector $\boldsymbol{X}$ is an unknown $p \times p$ positive definite $\boldsymbol{\Sigma}$ which is estimated by a $p \times p$ random symmetric matrix $\boldsymbol{S}$ being distributed as $\mathcal{W}_p(\boldsymbol{\Sigma}, n)$, $n \geq p$, independently of $\boldsymbol{X}$.

The second objective is to consider the problem of estimating the MSE matrix divided by σ^2, namely,

$$\begin{aligned} \boldsymbol{M}_I(\boldsymbol{\delta}_g) &\equiv \sigma^{-2}E[(\boldsymbol{\delta}_g - \boldsymbol{\theta})(\boldsymbol{\delta}_g - \boldsymbol{\theta})'] \\ &\equiv \sigma^{-2}\boldsymbol{M}(\boldsymbol{\delta}_g), \quad \text{say}, \end{aligned} \tag{2.4}$$

and to propose positive estimators improving on the UMVU estimator.

Note that the UMVU estimator for $\boldsymbol{M}_I(\boldsymbol{\delta}_g)$ is provided by

$$\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}_g) = \left\{1 - \frac{2g(W)}{W}\right\}\boldsymbol{I} + b(W)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2}, \tag{2.5}$$

where

$$b(W) = \frac{4g(W)}{W} + (n+2)\frac{g^2(W)}{W} - 4g'(W) - 4g(W)g'(W). \tag{2.6}$$

For the case of the James-Stein estimator $\boldsymbol{\delta}^{JS} = (1 - k/W)\boldsymbol{X}$, the UMVU estimator of $\boldsymbol{M}_I(\boldsymbol{\delta}^{JS})$ simplifies to

$$\begin{aligned}\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}^{JS}) &= \boldsymbol{I}_p - 2\frac{k}{W}\boldsymbol{I}_p + (p+2)\frac{k}{W}\frac{\boldsymbol{X}\boldsymbol{X}'}{\|\boldsymbol{X}\|^2} \\ &= \boldsymbol{\Gamma}'\left[\left(1+\frac{pk}{W}\right)\boldsymbol{E}_{11} + \left(1-2\frac{k}{W}\right)(\boldsymbol{I}_p - \boldsymbol{E}_{11})\right]\boldsymbol{\Gamma}, \end{aligned} \tag{2.7}$$

where $\boldsymbol{\Gamma}$ is a $p \times p$ orthogonal matrix such that $\boldsymbol{\Gamma}\boldsymbol{X} = (\|\boldsymbol{X}\|, 0, \dots, 0)'$ and $\boldsymbol{E}_{11}$ is a $p \times p$ matrix which has one for $(1,1)$-element and zero for the others. This expression indicates that the first(largest) eigenvalue is much larger than one for smaller W and others have possibilities of taking unreasonable negative values.

The problems of estimating $R_I(\boldsymbol{\delta}_g)$ and $\boldsymbol{M}_I(\boldsymbol{\delta}_g)$ are discussed in Sections 2.2 and 2.3, and the estimations of the MSE $R(\boldsymbol{\delta}_g)$ and the MSE matrix $\boldsymbol{M}(\boldsymbol{\delta}_g)$ are treated in Section 4.

2.2 Estimation of $R_I(\delta_g)$.

Let

$$\begin{aligned} a_{p,n} &= E_{\lambda=0}[a(W)] \\ &= \int_0^\infty \int_0^\infty a(u/v) f_p(u) f_n(v)\,du\,dv, \end{aligned}$$

where $\lambda = \|\boldsymbol{\theta}\|^2/\sigma^2$, and $f_p(u)$ and $f_n(v)$ denote densities of chi-square distributions with p and n degrees of freedom, respectively. Our aim in this section is to provide an improved positive estimator of $R_I(\boldsymbol{\delta}_g)$ over the UMVU estimator $\widehat{R}_I^{UB}(\boldsymbol{\delta}_g) = p - a(W)$ for $a(W)$ given by (2.2). We propose a class of estimators of the type

$$\widehat{R}_I(\phi; \boldsymbol{\delta}_g) = p - a(W)\phi(W), \tag{2.8}$$

where $\phi(w)$ is an absolutely continuous function. Under the conditions of Theorem 1, it will be shown that it has a smaller mean squared error

$$E[\{\widehat{R}_I(\phi;\boldsymbol{\delta}_g) - R_I(\boldsymbol{\delta}_g)\}^2]$$

than the UMVU estimator $\widehat{R}_I^{UB}(\boldsymbol{\delta}_g)$.

Theorem 1. *Assume that* $E[\{\widehat{R}_I^{UB}(\boldsymbol{\delta}_g)\}^2] < \infty$, $R_I(\boldsymbol{\delta}_g) < \infty$ *and* $p \geq 3$. *Assume that*

(a) $a(w)$ *is nonincreasing in* w,

(b) $\phi(w)$ *is absolutely continuous, nondecreasing in* w *and* $\lim_{w\to\infty}\phi(w) = 1$,

(c) $\phi(w) \geq \phi^{TR}(w)$, *where*

$$\phi^{TR}(w) = \min\{1,\ a_{p,n}/a(w)\}.$$

Then $\widehat{R}_I(\phi;\boldsymbol{\delta}_g)$ *dominates* $\widehat{R}_I^{UB}(\boldsymbol{\delta}_g)$.

Table 1. Values of $a_{p,n}$ for the Estimator $\boldsymbol{\delta}^{S+}$

n	p = 3	5	7	10	20
3	0.857799	2.268824	3.611723	5.683447	12.355029
5	1.025757	2.692283	4.269549	6.594915	14.436470
7	1.115159	2.892845	4.596360	7.153935	15.492499
10	1.182874	3.080593	4.892471	7.547653	16.438395

Kubokawa (1988) showed that the condition (a) implies the monotonicity of the risk function, so that (a) guarantees that $E_\lambda[a(W)] \geq E_{\lambda=0}[a(W)] = a_{p,n}$ (also see Casella (1990)). It can be seen that the condition (a) is satisfied for the James-Stein estimator $\boldsymbol{\delta}^{JS}$ with $g(w) = k = (p-2)/(n+2)$ and the positive-part Stein estimator $\boldsymbol{\delta}^{S+}$ with $g(w) = \min(k,\ w)$. For the James-Stein estimator $\boldsymbol{\delta}^{JS}$, it is seen that $a_{p,n} = (p-2)^2(n+2)^{-1}E[\chi_n^2/\chi_p^2] = n(p-2)(n+2)^{-1}$. For the positive-part Stein estimator $\boldsymbol{\delta}^{S+}$, we can numerically obtain the values of $a_{p,n}$, which are given in Table 1 for $p = 3, 5, 7, 20$, and $n = 3, 5, 7, 20$.

The monotonicity and boundedness of $g(w)$ stated in Section 2.1 implies that $\lim_{w\to\infty} g'(w) = 0$ or $\lim_{w\to\infty} a(w) = 0$, so that ϕ^{TR} satisfies the condition (b). Thus we get improved estimators

$$\begin{aligned} \widehat{R}_I^{TR}(\boldsymbol{\delta}_g) &= \widehat{R}_I(\phi^{TR}; \boldsymbol{\delta}_g) \\ &= p - a(W)\phi^{TR}(W) \\ &= \max\{p - a(W),\ p - a_{p,n}\}. \end{aligned} \tag{2.9}$$

Since $p - a_{p,n}$ is the value of risk $R_I(\boldsymbol{\delta}_g)$ at $\lambda = 0$, $\widehat{R}_I^{TR}(\boldsymbol{\delta}_g)$ is always positive. Hence $\widehat{R}_I^{TR}(\boldsymbol{\delta}_g)$ is eliminating the drawback of the unbiased estimator $\widehat{R}_I^{UB}(\boldsymbol{\delta}_g)$ as well as improving upon it. For the James-Stein estimator, $\widehat{R}_I^{TR}(\boldsymbol{\delta}^{JS})$ is given by

$$\widehat{R}_I^{TR}(\boldsymbol{\delta}^{JS}) = \max\left\{ \widehat{R}_I^{UB}(\boldsymbol{\delta}^{JS}),\ \frac{2(n+p)}{n+2} \right\}. \tag{2.10}$$

It is seen that $E[\{\widehat{R}_I^{TR}(\boldsymbol{\delta}^{JS})\}^2] < \infty$ for $p \geq 1$ while the condition that $p \geq 5$ is required for $E[\{\widehat{R}_I^{UB}(\boldsymbol{\delta}^{JS})\}^2] < \infty$. The truncated rule (2.9) is also reasonable because the parameter space of $R_I(\boldsymbol{\delta}_g)$ is restricted to the interval $[p - a_{p,n},\ p]$. This fact was indicated by Venter and Seel (1990), who considered such a truncated estimator for the risk reduction $E_\lambda[a(W)]$ of $\boldsymbol{\delta}^{S+}$ and revealed numerically the superiority of it.

The condition (a) about the monotonicity of $a(w)$ may be restrictive. If the conditions in Theorem 1 cannot be verified, then one can get the simple non-negative improved estimator

$$\left[\widehat{R}_I^{UB}(\boldsymbol{\delta}_g)\right]_c^+ = \begin{cases} \widehat{R}_I^{UB}(\boldsymbol{\delta}_g) & \text{if } \widehat{R}_I^{UB}(\boldsymbol{\delta}_g) \geq 0, \\ -\widehat{R}_I^{UB}(\boldsymbol{\delta}_g) & \text{if } -c \leq \widehat{R}_I^{UB}(\boldsymbol{\delta}_g) < 0, \\ c & \text{if } \widehat{R}_I^{UB}(\boldsymbol{\delta}_g) < -c, \end{cases}$$

where c is a non-negative constant suitably chosen. When $c = 0$, $[\widehat{R}_I^{UB}(\boldsymbol{\delta}_g)]_0^+$ is a procedure truncated at zero. When $c > 0$ and $P[\widehat{R}_I^{UB}(\boldsymbol{\delta}_g) = 0] = 0$, then $[\widehat{R}_I^{UB}(\boldsymbol{\delta}_g)]_c^+$ is almost surely positive. Practically small $c > 0$ should be chosen, but the choice includes the arbitrariness.

More generally, we consider the problem of estimating positive parameter α by estimator $\hat{\alpha}$. If $\hat{\alpha}$ take negative values with positive probability, we can consider non-negative estimator

$$[\hat{\alpha}]_c^+ = \begin{cases} \hat{\alpha} & \text{if } \hat{\alpha} \geq 0, \\ -\hat{\alpha} & \text{if } -c \leq \hat{\alpha} < 0, \\ c & \text{if } \hat{\alpha} < -c. \end{cases} \tag{2.11}$$

Proposition 1. *The non-negative estimator $[\hat{\alpha}]_c^+$ is better than $\hat{\alpha}$ in terms of the MSE criterion.*

This proposition can be shown easily. Denote the set of $-c \leq \hat{\alpha} < 0$ and the set of $\hat{\alpha} < -c$ by A_1 and A_2, respectively. Let I_{A_i} be the indicator function for $i = 1, 2$. Then the risk difference is written by

$$\begin{aligned} & E[(\hat{\alpha} - \alpha)^2] - E[([\hat{\alpha}]_c^+ - \alpha)^2] \\ &= E\left[(\hat{\alpha} - [\hat{\alpha}]_c^+)(\hat{\alpha} + [\hat{\alpha}]_c^+ - 2\alpha)\right] \\ &= E\left[(\hat{\alpha} - (-\hat{\alpha}))(\hat{\alpha} + (-\hat{\alpha}) - 2\alpha) I_{A_1}\right] \\ &\quad + E\left[(\hat{\alpha} - c)(\hat{\alpha} + c - 2\alpha) I_{A_2}\right], \end{aligned}$$

both of which can be seen to be nonnegative by taking the definitions of A_1 and A_2 into account.

We here provide the results of Monte Carlo simulation for the MSE of the estimators of the risks. For the estimation of the risk of the James-Stein estimator $\boldsymbol{\delta}^{JS}$, we first compute the values of $R_I(\boldsymbol{\delta}^{JS})$ based on 100,000 replications and then obtain the values of $E[\{\widehat{R}_I(\boldsymbol{\delta}^{JS}) - R_I(\boldsymbol{\delta}^{JS})\}^2]$ based on 50,000 replications. Three types of estimators $\widehat{R}_I^{UB}(\boldsymbol{\delta}^{JS})$, $\widehat{R}_I^{TR}(\boldsymbol{\delta}^{JS})$ and $\left[\widehat{R}_I^{UB}(\boldsymbol{\delta}^{JS})\right]_c^+$, denoted by UB, TR and NN(c), are treated for $c = 0.0, 0.5$ and 1.0. Figure 1 illustrates the MSE behaviors of the estimators where $p = 10$, $n = 3$, $\sigma^2 = 1$ and $\theta_i = (i/10)t$ for $i = 1, \ldots, p$ and $0 \leq t \leq 10$. This figure reveals that

(1) the MSE of each estimator is decreasing in t or the noncentrality parameter $\lambda = ||\boldsymbol{\theta}||^2/\sigma^2$,

(2) $\widehat{R}_I^{TR}(\boldsymbol{\delta}^{JS})$ has the best performance among the five estimators and a significant risk gain at $t = 0$,

(3) $\widehat{R}_I^{TR}(\boldsymbol{\delta}^{JS})$ dominates $\left[\widehat{R}_I^{UB}(\boldsymbol{\delta}^{JS})\right]_c^+$, being better than $\widehat{R}_I^{UB}(\boldsymbol{\delta}^{JS})$,

(4) three estimators $\left[\widehat{R}_I^{UB}(\boldsymbol{\delta}^{JS})\right]_c^+$ with $c = 0.0, 0.5$ and 1.0 has similar MSE performances.

These observations propose the use of the truncated estimator $\widehat{R}_I^{TR}(\boldsymbol{\delta}^{JS})$.

The same manner of the simulation is applied to get the MSE bahaviours of estimators $\widehat{R}_I^{UB}(\boldsymbol{\delta}^{S+})$, $\widehat{R}_I^{TR}(\boldsymbol{\delta}^{S+})$ and $\left[\widehat{R}_I^{UB}(\boldsymbol{\delta}^{S+})\right]_c^+$, denoted by UB, TR and NN(c), for $c = 0.0, 0.5$ and 1.0. Figures 2 and 3 provide their MSE behaviors for $p = 3$ and $p = 10$, respectively, and demonstrate that they perform similarly to the case of $\boldsymbol{\delta}^{JS}$. These indicate the superiority of $\widehat{R}_I^{TR}(\boldsymbol{\delta}^{S+})$.

The monotonicity of the MSEs of the estimators is one of remarkable properties observed from Figures 1, 2 and 3. This property can be also verified analytically for the MSE of $\widehat{R}_I^{UB}(\boldsymbol{\delta}^{JS})$ for the James-Stein estimator $\boldsymbol{\delta}^{JS}$. For the non-central chi-square random variable $\chi_p^2(\lambda)$, we note that

$$E[h(\chi_p^2(\lambda))] = E^J\left[E[h(\chi_{p+2J}^2)\,\big|J]\right], \tag{2.12}$$

where J follows a Poisson distribution with mean $\lambda/2$. The MSE of $\widehat{R}_I^{UB}(\boldsymbol{\delta}^{JS})$ can be easily rewritten by

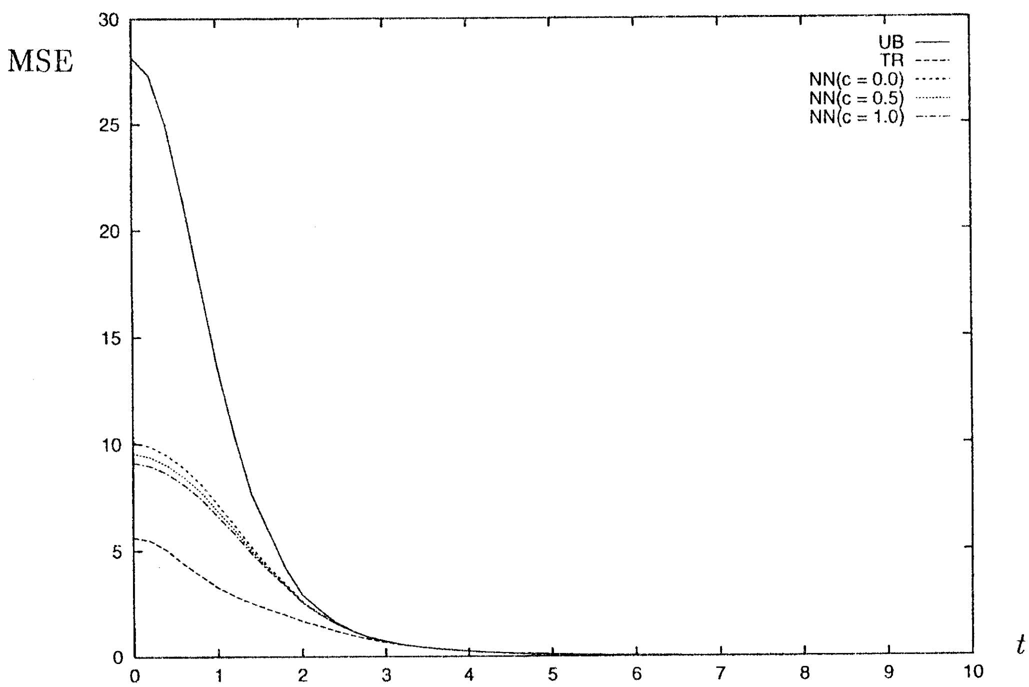

Figure 1. MSE of Estimators $\widehat{R}_I^{UB}(\boldsymbol{\delta}^{JS})$, $\widehat{R}_I^{TR}(\boldsymbol{\delta}^{JS})$ and $\left[\widehat{R}_I^{UB}(\boldsymbol{\delta}^{JS})\right]_c^+$ for $c = 0.0, 0.5, 1.0$ and $p = 10$
(These estimators are denoted by UB, TR and NN(c) in the figure.)

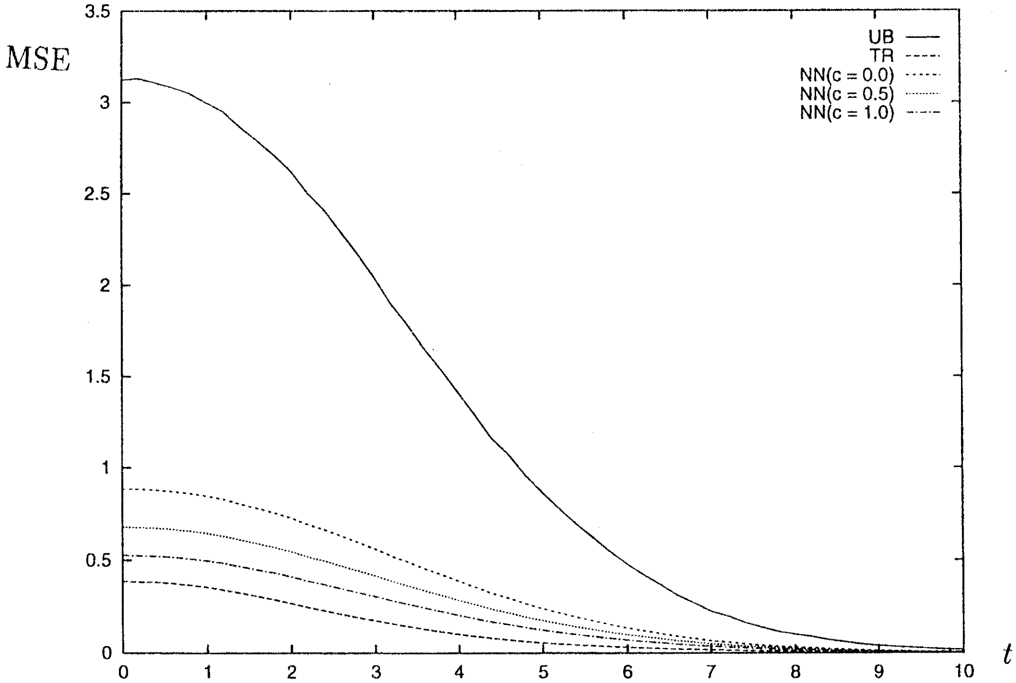

Figure 2. MSE of Estimators $\widehat{R}_I^{UB}(\boldsymbol{\delta}^{S+})$, $\widehat{R}_I^{TR}(\boldsymbol{\delta}^{S+})$ and $\left[\widehat{R}_I^{UB}(\boldsymbol{\delta}^{S+})\right]_c^+$ for $c = 0.0, 0.5, 1.0$ and $p = 3$
(These estimators are denoted by UB, TR and NN(c) in the figure.)

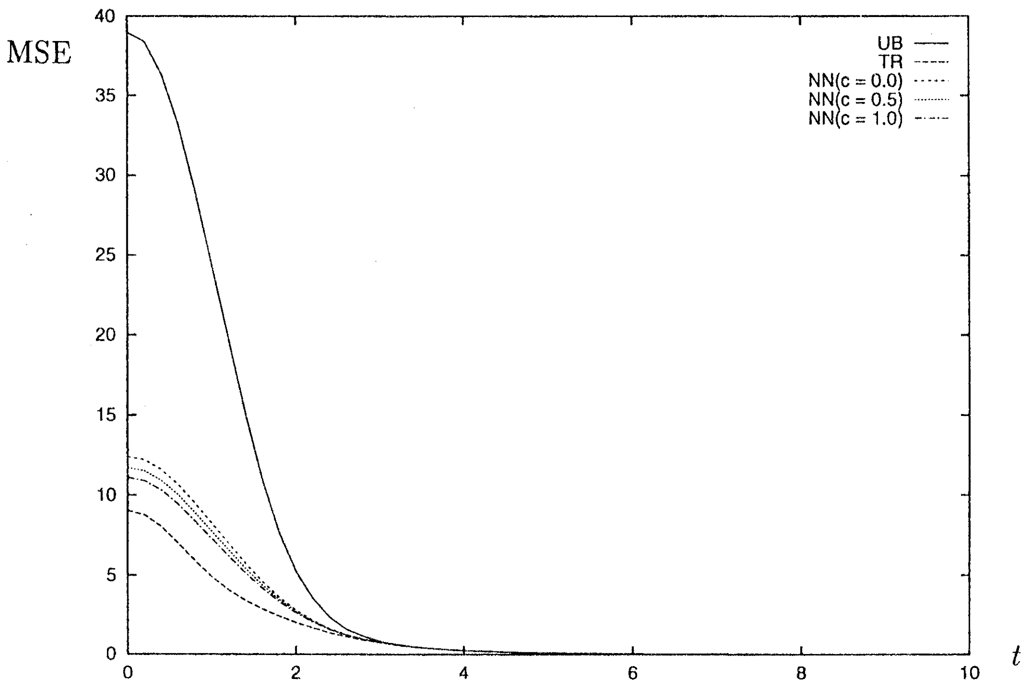

Figure 3. MSE of Estimators $\widehat{R}_I^{UB}(\boldsymbol{\delta}^{S+})$, $\widehat{R}_I^{TR}(\boldsymbol{\delta}^{S+})$ and $\left[\widehat{R}_I^{UB}(\boldsymbol{\delta}^{S+})\right]_c^+$ for $c = 0.0, 0.5, 1.0$ and $p = 10$

$$
\begin{aligned}
&MSE^{JS}(\lambda)\\
&= E\left[\left\{\widehat{R}_I^{UB}(\delta^{JS}) - R_I(\delta^{JS})\right\}^2\right]\\
&= c^2 n\left\{E[(n+2)(p-2+2J)^{-1}(p-4+2J)^{-1}] - n\left(E[(p-2+2J)^{-1}]\right)^2\right\}
\end{aligned}
$$

for $c = (p-2)^2/(n+2)$, which gives $MSE^{JS}(0) = 2n(p-2)^2(n+p-2)/\{(n+2)^2(p-4)\}$. From the proof of Lemma 2 given by Kubokawa(1988), differentiating $MSE^{JS}(\lambda)$ with respect to λ gives that

$$
\begin{aligned}
&\frac{c^2 n}{2}\left\{(n+2)E\left[(p-2+2J)^{-1}\left\{(p+2J)^{-1} - (p-4+2J)^{-1}\right\}\right]\right.\\
&\qquad \left. -2nE\left[(p-2+2J)^{-1}\right]E\left[(p+2J)^{-1} - (p-2+2J)^{-1}\right]\right\}\\
&\le 2c^2n^2\left\{-E\left[(p+2J)^{-1}(p-2+2J)^{-1}(p-4+2J)^{-1}\right]\right.\\
&\qquad \left. +E\left[(p-2+2J)^{-1}\right]E\left[(p+2J)^{-1}(p-2+2J)^{-1}\right]\right\}\\
&\le 0
\end{aligned}
$$

since $(p-2+2J)^{-1}$ and $(p+2J)^{-1}(p-2+2J)^{-1}$ are decreasing in J. This demonstrates that $MSE^{JS}(\lambda)$ is decreasing in λ.

Proof of Theorem 1. Letting

$$
\begin{aligned}
C(\widehat{R}_I(\phi)) &= E[\{\widehat{R}_I(\boldsymbol{\delta}_g;\phi) - R_I(\boldsymbol{\delta}_g)\}^2]\\
&= E[\{a(W)\phi(W) - E[a(W)]\}^2],
\end{aligned}
$$

we can apply the *IERD (Integral Expression of Risk Difference) method* given by Kubokawa (1994, 1998a,b). From the condition (b), we note that

$$
\begin{aligned}
&C(\widehat{R}_I^{UB}) - C(\widehat{R}_I(\phi))\\
&= E[\{a(W)\phi(\infty) - E[a(W)\}^2] - E[\{a(W)\phi(W) - E[a(W)]\}^2]\\
&= E\left[\{a(W)\phi(tW) - E[a(W)]\}^2\Big|_{t=1}^{\infty}\right].
\end{aligned}
$$

Then from the absolute continuity of $\phi(w)$, we have that

$$
\begin{aligned}
&C(\widehat{R}_I^{UB}) - C(\widehat{R}_I(\phi))\\
&= E\left[\int_1^\infty \frac{d}{dt}\{a(W)\phi(tW) - E[a(W)]\}^2\, dt\right]\\
&= 2E\left[\int_1^\infty \{a(W)\phi(tW) - E[a(W)]\}\, a(W)\phi'(tW)W dt\right]\\
&= 2\int\int\int_1^\infty \left\{a\left(\frac{u}{v}\right)\phi\left(t\frac{u}{v}\right) - E[a(W)]\right\} a\left(\frac{u}{v}\right)\phi'\left(t\frac{u}{v}\right)\frac{u}{v}dt\\
&\qquad \times f_p(u;\lambda)f_n(v)dudv, \qquad (2.13)
\end{aligned}
$$

where $f_p(u;\lambda)$ and $f_n(v)$ denote the densities of $||\boldsymbol{X}||^2/\sigma^2$ and S/σ^2, respectively. Making the transformations $w=(t/v)u$ with $dw=(t/v)du$ and $z=w/t$ with $dz=(w/t^2)dt$ in turn, we can rewrite (2.13) as

$$
\begin{aligned}
&2\int\int\int_1^\infty \left\{a\left(\frac{w}{t}\right)\phi(w)-E[a(W)]\right\}a\left(\frac{w}{t}\right)\phi'(w)\frac{vw}{t^2} \\
&\qquad\qquad\qquad\qquad \times f_p(\frac{vw}{t};\lambda)f_n(v)dtdvdw \\
&= 2\int\int\int_0^w \{a(z)\phi(w)-E[a(W)]\}a(z)\phi'(w) \\
&\qquad\qquad\qquad\qquad \times v f_p(vz;\lambda)f_n(v)dzdvdw.
\end{aligned}
$$

When $\phi'(w)=0$, $C(\widehat{R}_I^{UB})=C(\widehat{R}_I(\phi))$. When $\phi'(w)>0$, it suffices to show that

$$
\phi(w)\geq \frac{E_\lambda[a(W)]E^v[\int_0^w a(z)vf_p(vz;\lambda)dz]}{E^v[\int_0^w a^2(z)vf_p(vz;\lambda)dz]},
$$

where $E^v[\cdot]$ designates the expectation with respect to v. Hence from the condition (a), we note that

$$
\frac{\int_0^w a(z)vf_p(vz;\lambda)dz}{\int_0^w a^2(z)vf_p(vz;\lambda)dz}\leq \sup_{0<z<w}\left\{\frac{1}{a(z)}\right\}=\frac{1}{a(w)}.
$$

Also the condition (a) implies that

$$
\begin{aligned}
E_\lambda[a(W)] &= E_\lambda\left[a\left(\frac{\chi^2_{p+2J}}{\chi^2_n}\right)\right]=E_\lambda\left[a\left(\frac{\chi^2_p}{\chi^2_n}+\frac{\chi^2_{2J}}{\chi^2_n}\right)\right] \\
&\leq E_{\lambda=0}\left[a\left(\frac{\chi^2_p}{\chi^2_n}\right)\right]=a_{p,n},
\end{aligned}
$$

where J is a random variable following Poisson law $\mathcal{P}o(\lambda/2)$. Hence $C(\widehat{R}_I^{UB})-C(\widehat{R}_I(\phi))\geq 0$ if in the case of $\phi'(w)>0$, $\phi(w)\geq a_{p,n}/a(w)$, which is satisfied by the condition (c), and Theorem 1 is proved. □□

2.3 Estimation of $M_I(\delta_g)$.

We next consider the problem of estimating the MSE matrix divided by σ^2, namely, $\boldsymbol{M}_I(\boldsymbol{\delta}_g)=\sigma^{-2}E[(\boldsymbol{\delta}_g-\boldsymbol{\theta})(\boldsymbol{\delta}_g-\boldsymbol{\theta})']$. As seen in (2.7), the UMVU estimator $\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}^{JS})$ for the James-Stein estimator $\boldsymbol{\delta}^{JS}$ takes negative eigen values with positive probability. The object of this section is to obtain estimators not only eliminating this shortcoming but also improving upon the UMVU estimator.

More generally, we consider finding an estimator better than the unbiased estimator $\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}_g)$ given by (2.5) for the general type of estimator $\boldsymbol{\delta}_g$. For this purpose, the estimator we look into is of the form

$$
\begin{aligned}
\widehat{\boldsymbol{M}}_I(\psi;\boldsymbol{\delta}_g) &= \boldsymbol{I}_p-2\frac{g(W)}{W}\psi(W)\left(\boldsymbol{I}_p-\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2}\right)-2\frac{g(W)}{W}\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2}+b(W)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2} \quad (2.14)\\
&= \boldsymbol{\Gamma}'\left[\left\{1-2\frac{g(W)}{W}+b(W)\right\}\boldsymbol{E}_{11}+\left\{1-2\frac{g(W)}{W}\psi(W)\right\}(\boldsymbol{I}-\boldsymbol{E}_{11})\right]\boldsymbol{\Gamma},
\end{aligned}
$$

where

$$b(W) = 4\frac{g(w)}{W} + (n+2)\frac{g^2(W)}{W} - 4g'(W) - 4g(W)g'(W). \tag{2.15}$$

Since $(p-1)$ eigen values of $\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}_g)$ other than the first(largest) eigen value may take negative values with a positive probability, the shrinkage function $\psi(w)$ is incorporated in the $(p-1)$ eigenvalues except for the first. To investigate the dispersion of $\widehat{\boldsymbol{M}}_I(\psi;\boldsymbol{\delta}_g)$, the MSE criterion $E[\mathrm{tr}(\widehat{\boldsymbol{M}}_I(\psi;\boldsymbol{\delta}_g) - \boldsymbol{M}_I(\boldsymbol{\delta}_g))^2]$ is utilized. Let

$$\begin{aligned} b_{p,n} &= \sup_{j\geq 0}\{b_{p,n}(2j)\}, \qquad (2.16)\\ b_{p,n}(2j) &= E\left[\frac{2\chi_n^2}{\chi_{p+2j}^2} g\left(\frac{\chi_{p+2j}^2}{\chi_n^2}\right) - \frac{1}{p+2j} b\left(\frac{\chi_{p+2j}^2}{\chi_n^2}\right)\right], \end{aligned}$$

where χ_{p+2j}^2 and χ_n^2 are independently distributed as chi-square distributions with $p+2j$ and n degrees of freedom.

Theorem 2. *Assume that* $p \geq 3$, $E\left[\mathrm{tr}\left\{\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}_g)\right\}^2\right] < \infty$ *and* $\boldsymbol{M}_I(\boldsymbol{\delta}_g) < \infty$. *Assume the following conditions:*

(a) *$g(w)/w$ is nonincreasing in w,*
(b) *$E[b(\chi_{p+2j}^2/\chi_n^2)] \geq 0$ for any $j \geq 0$,*
(c) *$\psi(w)$ is nondecreasing in w and $\lim_{w\to\infty}\psi(w) = 1$,*
(d) *$\psi(w) \geq \psi^{TR}(w)$, where*

$$\psi^{TR}(w) = \min\left\{1, \ \frac{b_{p,n}}{2}\frac{w}{g(w)}\right\}.$$

Then $\widehat{\boldsymbol{M}}_I(\psi;\boldsymbol{\delta}_g)$ dominates $\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}_g)$.

It will be easy to check the condition (a). It may be troublesome to check the condition (b) and to get the value of $b_{p,n}$ for various shrinkage estimators $\boldsymbol{\delta}_g$. It is easily seen that the James-Stein estimator $\boldsymbol{\delta}^{JS}$ satisfies (a) and (b). For the derivation of $b_{p,n}$, the following expression may be helpful. For $\hat{R}_I^{UB}(\boldsymbol{\delta}_g)$ and $\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}_g)$, we note the relation given by

$$\begin{aligned} \hat{R}_I^{UB}(\boldsymbol{\delta}_g) &= \mathrm{tr}\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}_g)\\ &= p - 2p\frac{g(W)}{W} + b(W), \end{aligned}$$

which is used to get

$$\begin{aligned} b_{p,n}(2j) &= E\left[2\frac{g(W)}{W} - \frac{1}{p+2j}\left\{\hat{R}_I^{UB}(\boldsymbol{\delta}_g) - p + 2p\frac{g(W)}{W}\right\}\right]\\ &= \frac{2j}{p+2j}E\left[\frac{2g(W)}{W}\right] + \frac{1}{p+2j}\left\{p - R_I(\boldsymbol{\delta}_g)\right\}, \qquad (2.17) \end{aligned}$$

Table 2. Values of $b_{p,n}$ for the Estimator $\boldsymbol{\delta}^{S+}$

n	p = 3	5	7	10	20
3	0.801497	0.724463	0.730345	0.739707	0.749023
5	0.922034	0.849682	0.857039	0.849293	0.855735
7	0.943779	0.913598	0.908870	0.908389	0.911405
10	1.085830	0.975158	0.951344	0.958228	0.949081

for $W = \chi^2_{p+2j}/\chi^2_n$. For $\boldsymbol{\delta}^{JS}$, we can get

$$\begin{aligned} b_{p,n}(2j) &= \frac{2j}{p+2j}\frac{2kn}{p-2+2j} + \frac{1}{p+2j}\frac{(p-2)kn}{p-2+2j} \\ &= \frac{(p-2+4j)nk}{(p+2j)(p-2+2j)} \\ &\le \frac{nk}{p} = b_{p,n}, \end{aligned} \tag{2.18}$$

where the equality is attained at $j = 0$ and $j = 2$. For other estimators, however, it is hard to derive explicit values of $b_{p,n}$. We will need a computational help for maximizing $b_{p,n}(2j)$ numerically. For the positive-part Stein estimator $\boldsymbol{\delta}^{S+}$, we can numerically compute the values of $b_{p,n}$ and provide them in Table 2 for $p = 3, 5, 7, 10, 20$ and $n = 3, 5, 7, 10$. All the values of $b_{p,n}$ are attained at $j = 0$, that is, $b_{p,n} = b_{p,n}(0)$ in (2.16).

Under the conditions (a) and (b), we get the improved estimator

$$\begin{aligned} \widehat{\boldsymbol{M}}_I^{TR}(\boldsymbol{\delta}_g) &= \widehat{\boldsymbol{M}}_I(\psi^{TR}; \boldsymbol{\delta}_g) \\ &= \max\left\{1 - 2\frac{g(W)}{W},\ 1 - b_{p,n}\right\} \boldsymbol{I}_p + 2\frac{g(W)}{W}\psi^{TR}(W)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2} \\ &\quad -2\frac{g(W)}{W}\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2} + b(W)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2} \\ &= \boldsymbol{\Gamma}'\left[\left\{1 - 2\frac{g(W)}{W} + b(W)\right\}\boldsymbol{E}_{11}\right. \\ &\quad \left. + \max\left\{1 - 2\frac{g(W)}{W}, 1 - b_{p,n}\right\}(\boldsymbol{I} - \boldsymbol{E}_{11})\right]\boldsymbol{\Gamma}, \end{aligned} \tag{2.19}$$

In particular, for the James-Stein estimator, this estimator gives the form

$$\widehat{\boldsymbol{M}}_I^{TR}(\boldsymbol{\delta}^{JS}) = \boldsymbol{\Gamma}'\left[\left(1 + \frac{pk}{W}\right)\boldsymbol{E}_{11} + \max\left\{1 - \frac{2k}{W}, \frac{2(n+p)}{p(n+2)}\right\}(\boldsymbol{I}_p - \boldsymbol{E}_{11})\right]\boldsymbol{\Gamma}, \tag{2.20}$$

which is positive definite as well as better than the UMVU estimator (2.7).

Table 3. MSEs of the Estimators $\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}^{S+})$ and $\widehat{\boldsymbol{M}}_I^{TR}(\boldsymbol{\delta}^{S+})$ for $p = 3$ and $t = 0.0, 0.1, 0.5, 1.0, 2.0, 5.0, 10.0$

	0.0	0.1	0.5	1.0	2.0	5.0	10.0
$\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}^{S+})$	2.664	2.664	2.643	2.584	2.292	1.016	0.075
$\widehat{\boldsymbol{M}}_I^{TR}(\boldsymbol{\delta}^{S+})$	1.811	1.811	1.799	1.765	1.566	0.724	0.068

From (2.19) it is seen that the eigen values of $\widehat{\boldsymbol{M}}_I^{TR}(\boldsymbol{\delta}_g)$ other than the first eigen value are non-negative if $1 - b_{p,n} \geq 0$. We here state a note concerning the constant $b_{p,n}$. When $\boldsymbol{\theta} = \boldsymbol{0}$, the symmetry of the density of each X_i on zero implies that

$$E_{\theta=0}\left[b\left(\frac{||\boldsymbol{X}||^2}{S}\right)\frac{X_iX_j}{||\boldsymbol{X}||^2}\right] = 0 \quad \text{for} \quad i \neq j, \quad E_{\theta=0}\left[\frac{X_i^2}{||\boldsymbol{X}||^2}\right] = \frac{1}{p}$$

and $X_i^2/||\boldsymbol{X}||^2$, $i = 1, \ldots, p$, are independent of $||\boldsymbol{X}||$. Thus we have

$$\begin{aligned} & E_{\theta=0}\left[b\left(\frac{||\boldsymbol{X}||^2}{S}\right)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2}\right] \\ = \ & E_{\theta=0}\left[b\left(\frac{||\boldsymbol{X}||^2}{S}\right)\operatorname{diag}\left(\frac{X_1^2}{||\boldsymbol{X}||^2}, \ldots, \frac{X_p^2}{||\boldsymbol{X}||^2}\right)\right] \\ = \ & E\left[b\left(\chi_p^2/\chi_n^2\right)p^{-1}\boldsymbol{I}_p\right]. \end{aligned} \tag{2.21}$$

From (2.5), (2.16) and (2.21), it follows that

$$E_{\lambda=0}\left[\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}_g)\right] = \{1 - b_{p,n}(0)\}\boldsymbol{I}_p,$$

the value of the MSE matrix $M_I(\boldsymbol{\delta}_g)$ at $\lambda = 0$ or $\boldsymbol{\theta} = \boldsymbol{0}$. This means that

$$1 - b_{p,n}(0) > 0.$$

If we can choose $b_{p,n}$ in (2.16) as $b_{p,n} = b_{p,n}(0)$, then the $(p-1)$ eigen values of $\widehat{\boldsymbol{M}}_I^{TR}(\boldsymbol{\delta}_g)$ other than the first are always positive. The inequality (2.18) and Table 2 imply that $b_{p,n} = b_{p,n}(0)$ for $\boldsymbol{\delta}^{JS}$ and bde^{S+}.

The same manner of the simulation as used for giving Figure 1 is applied to investigate the MSE behaviors of the estimators $\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}^{S+})$ and $\widehat{\boldsymbol{M}}_I^{TR}(\boldsymbol{\delta}^{S+})$ of the risk matrix of the positive-part Stein estimator $\boldsymbol{\delta}^{S+}$. The values of their MSEs are given in Table 3 for $p = 3$ and $t = 0.0, 0.1, 0.5, 1.0, 2.0, 5.0, 10.0$, and it demonstrates that $\widehat{\boldsymbol{M}}_I^{TR}(\boldsymbol{\delta}^{S+})$ is much better than $\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}^{S+})$.

The conditions (a) and (b) of Theorem 2 may be restrictive, and the condition that $1-b_{p,n}>0$ may be hard to be checked for various estimators $\boldsymbol{\delta}_g$. In such cases, a simple non-negative definite and improved estimator is given by

$$\left\{\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}_g)\right\}_c^+ = \boldsymbol{\Gamma}'\Big[[1-2g(W)/W+b(W)]_c^+\boldsymbol{E}_{11} + [1-2g(W)/W]_c^+(\boldsymbol{I}-\boldsymbol{E}_{11})\Big]\boldsymbol{\Gamma},$$

where the truncation rule $[\cdot]_c^+$ is defined by (2.11).

Proposition 2. *The non-negative definite estimator* $\left\{\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}_g)\right\}_c^+$ *is better than* $\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}_g)$ *in terms of the MSE criterion.*

The proof is easily done as follows: For simplicity, write $\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}_g)$ by

$$\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}_g) = h_1(W)\boldsymbol{\Gamma}'\boldsymbol{E}_{11}\boldsymbol{\Gamma} + h_2(W)\boldsymbol{\Gamma}'\boldsymbol{E}_{22}\boldsymbol{\Gamma},$$

where $\boldsymbol{E}_{22}=\boldsymbol{I}-\boldsymbol{E}_{11}$, $h_1(W)=1-2g(W)/W+b(W)$ and $h_2(W)=1-2g(W)/W$. Also let $\boldsymbol{M}_i=E[h_i(W)\boldsymbol{\Gamma}'\boldsymbol{E}_{ii}\boldsymbol{\Gamma}]$ for $i=1,2$, and $\boldsymbol{M}_I=\boldsymbol{M}_I(\boldsymbol{\delta}_g)$. Then the variance of $\widehat{\boldsymbol{M}}_I^{UB}(\boldsymbol{\delta}_g)$ is written as

$$\begin{aligned}
&E\left[\operatorname{tr}\left(\widehat{\boldsymbol{M}}_I^{UB}-\boldsymbol{M}_I\right)^2\right] \\
&= E\left[\operatorname{tr}(h_1(W)\boldsymbol{\Gamma}'\boldsymbol{E}_{11}\boldsymbol{\Gamma}-\boldsymbol{M}_1)^2\right] + E\left[\operatorname{tr}(h_2(W)\boldsymbol{\Gamma}'\boldsymbol{E}_{22}\boldsymbol{\Gamma}-\boldsymbol{M}_2)^2\right] \\
&\quad +2E\left[\operatorname{tr}(h_1(W)\boldsymbol{\Gamma}'\boldsymbol{E}_{11}\boldsymbol{\Gamma}-\boldsymbol{M}_1)(h_2(W)\boldsymbol{\Gamma}'\boldsymbol{E}_{22}\boldsymbol{\Gamma}-\boldsymbol{M}_2)'\right]. \qquad (2.22)
\end{aligned}$$

For the first two terms in the r.h.s. of (2.22), the same arguments as in the proof of Proposition 1 gives that for $i=1,2$,

$$E\left[\operatorname{tr}(h_i(W)\boldsymbol{\Gamma}'\boldsymbol{E}_{ii}\boldsymbol{\Gamma}-\boldsymbol{M}_i)^2\right] \ge E\left[\operatorname{tr}\left([h_i(W)]_c^+\boldsymbol{\Gamma}'\boldsymbol{E}_{ii}\boldsymbol{\Gamma}-\boldsymbol{M}_i\right)^2\right].$$

Since $\boldsymbol{\Gamma}'\boldsymbol{E}_{11}\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\boldsymbol{E}_{22}\boldsymbol{\Gamma}=\boldsymbol{0}$, the third term is expressed as

$$-2E[h_1(W)\operatorname{tr}\boldsymbol{\Gamma}'\boldsymbol{E}_{11}\boldsymbol{\Gamma}\boldsymbol{M}_2 + h_2(W)\operatorname{tr}\boldsymbol{\Gamma}'\boldsymbol{E}_{22}\boldsymbol{\Gamma}\boldsymbol{M}_1] + 2\operatorname{tr}\boldsymbol{M}_1\boldsymbol{M}_2.$$

For the term $-2E[h_1(W)\operatorname{tr}\boldsymbol{\Gamma}'\boldsymbol{E}_{11}\boldsymbol{\Gamma}\boldsymbol{M}_2]$, the difference is written by

$$\begin{aligned}
&-2E[h_1(W)\operatorname{tr}\boldsymbol{\Gamma}'\boldsymbol{E}_{11}\boldsymbol{\Gamma}\boldsymbol{M}_2] - \left(-2E[[h_1(W)]_c^+\operatorname{tr}\boldsymbol{\Gamma}'\boldsymbol{E}_{11}\boldsymbol{\Gamma}\boldsymbol{M}_2]\right) \\
&= -4E[h_1(W)\operatorname{tr}\boldsymbol{\Gamma}'\boldsymbol{E}_{11}\boldsymbol{\Gamma}\boldsymbol{M}_2 I(-c\le h_1(W)<0)] \\
&\qquad -2E[(h_1(W)-c)\operatorname{tr}\boldsymbol{\Gamma}'\boldsymbol{E}_{11}\boldsymbol{\Gamma}\boldsymbol{M}_2 I(h_1(W)<-c)],
\end{aligned}$$

which is non-negative. Similarly,

$$-2E[h_2(W)\operatorname{tr}\boldsymbol{\Gamma}'\boldsymbol{E}_{22}\boldsymbol{\Gamma}\boldsymbol{M}_1] - \left(-2E[[h_2(W)]_c^+\operatorname{tr}\boldsymbol{\Gamma}'\boldsymbol{E}_{22}\boldsymbol{\Gamma}\boldsymbol{M}_1]\right) \ge 0,$$

and Proposition 2 is shown.

Proof of Theorem 2. For simplicity, let

$$\begin{aligned} \boldsymbol{A}(w) &= \frac{2}{w} g(w)\psi(w)\left(\boldsymbol{I}_p - ||\boldsymbol{X}||^{-2}\boldsymbol{X}\boldsymbol{X}'\right), \\ c(W) &= b(W) - 2W^{-1}g(W), \\ h(\lambda) &= E[2W^{-1}g(W)]. \end{aligned}$$

Then for $b(W)$ given by (2.15),

$$\begin{aligned} \widehat{\boldsymbol{M}}_I(\psi;\boldsymbol{\delta}_g) &= \boldsymbol{I}_p - \boldsymbol{A}(W) + c(W)||\boldsymbol{X}||^{-2}\boldsymbol{X}\boldsymbol{X}', \\ \boldsymbol{M}_I(\boldsymbol{\delta}_g) &= \boldsymbol{I}_p - h(\lambda)\boldsymbol{I}_p + E\left[b(W)||\boldsymbol{X}||^{-2}\boldsymbol{X}\boldsymbol{X}'\right]. \end{aligned}$$

Denoting $D(\widehat{\boldsymbol{M}}_I(\psi))$ by

$$\begin{aligned} D(\widehat{\boldsymbol{M}}_I(\psi)) &= E\left[\mathrm{tr}\{\widehat{\boldsymbol{M}}_I(\psi;\boldsymbol{\delta}_g) - \boldsymbol{M}_I(\boldsymbol{\delta}_g)\}^2\right] \\ &\quad -E\left[\mathrm{tr}\left\{c(W)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2} + h(\lambda)\boldsymbol{I}_p - E\left[b(W)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2}\right]\right\}^2\right], \end{aligned}$$

and noting that $\mathrm{tr}\boldsymbol{A}(W)c(W)\boldsymbol{X}\boldsymbol{X}'/||\boldsymbol{X}||^2 = 0$, we observe that

$$\begin{aligned} D(\widehat{\boldsymbol{M}}_I(\psi)) &= E\left[\mathrm{tr}\{\boldsymbol{A}(W)\}^2 - 2\mathrm{tr}\boldsymbol{A}(W)\left\{c(W)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2} + h(\lambda)\boldsymbol{I}_p - E\left[b(W)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2}\right]\right\}\right] \\ &= E\left[\mathrm{tr}\{\boldsymbol{A}(W)\}^2 - 2\mathrm{tr}\boldsymbol{A}(W)\left\{h(\lambda)\boldsymbol{I}_p - E\left[b(W)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2}\right]\right\}\right] \\ &= E\left[4(p-1)\frac{g^2(W)}{W^2}\psi^2(W) - 4(p-1)\frac{g(W)}{W}\psi(W)h(\lambda)\right. \\ &\quad \left. +4\frac{g(W)}{W}\psi(W)\left\{E[b(W)] - \frac{1}{||\boldsymbol{X}||^2}\boldsymbol{X}'E\left[b(W)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2}\right]\boldsymbol{X}\right\}\right]. \end{aligned} \tag{2.23}$$

Let $\boldsymbol{Y}$ be a random variable independent of $\boldsymbol{X}$, having $\mathcal{N}_p(\boldsymbol{\theta}/\sigma, \boldsymbol{I}_p)$ and let $\boldsymbol{P}$ be a $p \times p$ orthogonal matrix such that $\boldsymbol{P}\boldsymbol{\theta} = (||\boldsymbol{\theta}||, 0, \ldots, 0)'$. Also let $\boldsymbol{U} = (U_1, \ldots, U_p)' = \boldsymbol{P}\boldsymbol{X}$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_p)' = \boldsymbol{P}\boldsymbol{Y}$. Then,

$$\frac{1}{||\boldsymbol{X}||^2}\boldsymbol{X}'E\left[b\left(\frac{||\boldsymbol{Y}||^2}{S/\sigma^2}\right)\frac{\boldsymbol{Y}\boldsymbol{Y}'}{||\boldsymbol{Y}||^2}\right]\boldsymbol{X} = \frac{1}{||\boldsymbol{U}||^2}\boldsymbol{U}'E\left[b\left(\frac{||\boldsymbol{Z}||^2}{S/\sigma^2}\right)\frac{\boldsymbol{Z}\boldsymbol{Z}'}{||\boldsymbol{Z}||^2}\right]\boldsymbol{U}. \tag{2.24}$$

Noting that $Z_1, \ldots, Z_p$ are mutually independent and that Z_i has a density being symmetric on zero for $i = 2, \ldots, p$, we can see that

$$E\left[b\left(\frac{||\boldsymbol{Z}||^2}{S/\sigma^2}\right)\frac{Z_iZ_j}{||\boldsymbol{Z}||^2}\right] = 0 \quad \text{for} \quad i \neq j$$

so that the r.h.s. of the equality in (2.24) is rewritten as

$$\sum_{i=1}^{p} \frac{U_i^2}{||\boldsymbol{U}||^2} E\left[b\left(\frac{||\boldsymbol{Z}||^2}{S/\sigma^2}\right) \frac{Z_i^2}{||\boldsymbol{Z}||^2}\right]. \tag{2.25}$$

Here Z_1^2 has non-central chi-square distribution $\chi_1^2(\lambda)$ with noncentrality $\lambda = ||\boldsymbol{\theta}||^2/\sigma^2$ and one degree of freedom. Also Z_i^2 has central chi-square distribution χ_1^2 for $i = 2, \ldots, p$. For evaluating (2.25), we note that the following equality holds: for integrable function $h(\cdot)$,

$$\begin{aligned} E\left[(\chi_\nu^2)^r h(\chi_\nu^2)\right] &= \frac{\Gamma(\nu/2 + r)}{\Gamma(\nu/2)} 2^r E[h(\chi_{\nu+2r}^2)] \\ &= \begin{cases} (\nu-2)^{-1} E[h(\chi_{\nu-2}^2)] & \text{if } r = -1, \\ \nu E[h(\chi_{\nu+2}^2)] & \text{if } r = 1, \end{cases} \end{aligned}$$

where χ_ν^2 designates a random variable having a χ_ν^2-distribution. Taking this note and (2.12) into account, we observe that

$$\begin{aligned} E\left[b\left(\frac{||\boldsymbol{Z}||^2}{S/\sigma^2}\right) \frac{Z_1^2}{||\boldsymbol{Z}||^2}\right] &= E^J\left[E\left[b\left(\frac{\chi_{1+2J}^2 + \chi_{p-1}^2}{\chi_n^2}\right) \frac{\chi_{1+2J}^2}{\chi_{1+2J}^2 + \chi_{p-1}^2} \Big| J\right]\right] \\ &= E^J\left[E\left[b\left(\frac{\chi_{3+2J}^2 + \chi_{p-1}^2}{\chi_n^2}\right) \frac{1+2J}{\chi_{3+2J}^2 + \chi_{p-1}^2} \Big| J\right]\right] \\ &= E^J\left[E\left[b\left(\frac{\chi_{p+2+2J}^2}{\chi_n^2}\right) \frac{1+2J}{\chi_{p+2+2J}^2} \Big| J\right]\right] \\ &= E\left[b\left(\frac{\chi_{p+2J}^2}{\chi_n^2}\right) \frac{1+2J}{p+2J}\right]. \end{aligned}$$

Similarly, for $i \neq 1$,

$$\begin{aligned} E\left[b\left(\frac{||\boldsymbol{Z}||^2}{S/\sigma^2}\right) \frac{Z_i^2}{||\boldsymbol{Z}||^2}\right] &= E^J\left[E\left[b\left(\frac{\chi_1^2 + \chi_{p-1+2J}^2}{\chi_n^2}\right) \frac{\chi_1^2}{\chi_1^2 + \chi_{p-1+2J}^2} \Big| J\right]\right] \\ &= E\left[b\left(\frac{\chi_{p+2J}^2}{\chi_n^2}\right) \frac{1}{p+2J}\right]. \end{aligned}$$

Hence we can rewrite (2.25) as

$$\begin{aligned} \frac{1}{||\boldsymbol{U}||^2} E\left[b\left(\frac{||\boldsymbol{Z}||^2}{S/\sigma^2}\right) \frac{\sum_{i=1}^p U_i^2 Z_i^2}{||\boldsymbol{Z}||^2}\right] &= \frac{1}{||\boldsymbol{U}||^2} E\left[b\left(\frac{\chi_{p+2+2J}^2}{\chi_n^2}\right) \frac{(1+2J)U_1^2 + \sum_{i=2}^p U_i^2}{\chi_{p+2+2J}^2}\right] \\ &= \frac{1}{||\boldsymbol{U}||^2} E\left[b\left(\frac{\chi_{p+2J}^2}{\chi_n^2}\right) \frac{2J\ U_1^2 + ||\boldsymbol{U}||^2}{p+2J}\right]. \end{aligned}$$

Denoting $H(\lambda; \boldsymbol{U})$ by

$$
\begin{aligned}
H(\lambda; \boldsymbol{U}) &= (p-1)h(\lambda) - E[b(W)] + \frac{1}{\|\boldsymbol{U}\|^2}\boldsymbol{U}'E\left[b\left(\frac{\|\boldsymbol{X}\|^2}{S/\sigma^2}\right)\frac{\boldsymbol{Z}\boldsymbol{Z}'}{\|\boldsymbol{Z}\|^2}\right]\boldsymbol{U} \\
&= E\left[(p-1)\frac{2\chi_n^2}{\chi_{p+2J}^2}g\left(\frac{\chi_{p+2J}^2}{\chi_n^2}\right) - b\left(\frac{\chi_{p+2J}^2}{\chi_n^2}\right)\frac{(p-1)+2J(1-U_1^2/\|\boldsymbol{U}\|^2)}{p+2J}\right],
\end{aligned}
$$

we note that the following inequality holds by the condition (b):

$$
\begin{aligned}
H(\lambda; \boldsymbol{U}) &\leq (p-1)E\left[\frac{2\chi_n^2}{\chi_{p+2J}^2}g\left(\frac{\chi_{p+2J}^2}{\chi_n^2}\right) - \frac{1}{p+2J}b\left(\frac{\chi_{p+2J}^2}{\chi_n^2}\right)\right] \\
&\leq (p-1)b_{p,n} \qquad (2.26)
\end{aligned}
$$

for $b_{p,n}$ given by (2.16).

In this way, we can express the r.h.s. of the extreme equality in (2.23) as

$$
D(\widehat{\boldsymbol{M}}_I(\psi)) = E\left[4(p-1)\frac{g^2(W)}{W^2}\psi^2(W) - 4H(\lambda; \boldsymbol{U})\frac{g(W)}{W}\psi(W)\right].
$$

Applying the IERD method, using the inequality (2.26) and making the same transformations as in the proof of Theorem 1, we see that

$$
\begin{aligned}
& D(\widehat{\boldsymbol{M}}_I^{UB}) - D(\widehat{\boldsymbol{M}}_I(\psi)) \\
&= E\left[\int_1^\infty \frac{d}{dt}\left\{4(p-1)\frac{g^2(W)}{W^2}\psi^2(tW) - 4H(\lambda; \boldsymbol{U})\frac{g(W)}{W}\psi(tW)\right\}dt\right] \\
&= 4E\left[\int_1^\infty \left\{2(p-1)\frac{g^2(W)}{W^2}\psi(tW) - H(\lambda; \boldsymbol{U})\frac{g(W)}{W}\right\}\psi'(tW)W\,dt\right] \\
&\geq 4(p-1)E\left[\int_1^\infty \left\{2\frac{g^2(W)}{W^2}\psi(tW) - b_{p,n}\frac{g(W)}{W}\right\}\psi'(tW)W\,dt\right] \\
&= 4(p-1)\int_0^\infty E^v\left[\int_0^w \left\{2\frac{g^2(z)}{z^2}\psi(w) - b_{p,n}\frac{g(z)}{z}\right\}\psi'(w)vf_p(vz;\lambda)dz\right]dw,
\end{aligned}
$$

which is non-negative if, in the case of $\psi'(w) > 0$,

$$
\psi(w) \geq \frac{b_{p,n}}{2}\frac{E^v[\int_0^w \{g(z)/z\}vf_p(vz;\lambda)dz]}{E^v[\int_0^w \{g(z)/z\}^2 vf_p(vz;\lambda)dz]}. \qquad (2.27)
$$

From the condition (a),

$$
\frac{E^v[\int_0^w \{g(z)/z\}vf_p(vz;\lambda)dz]}{E^v[\int_0^w \{g(z)/z\}^2 vf_p(vz;\lambda)dz]} \leq \sup_{0<z<w}\left\{\frac{z}{g(z)}\right\} = \frac{w}{g(w)},
$$

so that (2.27) is satisfied by the condition (d), and the proof of Theorem 2 is complete. □□

2.4 Other situations of the covariance matrix

It can be shown that results similar to the ones stated in previous sections hold in the case of known $\sigma^2 = \sigma_0^2$. The results can be given by replacing $g(w)$, $\phi(w)$ and $\psi(w)$ in the above statements with $g(nw)/n$, $\phi(nw)$ and $\psi(nw)$ and by taking n to the infinity. More specifically, the MSE and the MSE matrix of the shrinkage estimator $\boldsymbol{\delta}_g = \{1 - g(||\boldsymbol{X}||^2)/||\boldsymbol{X}||^2\}\boldsymbol{X}$ with $\sigma_0^2 = 1$ are provided by

$$R(\boldsymbol{\delta}_g) = p - E[a(||\boldsymbol{X}||^2)],$$

where

$$a(||\boldsymbol{X}||^2) = 2(p-2)\frac{g(||\boldsymbol{X}||^2)}{||\boldsymbol{X}||^2} - \frac{g^2(||\boldsymbol{X}||^2)}{||\boldsymbol{X}||^2} + 4g'(||\boldsymbol{X}||^2),$$

and

$$\boldsymbol{M}(\boldsymbol{\delta}_g) = \boldsymbol{I}_p + E\left[-2\frac{g(||\boldsymbol{X}||^2)}{||\boldsymbol{X}||^2}\boldsymbol{I}_p + b(||\boldsymbol{X}||^2)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2}\right],$$

where

$$b(||\boldsymbol{X}||^2) = 4\frac{g(||\boldsymbol{X}||^2)}{||\boldsymbol{X}||^2} + \frac{g^2(||\boldsymbol{X}||^2)}{||\boldsymbol{X}||^2} - 4g'(||\boldsymbol{X}||^2).$$

By letting $a_p = E_{\lambda=0}[a(||\boldsymbol{X}||^2)]$ and

$$b_p = \sup_{j\geq 0} E\left[2\frac{g(\chi^2_{p+2j})}{\chi^2_{p+2j}} - \frac{b(\chi_{p+2j})}{p+2j}\right],$$

we can obtain similar results as in Theorems 1 and 2.

Also the results of Theorem 1 can be applied to the case where the covariance matrix $\boldsymbol{\Sigma}$ is fully unknown, that is, $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ and $\boldsymbol{S} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$, the Wishart distribution with $E[\boldsymbol{S}] = n\boldsymbol{\Sigma}$. The risk function of estimator $\boldsymbol{\delta}_g = \{1 - g(\boldsymbol{X}'\boldsymbol{S}^{-1}\boldsymbol{X})/\boldsymbol{X}'\boldsymbol{S}^{-1}\boldsymbol{X}\}\boldsymbol{X}$ for invariant loss $(\boldsymbol{\delta}_g - \boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\delta}_g - \boldsymbol{\theta})$ is provided by

$$\begin{aligned} R_I(\boldsymbol{\delta}_g) &= p - E[a(W)], \quad W = \boldsymbol{X}'\boldsymbol{S}^{-1}\boldsymbol{X}, \\ a(W) &= 2(p-2)\frac{g(W)}{W} + (n-p+3)\frac{g(W)}{W} \\ &\quad -4g'(W) - 4g(W)g'(W), \end{aligned}$$

so that results similar to that of Theorem 1 for this case can be obtained by replacing n with $n-p+1$.

3 Estimation of the MSE reduction matrix and its trace

We here consider estimating the quantity of how much the shrinkage estimator improves on the ML and UMVU estimator $\boldsymbol{X}$. One of such measures is the *Relative Risk Difference* $\{p\sigma^2 - R(\boldsymbol{\delta}_g)\}/(p\sigma^2) = p - R_I(\boldsymbol{\delta}_g)$, and its matricial version given by

$\{\sigma^2 \boldsymbol{I}_p - \boldsymbol{M}(\boldsymbol{\delta}_g)\}/\sigma^2 = \boldsymbol{I}_p - \boldsymbol{M}_I(\boldsymbol{\delta}_g)$. The unbiased estimators and the corresponding truncated improved estimators are similar to the ones given in Section 2.

In this section, we address the problem of estimating risk reductions (risk gains or risk differences) by the James-Stein estimator in the sense of the MSE and MSE matrix. The results given here are helpful for providing non-negative procedures for estimation of the MSE and MSE matrix which is discussed in the next section.

The MSE reduction and the MSE reduction matrix for $\boldsymbol{\delta}^{JS}$ are, respectively, denoted by

$$\begin{aligned} R^*(\boldsymbol{\delta}^{JS}) &= p\sigma^2 - R(\boldsymbol{\delta}^{JS}) \\ &= (p-2)kn\sigma^4 E\left[||\boldsymbol{X}||^{-2}\right], \\ \boldsymbol{M}^*(\boldsymbol{\delta}^{JS}) &= \sigma^2 \boldsymbol{I}_p - M(\boldsymbol{\delta}^{JS}) \\ &= kn\sigma^4 E\left[2||\boldsymbol{X}||^{-2}\boldsymbol{I}_p - (p+2)||\boldsymbol{X}||^{-4}\boldsymbol{X}\boldsymbol{X}'\right], \end{aligned}$$

for $k = (p-2)/(n+2)$. The UMVU estimators of them are given by

$$\begin{aligned} \widehat{R}^{*UB}(\boldsymbol{\delta}^{JS}) &= k^2\frac{S}{W}, \quad W = \frac{||X||^2}{S}, \\ \widehat{\boldsymbol{M}}^{*UB}(\boldsymbol{\delta}^{JS}) &= \frac{k}{n+2}\frac{S}{W}\left\{2\boldsymbol{I}_p - (p+2)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2}\right\}. \end{aligned}$$

Since $\boldsymbol{M}^*(\boldsymbol{\delta}^{JS})$ is not necessarily nonnegative definite, we do not require the non-negative definiteness for estimators of $\boldsymbol{M}^*(\boldsymbol{\delta}^{JS})$, but the estimator of $R^*(\boldsymbol{\delta}^{JS})$ should be positive. For improving upon the unbiased estimators, we consider shrinkage procedures similar to (2.8) and (2.14):

$$\begin{aligned} \widehat{R}^*(\phi;\boldsymbol{\delta}^{JS}) &= k^2\frac{S}{W}\phi(W) \\ \widehat{\boldsymbol{M}}^*(\psi;\boldsymbol{\delta}^{JS}) &= \frac{k}{n+2}\frac{S}{W}\left\{2\psi(W)\boldsymbol{I}_p - 2\psi(W)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2} - p\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2}\right\} \\ &= \frac{k}{n+2}\frac{S}{W}\boldsymbol{\Gamma}'\left\{-p\boldsymbol{E}_{11} + 2\psi(W)(\boldsymbol{I}_p - \boldsymbol{E}_{11})\right\}\boldsymbol{\Gamma}. \end{aligned}$$

The shrinkage function $\psi(w)$ operates the $(p-1)$ eigen values other than the first. Then the following theorems are obtained.

Theorem 3. *For $p \geq 5$, assume that*
(a) *$\phi(w)$ is nondecreasing and $\lim_{w\to\infty}\phi(w) = 1$,*
(b) *$\phi(w) \geq \phi^{TR}(w)$, where*

$$\phi^{TR}(w) = \min\left\{1, \frac{nw}{n+4}\right\}.$$

Then $\widehat{R}^*(\phi;\boldsymbol{\delta}^{JS})$ dominates $\widehat{R}^{*UB}(\boldsymbol{\delta}^{JS})$.

Theorem 4. *For $p \geq 5$, assume that*
(a) $\psi(w)$ *is nondecreasing and* $\lim_{w\to\infty}\psi(w) = 1$,
(b) $\psi(w) \geq \psi^{TR}(w)$, *where*

$$\psi^{TR}(w) = \min\left\{1,\ \frac{n(n+2)}{2p(n+4)}w\right\}.$$

Then $\widehat{\boldsymbol{M}}^*(\psi;\boldsymbol{\delta}^{JS})$ *dominates* $\widehat{\boldsymbol{M}}^{*UB}(\boldsymbol{\delta}^{JS})$.

The condition that $p \geq 5$ guarantees the existence of the MSEs of $\widehat{R}^*(\psi;\boldsymbol{\delta}^{JS})$ and $\widehat{\boldsymbol{M}}^*(\psi;\boldsymbol{\delta}^{JS})$. From the theorems, we get two improved truncated estimators:

$$\begin{aligned}
\widehat{R}^{*TR}(\boldsymbol{\delta}_g) &= \widehat{R}^*(\phi^{TR};\boldsymbol{\delta}^{JS}) \\
&= kS \times \min\left\{\frac{k}{W},\ \frac{n}{n+4}\right\}, \\
\widehat{\boldsymbol{M}}^{*TR}(\boldsymbol{\delta}^{JS}) &= \widehat{\boldsymbol{M}}^*(\psi^{TR};\boldsymbol{\delta}^{JS}) \\
&= \frac{k}{n+2}\frac{S}{W}\left[\min\left\{2, \frac{n(n+2)}{p(n+4)}W\right\}\boldsymbol{I}_p\right. \\
&\qquad \left. -\min\left\{p+2, p+\frac{n(n+2)}{p(n+4)}W\right\}\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2}\right].
\end{aligned}$$

Proof of Theorem 3. Let $C(\widehat{R}^*(\phi)) = k^{-2}\sigma^{-4}E[(\widehat{R}^*(\phi;\boldsymbol{\delta}^{JS}) - R(\boldsymbol{\delta}^{JS}))^2]$. Then, by the same arguments as in the proof of Theorem 1,

$$\begin{aligned}
&C(\widehat{R}^{*UB}) - C(\widehat{R}^*(\phi)) \\
&= E\left[\int_1^\infty \frac{d}{dt}\left\{\frac{S/\sigma^2}{W}\phi(tW) - E\left[\frac{n(n+2)\sigma^2}{||\boldsymbol{X}||^2}\right]\right\}^2 dt\right] \\
&= 2E\left[\int_1^\infty \left\{\frac{S/\sigma^2}{W}\phi(tW) - E\left[\frac{n(n+2)\sigma^2}{||\boldsymbol{X}||^2}\right]\right\}\frac{S/\sigma^2}{W}\phi'(tW)W dt\right] \\
&\geq 2\int_0^\infty E^v\left[\int_0^{vw}\left\{\frac{v^2}{x}\phi(w) - \frac{n(n+2)}{p-2}\right\}\frac{v^2}{x}\phi'(w)f_p(x;\lambda)dx\right]dw \\
&= 2\int_0^\infty E^v\Big[\int_0^{vw}\frac{1}{x}f_p(x;\lambda)dx v^2\Big\{\frac{\int_0^{vw}x^{-2}f_p(x;\lambda)dx}{\int_0^{vw}x^{-1}f_p(x;\lambda)dx}v^2\phi(w) - \frac{n}{k}\Big\}\Big]\phi'(w)dw \\
&\geq 2\inf_w\left\{E^v\left[\int_0^{vw}x^{-1}f_p(x;\lambda)dx \times v^2\left\{v\frac{\phi(w)}{w} - \frac{n}{k}\right\}\right]\phi'(w)\right\}
\end{aligned}$$

where the first inequality follows from the condition (a) and the fact that $E_\lambda[\sigma^2/||\boldsymbol{X}||^2] \leq 1/(p-2)$. Note that $\int_0^{vw}x^{-1}f_p(x;\lambda)dx$ is increasing in v and that $q(v) = v^2\{v\phi(w)/w - n/k\}$ has one sign change, that is, there is some v_0 such that $q(v) \leq 0$ for $0 < v \leq v_0$ and

$q(v) > 0$ for $v > v_0$. Then we can see that

$$E^v\left[\int_0^{vw} x^{-1} f_p(x;\lambda)dx \times v^2\left\{v\frac{\phi(w)}{w} - \frac{n}{k}\right\}\right]$$

$$\geq \left\{\int_0^{v_0 w} x^{-1} f_p(x;\lambda)dx\right\} \times E^v\left[v^2\left\{v\frac{\phi(w)}{w} - \frac{n}{k}\right\}\right]$$

(For instance, see Kubokawa(1998b)). This inequality implies that $C(\widehat{R}^{*UB}) - C(\widehat{R}^*(\phi)) \geq 0$ if, in the case of $\phi'(w) > 0$,

$$E^v\left[v^3\frac{\phi(w)}{w} - v^2\frac{n}{p-2}\right] \geq 0,$$

which is satisfied by the condition (b), and Theorem 3 is proved. □□

Proof of Theorem 4. Denote $D(\widehat{\boldsymbol{M}}^*(\psi)) = E[\mathrm{tr}\{\widehat{\boldsymbol{M}}^*(\psi;\boldsymbol{\delta}^{JS}) - M^*(\boldsymbol{\delta}^{JS})\}^2]/(k^2\sigma^4)$, and let $h(\lambda) = E[\sigma^2||\boldsymbol{X}||^{-2}]$ and $\boldsymbol{H}(\lambda) = \sigma^2 E[||\boldsymbol{X}||^{-4}\boldsymbol{X}\boldsymbol{X}']$. Then, similar to the proofs of Theorems 2 and 3,

$$\begin{aligned}
& D(\widehat{\boldsymbol{M}}^{*UB}) - D(\widehat{\boldsymbol{M}}^*(\psi)) \\
= \ & 4E\left[\int_1^\infty \frac{d}{dt}\left\{\frac{p-1}{(n+2)^2}\frac{S^2/\sigma^4}{W^2}\psi^2(tW) - \frac{n}{n+2}\Big[2(p-1)h(\lambda)\right.\right. \\
& \left.\left. -(p+2)\left\{h(\lambda) - ||\boldsymbol{X}||^{-2}\boldsymbol{X}'\boldsymbol{H}(\lambda)\boldsymbol{X}\right\}\Big]\frac{S/\sigma^2}{W}\psi(tW)\right\}dt\right] \\
= \ & 4E\left[\int_1^\infty \left\{2\frac{p-1}{(n+2)^2}\frac{S^2/\sigma^4}{W^2}\psi(tW) - \frac{n}{n+2}\Big[2(p-1)h(\lambda)\right.\right. \\
& \left.\left. -(p+2)\left\{h(\lambda) - ||\boldsymbol{X}||^{-2}\boldsymbol{X}'\boldsymbol{H}(\lambda)\boldsymbol{X}\right\}\Big]\frac{S/\sigma^2}{W}\right\}\psi'(tW)W\,dt\right] \\
\geq \ & 4E\left[\int_1^\infty \left\{2\frac{p-1}{(n+2)^2}\frac{S^2/\sigma^4}{W^2}\psi(tW) - \frac{n}{n+2}\frac{p-1}{p}\frac{S}{W}\right\}\psi'(tW)W\,dt\right],
\end{aligned}$$

since from (2.18) and (2.26),

$$\begin{aligned}
& E^Y\left[2(p-1)\frac{1}{||\boldsymbol{Y}||^2} - (p+2)\left\{\frac{1}{||\boldsymbol{Y}||^2} - \frac{1}{||\boldsymbol{X}||^2}\boldsymbol{X}'\frac{\boldsymbol{Y}\boldsymbol{Y}'}{||\boldsymbol{Y}||^4}\boldsymbol{X}\right\}\right] \\
\leq \ & (p-1)E^Y\left[\frac{2}{||\boldsymbol{Y}||^2} - \frac{p+2}{(p+2J)||\boldsymbol{Y}||^2}\right] \\
\leq \ & (p-1)\sup_j \frac{p-2+4j}{(p+2j)(p-2+2j)} = \frac{p-1}{p}.
\end{aligned}$$

Similar to the proof of Theorem 3, it can be verified that $D(\widehat{\boldsymbol{M}}^{*UB}) - D(\widehat{\boldsymbol{M}}^*(\psi)) \geq 0$ if, in the case of $\psi'(w) > 0$,

$$E^v\left[2(p-1)v^3\frac{\psi(w)}{w} - \frac{(n+2)n(p-1)}{p}v^2\right] \geq 0,$$

which is satisfied by the condition (b), and we get Theorem 4. □□

4 Estimation of the MSE and the MSE Matrix

In Section 2, the estimation of the scale invariant MSE and MSE matrix is treated. In the situation of constructing confidence sets based on the Stein-rule estimators, the non-scale-invariant measure, namely the usual MSE $R(\boldsymbol{\delta}^{JS})$ and MSE matrix $\boldsymbol{M}(\boldsymbol{\delta}^{JS})$ need to be estimated. This issue of the estimation is a bit different from that of estimating the scale-invariant MSE matrix $\boldsymbol{M}_I(\boldsymbol{\delta}^{JS})$ and its trace $R_I(\boldsymbol{\delta}^{JS})$, for the parameter space of the usual MSE is a whole space of positive real numbers and zero, that is, it is not restricted to any interval of positive numbers far from zero. Although the unbiased estimator of the MSE has an undesirable property of taking negative values, we could not apply the arguments used in Section 2 to develop any *positive* improved estimator of it. Of course, the unbiased estimator can be improved on by being truncated at zero. It is, however, not practical to estimate MSE by zero. A usual confidence set based on the James-Stein estimator may be of the form $(\boldsymbol{\delta}^{JS}-\boldsymbol{\theta})'\widehat{\boldsymbol{M}}^{-1}(\boldsymbol{\delta}^{JS}-\boldsymbol{\theta}) \leq c$ where c is a positive constant and $\widehat{\boldsymbol{M}}$ is an estimator of $E[(\boldsymbol{\delta}^{JS}-\boldsymbol{\theta})(\boldsymbol{\delta}^{JS}-\boldsymbol{\theta})'] = \sigma^2\boldsymbol{M}_I(\boldsymbol{\delta}^{JS}) = \boldsymbol{M}(\boldsymbol{\delta}^{JS})$. When a zero-truncated estimator is used for $\boldsymbol{M}(\boldsymbol{\delta}^{JS})$, it has a singularity with a positive probability, and so we could not construct any meaningful confidence sets.

In this section, we propose positive and positive definite estimators of the MSE $R(\boldsymbol{\delta}^{JS})$ and the MSE matrix $\boldsymbol{M}(\boldsymbol{\delta}^{JS})$ which are made from the estimation of the MSE reduction $R^*(\boldsymbol{\delta}^{JS})$ and reduction matrix $\boldsymbol{M}^*(\boldsymbol{\delta}^{JS})$. Although the preference of such procedures is not verified analytically, they have practical sense.

The UMVU estimators of $R(\boldsymbol{\delta}^{JS})$ and $\boldsymbol{M}(\boldsymbol{\delta}^{JS})$ are described, respectively, by

$$\begin{aligned}\widehat{R}^{UB}(\boldsymbol{\delta}^{JS}) &= p\frac{S}{n} - k^2\frac{S}{W}, \\ \widehat{\boldsymbol{M}}^{UB}(\boldsymbol{\delta}^{JS}) &= \frac{S}{n}\boldsymbol{I}_p - \frac{k}{n+2}\frac{S}{W}\left\{2\boldsymbol{I}_p - (p+2)\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2}\right\},\end{aligned}$$

which have drawbacks of taking negative values. As procedures suggested from the results of Theorems 3 and 4, we here propose the truncated estimators of the forms

$$\begin{aligned}\widehat{R}_{MSE}(\boldsymbol{\delta}^{JS}) &= \frac{S}{n}\times\max\left\{p - k^2\frac{n}{W},\ p - \frac{kn^2}{n+4}\right\}, \\ \widehat{\boldsymbol{M}}_{MSE}(\boldsymbol{\delta}^{JS}) &= \frac{S}{n}\times\max\left\{1 - \frac{2nk}{n+2}\frac{1}{W},\ 1 - \frac{kn^2}{p(n+4)}\right\}\boldsymbol{I}_p \\ &\quad + \frac{k}{n+2}\frac{S}{W}\min\left\{p+2, p+\frac{n(n+2)}{p(n+4)}W\right\}\frac{\boldsymbol{X}\boldsymbol{X}'}{||\boldsymbol{X}||^2},\end{aligned}$$

which are positive and positive-definite, since $p - kn^2(n+4)^{-1}$ is always greater than zero. Although their improvements are not guaranteed analytically, they will be good candidates for practical uses.

Another choice of getting positive or positive definite estimators is to apply Propositions 1 and 2, which yield almost surely positive or positive definite procedures, and they are guaranteed to improve upon the UMVU estimators.

Acknowledgments. The research of the first author was supported in part by a grant from the Mitsubishi Seimeikai (the Mitsubishi Bank co.), by a grant from the Center for International Research on the Japanese Economy, the University of Tokyo, and by the Ministry of Education, Japan, Grant No. 09780214. The research of the second author was supported in part by Natural Sciences and Engineering Research Council of Canada. This work was done while the first author was visiting Department of Statistics, the University of Toronto, 1995 summer, fall and 1998 summer. We are grateful to Mr. S. Aoki for his help in the simulation experiments.

REFERENCES

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.*, **83**, 28-36.

Bilodeau, M. and Srivastava, M.S. (1988). Estimation of the MSE matrix of the Stein estimator. *Canadian J. Statist.* **16**, 153-159.

Breiman, L. and Friedman, J.H. (1997). Predicting multivariate responses in multiple linear regression (with discussions). *J. Royal Statist. Soc., B* **59**, 3-54.

Carter, R.A.L., Srivastava, M.S., Srivastava, V.K. and Ullah, A. (1988). Unbiased estimation of the MSE matrix of Stein-rule estimators, confidence ellipsoids, and Hypothesis testing. *Econometric Theory* **6**, 63-74.

Casella, G. (1990). Estimators with nondecreasing risk: Application of a chi-squared identity. *Statist. Prob. Letters*, **10**, 107-109.

Efron,, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.* **70**, 311-319.

Efron,, B. and Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **4**, 11-21.

Fay, R.E. and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74**, 269-277.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, **1**, 361-379. University of California Press, Berkeley.

Johnstone, I. (1987). On inadmissibility of some unbiased estimates of loss. In *Statistical Decision Theory and Related Topics* IV (S.S. Gupta, J.O. Berger, eds.). Springer-Verlag, New York.

Kubokawa, T. (1988). Monotonicity of risk for a shrinkage estimator of a multivariate normal mean. *Comm. Statist.-Theory Methods*, **17**, 499-506.

Kubokawa, T. (1994). A unified approach to improving equivariant estimators. *Ann. Statist.*, **22**, 290-299.

Kubokawa, T. (1998a). The Stein phenomenon in simultaneous estimation: A review. *Applied Statistical Science III* (eds. S.E. Ahmed, M. Ahsanullah and B.K. Sinha), NOVA Science Publishers, Inc., New York, to appear.

Kubokawa, T. (1998b). Shrinkage and modification techniques in estimation of variance and the related problems: A review. *Comm. Statist. - Theory Methods*, to appear.

Lu, K.L. and Berger, J.O. (1989). Estimation of normal means: Frequentist estimation of loss. *Ann. Statist.*, **17**, 890-906.

Rukhin, A.L. (1987). Estimated loss and admissible loss estimators. In *Statistical Decision Theory and Related Topics IV* (s.s. Gupta and J.O. Berger, eds.), Springer, New York.

Stein, C. (1973). Estimation of the mean of a multivariate normal distribution. In *Proc. Prague Symp. Asymtotic Statist.*, 345-381.

Tsutakawa, R.K., Shoop. G.L. and Marienfeld, C.J. (1985). Empirical Bayes estimation of cancer mortality rates. *Statist. Medicine*, **4**, 201-212.

Venter, J.H. and Steel. S.J. (1990). Estimating risk reduction in Stein estimation. *Canadian J. Statist.*, **18**, 221-232.